SUPPLEMENTARY APPENDIX

Co-occurrence of cohesin complex and Ras signaling mutations during progression from myelodysplastic syndromes to secondary acute myeloid leukemia

Marta Martín-Izquierdo,[1*] María Abáigar,[1*] Jesús M. Hernández-Sánchez,[1] David Tamborero,[2,3] Félix López-Cadenas,[4] Fernando Ramos,[5] Eva Lumbreras,[1] Andrés Madinaveitia-Ochoa,[6] Marta Megido,[7] Jorge Labrador,[8] Javier Sánchez-Real,[5] Carmen Olivier,[9] Julio Dávila,[10] Carlos Aguilar,[11] Juan N. Rodríguez,[12] Guillermo Martín-Nuñez,[13] Sandra Santos-Mínguez,[1] Cristina Miguel-García,[1] Rocío Benito,[1] María Díez-Campelo[4*] and Jesús M. Hernández-Rivas[1,4#]

[1]Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC), Salamanca, Spain; [2]Research Program on Biomedical Informatics, Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain; [3]Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; [4]University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain; [5]Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain; [6]Hematology, Hospital Universitario Miguel Servet, Zaragoza, Spain; [7]Hematology, Hospital del Bierzo, Ponferrada, León, Spain; [8]Hematology, Hospital Universitario de Burgos, Burgos, Spain; [9]Hematology, Hospital General de Segovia, Segovia, Spain; [10]Hematology, Hospital Nuestra Señora de Sónsoles, Ávila, Spain; [11]Hematology, Hospital Santa Bárbara, Soria, Spain; [12]Hematology, Hospital Juan Ramón Jiménez, Huelva, Spain and [13]Hematology, Hospital Virgen del Puerto, Plasencia, Spain

**MA and MMI contributed equally as co-first authors.*
[#]MDC and JMHR contributed equally as co-senior authors.

 doi:10.3324/haematol.2020.248807

Received: February 3, 2020.

Accepted: July 14, 2020.

Pre-published: July 16, 2020.

Correspondence: *JESÚS M HERNÁNDEZ-RIVAS* -jmhr@usal.es

Supplementary materials

CO-OCCURRENCE OF COHESIN COMPLEX AND RAS SIGNALING MUTATIONS DURING PROGRESION FROM MYELODYSPLASTIC SYNDROMES TO SECONDARY ACUTE MYELOID LEUKEMIA

Marta Martín-Izquierdo, María Abáigar, Jesús M Hernández-Sánchez, David Tamborero, Félix López-Cadenas, Fernando Ramos, Eva Lumbreras, Andrés Madinaveitia-Ochoa, Marta Megido, Jorge Labrador, Javier Sánchez-Real, Carmen Olivier, Julio Dávila, Carlos Aguilar, Juan Nicolás Rodríguez, Guillermo Martín-Nuñez, Sandra Santos-Mínguez, Cristina Miguel-García, Rocío Benito, María Díez-Campelo, Jesús María Hernández-Rivas.

This section contains:

- Supplementary methods.
- 5 supplementary tables.
- 5 supplementary figures.

SUPPLEMENTARY METHODS

Patients

To study the mutational changes occurring during the evolution to secondary acute myeloid leukemia from a previous myelodysplastic/myelomonocytic phase, 486 samples from 437 patients, from different Spanish institutions, were included in the study. Diagnoses were established according to the 2008 World Health Organization criteria (1). For the purpose of analysis, the RARS, RCUD, RCMD, and MDS del(5q) morphological subtypes were considered to be low-risk MDS (LR-MDS), while RAEB-1 and RAEB-2 were considered high-risk MDS (HR- MDS). Conventional cytogenetic and FISH analyses were carried out in all samples, as previously described(2-4).

DNA isolation

Genomic DNA (gDNA) was obtained from all samples from BM/PB fixed pelleted cells or mononuclear cells using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's standard protocol. The concentration of extracted DNA was determined using a Qubit® 2.0 Fluorometer system (Life Techonologies, Carlsbad, CA, USA) and the adequate quality for the sequencing was tested using a TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA) and a nanodrop spectrophotometer (ND-1000, NanoDrop Technologies, Wilmington, DE, USA) by measuring the ratio of absorbance at 230/260 and 280 nm (A230/280 and A260/280).

Whole-exome sequencing summary: Construction of DNA libraries and data analysis

Sequencing libraries were prepared using high-quality native gDNA (not subjected to whole genome amplification) as the starting material, processed with the *TruSeq Exome Enrichment* Kit (Illumina, San Diego, CA, USA), covering a total of 62 Mb of the genome with target sequences encompassing exon, UTR and miRNA loci, according to the manufacturer's protocol(5-7). Enriched exome fragments were then sequenced on an Illumina-HiSeq 200 sequencer. Unfortunately, corresponding non-tumor samples were not available for these cases.

Quality assessment, alignment and variant calling of the sequencing data were performed using an in-house pipeline based on custom scripts and open-source software, as previously described(5-8). Since no germline filter could be applied, to identify somatically acquired deleterious changes, variants were selected according to their absence from the healthy population, using the information contained in the SNP database of human variants (dbSNP, MAF < 0.01) and knowledge of its putative effect on the protein (excluding synonymous variants and those whose structure was not correctly annotated)(9). In the next step, driver mutations were identified and selected over passenger variants using an *in silico* analysis with the oncodriveMUT method from the novel "Cancer Genome Interpreter" bioinformatic tool (https://www.cancergenomeinterpreter.org/home)(10).

Targeted-deep sequencing: custom gene panel and data analysis

Targeted-deep sequencing was performed using an in-house custom capture-enrichment panel (*Nextera Rapid Capture Enrichment*, Illumina) of 117 genes previously related to the pathogenesis of myeloid malignancies, according to a Nextera sequencing design using *Illumina DesignStudio*. Sequencing libraries were prepared according to the manufacturer's instructions, using unique barcodes for each sample, multiplexed and sequenced on Illumina NextSeq 500 and MiSeq sequencers.

All sequences were evaluated using *FastQC* and *NGSQCToolkit v2.3.3* software and aligned to the reference genome (GRCh37/hg19) using *BWA v0.7.12* and *GATK v3.5*. A minimum quality score of Q30 was required to ensure high-quality sequencing results. Variant calling and annotation were performed using an in-house pipeline, based on the *VarScan v2.3.9*, *SAMTools v1.3.1.,* and *ANNOVAR* bioinformatic tools. FLT3-ITD detection was performed using *ITDseek.* To visualize read alignments and variant calls, *Integrative Genomics Viewer version 2.3.68* (*IGV*, Broad Institute, Cambridge, MA, USA) was used.

For true oncogenic somatic variant calling, a severe criterion for variant filtering was applied. Thus, synonymous, noncoding variants and polymorphisms, present at a population frequency (MAF) ≥ 1% in *dbSNP138*, *1000G*, *EXAC*, *ESP6500* and our in-house databases, were excluded. Similarly, those variants recurrently observed and, from visual inspection on the *IGV* browser, suspected of being sequencing errors were removed. The remaining variants were considered candidate somatic mutations based on the following criteria: (i) variants with ≥10 mutated reads; (ii) described in *COSMIC* and/or *ClinVar* as being cancer-associated and known hotspot mutations; and (iii) classified as deleterious and/or probably damaging by PolyPhen-2 and SIFT web-based platforms. In addition, within each case from the discovery and control cohorts, variants found on only one occasion were carefully checked at the other disease stage, because the flow read depth might have caused them to be missed, but its variant allele frequency (VAF) at the different evolutionary stages was of clear interest.

Mutation validation

To validate SNVs and indels, all alterations detected in the discovery and control cohorts were resequenced using an amplicon-based approach (*Illumina Nextera XT*) on both paired samples at much higher coverage (mean depth of 5244X). In brief, genomic regions of interest (500-800 bp) were PCR-amplified using sequence-specific primers and purified with AMPure Beads. Libraries were prepared for sequencing following the *Nextera XT Illumina* protocol and sequenced on an Illumina MiSeq sequencer.

Pathway analysis

We compiled a list of seven biological pathways described in other previous studies as being related to MDS (11, 12). Moreover, pathway classification was determined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Gene Ontology Consortium databases (13, 14). Then, we classified our panel genes with respect to them and, consequently, the mutations

were also classified with respect to these biological pathways depending on the function of the affected gene.

Statistical analysis

Baseline characteristics were described as frequencies for categorical variables and as the medians and ranges for quantitative variables. Comparisons of categorical variables between patient subsets were performed using Chi-square or Fisher´s exact test, as appropriate, while the *t*-test, or Mann-Whitney U test and Wilcoxon signed-rank test were used to compare the means and medians of continuous variables of unpaired and paired data, respectively. The Kaplan-Meier method was used to analyze survival outcomes (sAML-progression-free and overall survival). Two-sided values of $p < 0.05$ were considered to be statistically significant.

Supplementary References

1. Swerdlow SH, Campo E, Harris NL, et al (eds). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. 4th edn. (IARC Publications, Lyon, 2008).

2. Hernandez JM, Gonzalez MB, Granada I, et al. Detection of inv(16) and t(16;16) by fluorescence in situ hybridization in acute myeloid leukemia M4Eo. Haematologica. 2000;85(5):481-485.

3. Shaffer LG, McGowan-Jordan J, Schmid M (eds). ISCN 2013: An International System for Human Cytogenetic Nomenclature (2013): Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature. (Karger Medical and Scientific Publishers, Switzerland, 2013).

4. Gonzalez MB, Hernandez JM, Garcia JL, et al. The value of fluorescence in situ hybridization for the detection of 11q in multiple myeloma. Haematologica. 2004;89(10):1213-1218.

5. Hernandez-Sanchez M, Kotaskova J, Rodriguez AE, et al. CLL cells cumulate genetic aberrations prior to the first therapy even in outwardly inactive disease phase. Leukemia. 2019;33(2):518-558.

6. Mossner M, Jann JC, Wittig J, et al. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. Blood. 2016;128(9):1246-1259.

7. da Silva-Coelho P, Kroeze LI, Yoshida K, et al. Clonal evolution in myelodysplastic syndromes. Nat Commun. 2017;8:15099.

8. Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet. 2017;49(2):204-212.

9. Ibanez M, Carbonell-Caballero J, Such E, et al. The modular network structure of the mutational landscape of Acute Myeloid Leukemia. PLoS One. 2018;13(10):e0202926.

10. Tamborero D, Rubio-Perez C, Deu-Pons J, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10(1):25.

11. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-247.

12. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616-3627; quiz 3699.

13. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.

14. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-29.

SUPPLEMENTARY TABLES

Table S1: Main clinical and biological characteristics of the three patient cohorts included in the study: demographics, WHO 2008 subtypes, IPSS/IPSS-R risk classification, peripheral blood parameters and cytogenetics.

Table S2: Coverage of all 32 samples studied by WES. An excel file.

Table S3: Panel of 117 myeloid-related genes used for TDS.

Table S4: List of all mutations found by WES in the discovery cohort. The table includes information on chromosome position (GRCh37/hg19), change at DNA level, type of change, VAF percentage at MDS and sAML stage, VAF ratio between sAML and MDS, recurrence and driver/passenger prediction from Cancer Genome Interpreter. An excel file.

Table S5: List of all mutations found by TDS in the discovery and control cohorts. The table includes information on chromosome position (GRCh37/hg19), change at DNA level, VAF percentage at MDS and sAML stages, VAF ratio between sAML and MDS and probabilities indicating a significant change in VAF during the evolution of the disease. An excel file.

Supplementary Table S1

	Discovery cohort	Control cohort	Validation cohort
Number of samples	84	14	388
Number of patients	42	7	388
Gender, male (%)	30 (70)	2 (29)	221 (57)
Diagnosis Age, median (range, years)	70 (50-82)	69 (68-77)	75 (29-92)
Deceased (%)	82	42.9	44.76
sAML Progression (%)	100	0	16.2
Classification (WHO 2008) at diagnosis (%)			
RCUD	1 (2)	1 (14.3)	18 (5)
RARS	1 (2)	0	40 (10)
RCMD	16 (38)	5 (71.4)	166 (43)
RAEB-1	13 (31)	1 (14.3)	43 (11)
RAEB-2	10 (24)	0	49 (13)
MDS del(5q)-	1 (2)	0	39 (10)
MDS-U	0	0	11 (3)
CMML	0	0	9 (2)
AML with dysplastic changes	0	0	3 (1)
Not available	0	0	10 (3)
IPSS classification (%)			
Low	11 (26)	4 (57.1)	90 (23)
Intermediate-1	17 (40)	3 (42.9)	66 (17)
Intermediate-2	9 (21)	0	7 (2)
High	1 (2)	0	2 (1)
Not available*	4 (10)	0	223 (57)
IPSS-R classification			
Very low	3 (7)	4 (57.1)	49 (13)
Low	10 (24)	3 (42.9)	76 (20)
Intermediate	10 (24)	0	30 (8)
High	4 (10)	0	11 (3)
Very high	5 (12)	0	3 (1)
Not available*	10 (24)	0	219 (56)
Blood count at diagnosis			
Hemoglobin, g/dL, median (range)	10.1 (5.5-14.6)	11.4 (9.8-12.8)	9.9 (3.8-15.3)
Platelet, $x10^9$/L, median (range)	95.5 (11.0-379.0)	171.0 (54.0-503.0)	155.0 (2.0-1067.0)
ANC, $x10^9$/L, median (range)	1.6 (0.1-6.2)	2.35 (1.36-4.11)	1.9 (0.1-56.0)
WBC, $x10^9$/L, median (range)	4.2 (1.8-12.8)	5.1 (4.4-5.9)	4.1 (1.2-67.9)
Cytogenetics at diagnosis			
Normal (%)	22 (52)	6 (85.7)	197 (51)
Complex, >3 abnormalities (%)	5 (12)	0	14 (4)
-5/del(5q) (%)	4 (10)	1 (14.3)	38 (10)
-7/del(7q) (%)	1 (2)	0	4 (1)
Trisomy 8	3 (7)	0	14 (4)
del(20q)	2 (5)	0	3 (1)
Other single abnormality	1 (2)	0	52 (13)
Not available	4 (10)	0	66 (17)

*: IPSS and IPSS-R risk classifications were not applicable in the case of chronic myelomonocytic leukemia.

Abbreviations: RCUD, refractory cytopenia with uni-lineage dysplasia; RARS, refractory anemia with ringed sideroblasts; RCDM, refractory cytopenia with multi-lineage dysplasia; RAEB, refractory anemia with excess blasts; MDS del(5q)-, myelodysplastic syndrome associated with isolated del(5q); MDS-U, myelodysplastic syndrome unclassified; CMML, chronic myelomonocytic leukemia; AML, acute myeloid leukemia; IPSS, International Prognostic Scoring System; WBC, white blood cell; ANC, absolute neutrophil count.

Supplementary Table S3

ABL1	*CBLC*	*EED*	*HRAS*	*MECOM*	*PHF19*	*SF1*	*TET2*
AEBP2	*CD177*	*EGFR*	*IDH1*	*KMT2A*	*PHF6*	*SF3A1*	*TGM2*
ARID2	*CDH13*	*EIF2AK2*	*IDH2*	*KMT2D*	*PHLPP1*	*SF3B1*	*TIMM50*
ASXL1	*CDH23*	*ENG*	*IKZF1*	*MPL*	*PTEN*	*SFPQ*	*TNFAIP3*
ATRX	*CDH3*	*EP300*	*IL3*	*MTOR*	*PTPN1*	*SH2B3*	*TP53*
BCAS1	*CDK2*	*ETV6*	*IRF1*	*NF1*	*PTPN11*	*SMC1A*	*TYK2*
BCOR	*CDKN2A*	*EZH2*	*JAK1*	*NOTCH1*	*RAD21*	*SMC3*	*U2AF1*
BCORL1	*CEBPA*	*FBXW7*	*JAK2*	*NPM1*	*RARA*	*SPARC*	*UMODL1*
BCR	*CREBBP*	*FLT3*	*JAK3*	*NRAS*	*RET*	*SRSF2*	*USB1*
BMI1	*CSF3R*	*G3BP1*	*JARID2*	*NR2F6*	*RPS14*	*STAG1*	*WASF3*
BRAF	*CSNK1A1*	*GATA1*	*JKAMP*	*NTRK1*	*RUNX1*	*STAG2*	*WT1*
CALR	*CTCF*	*GATA2*	*KDM6A*	*NUP98*	*SALL4*	*SUZ12*	*ZRSR2*
CBFB	*CTNNA1*	*GCAT*	*KIT*	*PBRM1*	*SBDS*	*TCL1B*	
CBL	*CUX1*	*GNAS*	*KRAS*	*PDGFRA*	*SETBP1*	*TERC*	
CBLB	*DNMT3A*	*GNB1*	*LUC7L2*	*PDGFRB*	*SETD2*	*TERT*	

SUPPLEMENTARY FIGURES

Figure S1: Overview of the study design and the distribution of cohorts of patients included in the study. Time of first and second sampling, where diagnosis and follow-up/sAML, respectively, are specified, as well as the number of patients and samples analyzed, and the sequencing strategy applied for each of the cohorts.

Abbreviations: sAML, secondary acute myeloid leukemia; LR, low-risk, HR, high-risk; HMs, hematological malignancies; pts, patients; WES, whole-exome sequencing.

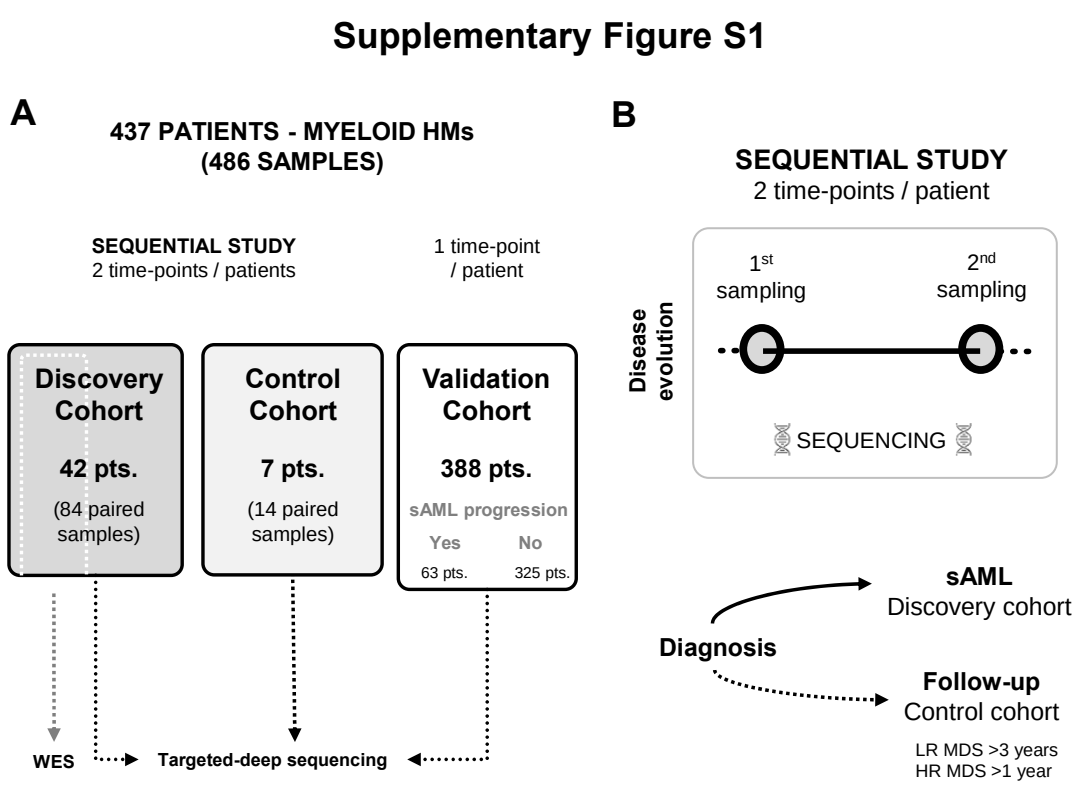


Figure S2: VAF comparison of the mutations detected by whole-exome sequencing *vs.* by targeted deed sequencing. The VAF correlation between these two platforms was high measured by Pearson coefficient (Pearson´s r = 0.90).

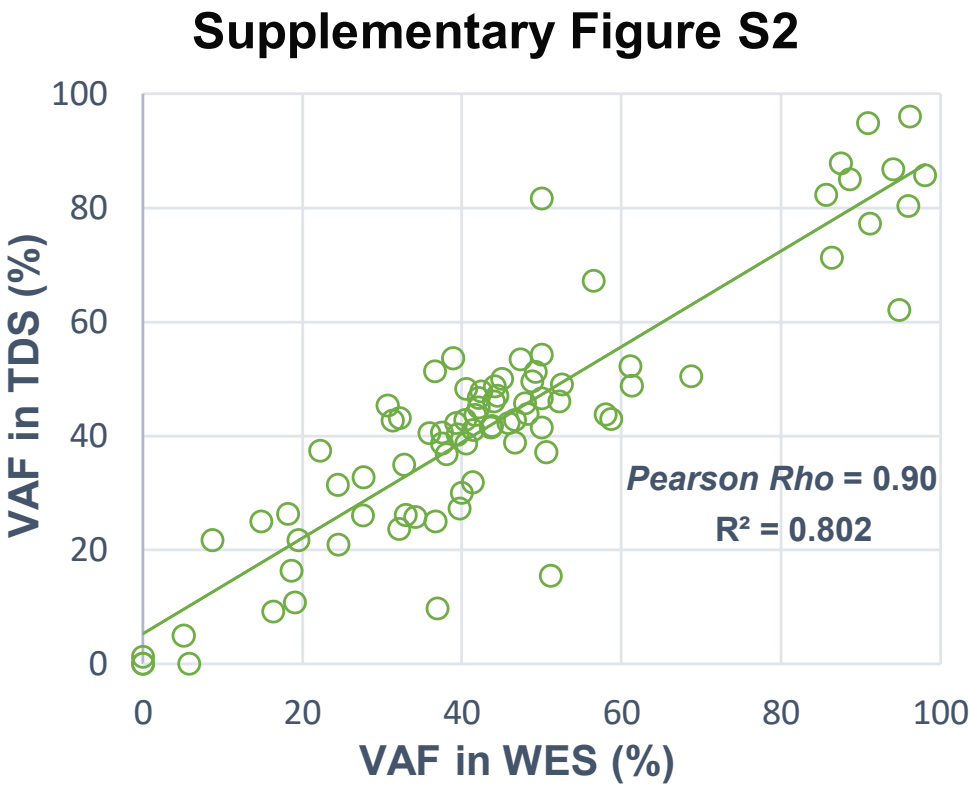
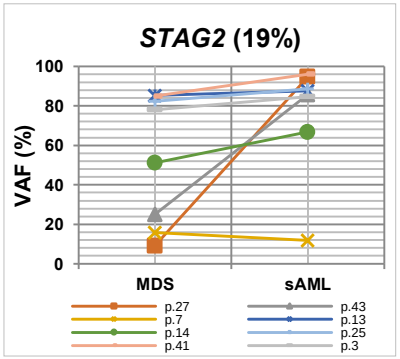


Figure S3: Representative examples of genes with different mutational dynamics in MDS patients who evolved to sAML. VAF at diagnosis and sAML of all mutations detected in *STAG2*, *NRAS*, *FLT3*, *SRSF2* and *DNMT3A* in our discovery cohort were represented showing type 1, type 3 and type 4 dynamics, respectively, during MDS progression.

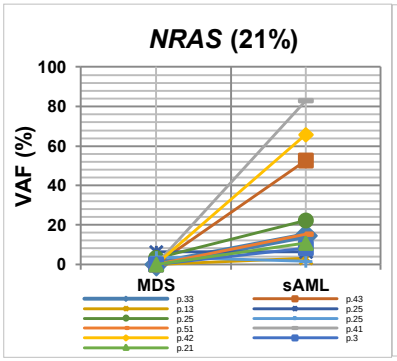
Supplementary Figure S3

Type-1 mutations

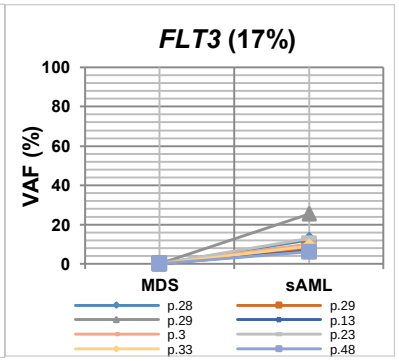


p-value= 0.0234

Type-3 mutations

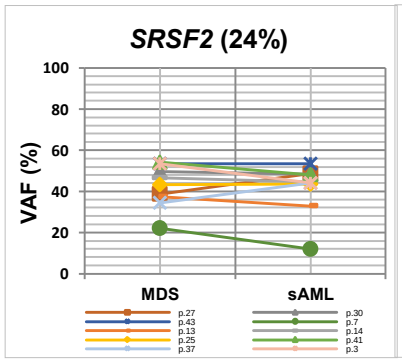


p-value= 0.0029

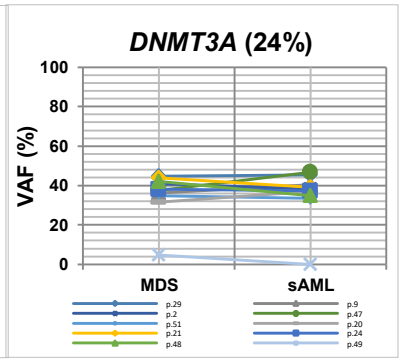


p-value= 0.0078

Type-4 mutations



p-value= 0.4922



p-value= 0.7695

Figure S4: Prognostic impact of single mutations and co-occurring mutations in the cohesin complex and Ras pathway. Kaplan-Meier curves for overall survival in double-mutant and cohesin and Ras single mutant patients in the entire validation cohort.

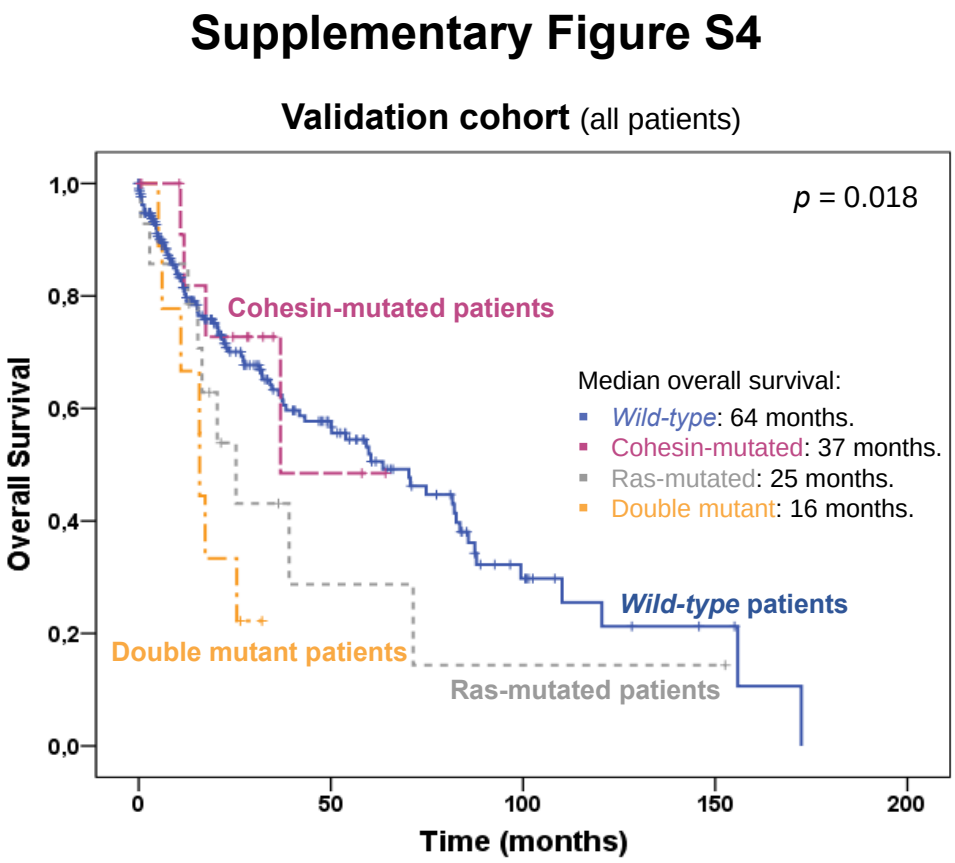


Figure S5: Landscape of mutational dynamics according to disease-modifying treatment on MDS patients who progressed to sAML. A) Patients in the discovery cohort were grouped according to whether they received treatment with a disease-modifying agent (38% 5-azacytidine or 10% lenalidomide) or supportive or no treatment, before they transformed into sAML. Genes are grouped by cellular functions and are represented in rows; patients are represented by separate columns. Dynamics are indicated by a color gradient: red/orange for newly acquired/increasing mutations, yellow for stable mutations, and blue/green colors for decreasing mutations. **B)** Graphs representing the proportion of patients harboring newly acquired/increasing (black color) and stable mutations (white color) in treated *vs.* non treated patients in the following cellular functions: chromatin modifiers, cohesin complex and Ras signaling.

Abbreviations: VAF, variant allele frequency; LR, low-risk; HR, high-risk; NS, not significant; *, $p < 0.05$.

Supplementary Figure S5