Department of **MEDICINE, HEMATOLOGY**

CANCER RESEARCH CENTER-IBMCC (USAL-CSIC)

Doctoral Dissertation

Molecular Mechanisms and Clonal Evolution underlying the Progression of Myelodysplastic Syndromes to Acute Myeloid Leukemia: Genomic characterization by Next Generation Sequencing

With the approval of Salamanca University, Department of Medicine, this thesis will be defended on 6th October 2021, in the Lecture Hall, Cancer Research Center-IBMCC, Salamanca

Supervisors:

Prof. Dr. Jesús María Hernández Rivas

Dra. M. del Rocío Benito Sánchez

Dra. María Abáigar Alvarado

Marta Martín Izquierdo

2021

D. **Jesús María Hernández Rivas,** Doctor en Medicina, Catedrático del Departamento de Medicina de la Universidad de Salamanca, Médico Adjunto del Servicio de Hematología del Hospital Clínico Universitario de Salamanca e Investigador del Centro de Investigación del Cáncer de Salamanca.

Dª. **María del Rocío Benito Sánchez,** Doctora en Ciencias Biológicas e Investigadora del Centro de Investigación del Cáncer de Salamanca.

Dª. **María Abáigar Alvarado,** Doctora por la Universidad de Salamanca e Investigadora del centro de Investigación del Cáncer de Salamanca.

CERTIFICAN

Que Dª **Marta Martín Izquierdo**, graduada en Bioquímica por la Universidad Autónoma de Madrid, ha realizado bajo nuestra dirección el trabajo de Tesis Doctoral titulado "*Molecular mechanisms and clonal evolution underlying the progression of Myelodysplastic Syndromes to Acute Myeloid Leukemia: Genomic characterization by Next Generation Sequencing*", y que éste reúne, a nuestro juicio, las condiciones de originalidad y calidad científica requeridas para su presentación y defensa ante el tribunal correspondiente para optar al grado de Doctor, con mención "Doctor International", por la Universidad de Salamanca.

Y para que así conste a los efectos oportunos, firmamos el presente certificado en Salamanca, a 2 de Septiembre de 2021.

Fdo.:

Prof. Dr. Jesús María Hernández Rivas

Dra. Mª Rocío Benito Sánchez

Dra. María Abáigar Alvarado

This thesis was performed being Marta Martín Izquierdo partially supported by "Beca Predoctoral de personal investigador", co-funded by the Fondo Social Europeo and by the Junta de Castilla y León (JCYL-EDU/556/2019).

Acknowledgments:

- Proyectos de Investigación en Salud del Instituto de Salud Carlos III:
  PI17/01741: Caracterización de los SMD de bajo riesgo sin sideroblastos en anillo: de las CPH a las vesículas extracelulares circulantes. 2018-2021.
  PI20/00970: UMBRELLA PROJECT: Unified platforM for a Better integRal Evaluation of MyeLodyspLastic Syndromes in SpAin. 2021-2024.
- Conserjería de Sanidad y Gerencia Regional de Salud de la Junta de Castilla y León (SACYL):
  GRS 2155/A/2020: Estudio de los Mecanismos Moleculares Implicados en la Transformación de los Síndromes Mielodisplásicos a Leucemia Aguda Mieloblástica.
- Fundación Memoria D. Samuel Solórzano Barruso: Estudio de los mecanismos moleculares que subyacen al envejecimiento y las neoplasias hematológicas: hematopoyesis clonal de potencial indeterminado (CHIP).
- Convocatoria ERA PerMed "1 st Joint Call for Proposals (JCT) on Research Project on Personalized Medicine-Smart Combination of Pre-Clinical and Clinical Research with data and OCT solutions":
  ERAPERMED2018-275: SYNtherapy: Synthetic Lethality for Personalized Therapy-base Stratification in Acute Leukemia.
- Sociedad Española de Hematología y Hemoterapia (SEHH) – Ayuda Económica para Innovación Tecnológica.

Scientific contribution during the thesis performed by **Marta Martín Izquierdo**:

Published papers

1. Quijada-Álamo M, Hernández-Sánchez M, Rodríguez-Vicente AE, Pérez-Carretero C, Rodríguez-Sánchez A, **Martín-Izquierdo M**, Alonso-Pérez V, García-Tuñón I, Bastida JM, Vidal-Manceñido MJ, Galende J, Aguilar C, Queizán JA, González-Gascón Y Marín I, Hernández-Rivas JÁ, Benito R, Ordóñez JL, Hernández-Rivas JM. Biological significance of monoallelic and biallelic BIRC3 loss in del(11q) chronic lymphocytic leukemia progression. *Blood Cancer J.* 2021 Jul 9;11(7):127. doi: 10.1038/s41408-021-00520-5. PMID: 34244476.

2. Quijada-Álamo M, Pérez-Carretero C, Hernández-Sánchez M, Rodríguez-Vicente AE, Herrero AB, Hernández-Sánchez JM, **Martín-Izquierdo M**, Santos-Mínguez S, Del Rey M, González T, Rubio-Martínez A, García de Coca A, Dávila-Valls J, Hernández-Rivas JÁ, Parker H, Strefford JC, Benito R, Ordóñez JL, Hernández-Rivas JM. Dissecting the role of TP53 alterations in del(11q) chronic lymphocytic leukemia. *Clin Transl Med*. 2021 Feb;11(2):e304. doi: 10.1002/ctm2.304. PMID: 33634999; PMCID: PMC7862176.

3. Janusz K, **Izquierdo MM**, Cadenas FL, Ramos F, Sánchez JMH, Lumbreras E, Robledo C, Del Real JS, Caballero JC, Collado R, Bernal T, Pedro C, Insunza A, de Paz R, Xicoy B, Salido E, García JS, Mínguez SS, García CM, Muñoz AMS, Barba MS, Rivas JMH, Abáigar M, Campelo MD. Clinical, biological, and prognostic implications of SF3B1 co-occurrence mutations in very low/low- and intermediate-risk MDS patients. *Ann Hematol.* 2021 Aug;100(8):1995-2004. doi: 10.1007/s00277-020-04360-4. Epub 2021 Jan 6. PMID: 33409621.

4. Vega-Garcia N, Benito R, Esperanza-Cebollada E, Llop M, Robledo C, Vicente-Garcés C, Alonso J, Barragán E, Fernández G, Hernández-Sánchez JM, **Martín-Izquierdo M**, Maynou J, Minguela A, Montaño A, Ortega M, Torrebadell M, Cervera J, Sánchez J, Jiménez-Velasco A, Riesco S, Hernández-Rivas JM, Lassaletta Á, Fernández JM, Rives S, Dapena JL, Ramírez M, Camós M, On Behalf Of The Group Of Leukemia Of The Spanish Society Of Pediatric Hematology And Oncology Sehop. Helpful Criteria When Implementing NGS Panels in Childhood Lymphoblastic Leukemia. *J Pers Med.* 2020 Nov 26;10(4):244. doi: 10.3390/jpm10040244. PMID: 33255984; PMCID: PMC7711852.

5. Santos-Briz A, Medina-Miguelañez M, Moyano-Bueno D, Viñolas-Cuadros A, Martínez TG, **Izquierdo MM**, García-Sánchez MS. Indeterminate Dendritic Cell Tumor as Cutaneous Involvement of Chronic Myelomonocytic Leukemia Successfully Treated With Phototherapy. *Am J Dermatopathol.* 2020 Nov;42(11):876-880. doi: 10.1097/DAD.0000000000001703. PMID: 32568837.

6. Pérez-Carretero C, Hernández-Sánchez M, González T, Quijada-Álamo M, **Martín-Izquierdo M**, Hernández-Sánchez JM, Vidal MJ, de Coca AG, Aguilar C, Vargas-Pabón M, Alonso S, Sierra M, Rubio-Martínez A, Dávila J, Díaz-Valdés JR, Queizán JA, Hernández-Rivas JÁ, Benito R, Rodríguez-Vicente AE, Hernández-Rivas JM. Chronic lymphocytic leukemia patients with IGH translocations are characterized by a distinct genetic landscape with prognostic implications. *Int J Cancer.* 2020 Nov 15;147(10):2780-2792. doi: 10.1002/ijc.33235. Epub 2020 Sep 4. PMID: 32720348.

7. Forero-Castro M, Montaño A, Robledo C, García de Coca A, Fuster JL, de Las Heras N, Queizán JA, Hernández-Sánchez M, Corchete-Sánchez LA, **Martín-Izquierdo M**, Ribera J, Ribera JM, Benito R, Hernández-Rivas JM. Integrated Genomic Analysis of Chromosomal Alterations and Mutations in B-Cell Acute Lymphoblastic Leukemia Reveals Distinct Genetic Profiles at Relapse. *Diagnostics (Basel)*. 2020 Jul 4;10(7):455. doi: 10.3390/diagnostics10070455. PMID: 32635531; PMCID: PMC7400270.

8. Quijada-Álamo M, Hernández-Sánchez M, Alonso-Pérez V, Rodríguez-Vicente AE, García-Tuñón I, **Martín-Izquierdo M**, Hernández-Sánchez JM, Herrero AB, Bastida JM, San Segundo L, Gruber M, García JL, Yin S, Ten Hacken E, Benito R, Ordóñez JL, Wu CJ, Hernández-Rivas JM. CRISPR/Cas9-generated models uncover therapeutic vulnerabilities of del(11q) CLL cells to dual BCR and PARP inhibition. *Leukemia*. 2020 Jun;34(6):1599-1612. doi: 10.1038/s41375-020-0714-3. Epub 2020 Jan 23. PMID: 31974435; PMCID: PMC7266745.

9. García-Tuñón I, Alonso-Pérez V, Vuelta E, Pérez-Ramos S, Herrero M, Méndez L, Hernández-Sánchez JM, **Martín-Izquierdo M,** Saldaña R, Sevilla J, Sánchez-Guijo F, Hernández-Rivas JM, Sánchez-Martín M. Splice donor site sgRNAs enhance CRISPR/Cas9-mediated knockout efficiency. *PLoS One.* 2019 May 9;14(5):e0216674. doi: 10.1371/journal.pone.0216674. PMID: 31071190; PMCID: PMC6508695.

10. Bastida JM, Benito R, Lozano ML, Marín-Quilez A, Janusz K, **Martín-Izquierdo M**, Hernández-Sánchez J, Palma-Barqueros V, Hernández-Rivas JM, Rivera J, González-Porras JR. Molecular Diagnosis of Inherited Coagulation and Bleeding Disorders. *Semin Thromb Hemost*. 2019 Oct;45(7):695-707. doi: 10.1055/s-0039-1687889. Epub 2019 Apr 30. PMID: 31041795.

11. Bastida JM, Morais S, Palma-Barqueros V, Benito R, Bermejo N, Karkucak M, Trapero-Marugan M, Bohdan N, Pereira M, Marin-Quilez A, Oliveira J, Yucel Y, Santos R, Padilla J, Janusz K, Lau C, **Martin-Izquierdo M**, Couto E, Francisco Ruiz-Pividal J, Vicente V, Hernández-Rivas JM, González-Porras JR, Luisa Lozano M, Lima M, Rivera J. Identification of novel variants in ten patients with Hermansky-Pudlak syndrome by high-throughput sequencing. *Ann Med.* 2019 Mar;51(2):141-148. doi: 10.1080/07853890.2019.1587498. Epub 2019 Apr 16. PMID: 30990103; PMCID: PMC7857454.

12. Hernández-Sánchez M, Rodríguez-Vicente AE, González-Gascón Y Marín I, Quijada-Álamo M, Hernández-Sánchez JM, **Martín-Izquierdo M**, Hernández-Rivas JÁ, Benito R, Hernández-Rivas JM. DNA damage response-related alterations define the genetic background of patients with chronic lymphocytic leukemia and chromosomal gains. *Exp Hematol.* 2019 Apr;72:9-13. doi: 10.1016/j.exphem.2019.02.003. Epub 2019 Feb 23. PMID: 30807786.

Submitted papers

JOCES/2021/259337: Quwaider D; Corchete LA; **Martín-Izquierdo M**; Hernández-Sánchez JM; Rojas EA; Cardona-Benavides IJ; García-Sanz R; Herrero AB; Gutiérrez NC. RNA sequencing identifies novel regulated IRE1-dependent decay targets that affect multiple myeloma survival and proliferation. *Journal of Cell Science*.

Table of Contents

List of Abbreviations

allo-HSCT	allogenic hematopoietic stem cell transplantation
AML-MRC	AML with myelodysplasia-related changes
AZA	Azacitidine
BM	Bone marrow
CFU-E	Early erythroid progenitor
CFU-GM	Granulocyte-monocyte progenitor
CFU-Mega	Megakaryocyte progenitor
CGH array	Comparative Genomic Hybridization array
CHIP	Clonal Hematopoiesis of Indeterminate Potential
ChIP-seq	chromatin immunoprecipitation sequencing
CI	Confidence interval
CMML	chronic myelomonocytic leukemia
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRISPR	Clustered Regions of Interspersed Palindromic Repeats
DNA	deoxyribonucleic acid
DNA-seq	DNA sequencing
DSB	double strand break
EMA	European Medicines Agency
ESA	erythropoiesis-stimulating agents
FAB	French-American-British
FACS	fluorescence-activated cell sorting
FDA	Food and Drug Administration
FISH	Fluorescence *in situ* hybridization
G-CSF	Granulocyte colony-stimulating factor
gDNA	genomic DNA
GEP array	Gene Expression Profile array
GTP	guanosine triphosphate
HDR	homology directed repair
HMAs	DNA hypomethylating agents
HR	Hazard ratio
HR-MDS	High-risk MDS
HSC	Hematopoietic Stem Cells
IPSS	International Prognostic Scoring System
IPSS-R	Revised International Prognostic Scoring System
ITD	internal tandem duplication
LFS	Leukemia free survival
LR-MDS	Low-risk MDS
LR-PSS	Lower-risk MDS Prognostic Scoring System
MAF	Minor Allele Frequency
MDA-CSS	MD Anderson Comprehensive Scoring System
MDS	Myelodysplastic syndromes
MDS-EB-1	MDS with excess of blasts type 1
MDS-EB-2	MDS with excess of blasts type 2
MDS-MLD	MDS with multilineage dysplasia
MDS-RS-MLD	MDS with ring sideroblasts and multilineage dysplasia
MDS-RS-SLD	MDS with ring sideroblasts and single lineage dysplasia

MDS-SLD	MDS with single lineage dysplasia
MDS-U	MDS unclassifiable
MPN	myeloproliferative neoplasms
NGS	Next Generation Sequencing
NHEJ	non-homologous end joining
OR	Odds ratio
OS	Overall survival
PB	Peripheral blood
PCR	Polymerase Chain Reaction
Q score	Phred quality score
RNA	ribonucleic acid
RNA-seq	RNA transcriptome sequencing
RNP	ribonucleoprotein
sAML	secondary Acute Myeloid Leukemia
sgRNA	guide RNA
SMC	Structural maintenance of chromosomes
SNP	single nucleotide polymorphism
TDS	Targeted-deep sequencing
TKD	tyrosine kinase domain
TTL	Time to leukemic progression
UTRs	untranslated region
VAF	Variant Allele Frequency
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHO	World Health Organization
WPSS	WHO-classification-based Prognostic Scoring System

Introduction

1. Myelodysplastic Syndromes

1.1. Disease overview

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic disorders characterized by ineffective hematopoiesis with resulting peripheral blood (PB) cytopenia, a hyper-cellular bone marrow (BM), dysplastic morphology in at least one myeloid lineage and an increased risk of progression to secondary acute myeloid leukemia (sAML) [1-4].

MDS present a highly variable clinical course, ranging from indolent conditions with a near-normal life expectancy spanning years, to aggressive forms of the disease rapidly progressing to sAML [5-7].

The process of normal blood cell development and how this is disrupted in MDS is still on debate. The mutated MDS cell clone can suppress normal hematopoiesis and alters ability to develop along all lineages, resulting in dysplasia and deficiencies (PB cytopenia) of multiple cell lineages (**Figure 1**).

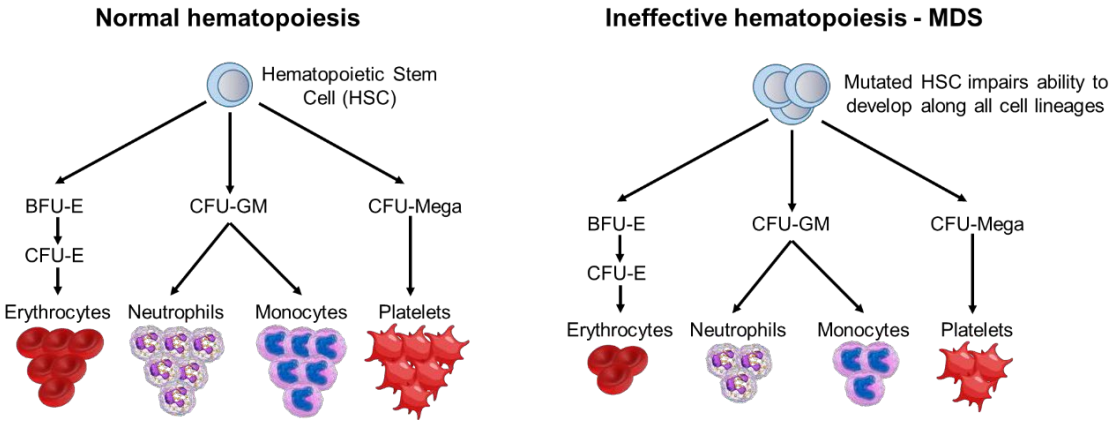


Figure 1. Myelodysplastic syndromes alter normal hematopoiesis (adapted from Corey *et al.*, *Nat Rev Cancer*, 2007) [8]. In MDS, the mutated Hematopoietic Stem Cells (HSC) clones impair ability to develop along all cell lineages, suppressing normal hematopoiesis. Thereby, cell lineages progenitors, such as early erythroid progenitor (CFU-E), granulocyte-monocyte progenitor (CFU-GM) and megakaryocyte progenitor (CFU-Mega), can be altered, resulting in dysplasia and deficiencies or cytopenia along all cell lineages.

1.2. Epidemiology and etiology

MDS are some of the most prevalent hematological malignancies and are often described as a disease of the ageing population, with a median age at diagnosis of 65-70 years and with <10% of patients younger than 50 years [8-10].

The annual incidence of MDS ranges from 2-12 cases in 100.000 people per year, increasing with age, and reaching 40-50 per 100.000 among patients aged over 70 years [8,9,11]. There are not known ethnic differences in the incidence, however, a different pattern is seen in Asia, where MDS tend to occur at an earlier age with a median age at diagnosis of 53 years [12,13]. MDS affects more frequently males than females, except for in the form which isolated 5q deletion in which women predominate [14,15].

The cause of MDS is known in only 15% of cases, but the association of MDS with ageing population suggests genetic damage by hazardous exposure or inherited susceptibility. Environmental risk factors include exposure to chemotherapy, radiation, benzene and its derivatives, diesel fuel, excess alcohol intake, smoking, viral infections and immunosuppressive agents [8,16-20]. In fact, the incidence of therapy-related MDS has been increasing in recent years, perhaps in relation to the use of chemotherapy and radiotherapy in cancer [10]. Although MDS in older adults typically presents with age-associated acquisition of somatic mutations driving clonal evolution, children and young adults are more likely to harbor a germline genetic condition leading to MDS. Genetic susceptibility to MDS occurs in a third of pediatric cases, including in children with Down's syndrome, Fanconi's anemia, Schwachman-Diamond syndrome and neurofibromatosis [9]. In adults, inherited predisposition is less common, but in the last years when genetic sequencing has become increasingly integrated into clinical practice, multiple hereditary predispositions to MDS have been recently discovered [21].

1.3. Diagnosis

The diagnosis of MDS is often very challenging because they are a very heterogeneous group of diseases. It is mainly based on the presence of cytopenias in a routine analysis of the PB and the predominant clinical manifestations are fatigue, infections and bleeding as a result from the PB cytopenias [11,16]. According to the World Health Organization (WHO) and International Prognostic Scoring System (IPSS), the threshold levels consider as cytopenias are: anemia, when hemoglobin is lower than 10 g/dL; thrombocytopenia, when platelet count is lower than 100 $x10^9$/L; and neutropenia, when absolute neutrophil count is lower than 1.6 (WHO criteria) or 1.8 $x10^9$/L (IPSS criteria) [22].

Morphological examination of the peripheral blood smear and bone marrow aspirate should be performed in order to confirm that the cytopenias are a consequence of bone marrow failure [23]. The BM in MDS patients is usually hyper- or normocellular with dysplasia in one or several myeloid hematopoietic lineages (erythroid, granulocytic, megakaryocytic), considering as dysplasia when higher than 10% of the cells of at least one myeloid lineage show unequivocal morphological

changes [16,24,25]. Moreover, the proportion of blasts in the BM must also be assessed to provide a correct classification, since the forms with <2% BM blasts are to be distinguished from those with >2% blasts, as they have a better prognosis, and further, according to WHO, a 20% BM blasts is the threshold for the diagnosis of AML. By contrast, BM biopsy is not mandatory, although is important for identifying fibrotic MDS or hypocellular MDS [9].

A careful study must be performed to exclude patients with cytopenia but no dysplasia, which could have similar characteristics to those of MDS. Moreover, the key feature of MDS is the clonal nature of the dysplasia. Thus, the detection of a chromosomal abnormality (karyotype and/or FISH), flow cytometry analysis, as well as mutational analysis must be considered as complementary technique to better define MDS subtypes and prognosis of these patients [9,11,26,27].

1.4. Classification

MDS is a heterogeneous disease and thus it is necessary to categorize patients to identify groups of patients with similar morphological features, molecular etiology, prognosis and likelihood of progression to sAML and of response to common therapies [28].

Morphological classification

In 1982 the French-American-British (FAB) cooperative group proposed the first classification system according to cytomorphology and blasts percentage at diagnosis [29]. The FAB system has gone through several stages of revision integrating blood cytopenias, morphological features, cytogenetic and genetic information in order to refine the diagnostic classification. In this sense, the WHO 2001, the WHO 2008 and currently the WHO 2017 classification systems have also been successfully used to classify MDS [30-32].

The latest used system of classification of MDS is the WHO 2017 classification. This system divides MDS into several disease entities based on the morphological findings in PB and BM such as the type and degree of dysplasia, the number of cytopenias, the presence of ring sideroblasts, the proportion of blasts in the BM and PB, the presence of specific chromosomal abnormality (deletion of chromosome 5q) and, for the first time, genetic information (presence of *SF3B1* mutations) [32-36].

The last edition of WHO classification identify 9 MDS subtypes: MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts and single lineage dysplasia (MDS-RS-SLD), MDS with ring sideroblasts and multilineage dysplasia (MDS-RS-MLD), MDS with isolated del(5q), MDS with excess of blasts type 1 and type 2 (MDS-EB-1 and MDS-EB-2), MDS unclassifiable (MDS-U) and refractory cytopenia of childhood (**Table 1**).

Table 1. WHO 2017 criteria for classifying MDS patients (adapted from Arber *et al.*, *Blood*, 2016) [32].

Name	Dysplastic lineages	Cytopenias*	Ring sideroblasts in BM (%)	BM and PB blasts (%)	Cytogenetics
MDS with single lineage dysplasia (MDS-SLD)	1	1 or 2	<15% / <5%†	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with multilineage dysplasia (MDS-MLD)	2 or 3	1-3	<15% / <5%†	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with ring sideroblasts and single lineage dysplasia (MDS-RS-SLD)	1	1 or 2	≥15% / ≥5%†	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with ring sideroblasts and multilineage dysplasia (MDS-RS-MLD)	2 or 3	1-3	≥15% / ≥5%†	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with isolated del(5q)	1-3	1-2	None or any	BM <5%, PB <1%, no Auer rods	Del(5q) alone or with 1 additional abnormality except -7 or del(7q)
MDS with excess of blasts type 1 (MDS-EB-1)	0-3	1-3	None or any	BM 5%-9% or PB 2%-4%, no Auer rods	Any
MDS with excess of blasts type 2 (MDS-EB-2)	0-3	1-3	None or any	BM 10%-19% or PB 5%-19% or Auer rods	Any
MDS Unclassifiable (MDS-U)					
With 1% blood blasts	1-3	1-3	None or any	BM <5%, PB = 1%, ‡ no Auer rods	Any
With single lineage dysplasia and pancytopenia	1	3	None or any	BM <5%, PB <1%, no Auer rods	Any
Based on defining cytogenetic abnormality	0	1-3	<15%§	BM <5%, PB <1%, no Auer rods	MDS-defining abnormality
Refractory cytopenia of childhood	1-3	1-3	None	BM <5%, PB <2%	Any

***** Cytopenias defined as: hemoglobin, <10 g/dL; platelet count, <100 x 10^9/L; and absolute neutrophil count <1.8 x 10^9/L. Rarely, MDS may present with mild anemia or thrombocytopenia above these levels. PB monocytes must be <1 x 10^9/L.

† *SF3B1* mutation is present.

‡ One percent PB blasts must be recorded on at least 2 separate occasions.

§ Cases with ≥15% ring sideroblasts by definition have significant erythroid dysplasia, and are classified as MDS-RS-SLD.

Risk stratification systems

MDS patients have a highly variable clinical course, with large differences in overall survival (OS) and risk of transformation to AML, ranging from indolent conditions over many years to forms rapidly progressing to leukemia [5,6,28]. For this reason, prognostic scoring systems have been developed in order to risk-stratify patients according to their predicted survival and risk of evolution to AML, and, consequently, to help clinical decision-making and to assess a better prognosis estimation of MDS.

Some of these prognostic scoring systems are the International Prognostic Scoring System (IPSS), the WHO classification-based Prognostic Scoring System (WPSS), the Lower-Risk MDS Prognostic Scoring System (LR-PSS), the MD Anderson Comprehensive Scoring System (MDA-CSS) and the Revised International Prognostic Scoring System (IPSS-R) [6,28,37-41].

The most widely adopted predictor of prognosis for patients with MDS has been the IPSS, first published in 1997 by Greenberg and his colleagues [38]. However, over time, IPSS has been noted as not very precise predictor of prognosis in lower-risk patients [42]. Therefore, the IPSS has been revised (IPSS-R) in order to refine it incorporating larger and more differentiated cytogenetic subgroups, more stringent blast description and more detailed cytopenia information [41]. The revised prognostic scoring system IPSS-R is detailed below (**Table 2**).

The IPSS-R include five risk groups that are significantly different in overall survival and risk of progression to AML. The two lowest and the two highest risk groups are often called 'low risk MDS' and 'high risk MDS', respectively. However, the intermediate group, which represents nearly 20% of the MDS patients, is a heterogeneous category with indolent patients similar to lower risk MDS and other patients with aggressive disease [43,44].

One of the most important limitations of the IPSS-R is the lack of genetic information. In the recent years, huge amounts of molecular data from MDS patients have been generated providing significant knowledge about some gene mutations have a significant influence in the prognosis of the patients. In fact, *SF3B1* mutations are associated with favorable outcome while *TP53*, *SRSF2*, *DNMT3A*, *IDH2*, *RUNX1* and *ASXL1* mutations are associated with poor outcomes [36,45-51]. Therefore, a proposal to include this molecular knowledge into the IPSS-R stratification system is being carried out.

Table 2. Revised International Prognostic Scoring System (IPSS-R) (adapted from Greenberg *et al.*, *Blood*, 2012) [41].

	Score Values						
	0	0.5	1	1.5	2	3	4
Cytogenetics	-Y, del(11q)	-	Normal, del(5q), del(12p), del(20q), doublé including del(5q)	-	Del(7q), +8, +19, i(17q), any other single or double independent clones	-7, inv(3)/t(3q)/del(3q), double including -7/del(7q), Complex: 3 abnormalities	Complex: >3 abnormalities
BM Blasts %	≤2	-	>2% to <5%	-	5% to 10%	>10%	
Hemoglobin (g/dL)	≥10	-	8 to 10	<8			
Platelets ($x10^9$/L)	≥100	50 to <100	<50				
ANC ($x10^9$/L)	≥0.8	<0.8					

	Categories and Clinical Outcomes				
	Very Low	**Low**	**Intermediate**	**High**	**Very High**
Risk Score	≤1.5	>1.5 to 3	>3 to 4.5	>4.5 to 6	>6
Patients (%)	19%	38%	20%	13%	10%
Survival (median, years)	8.8	5.3	3.0	1.6	0.8
Time to 25% AML evolution (median, years)	Not reached	10.8	3.2	1.4	0.7

1.5. Outcome

MDS patients present a wide range of clinical courses, from indolent conditions, having a near-normal life expectancy spanning years, to forms rapidly progressing to acute leukemia. Thereby, this clinical heterogeneity complicates decision making regarding therapeutic the choice of treatment and therapeutic modalities, as well as the timing of intervention [5]. In addition to disease-specific variables, patient-related factors, such as age and other pathologies, are also essential for risk estimation. In fact, due to the median age at diagnosis of MDS is around 70 years, patients frequently have comorbidities that may influence outcomes and the spectrum of available therapies [43].

MDS treatment

Therapeutic options for patients with MDS vary from supportive care and symptomatic treatment of blood cytopenias, mainly by transfusions, to ultimately allogenic hematopoietic stem cell transplantation (allo-HSCT), depending on disease and patient-related risk factors [7]. Moreover, MDS mainly affect the elderly population, which means that the majority of patients cannot tolerate intensive therapeutic approaches such as intensive chemotherapy and allogeneic transplant [52].

Over the years, three drugs have been approved by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for their use in MDS-related indications: lenalidomide (approved in 2005), an orally administered immunomodulatory drug; and azacitidine and decitabine (approved in 2004 and 2006, respectively), two nucleoside analogs that are DNA hypomethylating agents (HMA). In addition to these agents, there is extensive off-label use of the erythropoiesis-stimulating agents (ESAs), such as epoetin and darbepoetin [43]. However, since decitabine´s approval in 2006, more than 15 years have passed without any new drug for MDS patients [53], except for luspatercept, approved for treating anemia in MDS patients in 2020 [54].

Therefore, MDS therapy is selected based on their risk classification and the goals of therapy are different in low-risk patients than in high-risk individuals, and even in those after HMA failure.

In low-risk MDS, the goal is to reduce transfusion needs with erythropoiesis-stimulating agents and luspatercept treatment, and prevent the transformation to higher-risk disease or sAML, as well as to improve survival. Additionally, low-risk patients with del(5q) display high response rates to lenalidomide, reaching transfusion independence and, even, cytogenetic remission [52,55,56].

In high-risk MDS, the goal is to prolong survival and current available therapies include growth factor support, lenalidomide, HMAs, intensive chemotherapy and allo-HSCT. Before starting any treatment in high-risk patients, the eligibility of the patient for allo-HSCT should be evaluated, because it is the only available curative treatment and, consequently, this will have an important effect on patient surveillance. However, for those patients who are not transplant candidates, HMA therapy is most appropriate [55,56].

Although HMA are active in roughly one-half of MDS patients, the majority either fail to respond or lose an initial response. For these patients with progressive or refractory disease after HMA-based therapy, at the present time there are no approved interventions and the options include participation in a clinical trial, cytarabine-based therapy or allo-HSCT [52,55,56].

Notwithstanding these options, there remains a substantial number of patients who display disease progression and evolve to sAML. Thus, there is a tremendous clinical need for novel therapies in MDS patients.

MDS progression to sAML

MDS patients are at high risk of eventually progression to MDS-related acute myeloid leukemia (secondary AML, sAML), occurring this transformation in approximately one third of patients. By definition, progression to sAML is clinically confirmed when a rise in the proportion of BM blasts is observed in a morphological evaluation, accounting for higher than 20% [32,57].

The progression of the disease is associated with a dismal prognosis, partly because most of these patients are resistant to currently available treatments, have inferior rates of complete remission and the long-term survival rate of treated patients is less than 10% [42,58]. Similarly, this risk pattern to sAML progression should be taken into account when making critical treatment decisions, such as whether and when to suggest hematopoietic stem cell transplantations, as in done in other high risk premalignant conditions [58].

More detailed information about MDS progression to sAML is covered below, introduction section 3.

2. Genetic and molecular pathways in MDS

It is well recognized that MDS is, like other cancers, shaped by recursive rounds of positive selections, where gene mutations and other genetic alterations play central roles [59-61]. Moreover, the presence of a only genetic lesion is uncommon, and a combination of gene mutations and/or cytogenetic abnormalities is usual [62,63]. Thus, to decipher what alterations and by what mechanisms are involved in the positive selections ('driver' alterations) is central to the understanding of the pathogenesis and disease phenotypes of MDS [64].

Early genetic studies on MDS were mainly focused on cytogenetic abnormalities detected by conventional karyotyping. The cytogenetic alterations are found in approximately 50% of MDS cases and are one of the most important factors associated with the patient's outcome [33,38,41,65]. Most of these alterations in MDS are unbalanced changes, which result in gains and losses of chromosomal materials. The most frequent among these are -5/del(5q) and -7/del(7q), followed by +8, del(20q), -18/del(18q), -Y, +21q gains, del(17p)/iso(17q) and inv/t(3q). Many of these lesions often co-occur being part of complex karyotypes [33,65]. The recurrent chromosomal abnormalities and the incidence in MDS are shown in **Figure 2**.

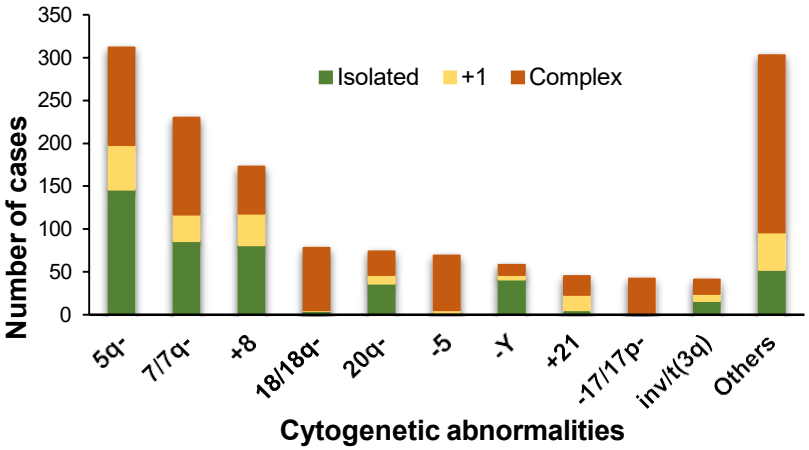


Figure 2. Frequencies of the most common cytogenetic anomalies subdivided into isolated, with 1 additional anomaly and complex anomalies (adapted from Haase *et al.*, *Blood*, 2007) [33]. This graph shows the most frequent cytogenetics alterations in MDS patients. The presence of the alterations with other anomalies or isolated is represented by color: green for isolated cytogenetic alterations, yellow for cases with one additional anomaly, and orange for the presence in complex karyotypes.

However, more than 50% of MDS patients have normal cytogenetics and, additionally, several studies have demonstrated the presence of gene mutations in practically all MDS [62-64,66].

2.1. Mutational landscape of MDS

Patient's genetic profile critically impacts on clinical phenotype, prognosis and response to therapy, highlighting the importance of updated knowledge about MDS genetics for predicting clinical course of patients as well as optimizing therapy and management [64].

Until 2010 only a handful of genes, such as *TP53*, *EZH2*, *RUNX1*, *TET2*, *NRAS*, *KRAS*, *FLT3*, *ASXL1*, had been known to be recurrently mutated in MDS due to the application of SNP array karyotyping and Sanger sequencing [67-69]. However, during the following decade, our knowledge about MDS genomes has exponentially improved through a comprehensive detection of somatic mutations in MDS thanks to the application of novel advanced high-throughput sequencing technologies [62,63,70].

In the last years, it has been described that 78 to 90% of MDS patients carries at least one genetic mutation [62,63,71], and it is noted that most of these genetic mutations (approximately two thirds) are found in patients with normal karyotype [3,66]. Furthermore, the number of mutations was shown to vary depending on the MDS subtype. In this sense, while a typical low-risk MDS patient, including MDS-SLD with or without sideroblasts and MDS with isolated del(5q), harbors a median of 2-3 driver mutations detected by target-deep sequencing and 6 alterations by whole exome sequencing, high-risk MDS (MDS-EB and MDS-MLD) and sAML patients tend to show higher number of mutations, ranging from 9 to 13 alterations [64].

To date, more than 40 driver genes have been identified involved in the pathogenesis of MDS. These can be categorized into several functional pathways: RNA splicing (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*, among others), DNA methylation (*TET2*, *DNMT3A*, *IDH1/IDH2*), transcription (*RUNX1*, *ETV6*, *TP53*, *BCOR*, among others), chromatin modification (*EZH2*, *ASXL1*, *ATRX*, among others), cohesin complex (*STAG2*, *STAG1*, *RAD21*, *SMC1A*, *SMC3*, *CTCF*) and signal transduction (*NRAS*, *KRAS*, *FLT3*, *CBL*, *JAK2*, along with others). Among these, genes implicated in the epigenetic regulation, including DNA methylation and chromatin modification, as well as RNA splicing are the most frequently mutated. However, most genes are mutated at low frequency, with only the 6 most frequently genes mutated in >10% of cases [62,63,71-75]. The frequencies of driver alterations identified by sequencing or by cytogenetics and the most frequent altered functional pathways are shown in **Figure 3**, stratified by MDS subtypes.

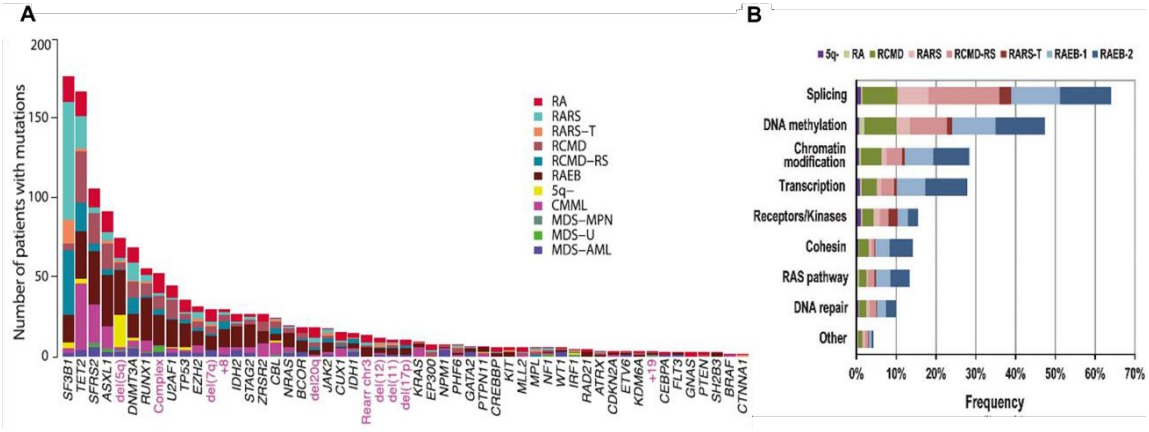


Figure 3. Frequencies of driver alterations identified by sequencing or by cytogenetics and the most frequent altered functional pathways considering the MDS subtypes (adapted from Papaemmanuil *et al.*, *Blood*, 2013 and from Haferlach *et al.*, *Leukemia*, 2014) [62,63]. a) The most frequent alterations in MDS patients are *SF3B1*, *TET2*, *SRSF2*, *ASXL1*, del(5q), *DNMT3A*, *RUNX1*, complex karyotype, *U2AF1* and *TP53*. **b)** Regarding the functional pathway which could be altered in MDS patients, the most frequent target is RNA splicing, followed by genes associated with DNA methylation.

Under this scenario, where multiple driver genes are mutated, the combination of mutations is not totally random. Some of them tend to appear together (co-occurrence) more frequently than expected only by chance, while others are defined as mutually exclusive, since they are rarely observed together. Overall, this suggests that gene-gene interactions, mutationally speaking, are behind the positive and negative clone selections during the whole course of the disease [64,75]. As few examples, gene mutations in the spliceosome components as well as in the cohesin complex are considered almost mutually exclusive, likely due to synthetic lethality of mutations in genes acting in the same cellular pathway [76,77]. By contrast, RNA splicing and DNA methylation mutations appear to occur together at an early state of disease, being reported as 'founder' mutations in over 50% MDS patients [78], *RUNX1* mutations are associated with *ASXL1* alterations [79] or *STAG2* mutations co-occur with others in *RUNX1*, *SRSF2*, *ASXL1* or *EZH2* [77,80,81].

Thereby, MDS is characterized by mutations in multiple driver genes, a complex structure of gene-gene interactions and extensive subclonal diversification, showing a great heterogeneity, inter-patient and intratumoral. Nevertheless, despite low mutation frequency and inter-patient variation, some of the mutations have demonstrated correlation with disease phenotype and prognostic value.

For instances, *TP53* mutations are associated with complex karyotype [45,82] while *SF3B1* mutations are strongly associated with increased ring sideroblasts in BM and generally predict a favorable prognosis [34,36,51]. Moreover, the presence of some mutations has been associated with a worse clinical outcome and the number of oncogenic mutations negatively correlates with leukemia-free survival, providing independent prognostic information after stratification by IPSS-R classification [62]. In contrast to generally favorable prognosis in *SF3B1*-mutated cases, patients with other splicing mutations, such as *SRSF2* and *U2AF1*, have a poor overall survival [83-85]. Similarly, mutations in *TP53*, *DNMT3A*, *RUNX1*, *EZH2* and *ASXL1* have been described as predictors of poor overall survival [62,82,86-89]. Furthermore, some studies have recently reported the impact of genetic aberrations in the context of other mutations on the disease phenotype and outcome of MDS patients. In this line, the co-mutation pattern between *SRSF2* and RAS-pathway mutations is associated with leukocytosis while the co-occurrence of *SRSF2* mutations with *STAG2*, *RUNX1* and *IDH1/2* mutations is associated with blast phenotype [81]. Similarly, the co-occurrence of *SF3B1* mutations with specific genes, such as *SRSF2*, *IDH2*, *BCOR*, *NUP98* and *STAG2*, is linked to a dismal prognosis [90].

A summary of the most frequent mutations found in MDS patients and their associated prognostic impact is shown in **Figure 4**.

Therefore, the identification of recurrent mutations in several mechanistic pathways has improved our knowledge about the clinical impact of these mutations, suggesting that the diagnostic and prognostic categorization of patients can be foster by the inclusion of genetic data. However, large data bases are required for understanding the patho-mechanisms and tailoring therapy for specific mutations accordingly. Furthermore, lack of clinical validation limits adoption of sequencing of these identified genes into real-time clinical practice.

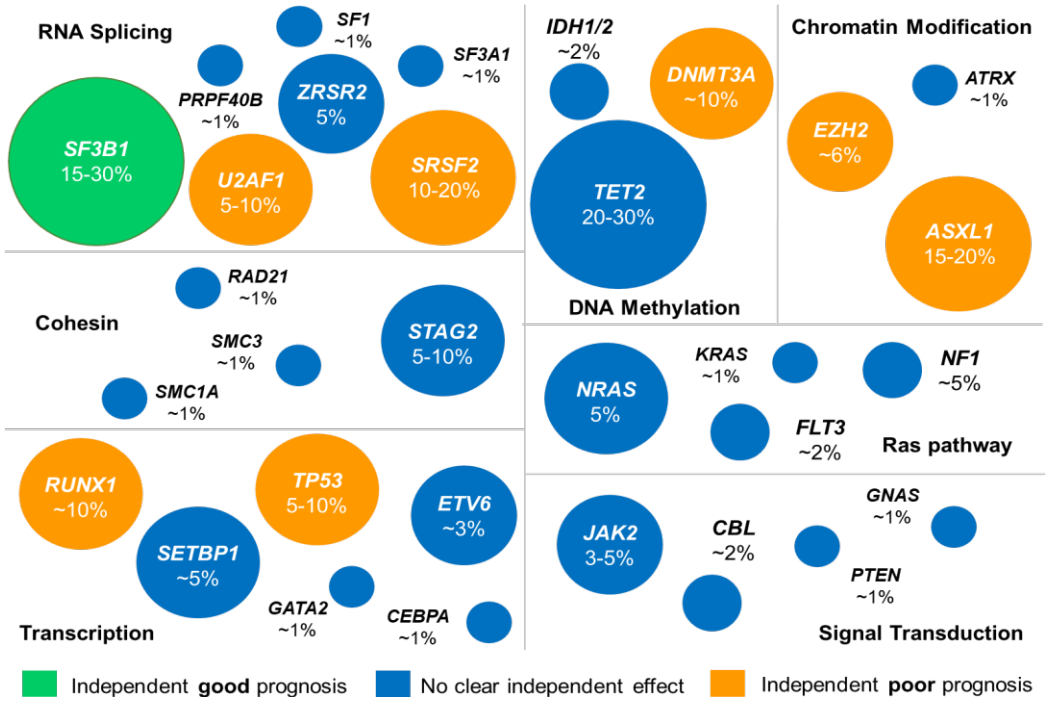


Figure 4. Summary of the most frequent mutations found in MDS patients and their prognostic impact (adapted from Steensma, *Mayo Clin Proc*, 2015)[23]. Mutational frequencies of the genes affected in MDS patients are described, as well as their prognostic significance is represented by color: green, genes whose mutations predict a favorable prognosis (such as *SF3B1*); orange, mutations in genes associated with poor prognosis (such as *SRSF2*, *ASXL1*, *DNMT3A*, *RUNX1*, *TP53*, *U2AF1* and *EZH2*); and blue, genes whose clinical impact is not defined.

2.2. The cohesin complex

Cohesin is a multimeric protein complex that is conserved across species and is composed of two structural maintenance of chromosomes (SMC) subunits SMC1A and SMC3 and an α-kleisin subunit RAD21 which form a ring that binds to either a stromal antigen one or two (STAG1/2) subunit [91,92] (**Figure 5**).

Cohesin complex is canonically known to align and stabilize sister chromatids during metaphase in cell division [92]. However, cohesin mutations have been described in several neoplasms, not being associated with aneuploidy or complex cytogenetics [93,94], thereby suggesting an alternative mechanism by which altered cohesin function leads to clonal expansion and malignant transformation. Recent studies have characterized the role of the cohesin complex in post-replicative DNA repair [95-97] and transcriptional activation through long-range *cis* interactions, specifically in DNA-loop formation between enhancers and promoters within topologically associated domains bounded by the CCCTC-binding factor (CTCF) [92,98-102].

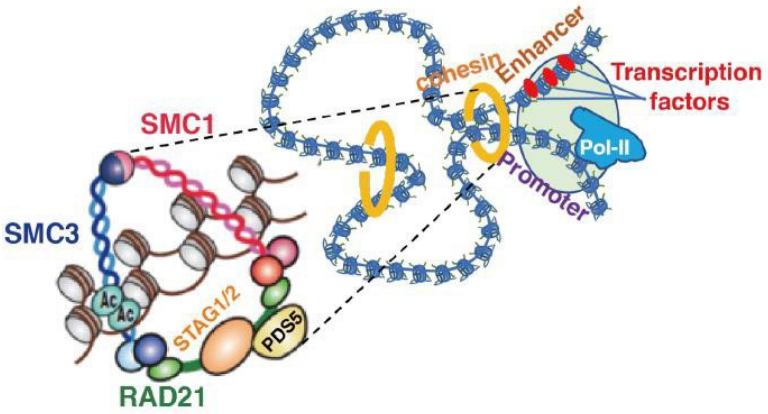


Figure 5. Representation of the cohesin protein complex (adapted from Ogawa, *Blood*, 2019) [64]. Cohesin is composed of two structural maintenance of chromosomes (SMC) subunits SMC1A and SMC3 and an α-kleisin subunit RAD21 which form a ring that binds to either a stromal antigen one or two (STAG1/2) subunit.

In the case of myeloid malignancies, mutations in the genes coding to all aforementioned cohesin complex subunits (*STAG2*, *STAG1*, *SMC3*, *SMC1A*, *RAD21* and *CTCF*) have been found in 10-15% of patients, being most prevalent in high-risk MDS and AML, where their incidence raised to 20% [77]. Clonal analysis has shown that they are early events [103] and discrimination by variant allele frequency (VAF) suggests that mutations are in the dominant clone [94]. Retrospective clinical outcomes by several groups found that, although not independently prognostic, cohesin mutations trend towards a worse prognosis in patients with myeloid diseases [63,77,94]. Regarding their functional and pathophysiological role in MDS, as described previously in other neoplasms, some studies have identified that cohesin mutations do not contribute to hematopoietic transformation by affecting chromosomal instability [94,103], instead they might alter the transcriptional orchestration of differentiation/proliferation, as well as the DNA damage repair, and therefore, may have a role in clonal evolution and tumorigenesis [77,97]. In addition, some authors suggest that cohesin mutations may have a negative impact on functional hematopoiesis, being an ancestral clone susceptible to a 'second-hit' resulting in leukemic transformation [104].

With regard to the co-occurrence of cohesin mutations in myeloid patients, these mutations were described to co-occur with others known to be drivers of clonal evolution, such as splicing mutations (*SRSF2*) [80,105], epigenetic chromatin modifiers (*BCOR* and *ASXL1*), transcription regulation (*RUNX1*, *NPM1*) [77,94] and Ras pathway alterations [77]. Thereby, these data indicate that epigenetic dysregulation may be synergistic in the mechanism of action exerted by cohesin haploinsufficiency. Likewise, activated Ras signaling may alter the expression of key identify genes, which are then further dysregulated by cohesin inactivity [77].

Nevertheless, the molecular mechanisms by which cohesin mutations promote leukemogenesis are not well understood and the clinical significance of these mutations is not clear.

2.3. The Ras signaling pathway

RAS is a proto-oncogene that belongs to the family of small GTP-ases and is involved in cellular signal transduction [106,107]. This gene family is composed of *NRAS*, *KRAS* and *HRAS* genes encoding homologous proteins that have the biochemical property of binding guanine nucleosides and exhibit intrinsic GTPase activity. Therefore, it is implicated with the transduction of receptor-mediated external signals into the cell [108].

Activating *RAS* mutations, which means that RAS is an active state, and thus preventing hydrolysis of GTP, are commonly found in a diverse range of human malignancies, including MDS, AML, chronic myelomonocytic leukemia (CMML) and juvenile myelomonocytic leukemia [109-113]. RAS activation leads to increase signaling through the *RAS*/*RAF*/*MEK* and *RAS*/*PI3K* pathways [114] and therefore, activating *RAS* mutations lead to increased proliferation of hematopoietic progenitor cells [115,116]. Moreover, besides the direct effect of activating *RAS* mutations, on this pathway, the RAS-dependent signaling cascade can be induced in response to mutations in other genes that encode RAS signaling upstream factors (*FLT3*, *KIT*, *CBL*), genes that are involved in RAS signaling regulation (*NF1*, *PTPN11*, *PTPN1*), as well as in genes coding RAS downstream effectors (*BRAF*) [116,117].

In the case of myeloid neoplasms, *RAS*-pathway mutations have been reported in 10-15% of cases in AML but they appear to be less frequent among patients with MDS, being around 5-10% and mainly occurring in the *NRAS* gene [112,118,119]. Although the prognostic impact of these gene mutations has been extensively studied in patients with myeloproliferative neoplasms (MPN) and AML, limited and contradictory information is available regarding the prognosis of MDS patients carrying *RAS*-pathway mutations [112].

Some studies have described that *RAS*-pathway mutations do not appear to correlate with a poor prognosis [112,120]. By contrast, other recent studies have described that *RAS*-pathway mutations may promote cell proliferation and be associated with a high risk of progression to AML and worse prognosis [121-123]. Similarly, patients harboring *RAS*-pathway mutations were most frequently diagnosed with MDS with excess blast and with IPSS-R high-risk categories, and also they display higher white blood cell counts and bone marrow blasts percentages [112,124]. Regarding their functional and pathophysiological role in MDS, mutations in *RAS*-pathway can lead to constitutive signaling activation or hyper-responsiveness to endogenous

stimuli and, consequently, they might enhance proliferation, decreased apoptosis or alteration of cell growth and survival [4]. Nonetheless, functional studies in MDS cells and *in vivo* models are required to confirm these effects.

In MDS patients, *RAS*-pathway mutations have been described as late events in the disease course, being sub-clonal events that co-occur with other 'founder' mutations [70]. Regarding the co-occurrence of these mutations, patients with *RAS*-pathway mutations were enriched in *U2AF1*, *RUNX1*, *WT1*, *PTPN11*, *ETV6* and *NPM1* mutations [124], but patients carried these mutations do not overlap with *FLT3* mutations (ITD or TKD) [112].

Hence, information about the clinical impact of *RAS*-pathway mutations is limited and contradictory and the molecular mechanisms by what these mutations are involved in MDS pathogenesis and disease progression are still partly unknown.

3. Progression of MDS to secondary Acute Myeloid Leukemia

The clinical course of MDS is highly variable, but approximately one third of MDS patients eventually progress to AML, commonly termed secondary AML or sAML, involving a dismal prognosis [42,58,125].

sAML represents up to 25-35% of total AML cases, with most (60-80%) arising from an antecedent of MDS [126-128], and it presents different clinicopathologic characteristics than *de novo* AML. In fact, in the WHO 2017 classification, AML developing from previous MDS is classified as a distinct entity termed AML with myelodysplasia-related changes (AML-MRC) [32], which displays inferior rates of complete remission, relapse-free and overall survival compared to *de novo* AML[128-130].

While MDS and secondary AML are classified as distinct entities, they represent a disease continuum that undergoes genetic clonal evolution. In fact, as defined by the IPSS-R, the overall survival of high-risk and very high-risk MDS [41] is similar with that of AML-MRC [131], indicating that higher-risk MDS and secondary AML are clinically similar, and they would represent cancer states on a disease continuum. However, these two identities show some cellular differences. AML is characterized by a complete block in hematopoietic cell maturation, leading to an accumulation of myeloblasts in the BM (≥20% of nucleated cells), while MDS patients produce mature, but defective, blood cells, because maturation and morphology are aberrant, and, consequently, low number of mature cells is observed. In addition, MDS is usually characterized by increased apoptosis and inflammatory cell death processes, but sAML, and AML in general, are associated with increased cell survival and proliferation, suggesting that the increased cell death phenotype in MDS gives way to a pro-survival phenotype during sAML progression [57].

By definition, blast count in BM is the parameter used to differentiate MDS from sAML. However, it is a subjective measurement, relying on a morphological assessment which is sensitive to interobserver variability [132,133], and has some limitations due to disease progression is a continuum and to define a boundary is quite complex. In fact, the definition of sAML transformation has little evolved during time and it is currently set at a threshold of ≥20% BM blasts by the WHO, but in the earlier FAB classification this threshold was ≥30% [134]. Moreover, recent genetic studies have identified clonality in MDS, since nearly all MDS patients have a clonal mutation, providing a measure of tumor burden that often greatly exceeds the blast percentage [70,135-137]. For these reasons, incorporating serial sequencing to monitor dynamic changes in MDS tumor burden over time may be useful, in combination with blast count among other clinical parameters, to anticipate and to identify progression earlier.

3.1. The genetics of progression from MDS to sAML: Clonal Evolution

MDS and sAML show similar cytogenetic abnormalities, such as del(5q), del(7q), del(20q) or complex karyotype, which are considered as 'MDS-associated' cytogenetic abnormalities. In fact, the presence of these alterations defines the WHO-diagnosis of AML-MRC, even in the absence of a previous MDS diagnosis or morphological dysplasia. By contrast, balanced rearrangements are more common in *de novo* AML, such as t(15;17), t(8;21), inv(16) or *KMT2A* rearrangements [32,65].

Likewise, large-scale next generation sequencing studies have identified mutations in genes that are shared between MDS and sAML, such as mutations in spliceosome genes (*SRSF2*, *SF3B1*, *U2AF1*), *EZH2*, *BCOR* and *STAG2* genes, whose presence is highly suggestive of an AML evolved from MDS, even without a known history of previous MDS diagnosis [80,138-140]. Thus, these mutations occur early in MDS pathogenesis prior to sAML progression.

Next generation sequencing studies have also shown that a MDS patient´s BM can be clonally complex with multiple genetically-related but unique clones, including a founding clone and subclone(s) derived from the founding clone [70,135,141]. Besides complexity, sAML progression has been defined as a dynamic process, where the founding clone of MDS persists, but new subclones with different genetic alterations grow or appear, triggering an increase of intraclonal heterogeneity [70]. Thus, sAML evolution is a process of clonal expansion, which includes genetic and epigenetic diversification, and clonal selection, requiring dynamic adaptability, resulting in a complex clonal architecture, namely a BM composition with great intraclonal heterogeneity.

In this context of intraclonal heterogeneity, the co-occurrence of gene mutations as well as the order of mutation acquisition are critical for disease pathogenesis, though deciphering the exact order of mutation acquisition in a patient can be difficult.

Recently, some studies have shown that around 10% of people older than 70 years have somatic mutations in their blood cells, defined as Clonal Hematopoiesis of Indeterminate Potential (CHIP) [59,142,143]. Many of these mutations (*DNMT3A*, *TET2*, *ASXL1* and spliceosome genes) are commonly mutated with high VAFs in both MDS and sAML, suggesting that mutations in these genes occur early in the development of the disease and are present in the majority of cells, which means that they are present in the founding clone [70,135,137]. In contrast, other mutations are more common in sAML, such as mutations in transcription factor (*RUNX1*, *GATA2*, *CEBPA*) and activating signaling related genes (*RAS* family genes, *FLT3*), suggesting that these mutations are acquired later during disease progression in a subset of expanding cells [121,137,144,145].

In this line, Makishima et al. identified two sets of gene mutations associated with sAML evolution from MDS: mutations in genes that were enriched in sAML compared to high-risk MDS, designated as type-1 mutations (mutations in genes including *FLT3*, *PTPN11*, *WT1*, *IDH1*, *NPM1*, *IDH2* and *NRAS*); and type-2 mutations that were enriched in high-risk compared to low-risk MDS (mutations in *TP53*, *GATA2*, *KRAS*, *RUNX1*, *STAG2*, *ASXL1*, *ZRSR2* and *TET2*) [70].

Altogether, these results suggest that mutations follow a typical order of mutations acquisition during progression. Nevertheless, our understanding of the order of mutation acquisition remains incomplete because most data are based on the frequency of gene mutations in cohorts of unpaired MDS and sAML rather than serial samples from the same patient.

In order to solve this limitation, recent studies have included some paired serial MDS/sAML samples from the same patient and have confirmed that a typical order of mutation acquisition may exist during MDS progression to sAML. Mutations in spliceosome genes and epigenetic modifiers had already present in the MDS sample with high VAFs, indicating that they are acquired early in the disease course, while mutations in signaling genes were only rarely detectable at MDS, indicating that they are acquired later in disease progression [70,135-137]. These results are in line with the traditional simplistic model of sAML evolution that postulated to require mutations in two classes of cancer associated genes [146,147]. This model proposed that loss-of-function mutations in transcription factors, such as *RUNX1*, *CEBPA* and *RARA* (denominated as 'Class II' mutations), impair hematopoietic differentiation and led to the accumulation of more primitive cells. However, these variations alone are not sufficient to cause acute leukemia and a second mutations ('Class I' mutations) in activated signaling genes, such as *FLT3*, *RAS* and *KIT,* are required to promote proliferative and/or survival advantage to hematopoietic stem cells [148-150]. Nonetheless, these serial studies have also demonstrated that the mutational spectrum is substantially more complicated than a two-class model could explain and that sAML progression is a very complex and heterogeneous process, including other genes frequently mutated that do not fit into either of Class I and Class II categories [134].

In addition, the wide-range of sequencing approaches used across these studies limited the ability to consistently detect low-level mutations and define clonality. Thus, imputing accurate clonal architecture is limited when mutations are present at low VAFs, making it challenging to determine whether mutations co-occur in the same cell or exist in parallel subclones. In these cases, single-cell sequencing technologies may be needed, however, they can be limited by allele drop-out, namely loss of the mutated allele during PCR amplification of DNA caused by sequence variations or mutations [57].

Taking into account all of these results, progression from MDS to sAML is clearly a process of clonal evolution, which can occur over a relatively short period of time (e.g., weeks to months). Analysis of clonal evolution has shown that evolution can display two patterns: linear, in which each successive subclone occurs within the previous parental clone, or branched, in which a founding clone can generate several parallel subclones [70,136,137]. In addition, a process termed 'clone sweeping' has been identified when a subclone expand to occupy the entire tumor, forcing the collapse of other subclones [70] (**Figure 6**).

Furthermore, recently Chen et al. performed sequencing experiments in a limited number of phenotypically defined stem cell populations in MDS/sAML patients. They have described that subclonal diversity may be higher in stem cells than in blasts at both MDS and sAML and that the dominant clone in the stem cells at MDS is not always the same as the dominant clone at AML, indicating a potential non-linear path of clonal evolution at the stem cell level [141]. Thus, these results suggest a greater level of clonal diversity than appreciated using bulk BM samples, raising interesting questions about the clonal diversity of stem cells in MDS and sAML.

Therefore, in this scenario of 'game of clones' that occurs during MDS progression to sAML, additional studies of paired MDS and sAML samples are needed to provide more information regarding the genetic patterns of clonal expansion during disease progression. Moreover, these sequential studies could provide insight into a critical order of mutation acquisition that drives disease evolution and is predictive of MDS progression, potentially allowing for early intervention.

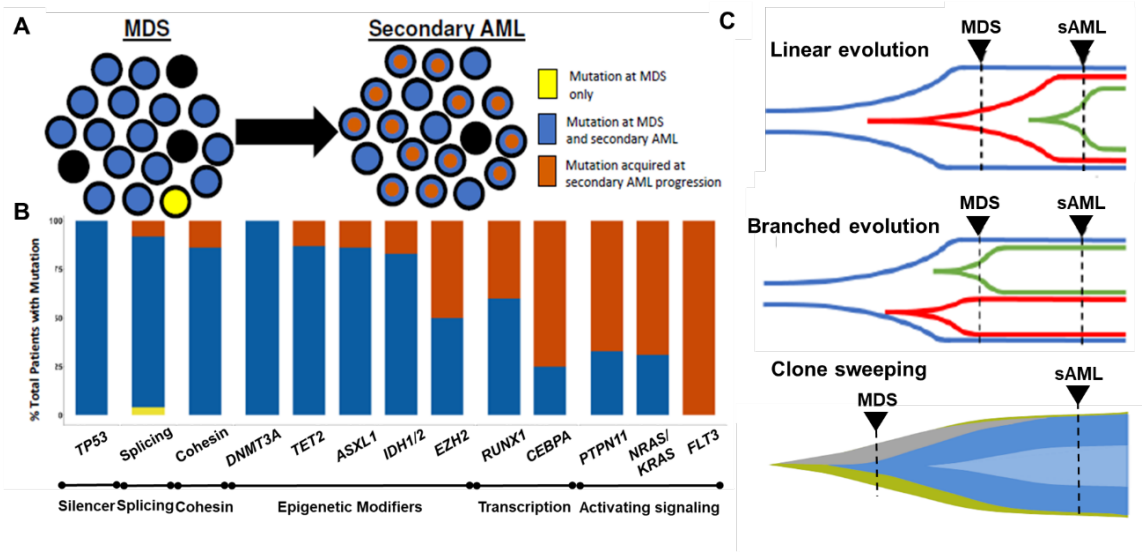


Figure 6. Clonal evolution during progression from MDS to sAML (adapted from Menssen, *Blood*, 2020) [57]. A) A model for sequential accumulation of mutations during MDS progression to sAML is shown. Black cells are normal at MDS and sAML stages. Mutations in blue and yellow are acquired early, being present in the founding clone. Blue mutations expand in the marrow at sAML stage, while yellow mutations disappear at sAML stage for clone sweeping. Some cells acquire orange mutations, form a subclone and expand at the time of sAML progression. **B) Percent of patients with a gene mutation detectable at MDS and/or sAML states**. Mutation detectable at only MDS time are represented in yellow, detected only at sAML state are in orange and detected at MDS and persist during disease progression are represented in blue. **C) Patterns of clonal evolution during the MDS progression to sAML.** Three patterns of subclone expansion are observed during progression from MDS to sAML: linear evolution, branched evolution and clone sweeping.

3.2. MDS treatment and clonal evolution

Clonal evolution during progression to sAML is characterized by the persistence of founding clone mutations and typically the expansion of a subclone with unique mutations. Although this pattern is similar for patients receiving supportive care or some treatments such as erythropoietin, thrombopoietin receptor agonists or G-CSF, other specific therapies can influence the pattern of clonal evolution [57].

One of the most common chromosomal abnormalities in MDS and sAML is del(5q). Lenalidomide is associated with a decrease in clonal del(5q) cells in the BM and patients harboring del(5q) treated with this drug can achieve a complete remission [151,152]. However, patients who eventually relapse present usually an expansion of the del(5q) cells with additional abnormalities, such as *TP53* mutations [136,153]. These *TP53* mutations have been found at low levels months prior to disease progression [154], suggesting that lenalidomide treatment can exert a selective pressure that may impact clonal evolution during progression to sAML.

Likewise, hypomethylating agents (HMAs) are one of the most common drugs to treat MDS patients and they have been considered standard of care for MDS for more than a decade. Though they are associated with longer overall survival in controlled clinical trials [155-158], a high number of patients either fail to respond or lose an initial response [52,55,56,159]. Predicting response to HMAs based on a patient's gene mutations remains challenging and, to date, results are variable across studies. Interestingly, some studies have shown that HMA therapy can induce C>G transversion in genes implicated in MDS and AML pathogenesis [160,161]. Moreover, these C>G transversion mutations can occur in a subclone that emerge at relapse [162], indicating that HMAs may influence clonal evolution and raising questions of whether HMAs may induce pathogenic mutations.

Regarding allogeneic hematopoietic stem cell transplantation, it is the only curative therapy for MDS and sAML patients and it significantly improve their outcome. However, relapse is still common [52,163]. A recent study of MDS patients who progressed after transplantation has shown that founding clone mutations were detectable at relapse in all cases and that subclones emerged at progression often harbored a new structural variant [162], suggesting a pattern of clonal evolution and disease progression following transplant that can be useful to monitor patients for molecular residual disease.

Taken altogether, current available therapeutic options for MDS patients do not prevent that a substantial number of patients display disease progression and evolve to sAML. Furthermore, some treatments, such as lenalidomide, HMAs or allo-HSCT, can influence the pattern of clonal evolution underlying the sAML progression.

4. Next generation sequencing (NGS) in the MDS study

DNA sequencing is a critical tool in biological research. Traditional Sanger sequencing has been the most used method for more than three decades since it was first introduced in 1977 [164]. It is based on the use of radioactively or fluorescently labeled chain-terminating nucleotides (dideoxy nucleotides) [164-166]. Nevertheless, traditional Sanger sequencing is restricted to the discovery of substitutions and small insertions and deletions (indels) and, thus, new and exciting technology, enabling faster and more accurate sequencing to occur in parallel reactions, was emerged [167].

Next Generation Sequencing (NGS), massively parallel or deep sequencing are related terms that describe a DNA sequencing technology which has revolutionized genomic research [165,168]. Most so-called 'massively parallel' or 'next-generation' sequencing methods are based on the standpoint of Sanger fundamental enzymological support, but in an orchestrated and stepwise fashion, enabling sequence data to be generated from tens of thousands to billions of templates simultaneously [169].

There are three main types of NGS sequencing of DNA that can be used for the identification of genomic mutations: whole-genome sequencing (WGS), whole-exome sequencing (WES) and targeted-deep sequencing (TDS) [170]. WGS provides data of the entire human genome, but as approximately 98% of the genome is non-coding and little is known about involvement of non-coding DNA in MDS pathogenesis, this method is rarely used clinically in MDS patients [171]. WES has played an important role in the study of myeloid neoplasms, particularly in unveiling the clonal evolution in MDS and AML [135,136]. Although WES and WGS may be considered *a priori* the best options due to the high number of assessed regions, they actually have limitations to incorporate into routine clinical practice. Some of these limitations are the high sequencing costs, the time-consuming bioinformatics analysis and the laborious clinical interpretation of the sequencing results [172-174]. In addition, WGS and WES can miss some subclonal mutations with low VAF, but with clinical relevance, such as *TP53* mutations [175]. For these reasons, TDS, that allows sequencing only a target regions of interest, is a powerful approach that can fulfil the best balance between the accurate identification of targeted events with great sensitivity and the overall cost and data burden for large-scale executions. Additionally, read depths of TDS are greater and allow detection of VAF at lower fractions, being a key consideration in MDS study because it displays a very high intratumoral heterogeneity and the vast majority of known pathogenic variants are found in a relatively small number of genes [62,63,166].

TDS comes in two main forms, amplicon or capture-based. Amplicon-based enrichment utilizes specifically designed primers to amplify only the regions of interest prior to library preparation. In contrast, in capture-based approaches, the DNA is fragmented and targeted regions are enriched via hybridization oligonucleotide bait sequences attached to biotinylated probes, allowing for isolation from the remaining genetic material [176-178]. Amplicon-based enrichment is the cheaper of the two technologies, shows a greater number of on target reads and requires much less starting material than capture, making it ideal if there is little DNA available for TDS. However, the coverage of the target regions is more uniform with hybrid sequencing [170]. Moreover, capture has been shown to produce fewer PCR duplicates than amplicon enrichment (<40% *vs.* ~80%, respectively) and they can be removed computationally easier, because the random shearing of the DNA in capture platforms reduces the likelihood of two unique fragments aligning to the same genomic coordinates. Similarly, the long bait sequences used in capture allow the sequencing of difficult regions, such as repeated sequences, and a greater level of specificity in region selection [170]. Therefore, capture based techniques provide more accurate and uniform target selection, while amplicon-based methodologies are often used in small scale experiments where sample quantity or cost are a factor.

A summary of the main types of NGS sequencing of DNA and of the two TDS approaches that can be used for the identification of genomic mutations is shown in **Figure 7**.

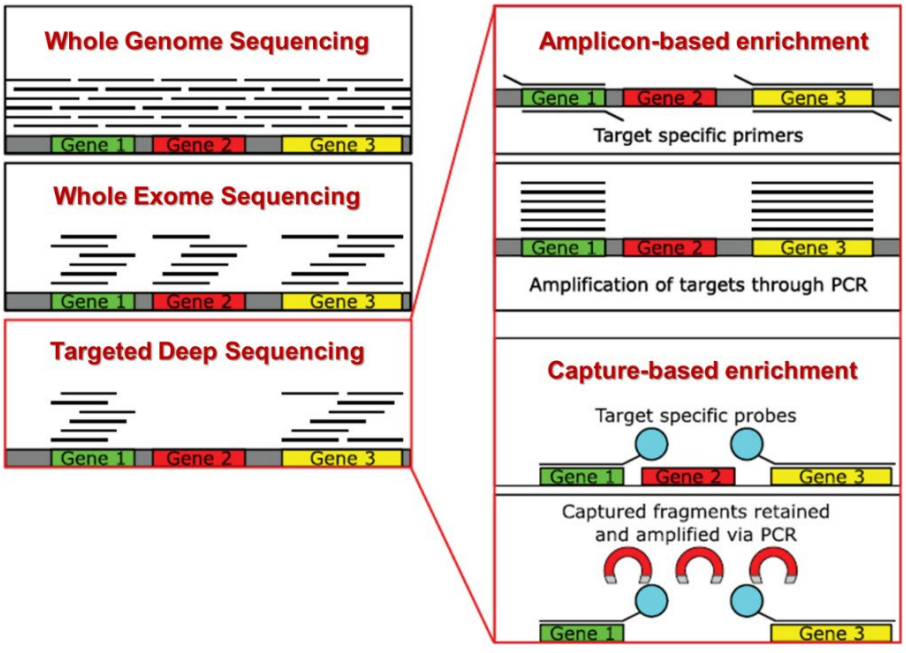


Figure 7. The main types of next generation sequencing of DNA (adapted from Bewicke-Copley, *Comput Struct Biotechnol J*, 2019) [170]. Whole genome, whole exome and targeted deep sequencing are shown, as well as the two targeted deep sequencing approaches (amplicon and capture-based enrichment) that can be used for the identification of genomic mutations.

All approaches involve a series of general steps that are common to most NGS platforms: 1) sample preparation and choice of DNA sequencing strategy; 2) library preparation; 3) sequencing; and 4) data analysis.

Sample preparation and choice of DNA sequencing strategy

The first step in NGS is the extraction of genetic material, DNA or RNA. Although this step can be considered trivial, quantity and quality of the input material is crucial because the generation of micrograms of DNA from input material by PCR is a sensitive process subject to many variations.

Ideally, a NGS library would perfectly represent the DNA present in the biological specimen from which it was derived and it cannot create more information than was present in the original template. NGS libraries can be made from small amounts of DNA, but reducing the input may comprise assay sensitivity due to the loss of sequence heterogeneity by the overamplification of particular DNA molecules during the process of library construction (duplicate reads). In fact, these duplicate reads cause failure to detect sequence variants that were present in the original specimen, over- or under-representation of particular variants, and false-positive variants calls resulting from PCR errors that are propagated through library preparation and sequencing [179,180].

Likewise, the choice of DNA sequencing strategy is one of the first steps that should be assessed. As it is described previously, there are several types of NGS approaches that differ in total sequence capacity, sequence read length, sequence run time and the quality and accuracy of the data. Therefore, these characteristics should be considered in order to choose the best strategy to be used regarding the specific clinical applications [174].

Library preparation

Library preparation is a process that consists of the fragmentation of genomic DNA and the generation of a random set of short DNA fragments of a particular size range, usually 100 to 500 base pairs, flanked by platform-specific adapters at both ends.

As detailed previously, two main approaches are used for target enrichment during the library preparation: capture-based and amplicon-based methods. In the case of capture-based assays, library preparation begins with random shearing of the genomic DNA, by mechanic or enzymatic fragmentation, followed by capture of the regions of interest using biotinylated oligonucleotides probes complementary to these sequences. Then, targeted fragments are isolated using streptavidin magnetic beads. On the other hand, amplicon-based methods involve PCR amplification of the selected genome regions using primers, thus, the sequencing reads generated will have the same start and stop coordinates dictated by the primer design [166,169,174].

In both cases, platform-specific adapters are attached at each end of the template fragments. These adapter sequences are complementary to oligonucleotides bound to beads or another solid surface, called 'flow cell', so that libraries can undergo bridge hybridization in the sequencer platform [166]. Thereby, sequencing templates are immobilized on a proprietary flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template and low non-specific binding of fluorescently labeled nucleotides.

Sequencing

Most commercial platforms using massively parallel sequencing are based on the concept of sequencing by synthesis. In essence, these methods allow nucleotide incorporation using a variety of enzymes and detection schemes that permit the corresponding instrument platform to collect data in lockstep with enzymatic synthesis on a template [169].

Sequencing by synthesis consists of the amplification in parallel reactions of DNA fragments using bridge hybridization and PCR in the presence of fluorescently-labelled nucleotides. In details, after the hybridization of the amplified libraries on the flow cell, library fragments act as a template, from which a new complementary strand is synthesized by PCR, and template strand is removed. As complementary strands undergo bridge hybridization with neighboring adaptor binding sites, repeated PCR cycles lead to cluster amplification [166,169,181]. Finally, the sequencer detects the fluorescence generated by the incorporation of fluorescently labeled nucleotides by PCR into the growing strand of DNA [181,182]. During each sequencing cycle, a single labeled nucleotide is added to the nucleic acid chain, and it serves as a terminator for polymerization, so after each nucleotide incorporations, the fluorescence is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. The result is highly accurate base-by-base sequencing that eliminates sequence-context specific errors, enabling robust base calling across the genome, including repetitive sequence regions and within homopolymers.

Nevertheless, sequencing by synthesis presents some limitations. For example, it is ultimately limited in its length of sequence read ('read length') because of the increasing noise over sequential incorporation and imaging cycles and, still, their read lengths remain shorter than Sanger read lengths. These limitations should be considered because they have impacted on how the sequence data are analyzed [169].

The main steps of library preparation, bridge hybridization, cluster amplification and sequencing by synthesis are shown in **Figure 8**.

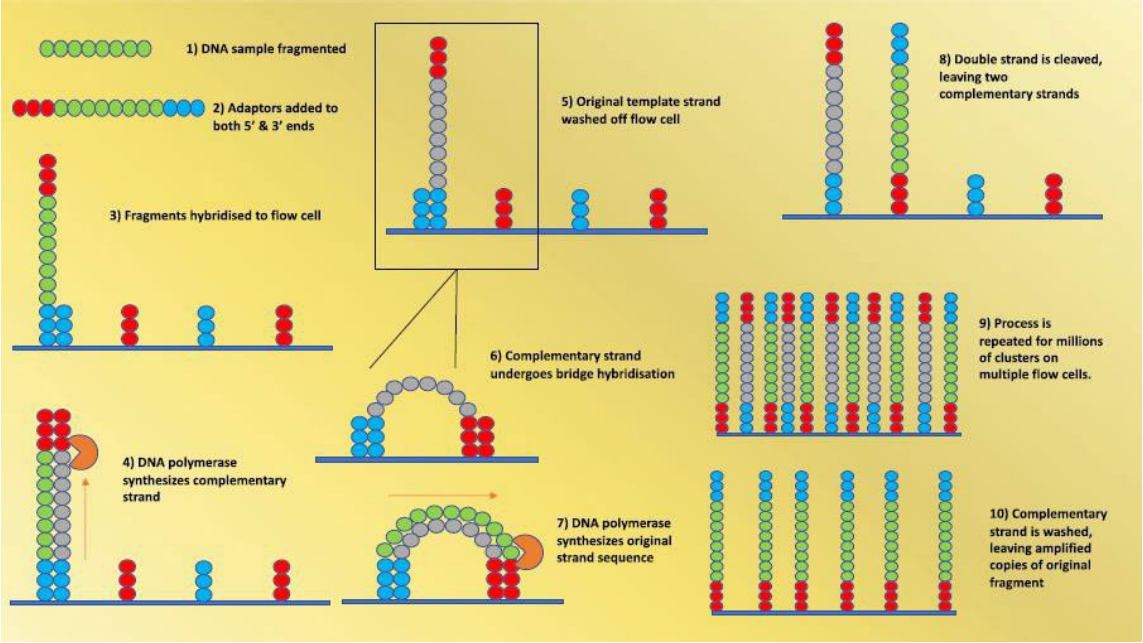


Figure 8. The steps of next generation sequencing (adapted from Spaulding, *Br J Haemat*, 2020) [166]**.** The library preparation, bridge hybridization, cluster amplification and sequencing by synthesis steps are shown.

Data analysis

NGS generates large amounts of data that require substantial computational infrastructure for storage, analysis and interpretation. Thus, data analysis of NGS can be an extremely complex, time-consuming and laborious process.

The workflow of NGS data analysis includes four phases: 1) base calling, assignment of base quality scores and de-multiplexing; 2) alignment of reads to a reference assembly or sequence(s) and variant calling; 3) annotation of the variants; and 4) interpretation of clinically relevant variants [183]. Nowadays, there is a huge number of bioinformatics tools and software to analyze NGS data. Most of them are algorithms specialized on one phase of the analysis, but there are also multiple *pipelines* that group several algorithms which analyze the data serially.

The first phase of data analysis occurs into the sequencer platform, where sequence reads are produced. It is consisted of: i) the **base calling**, which is the identification of the specific nucleotide present at each position in a single sequencing read; ii) the **base calling accuracy** measured by the Phred quality score (Q score) that indicates the probability that a given base is called incorrectly by the sequencer; and iii) the **de-multiplexing** that is the computational association of reads

with a patient when multiple samples have been multiplexed (pooled) before sequencing [183,184].

The second phase of NGS data analysis is the **mapping and alignment** of each read to the reference human genome assembly, which is available from the Genome Reference Consortium (the latest versions can be found at http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/). The alignment tools used by clinical laboratories include multiple algorithms that sometimes trade speed for accuracy and can incorporate different levels of sensitivity and specificity. However, it is recommended that alignment includes additional processing for local realignment, removal of PCR duplicates and recalibration of base-call quality scores. As a consequence of the alignment analysis, the next step is the variant calling that is the process of comparing the aligned reads to a reference genome to identify base pair variations. A variety of software tools and strategies are also available for variant calling, but no single software setting is currently able to identify all variant classes with equal accuracy. Thus, several variant callers and/or parameters setting should be evaluated to optimize the detection of different variant types during assay development [170,183,185].

Following variant calling, the next phase is to **annotate the variants** in relation to genes (within or outside a gene), codon, amino acid positions, and classify types of variants, such as nonsense, missense, exonic deletions and synonymous variants. This allows for greater understanding of their functional consequences on genes they relate to. In many studies, only non-silent exonic or splicing mutations are selected for further analysis, focusing on functional coding variants and mutations only. However, these criteria may change depending on the purpose of the study, for example variants in UTRs or in promoter regions can be interesting [170].

The final phase of NGS data analysis is a **clinical assessment** to determine which variants should be reported and how they should be described in the laboratory report. There are some algorithms that optimize settings to predict deleterious variants located in clinically relevant genes and to filter out selected variants, such as those with a high population allele frequency. Similarly, existing database, such as ClinVar, COSMIC, ClinGen, protein predictors or Human Variome Project, are valuable tools for assessing the relevance of variants and for categorizing variants according to predicted functional impact (‘often benign’, ‘likely benign’, ‘uncertain significance’, ‘likely pathogenic’ or ‘pathogenic’). Nonetheless, the interpretation of variants is a greater challenge because there is no standard format and a common understanding of variant data between laboratories. Therefore, collaborations and ongoing discussions among laboratory researchers, clinicians, manufacturers, bioinformatics, software developers and professional organizations are needed to ensure the quality of NGS tests and to guide clinical decision-making in myeloid malignancies [183,185].

Despite the potential to inform and guide clinical practice, NGS has some limitations. Although the costs for NGS have recently become more affordable, they remain be high. In addition, it can take one to three weeks to receive a finalized data report, which may be too long to affect an initial treatment decision for a patient with more aggressive disease. Regarding practical considerations, NGS are not standardized and there is still a lack of gold-standard pipelines for analysis, which can lead to poor reproducibility between laboratories, even for the same data sets. Thus, there is a need for collaboration to form larger data pools in order to improve the accuracy and efficacy in variant calling for prognostic and therapeutic signatures and to infer meaningful conclusions [166,170]. Moreover, though the ability of NGS to detect down to 1% VAF is an improvement from Sanger sequencing (threshold >20% VAF), interpreting VAF in a meaningful way requires context. Not all somatic gene mutations detected by NGS are guaranteed to be pathogenic mutations contributing to myeloid malignancy development and further studies are needed to establish a meaningful threshold for VAF sensitivity [166].

Therefore, NGS has revolutionized genomic research and it is one of the most rising methodologies that will change the management of patients. However, it will be important to standardize sequencing and variant calling methods and continue to expand knowledge of the pathogenic and prognostic role of specific variants in order to improve the clinical utility of NGS.

Hypothesis

MDS are a heterogeneous group of clonal hematopoietic disorders that present a highly variable clinical course, being their diagnosis, prognostic stratification and clinical decision making very challenging. In last years, the implementation of high-throughput techniques, such as NGS, has revolutionized the genomic research, expanding our knowledge about the molecular abnormalities underlying MDS pathogenesis and improving our understanding of the mechanisms involved in disease progression. However, the lack of efficient combination of genetics and clinical information has restrained the clinical interpretation of the sequencing results and, consequently, it has limited the adoption of NGS data into real-life clinical practice.

Although the clinical course of MDS is variable, approximately a third of patients diagnosed with MDS eventually evolve to sAML, displaying these patients a dismal prognosis. The application of high-throughput techniques has enabled to define that this disease evolution is clearly a process of clonal evolution and also to identify some of the molecular alterations behind this progression. Nevertheless, most data are based on studies using cohorts of unpaired MDS-sAML samples rather than sequential samples from the same patients. Accordingly, the order of mutations acquisition as well as the mutational dynamics remain incomplete and, thus, the mechanisms involved in leukemic transformation have just begun to become clear. Hence, longitudinal sequencing genomic studies are still required to determine which mutations, alone or in combinations, are involved in MDS progression to sAML as well as to define the clonal evolution patterns underlying this progression.

Over the last several years, our knowledge about the MDS mutational landscape has improved due to the application of NGS methodologies. It has identified that approximately 90% of MDS patients carry at least one mutation, being some of them correlated with clinical features and being of prognostic value. However, multiple driver genes are affected in MDS patients and combinations of mutations are demonstrated not to be totally random, suggesting the presence of gene-gene interactions that could underlay clonal evolution. Moreover, the lack of clinical validation has restrained the diagnostic and prognostic categorization of MDS patients based on sequencing results. Therefore, large studies combining genomic data and reliable clinical information are required to define the prognostic impact of mutations and translate them into clinical practice.

Cohesin mutations have been found in 10-15% of MDS patients and have been described to alter the transcriptional orchestration of differentiation/proliferation processes, suggesting that cohesin complex plays a relevant role in the MDS pathogenesis. Nonetheless, given the molecular heterogeneity of MDS, the clinical impact of cohesin mutations is still undetermined. Thus, it is necessary to gain insight into the relevance of cohesin mutations to clinical phenotypes, as well as clinical

outcome, thereby further studies in large cohorts of patients with detailed clinical information are still required.

Similarly, *RAS*-pathway mutations have been reported in 5-10% of MDS cases, being most prevalent in high-risk MDS and AML patients. *RAS*-pathway pathway is one of the several functional pathways that are involved in the pathogenesis of MDS, and activating *RAS*-pathway mutations are shown to lead to increased proliferation of hematopoietic progenitor cells. Although the prognostic impact of these mutations has been extensively studied in patients with MPN and AML, information regarding their impact on MDS patients is limited. This implies that there is a lack of clear consensus about which the prognosis of patients carrying *RAS*-pathway mutations could be. Hence, and similar as before, further investigation using large patient cohorts is needed.

In summary, the implementation of the new high-throughput technologies will help us to elucidate the mechanisms involved in both MDS pathogenesis and their leukemic progression. Moreover, the integration of mutation information with the clinical-biological parameters, already included in current prognostic classifications, would allow us to advance towards a more accurate integrated diagnosis, a more precise prognostic stratification and even a more adequate therapeutic orientation, due to it would enable the detection of somatic mutations against which target drugs are being developed.

Aims

General Aim:

To characterize the molecular mechanisms and the clonal evolution patterns that are involved in the pathogenesis of MDS and the progression from MDS to sAML, as well as to use the molecular data in order to reach a more accurate integrated diagnosis and a more precise prognostic stratification.

Specific aims:

1. To characterize the mutational landscape of MDS patients evolving to sAML.

    1.1. To assess the mutational dynamics and the clonal evolution patterns underlying the MDS progression to sAML.

    1.2. To explore the effect of disease-modifying treatment on clonal evolution profiles during the leukemic transformation of MDS patients.

2. To define the relevance of cohesin mutations in MDS patients:

    2.1. To analyze the incidence of somatic mutations in cohesin genes by next generation sequencing in a large cohort of MDS patients, as well as the co-occurrence between these mutations with other myeloid-related genes.

    2.2. To correlate the presence of mutations in cohesin genes, alone or in co-occurrence with additional mutations, with the clinical characteristics and outcome of MDS patients.

    2.3. To analyze the effect of the integration of molecular data from cohesin genes into the current diagnostic and prognostic classifications.

3. To study the significance of somatic mutational profiling in oncogenic RAS signaling pathway genes in MDS pathogenesis:

    3.1. To establish the incidence of somatic *RAS*-pathway mutations in MDS patients through next generation sequencing techniques, as well as to assess the mutational profile of MDS patients with *RAS*-pathway mutations.

    3.2. To define the correlation of activating *RAS*-pathway mutations with disease phenotype and clinical impact on overall survival and time to progression to sAML.

Results

This section includes the experimental work performed on this thesis, including Patients and Methods, Results and Discussion. This section has been divided into three chapters:

Chapter 1. Martín-Izquierdo M, Abáigar M, Hernández-Sánchez JM, Tamborero D, López-Cadenas F, Ramos F, Lumbreras E, Madinaveitia-Ochoa A, Megido M, Labrador J, Sánchez-Real J, Olivier C, Dávila J, Aguilar C, Rodríguez JN, Martín-Nuñez G, Santos-Mínguez S, Miguel-García C, Benito R, Díez-Campelo M, Hernández-Rivas JM. **Co-occurrence of cohesin complex and Ras signaling mutations during progression from myelodysplastic syndromes to secondary acute myeloid leukemia.** *Haematologica*. 2021 Aug 1;106(8):2215-2223. doi: 10.3324/haematol.2020.248807. PMID: 32675227.

Chapter 2. Martín-Izquierdo M, Díez-Campelo M, Sánchez-Real J, Hernández-Sánchez A, Hernández-Sánchez JM, Janusz K, López-Cadenas F, Tormo M, Megido M, Olivier C, Madinaveitia-Ochoa A, Dávila J, Sierra M, Vargas M, Santos-Mínguez S, Miguel-García C, Benito R, Hernández Rivas JM, Ramos F, Abáigar M. **Incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of very low/low-risk myelodysplastic syndromes**. *Am J Hematol*. Second review.

Chapter 3. Martín-Izquierdo M, Ramos F, Abáigar M, Hernández-Sánchez JM, González T, López-Cadenas F, Quijada-Álamo M, Santos-Mínguez S, Miguel-García C, Lumbreras E, Janusz K, del Rey M, Toribio-Castelló S, Hernández-Rivas JM, Díez-Campelo M and Benito R. ***RAS*-pathway mutations are associated with a worse clinical outcome in myelodysplastic syndromes patients** *Cancers*. Submitted.

All of them have been developed to accomplish the general aim of this PhD work and give the title to this doctoral dissertation: "*Molecular mechanisms and clonal evolution underlying the progression of Myelodysplastic Syndromes to Acute Myeloid Leukemia: Genomic characterization by Next Generation Sequencing*".

A General Discussion, with additional data and which comprises all research, is addressed in a separate section of this thesis.

In addition, the supplementary material corresponding to each of the above chapters is collected at the end of the digital version of the thesis.

Chapter 1

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).
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.
13. Hematology, Hospital Virgen del Puerto, Plasencia, Spain.

*MA and MMI contributed equally to this work.
*MDC and JMHR contributed equally to this work and both are co-senior authors.

Haematologica (2020). doi: 10.3324/haematol.2020.248807. PMID: 32675227.

Myelodysplastic Syndromes

ARTICLE

• Ferrara Storti Foundation

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,' Andrés Madinaveitia-Ochoa,[6] Marta Megido,[7] Jorge Labrador," Javier Sánchez-Real,5 Carmen Olivier,[9] Julio Dávila,1° 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]..

[1]1nstitute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC), Salamanca, Spain; '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 MM/ contributed equa//y* as *co-first authors.*
'MDCand JMHR contributed equa/ly as *ca-senior authors.*

Haematologica 2021
Volume 106(8):2215-2223

ABSTRACT

Myelodysplastic syndromes (MDS) are hematological disorders at high risk of progression to secondary acute myeloid leukemia (sAML). However, the mutational dynamics and clona! evolution underlying disease progression are poorly understood at present. In arder to elucidate the mutational dynamics of pathways and genes occurring during the evolution to sAML, next-generation sequencing was performed on 84 serially paired samples of MDS patients who developed sAML (discovery cohort) and 14 paired samples from MDS patients who did not progress to sAML during follow-up (control cohort). Results were validated in an independent series of 388 MDS patients (validation cohort). We used an integrative analysis to identify how mutations, alone or in combination, contribute to leukemic transformation. The study showed that MDS progression to sAML is characterized by greater genomic instability and the presence of severa! types of mutational dynamics, highlighting increasing *(STAG2)* and newly-acquired *(NRAS* and *FLT3)* mutations. Moreover, we observed co-operation between genes involved in the cohesin and Ras pathways in 15-20% of MDS patients who evolved to sAML, as well as a high proportion of newly acquired or increasing mutations in the chromatin-modifier genes in MDS patients receiving a disease-modifying therapy befare their progression to sAML.

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

Received: February 3, 2020.
Accepted: July 14, 2020.
Pre-published: July 16,2020.

https:// doi.org/10.3324/haematol.2020.248807

Introduction

Myelodysplastic syndromes (MDS) are a heterogeneous group of clona! hematopoietic disorders characterized by peripheral blood (PB) cytopenia with dysplastic bone marrow (BM) morphology and an increased risk of progression to secondary acute myeloid leukemia (sAMl).[13] Approximately one third of patients diag-

nosed with MDS eventually transfarm into sAML.[4] Disease progression is associated with a dismal prognosis, partly because most of these patients are resistant to currently available treatrnents, and the long-term survival rate of treated patients is less than 10% after a couple of years.[1,7]

In recent years, new high-throughput genomic technologies, such as next-generation sequencing (NGS), have enabled a large number of studies to elucidate sorne of the mechanisms involved in MDS pathogenesis such as epigenetic regulation, transcription, signaling pathways, splicing, cohesin complex, apoptosis and angiogenesis. However, MDS exhibit great genetic and clinical heterogeneity, so the nature of their pathogenesis is still not fully understood.[8,10] The mutational dynamics and clona! evolution underlying disease progression have just begun to become clear. Previous studies have identified multiple genes recurrently mutated in MDS and sAML and these have provided insight into the great intratumoral heterogeneity typical of progression from MDS to sAML.[a,13] These studies have shown that the evolution of the disease is a complex process involving new additional alterations co-existing with the MDS faunder clone.[14] Moreover, recent studies have described the association of mutations in genes such as *TET2, RUNX1, ASXL1,* and *STAG2* with high-risk MDS, as well as the presence of mutations in genes activating signaling pathways, such as *FLT3, PTPN11, NPM 1,* and *NRAS,* which are newly ac uired in sAML and associated with faster progression.[15,1] However, this complexity and the lack of large cohorts of serial samples means that molecular mechanisms of disease progression are only partly understood. Thus, longitudinal sequencing genomic studies are still required to determine which mutations or combinations of them are important in leukemic transfarmation.

In this study, we perfarmed whole-exome sequencing (WES) and/or targeted deep sequencing (TDS) on serial samples from MDS patients who evolved to sAML (discovery cohort) befare and after progression, as well as TDS on additional MDS patients who did not progress to sAML during fallow-up (control cohort). The results were validated in an independent series of MDS patients (validation cohort). Interestingly, we undertook an integrative analysis to determine the mutational dynarnics of the pathways and genes and to identify how mutations, alone or in combination, contribute to leukemic transfarmation. The study showed involvement of co-occurrence of alterations in the cohesin and Ras pathways in the MDS transfarmation to sAML, as well as a high proportion of newly acquired or increased clona! selection of mutations in the chromatin-modifier genes in MDS patients who received a disease-modifying therapy befare their progression to sAML.

Methods

Study design

In arder to study the mutational changes occurring during the evolution to sAML from a previous myelodysplastic phase, 486 samples from 437 patients were included in the study. The patient series was divided into three cohorts *(Online Supplementary Figure S*1): i) discovery cohort: a cohort of MDS -+ sAML progressing patients that included 42 patients diagnosed with MDS who progressed to sAML; according to the study design, 84 BM serial patient-matched samples were collected and sequenced on two occasions with the first sampling, at initial presentation of the disease (diagnosis, MDS stage), and the second sampling, after progression to sAML (disease evolution, leukemic phase); ali samples were analyzed by a TDS strategy; furtherrnore, 16 of those progressing patients (32 samples) were initially studied by WES; infarmation about the treatrnent received befare progression was available far ali 42 patients: azacytidine (n=16), lenalidomide (n=4) and no treatrnent or supportive care (n=22); ii) control cohort: a cohort of MDS non-progressing patients consisted of 14 BM paired samples from seven MDS patients who did not progress to sAML after a mínimum of 3-year fallow-up far low-risk MDS (LR-MDS) and 1 year far high-risk MDS (HR-MDS) (median fallow-up of 52 months; range, 20-89 months); according to the study design, the second sampling in this control cohort corresponded to a time when the disease was stable and TDS was perfarrned on ali these samples; iii) validation cohort: a cohort of 388 BM or PB samples from patients suffering MDS at diagnosis and far which only one time-point (sample) was studied by TDS;. notably, 63 of these patients eventually evolved to sAML, while 325 had not progressed to sAML after a median fallow-up of 19.6 months. The main patient clinical characteristics are summarized in the *Online S11pple111entary Table S1.*

This research was perfarmed in accordance with the Declaration of Helsinki guidelines, and was approved by the Local Ethics Committee ("Comité Ético de Investigación Clínica, Hospital Universitario de Salamanca"). Ali patients provided written informed consent.

Sequencing analysis

Who/e-exome *sequencing*

WES was perfarmed on matched diagnosis-progression samples from 16 patients of the discovery cohort. The mean coverage of WES was 77.6x (range, 36-124) and at least 73% of the captured regions had a coverage of 30x or more far ali 32 samples *(Online S11pple111entary Table S2).* See the *Online Supplementary Appendix* far ful! details.

Targeted-deep sequencing

Ali genomic DNA samples underwent TDS using an in-house custom capture-enrichment panel of 117 genes previously related to the pathogenesis of myeloid malignancies (*Online Supplementary Table S3).* The mean coverage ofTDS was 665x (range, 251-1,198) where 99.5% of target regions were captured at a leve! greater than l00x. See the *Online S11pple111entary Appendix* far full details.

Analysis of mutational dynamics

The main aim of this study was to analyze the mutational changes occurring between the first sampling (MDS stage) and the second sampling (stable disease/sAML stage) in the discovery and control cohorts. To this end, variant allele frequency (VAF) at these two stages were compared using two approaches: i) VAF ratio between second and first sampling, where thresholds of >1.2 and <Ü.8 were used to classify mutations as increasing or decreasing, respectively, while ratios between these thresholds were considered to be stable; and ii) Fisher' s exact test where values of $P<0.05$ were taken to indicate statistically significant changes during progression.

Results

Molecular landscape of the progression from myelodysplastic syndromes to secondary acute myeloid leukemia

In order to characterize the main cohort of the study, the discovery cohort, 16 patients (patients #27 - #45) were analyzed by WES at the time of diagnosis and at leukemic

transformation, as previously explained. After a stringent analysis[18] as described in the *Online Supplementary Appendix,* 61 variants were identified as likely somatic mutations: 40 were called driver and 21 were called passenger using the novel bioinformatics too! "Cancer Genome Interpreter"[19] *(Online Supplementary Table S4).* A total of 47 variants in genes known to drive myeloid malignancies were further validated with a true positive rate of >89% using TDS and VAF correlation between two platforms was high (Pearson' s r=0.90) *(Online 5upplementary Figure 52).* However, the application of TDS revealed that severa! driver mutations were not detected by WES as they were poor-!y covered and displayed a low VAF. Then, we decided to more comprehensively study disease progression by applying the TDS panel in a larger cohort of serially collected samples.

We performed TDS on these 16 patients of the initial discovery cohort and in additional 26 patients. A total of 159 mutations were identified at diagnosis of the 42 patients *(Online 5upplementary Table 55).* The most recurrently mutated genes were *TET2* (14 of 42, 33%), *5F3B1* (13 of 42, 31%), *5R5F2* (ten of 42, 24%), *DNMT3A* (10 of 42, 24%), *STAG2* (8 of 42, 19%), *TP53* (8 of 42, 19%) and *A5XU* (6 of 42, 14%). At the time of the second sampling, the sAML

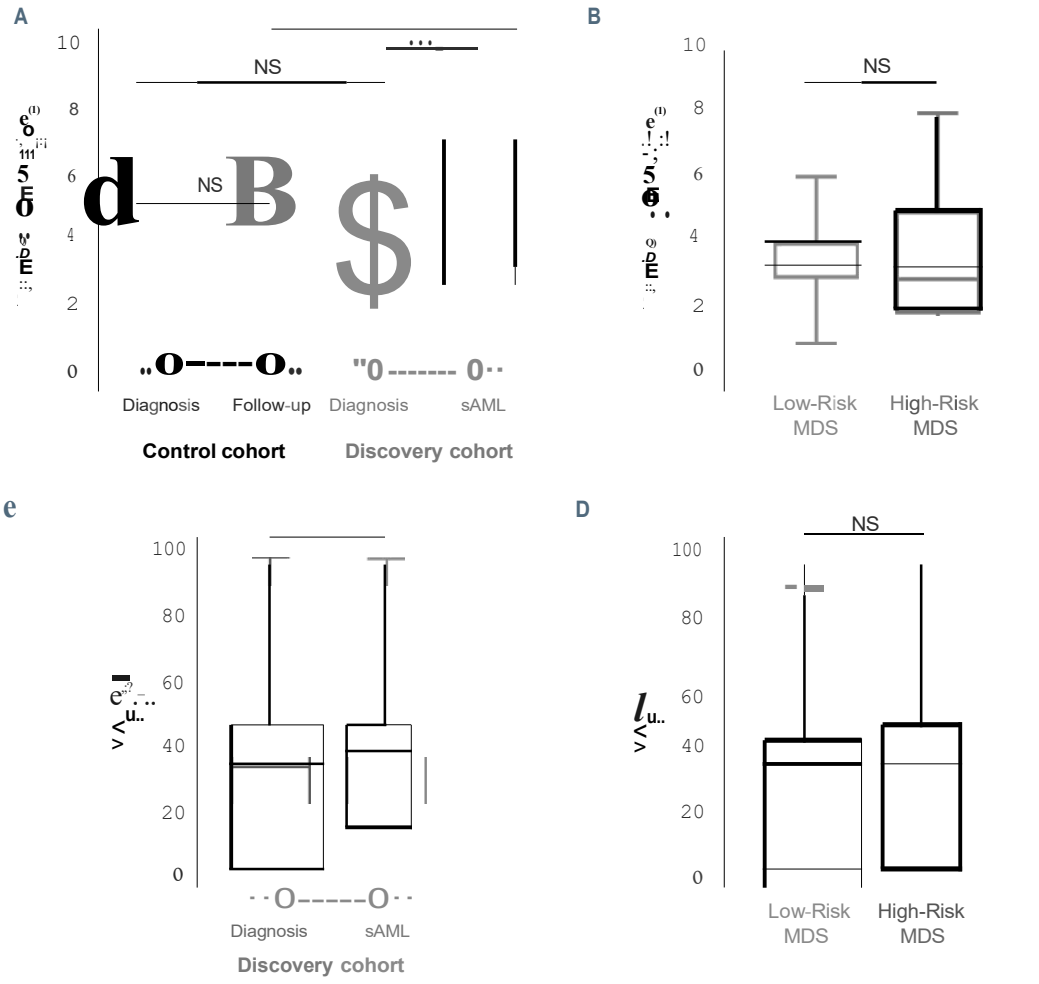


Figure 1. Boxplots showing the differences in the number of mutations (A-B) and variant allele frequency (C-D) between the two times analyzed during the evolution of the disease and between the different French-American-British/World Health Organization (FAB/WHO) subtypes at the time of diagnosis. (A) Differences in the number of mutations between diagnosis and follow-up/secondary acute myeloid leukemia (sAML) stages within and between the control and discovery cohorts. These graphs show a statistically significant increase in the number of mutatlons during disease evolution in patients who progressed to sAML (P<0.0001). (B) No significant differences in the number of mutations between the FAB/WHO subtypes at time of diagnosis (number of mutations low-risk myelodysplastic syndromes [LR-MDS] vs. high-risk MDS [HR-MDS], P=0.588). (C) Differences in variant allele frequency (VAF) of detected mutations between diagnosis and sAML stage in the discovery cohort. The VAF was higher at the time of sAML progression (P<0.0001). (O) No significant difference in VAF between the FAB/WHO subtypes at time of diagnosis (VAF: LR-MDS vs. HR MDS, P=0.528). NS: not significant; *P<0.05; ****P<0.0001.

M. Martín-Izquierdo *et al*

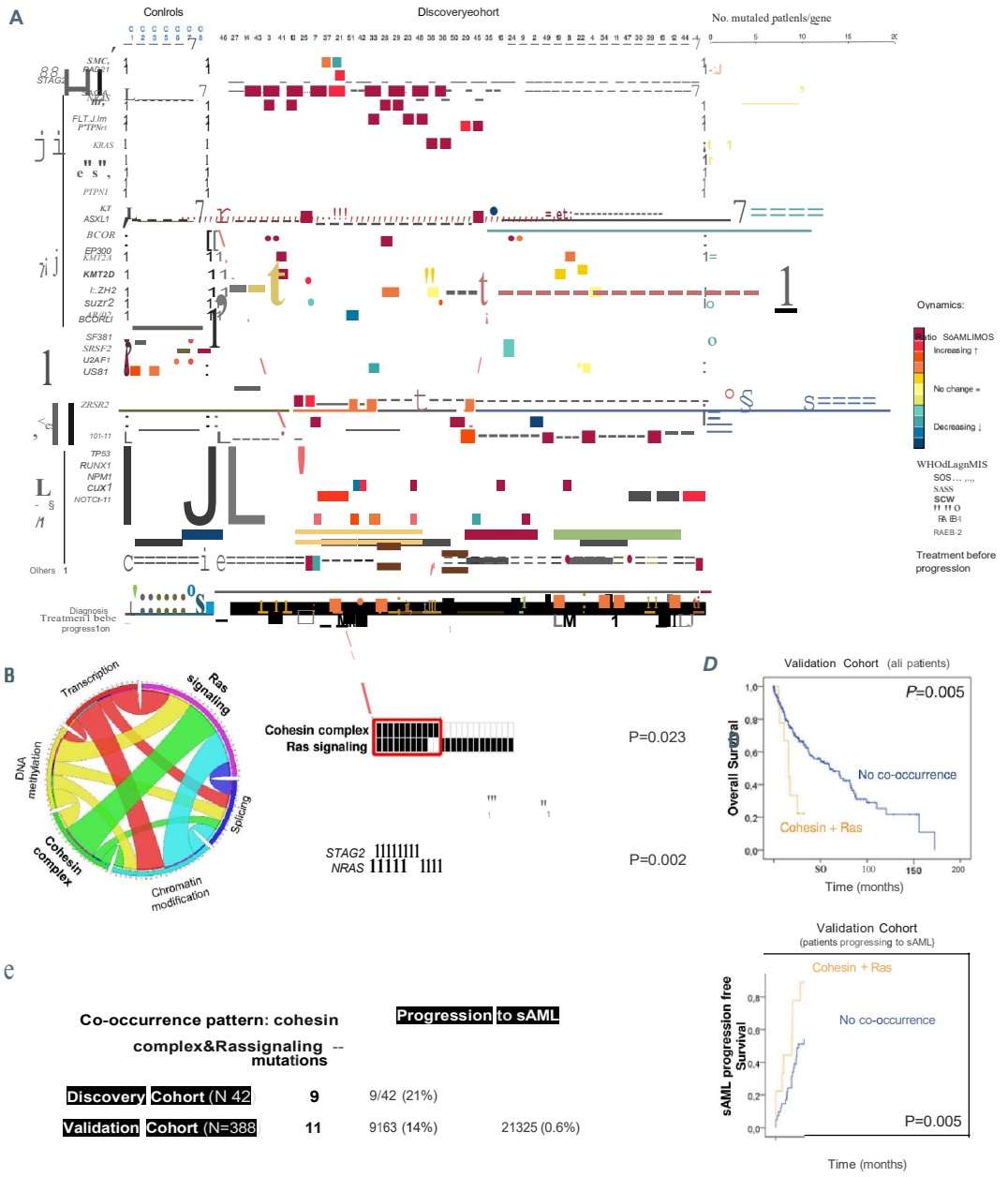


Figure 2. Dynamics of gene mutations in the myelodysp1astic syndromes to secondary acute myeloid leukemia progression axis. (A) Comprehensive landscape of mutational dynamics in the discovery and control cohorts. Genes are grouped by cellular functions and are represented in rows; each column represents a patient. Dynamics are represented by a color gradient: red/orange for newly acquired/increasing mutations, yellow far stable mutations, and blue/green colors for decreasing mutations. (B) Co-occurrence of cohesin complex and Ras signaling mutations in the discovery cohort. Circos plot of statistically significant associations between mutations detected in the discovery cohort, grouped by functional pathways. Graphs represent patients with mutations in the cohesin complex and Ras signaling, and the most frequently mutated genes In these pathways, *STAG2* and *NRAS*, showing a statistically significant association (P=0.023 and P=0.002, respectively). (C) Incidence of this co-occurrence pattern in the discovery and validation cohorts. The table contains the number of patients with the combination of cohesin and Ras signaling mutations in the discovery and validation cohorts andan indication of whether they evolved to secondary acute myeloid leukemia (sAML). (O) Prognostic impact of the co-occurring mutations in the cohesin complex and Ras pathway. Kaplan-Meier curves for overall survival and sAML progression-free survival in patlents bearing co-occurring cohesin and Ras pathway mutations in the entire validation cohort. VAF: variant allele frequency; LR: low-risk; HR: high-risk; NS: not significant; *P<0.05.

stage, 210 mutations were detected, 159 of which were already known to be present in the MDS stage, at clona! or subclonal levels (128 were retained and 31 evolved during disease progression), while 51 were detected only at the second sampling. The most recurrently mutated genes at the sAML stage were similar to those noted at the MDS stage: *TET2* (17 of 42, 40%), *SF3B1* (13 of 42, 31%), *TP53* (12 of 42, 29%), *SRSF2* (ten of 42, 24%), *DNMT3A* (ten of 42, 24%), *ASXL1* (nine of 42, 21%) and *STAG2* (eight of 42, 19%). However, the *NRAS* gene (9 of 42, 21%) also stood out at this stage. It should be noted that 32 of 54 genes (59%) were mutated only in fewer than three (<7%) patients, highlighting the great heterogeneity in the mechanisms of disease evolution.

Regardless of World Health Organization diagnosis subtypes, patients progressing to secondary acute myeloid leukemia present a higher number of mutations than those that do not progress

In arder to analyze the changes in clona! size and distribution during evolution to sAML, we compared the number of mutations identified at diagnosis and at the second sampling in the discovery and control cohorts. The control cohort presented a median of three mutations at both sampling times (p10-p90: 2-4 in both), indicating no significant differences (P=0.449). By contrast, the discovery cohort had a median of four (p10-p90: 1-6) and five (p10-p90: 2-9) mutations at the first and second samplings, respectively, representing a highly significant increase in the number of mutations during disease progression (P<0.0001). Remarkably, the control and discovery cohorts hada similar number of mutations at the time of diagnosis (P=0.097), although a slight trend was observed, while patients who progressed showed a significantly higher number of mutations at the time of sAML than the control patients at the second sampling (P=0.027) (Figure lA). Considering the discovery cohort patients by World Health Organization diagnosis subtype (LR-MDS and HR-MDS) <lid not revea! any significant differences in the number of mutations in patients progressing to sAML (P=0.588) (Figure lB).

In arder to further study what characterizes disease evolution, we compared the VAF of mutations at both times. Patients who evolved to sAML presented a significantly higher VAF median at second sampling (29.11% *vs.* 36.76%, P<0.0001) (Figure lC). However, no differences were identified in the median VAF between each subtype at the time of diagnosis (P=0.528). (Figure lD).

Therefore, tak.ing ali these results together, MDS patients, irrespective of their diagnostic subtype, displayed a greater genomic instability during disease progression than patients who <lid not evolve to sAML.

Mutational dynamics during the progression to secondary acute myeloid leukemia: clonal evolution

In arder to study the mutational dynamics and identify which mutations could be involved in clona! evolution and play an important role during disease progression, the VAF of mutations detected at both times (follow-up/sAML *vs.* diagnosis) were compared in ali patients of the discovery and control cohorts.

Four types of clona! dynamics were identified: type 1, in which mutations were initially present in the MDS stage, but whose VAF increased significantly in the sAML stage; type 2, mutations whose VAF significantly decreased; type 3, mutations that were newly acquired at the sAML stage; type 4, mutations that persisted with a similar allelic burden at both stages.

Stable mutations (Figure 2A, type 4, depicted in yellow) were detected in genes involved in the spliceosome and DNA methylation pathways, such as the splicing factor *SRSF2* (diagnosis *vs.* sAML median VAF, P=0.4922) and the DNA methylation gene *DNMT3A* (diagnosis *vs.* sAML median VAF, P=0.7695) *(Online Supplementary Figure* S3).

Only a minority of the mutations detected at diagnosis showed a decrease in their allelic burden (Figure 2A, type

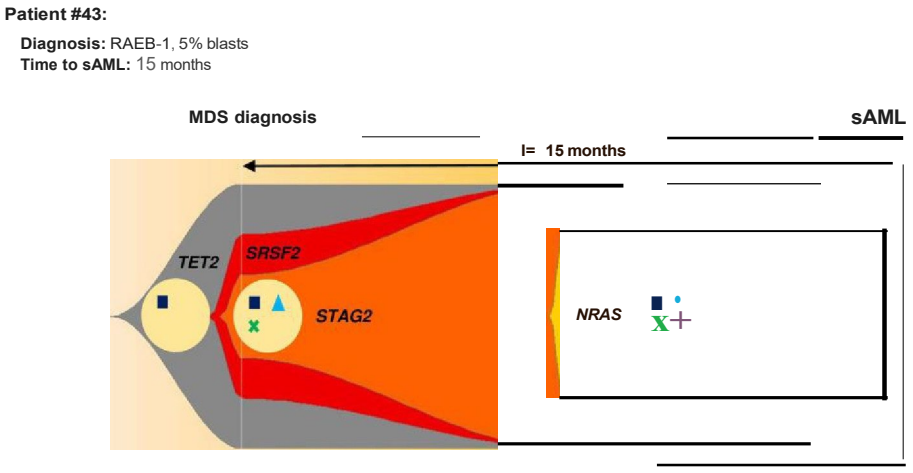


Figure 3. Model of clonal evolution during myelodysplastic syndrome progression to secondary acute myeloid leukemia using patient #43 asan example and applying the Fishplot R package.[29] In this patient diagnosed as RAEB-1, an myelodysplastic syndrome (MDS) founder clone was present at the time of diagnosis with typical myeloid mutations such as *TET2* and *SRSF2*. This clone also harbored a mutation in *STAG2*, thus it triggered the acquisition of a subsequent mutation in a Ras pathway gene, namely *NRAS*. This clone expanded, driving the evolution of the disease.

2, in green-blue), and these were randomly distributed throughout all the genes without showing a pattern.

The most interesting dynamic patterns were those of newly acquired mutations (Figure 2A, type 3, represented in red) or increased in clona! size (Figure 2A, type 1, in orange) at the time of sAML progression. These mutational patterns were mainly found in the cohesin complex and Ras signaling, where they were clustered in the *STAG2* and *NRAS, KRAS* and *FLT3* genes. These profiles were also detected in transcription factors and epigenetic modifiers, but in these cases they were randomly distributed among the genes. In fact, the *STAG2* VAF median was significantly higher at sAML stage (diagnosis *vs.* sAML median VAF, P=0.023) and this gene was mutated in eight patients of the discovery cohort, in five of which the VAF had increased by the time they had become sAML *(On/ine Supplementary Figure* S3). The increase was statistically significant in three of these five patients, while it was not significant in the other two, although a trend was observed (P<0.08), probably because the VAF at diagnosis was already very high. Moreover, most of the mutations (nine of 12) in the cohesin complex genes were of the frameshift or stop gained (loss of function) type and the cohesin-mutated patients showed a higher number of mutations than wild-type patients (median number of mutations: seven *vs.* four, P=0.0179). On the other hand, *NRAS* and *FLT3* mutations were newly acquired (diagnosis *vs.* sAML median VAF, *P=0.0029* and *P=0.0078,* respectively) during the evolution and so were detected at the sAML stage *(Online Supplementary Figure S3).*

Co-occurrence of cohesin complex and Ras signaling mutations in patients after progressing to secondary acute myeloid leukemia

Within this heterogeneous landscape of mutational dynamics, we focused our study on increasing (type 1) and newly acquired (type 3) mutations because their dynamic patterns suggested that they were positively selected during disease evolution. Moreover, in order to better characterize the mechanisms driving sAML progression, we studied which pathways and combination of them were affected by these types of mutations.

In the discovery cohort, a high proportion of Ras signaling-mutated patients at the sAML stage, already harbored cohesin complex mutations. In fact, 26% (11 of 42) of the discovery cohort patients carried mutations in the cohesin complex at diagnosis. On the other hand, 52% (22 of 42) of the patients had at least one Ras pathway mutation at the sAML stage, mainly acquired during the evolution of the disease. Of interest, nine of these cohesin-mutated patients (nine of 11, 82%) carried a co-occurring Ras signaling mutation at the sAML stage. Considering only the most recurrently mutated gene, *STAG2* (n=eight of 11), seven patients (seven of eight, 88%) carried another mutation in the Ras pathway, this being a *NRAS* mutation in five patients. Therefore, there was a statistically significant co-occurrence of these two pathways (P=0.023) and of the most recurrently mutated genes of these pathways, *STAG2* and *NRAS (P=0.002)* (Figure 2B).

In order to confirm these observations and their impact on MDS progression to sAML, the combination of the cohesin complex and Ras pathway mutations was sought in the validation cohort, an independent cohort of 388 patients in which the disease was studied on only one occasion, at diagnosis. In fact, these co-occurring mutations were detected in eleven additional patients: nine of which finally transformed into sAML (nine of 63), while two patients did not evolve during the median follow-up of 19.6 months (two of 325) (Figure 2C). Although all samples of this cohort were studied at diagnosis, these nine patients carried cohesin and Ras co-occurring mutations at an advanced stage of the disease, indeed these were detected in sAML sampling or in patients who transformed in a median time of 11 months from sampling.

The discovery cohort included only patients who evolved to sAML, and therefore displayed a very poor outcome. This made it difficult to measure the impact of this co-occurrence in these patients. For that reason and also to further study the clinical consequences of this co-occurrence on outcome, the effects on overall survival and progression-free survival in the validation cohort (median follow-up of 19.6 months) were analyzed. In our validation cohort, where 16.2% of patients evolved to sAML and 44.76% died *(Online Supplementary Table S1),* those patients harboring both the cohesin complex and Ras signaling mutations had significantly shorter overall survival (16 *vs.* 60 months, *P=0.005)* and significantly earlier progression to sAML (10 *vs.* 15 months, *P=0.005)* (Figure 2D). Moreover, in order to study the contribution of the cohesin and Ras mutations alone to these effects, comparison of median overall survival of the double-mutant and cohesin and Ras single mutant patients was performed and patients harboring double mutations showed shorter overall survival than patients with Ras or cohesin single mutations (16 *vs.* 25 *vs.* 37 months, respectively, P=0.018, *Online Supplementary Figure S4).*

Higher proportion of newly acquired or increasing mutations in chromatin modifiers in treated myelodysplastic syndrome patients

As previously mentioned, 48% of the patients in the discovery cohort of this study were treated with 5-azacytidine (AZA) (n=16) or lenalidomide (n=4), and progressed to sAML after therapy, whereas the other 52% received no treatment (only supportive care). Thus, we investigated whether the mechanisms of progression could be slightly different between patients who were treated with disease-modifying agents (AZA and lenalidomide) before transformation into sAML and non-treated patients.

In order to achieve this aim, the proportions of the different mutational dynamics were compared between treated and untreated patients. Thereby, the mutational dynamics featured a significantly higher proportion of newly acquired or increasing mutations in chromatin modifiers at the time of sAML in treated patients (eight of 15 mutations), while in untreated patients the majority of mutations were stable (53% [eight of 15] *vs.* 19% [four of 21], P=0.031). By contrast, and with respect to the treatment, no differences were detected in the dynamics of the cohesin complex (50% [three of six] of newly acquired or increasing mutations in treated patients *vs.* 50% [three of six] in untreated, *P=l.00)* or Ras pathway mutations (91% [ten of 11] in treated *vs.* 76% [13 of 17] in untreated patients, P=0.3299) *(Online Supplementary Figure S5).* Thus, our study suggests that mutations in chromatin-modifier genes could be related to the evolution of patients who receive disease-modifying treatment before progression to sAML.

Discussion

Our study characterizes the landscape of mutations during progression toward sAML and how mutations, alone or in combination, contribute to leukemic transfarmation, based on the analysis of a large number of serial samples from patients who progressed to sAML. Moreover, the comparison of a control and a validation cohort supports the identification of mechanisms mostly related to evolution to sAML.

Previous studies have documented the clona! evolution and greater clona! heterogeneity during cancer development and, specifically, as MDS evolves to advanced stages and transfarms into sAML,[8,9,15,16,20] and according to this fact, the study demonstrates that MDS patients, irrespective of their diagnostic subtype, gain more mutations and a higher VAF during disease progression. Therefare, these results are evidence that progression toward sAML is associated with pronounced genomic instability and a heavy mutational burden.

The mechanisms of disease evolution showed a great heterogeneity.[1,17] However, this study identified faur distinct types of mutational dynamics and their roles during sAML progression. It is of particular note that this study shows a higher incidence of mutations in the cohesin complex and Ras si naling genes than in previously published MDS series, [9,21,22] probably because this cohort consisted only of patients who progressed to sAML. Moreover, increasing mutations (type 1) were mainly faund in genes of this pathway, such as *STAG2,* most of which were loss of function mutations. Previous studies of cohesin mutations had already shown that these mutations could be related to sAML progression, [15,21,22] but the dynamics observed in this study confirm that mutations of cohesin complex genes could be an early event in sAML progression, and the loss of function of these genes could play an important role in the leukemic transfarmation. Similarly, the dynamics of Ras signaling genes observed here, mainly newly acquired mutations (type 3), highlight the importance of this pathway in sAML progression. Earlier studies have described how alterations in Ras pathway genes, such as *NRAS* and *FLT3,* could drive the progression to sAM1[23,25] but these results confirm that they are late events that may drive leukemic transfarmation. Furthermore, this study reveals a significant co-occurrence of mutations in these two pathways and, also, in the main genes of these pathways *(STAG2* and *NRAS"),* excluding the *FLT3-ITD* mutations which could have a different behavior than other mutations in Ras pathway and they could be involved in an independent mechanism of sAML progression.[26,27] Although Walter ,;¡ al.[28] described that *NRAS* and cohesin mutations tend to be mutually exclusive, this work included a low number of patients and, conversely, this co-occurrence was briefly described in another study with a higher number of patients.[21] Thus, this study, dueto detecting in a cohort of sAML-progressing patients, demonstrates that this progressive combination of the cohesin complex, mainly *STAG2,* and Ras signaling mutations, mostly *NRAS,* could play an important role in the progression of MDS to sAML. In addition, the results from the validation cohort confirmed the impact of this co-occurrence on sAML progression not only in the discovery cohort, but also in the validation cohort. Therefare, these findings support a hypothesis of genetic "predisposition", that early mutations shape the future trajectories of clona! evolution from MDS to sAML. Therefare, a new model of genetic evolution could be suggested consisting of cohesin mutations as an early event in the evolution of the disease that trigger to acquire new mutations, mainly Ras signaling mutations. Consequently, this clone expands, driving the disease evolution (Figure 3, model of clona! evolution using Fishplot R package).[29] Recent studies have described that cohesins are involved in DNA damage repair, chromatin accessibility and transcription factor activity3°,[32] and our results show than cohesin-mutated patients displayed a higher number of mutations, thereby cohesins could cause an instability where new mutations are generated, mainly Ras mutations, leading the disease progression. These are not absolute rules and, unfartunately, this novel model does not fully explain the progression to sAML, but it could explain the evolution in 15-20% of ali sAML transfarmations. Moreover, cohesin mutations potentiate the subsequent acquisition of Ras mutations, so these mutations could be used to identify patients whose disease is progressing befare symptoms associated with progression to sAML are manifested.

On the other hand, the mutations that were stable during the evolution (type 4) were faund in splicing and DNA methylation genes. This is in line with the finding of sorne recent reports showing that variants affect these pathways in this steady-state pattern.[16] Moreover, mutations in DNA methylation and RNA splicing pathways are well known to have a heavier mutational burden than those in other?enes, suggesting an early event in MDS development.a, Considering these findings together, these results showed that mutations in these pathways, which have high VAF and are stable during the disease evolution, could be directly involved in MDS pathogenesis (driver role) but not in sAML progression (passenger role). The mutations whose VAF decreased during progression (type 2) were distributed randomly throughout ali the genes without showing a particular pattern. This could be the result of clone sweeping, a previously described event,[15] that is specific to each patient rather than to a specific pattern of each gene or pathway.

Furthermore, a mechanism that could be linked to the evolution of patients who receive disease-modifying treatment befare progression to sAML was detected. Severa! studies have described the mutational dynamics in treated MDS patients and have demonstrated that therapy alters clona! distribution, but the predictive impact of the dynamics is still unclear.[33,37] A significantly higher proportion of newly acquired or increasing mutations in chromatin modifiers at the time of sAML was identified in treated patients. Thus, these results suggest that mutations in chromatin-modifier genes could be related to the evolution of treated patients. However, more studies with larger numbers of patients are required to validate this result.

In summary, MDS progression to sAML is characterized by greater genomic instability, irrespective of the MDS subtypes at diagnosis, and there are faur types of mutational dynamics during the disease evolution, increasing and newly acquired mutations (type 1 and type 3, respectively) being of particular importance. Moreover, a co-occurrence of cohesin complex and Ras signaling mutations could play an important role in the 15-20% of MDS patients who evolved to sAML. With regard to treatment, we faund that mutations in chro-

matin-modifier genes could be related to the evolution of MDS patients who received disease-modifying treatment befare progression to sAML.

Disclosures

No conflicts o(interest to disclose.

Co11trib11tio11s

MMI designed the experiments, pe,formed targeted-deep sequencing experiments, analyzed the data and wrote the pape,; MA designed the experiments, performed whole-exome sequencing experiments, contributed to inte1pret the resu!ts and wrote the pape,;]MHS performed NGS data analysis and contributed to the experiment design; DT analyzed the whole-exome sequencing data; FLC, FR, EL, AMO, MM, JL, JSR, CO,]D, CA, JNR and GMN provided patient samples and clinical information and SSM and CMG contributed to perform the NGS experiments; RB contributed to data analysis, interpretation o(the resu!ts and critically reviewed the manuscript and MDC and]MHR conceived the study, designed the experiments and wrote the manuscript. Ali authors discussed the resu!ts and revised the manuscript.

Ack11owledg111e11ts

The authors would like to thank to Sara González, Irene Rodríguez, Teresa Prieto, Mª Ángeles Ramos, Filomena Corral, Mª Almudena Martín, Ana Díaz, Ana Simón, María del Pozo, Isabel M Isidro, Vanesa Gutiérrez, Sandra Pujante and Mª Ángeles Hernández f,·om the Cancer Research Center o(Salamanca, Spain, for their technical support. We a/so thank to Teresa González and Alba Redondo-Guijo for providing patients samples and clinical information and we are deeply grateful to Miguel Quijada Álamo for his helpful suggestions and personal support.

Ftmdillg

This work was supported by grants f,·om the Spanish Fondo de Investigaciones Sanitarias FIS PI18/01500, PI17/01741, Instituto de Salud Carlos III (ISCIII), Fondo de Investigación Sanitaria (Instituto de Salud Carlos III - Contratos Río Hortega (CM17/0017), European Regional Development Fund (ERDF), Una manera de hacer Europa, European Union Seventh Framework Programme [FP7/2007-2013} under Grant Agreement n°306242-NGS-PTL, SYNtherapy: Synthetic Letha!ity for Personalized Therapy-based Stratification in Acute Leukemia (ERAPERMED2018-275}; ISCIII (AC18/00093), Proyectos de Investigación del SACYL, Gerencia Regional de Salud de Castilla y León: GRS1850/A18, GRS1653/A17, and Centro de Investigación Biomédica en Red de Cáncer (CIBERONC CB16/12/00233}. MMI is supported by a pre-doctoral grant f,-om the junta de Castilla y León, and by the Fondo Social Europeo (JCYL-EDU/556/2019 PhD scholarship) and JMHS is supported by a research grant from Fundación Española de Hematología y Hemoterapia.

References

1. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood. 2013;122(25):4021-4034.
2. Lindsley RC, Ebert BL. Molecular pathophysiology of myelodysplastic syndromes. Annu Rev Pathol. 2013;8:21-47.
3. Tefferi A, Vardiman JW. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. Leukemia. 2008;22(1):14-22.
4. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system far myelodysplastic syndromes. Blood. 2012;120(12):2454-2465.
5. Shukron O, Vainstein V, Kundgen A, Germing U, Agur Z. Analyzing transfarmation of myelodysplastic syndrome to secondary acute myeloid leukemia using a large patient database. Am J Hematol. 2012 87(9):853-860.
6. Montalban-Bravo G, García-Manero G. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. Am J Hematol. 2018;93(1):129-147.
7. Steensma DP, Bennett JM. The myelodysplastic syndromes: diagnosis and treatment. Mayo Clin Proc. 2006;81(1):104-130.
8. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616-3627.
9. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-247.
10. Abaigar M, Robledo C, Benito R, et al. Chromothripsis is a recurrent genomic abnormality in high-risk myelodysplastic syndromes. PLoS One. 2016; 11(10): e0164370.
11. Lindsley RC, Mar BG, Mazzola E, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood. 2015; 125(9):1367-1376.
12. da Silva-Coelho P, Kroeze 11, Yoshida K, et al. Clona! evolution in myelodysplastic syndromes. Nat Commun. 2017;8:15099.
13. Sperling AS, Gibson CJ, Ebert BL. The genet· ics of myelodysplastic syndrome: from clona[haematopoiesis to secondary leukaemia. Nat Rev Cancer. 2017;17(1):5-19.
14. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med. 2012;366(12):1090-1098.
15. Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clona! evolution in myelodysplastic syndromes. Nat Genet. 2017; 49(2):204-212.
16. Kim T, Tyndel MS, Kim HJ, et al. The clona! origins of leukemic progression of myelodysplasia. Leukemia. 2017;31(9):1928-1935.
17. Stosch JM, Heumuller A, Niemoller C, et al. Gene mutations and clona! architecture in myelodysplastic syndromes and changes upan progression to acute myeloid leukaemia and under treatment. Br J Haematol. 2018;182(6):830-842.
18. Ibanez M, Carbonell-Caballero J, Such Ei et al. The modular network structure of the mutational landscape of acute myeloid leukemia. PLoS One. 2018;13(10):e0202926.
19. 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.
20. Reiter JG, Baretti M, Gerold JM, et al. An analysis of genetic heterogeneity in untreated cancers. Nat Rev Cancer 2019;19(11):639-650.
21. Thota S, Viny AD, Makishima H, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. Blood. 2014;124(11):1790-1798.
22. Viny AD, Levine RL. Cohesin mutations in myeloid malignancies made simple. Curr Opin Hematol. 2018;25(2):61-66.
23. Pellagatti A, Roy S, Di Genua C, et al. Targeted resequencing analysis of 31 genes commonly mutated in myeloid disorders in serial samples from myelodysplastic syndrome patients showing disease progression. Leukemia. 2016;30(1):247-250.
24. Takahashi K, Jabbour E, Wang X, et al. Dynamic acquisition of FLT3 ar RAS alterations drive a subset of patients with lower risk MDS to secondary AML Leukemia. 2013;27(10):2081-2083.
2.5. Meggendorfer M, de Albuquerque A, Nadarajah N, et al. Karyotype evolution and acquisition of FLT3 ar RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. Haematologica. 2015;100(12):e487-490.
26. Badar T, Patel KP, Thompson PA, et al. Detectable FLT3-ITD ar RAS mutation at the time of transfarmation from MDS to AMl predicts far very poor outcomes. Leuk Res. 2015;39(12):1367-1374.
27. Xu F, Han R, Zhang J, et al. The Role of FLT3-ITD Mutation on de novo MDS in Chinese population. Clin Lymphoma Myeloma Leuk. 2019;19(2):el07-el15.
28. Walter MJ, ShenD, ShaoJ, etal. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia. 2013;27(6):1275-1282.
29. Miller CA, McMichael J, Dang HX, et al. Visualizing tumor evolution with the fishplot package far R. BMC Genomics. 2016; 17(1):880.
30. Mazumdar C, Shen Y, Xavy S, et al. leukemia-associated cohesin mutants dominantly enfarce stem cell programs and impair human hematopoietic Progenitor differentiation. Cell Stem Cell. 2015; 17(6):675-688.
31. Monda! G, Stevers M, Goode B, Ashworth A, Solomon DA. A requirement far STAG2 in

replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. Nat Commun. 2019;10(1):1686.
32. Meisenberg C, Pinder SI, Hopkins SR, et al. repression of transcription at DNA breaks requires cohesin throughout interphase and prevents genome instability. Mol Cell. 2019; 73(2):212-223.e7.
33. Kuendgen A, Muller-Thomas C, Lauseker M, et al. Efficacy of azacitidine is independent of molecular and clinical characteristics - an analysis of 128 patients with myelodysplastic syndromes ar acute myeloid leukemia and a review of the literature. Oncotarget. 2018; 9(45):27882-27894.
34. Polgarova K, Vargova K, Kulvait V, et al. Somatic mutation dynamics in MDS patients treated with azacitidine indicate clona! selection in patients-responders. Oncotarget. 2017;8(67):111966-111978.
35. Tobiasson M, Mclornan DP, Karimi M, et al. Mutations in histone modulators are associated with prolonged survival during azacitidine therapy. Oncotarget. 2016;7(16):22103-22115.
36. Cabezon M, Bargay J, Xicoy B, et al. Impact of mutational studies on the diagnosis and the outcome of high-risk myelodysplastic syndromes and secondary acute myeloid leukemia patients treated with 5-azacytidine. Oncotarget. 2018;9(27):19342-19355.
37. Dohner H, Dolnik A, Tang L, et al. Cytogenetics and gene mutations influence survival in older patients with acute myeloid leukemia treated with azacitidine or conventional care. Leukemia. 2018; 32(12):2546-2557.

Chapter 2

Incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of very low/low-risk myelodysplastic syndromes

Marta Martín-Izquierdo[1], María Díez-Campelo[2], Javier Sánchez-Real[3], Alberto Hernández-Sánchez[2], Jesús M Hernández-Sánchez[1], Kamila Janusz[4], Félix López-Cadenas[2], Javier Martínez-Elicegui[1], Mar Tormo[5], Marta Megido[6], Carmen Olivier[7], Andrés Madinaveitia-Ochoa[8], Julio Dávila[9], Magdalena Sierra[10], Manuel Vargas[11], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Rocío Benito[1], Jesús M Hernández-Rivas[1,2], Fernando Ramos[3*] and María Abáigar[1*]

1. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
2. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
3. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
4. IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.
5. Hematology, Hospital Clínico Universitario de Valencia, INCLIVA Research Institute, Valencia, Spain.
6. Hematology, Hospital del Bierzo, Ponferrada, León, Spain.
7. Hematology, Hospital General de Segovia, Segovia, Spain.
8. Hematology, Hospital Universitario Miguel Servet, Zaragoza, Spain.
9. Hematology, Hospital Nuestra Señora de Sonsoles, Ávila, Spain.
10. Hematology, Hospital Virgen de la Concha, Zamora, Spain.
11. Hematology, Hospital de Jarrio, Asturias, Spain

*FR and MA contributed equally to this work and both are co-senior authors.

American Journal of Hematology. Second review.

RESEARCH ARTICLE

Incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of very low/low-risk myelodysplastic syndromes

Marta Martín-Izquierdo[1] | María Díez-Campelo[2] | Javier Sánchez-Real[3] | Alberto Hernández-Sánchez[2] | Jesús M Hernández-Sánchez[1] | Kamila Janusz[4] | Félix López-Cadenas[2] | Javier Martínez-Elicegui[1] | Mar Tormo[5] | Marta Megido[6] | Carmen Olivier[7] | Andrés Madinaveitia-Ochoa[8] | Julio Dávila[9] | Magdalena Sierra[10] | Manuel Vargas[11] | Sandra Santos-Mínguez[1] | Cristina Miguel-García[1] | Rocío Benito[1] | Jesús M Hernández-Rivas[1,2] | Fernando Ramos[3*] | María Abáigar[1*]

[1]Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).

[2]University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.

[3]Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.

[4]IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.

[5]Hematology, Hospital Clínico Universitario de Valencia, INCLIVA Research Institute, Valencia, Spain.

[6]Hematology, Hospital del Bierzo, Ponferrada, León, Spain.

[7]Hematology, Hospital General de Segovia, Segovia, Spain.

[8]Hematology, Hospital Universitario Miguel Servet, Zaragoza, Spain.

[9]Hematology, Hospital Nuestra Señora de Sonsoles, Ávila, Spain.

[10]Hematology, Hospital Virgen de la Concha, Zamora, Spain.

[11]Hematology, Hospital de Jarrio, Asturias, Spain.

*FR and MA contributed equally to this work and both are co-senior authors.

Correspondence

Jesús María Hernández-Rivas
IBMCC, CIC Universidad de Salamanca-CSIC, Hospital Universitario de Salamanca
Paseo de San Vicente s/n, 37007, Salamanca, Spain
Tel: +34923291316; Fax: +34923294624
E-mail: jmhr@usal.es

Funding information

Spanish Fondo de Investigaciones Sanitarias FIS PI20/00970, PI18/01500, PI17/01741; Instituto de Salud Carlos III – Contratos Río Hortega (CM17/0017); European Regional Development Fund (ERDF), Grant Agreement nº306242-NGS-PTL; SYNtherapy: Synthetic Lethality for Personalized Therapy-based Stratification in Acute Leukemia (ERAPERMED2018-275), ISCIII (AC18/00093); Erasmus+ programme NEMHESYS_612639-EPP-1-2019-ES-EPPKA2-KA; SACYL: GRS2155/A/2020, GRS1850/A18, GRS1653/A17; CIBERONC CB16/12/00233).
MMI is supported by a predoctoral grant from the Junta de Castilla y León, and by the Fondo Social Europeo (JCYL-EDU/556/2019 PhD scholarship).

Abstract

Mutations in cohesin genes have been described as novel mutations occurring in 10-15% of myelodysplastic syndromes (MDS), suggesting that the cohesins plays an important role in the MDS pathogenesis. However, the clinical impact of cohesin mutations is still in discussion. We studied a cohort of 418 MDS patients and we aimed to elucidate the clinical characteristics of MDS patients harboring cohesin mutations (*STAG1*, *STAG2*, *SMC1A*, *SMC3*, *RAD21* and *CTCF*) with respect to genetic profile, disease phenotype and outcome, as well as to assess the prognostic impact of integration of these mutations with risk factors already defined in MDS. The study identified cohesin mutations in 11.5% of cases, which likely result in loss-of-function. The most frequently mutated gene was *STAG2* (n=34, 69% of all cohesin-mutated patients), followed by *SMC3* (n=7), *RAD21* (n=5), *SMC1A* (n=3) and *CTCF* (n=1). Strong significant co-occurrences between

cohesin and splicing mutations, mainly *SRSF2*, Ras pathway alterations and chromatin modifiers mutations were observed. Interestingly, patients harboring these mutations displayed a poor prognosis disease phenotype, with severe cytopenias and higher number of bone marrow blasts, and a worse clinical outcome, characterized by a shorter leukemia-free survival (1.3 *vs.* 3.5 years, $p<0.0001$) and an earlier leukemic progression, mainly, in very low, low and intermediate-risk IPSS-R patients (multivariate analysis: HR 2.44, 95%CI 1.19-4.98; $p=0.015$). Furthermore, our study proved that cohesin mutations also have impact on the IPSS-R stratification of very low and low-risk patients. Thus, incorporation of mutational profiles into current IPSS-R classification may improve the prognostic stratification of MDS patients.

1 | INTRODUCTION

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic disorders characterized by peripheral blood (PB) cytopenia with dysplastic bone marrow (BM) morphology and an increased risk of progression to secondary acute myeloid leukemia (sAML)(1-4). Over the last several years, the use of high-throughput next generation sequencing has enabled great strides in the MDS molecular characterization identifying some of the mechanisms involved in their pathogenesis such as epigenetic regulation, splicing, transcription, and cohesin complex(5, 6). Moreover, it has also identified that approximately 90% of MDS patients have at least one mutation and that some of them are associated with distinct clinical features and can predict survival(5, 7, 8). However, due to their great genetic and clinical heterogeneity, the molecular data has not yet been included in the current prognostic classifications(2, 9, 10), except the status of *SF3B1* mutation in the updated 2017 WHO classification, which clearly defines a distinct subgroup of MDS with ring sideroblasts(11).

In recent years, mutations in the cohesin complex have been described as novel mutations occurring in 10-15% of MDS patients, suggesting that the cohesin complex plays an relevant role in the MDS pathogenesis(12-14). Cohesin is a multimeric protein complex composed of 2 long structural maintenance proteins, *SMC1A* and *SMC3*, which heterodimerize at the hinge domain, forming a closed loop by binding at the α-klesin end to *RAD21* and adapter proteins *STAG1*/*STAG2*(15-17). This complex is canonically known to align and stabilize sister chromatids during the metaphase regulating chromosome segregation during meiosis and mitosis(16, 18). Interestingly, more recent studies suggest that cohesin genes have additional functions such as double-strand DNA repair, chromatin accessibility and transcriptional activation, specifically in DNA-loop formation between enhancers and promoters bounded by CCCTC-binding factor (*CTCF*), contributing to self-renewal and leukemogenesis (19-23).

However, given the molecular heterogeneity of MDS, the clinical impact of cohesin mutations is still undetermined and further investigation in large patient cohorts with detailed clinical information is required to define their relevance to clinical phenotype,

genetic background and clinical outcome. Therefore, we studied a large cohort of 418 patients diagnosed with MDS and we aimed to elucidate the clinical characteristics of MDS patients harboring cohesin mutations with respect to genetic profile, disease phenotype and clinical outcome, as well as to assess the prognostic impact of integration of these mutations with the MDS risk factors (IPSS and IPSS-R).

2 | MATERIALS AND METHODS

2.1 | Patients

To study the mutational status of cohesin genes (*STAG2*, *STAG1*, *SMC1A*, *SMC3*, *RAD21* and *CTCF*), diagnostic BM or PB samples from 418 patients with MDS from different Spanish institutions were included. Clinical parameters studied included age, gender, diagnosis, bone marrow morphology, blood counts, sAML progression and overall survival (OS). Conventional cytogenetic and FISH analyses were carried out in all samples, as previously described(24-26). Diagnoses were established according to the 2008 and 2017 World Health Organization criteria and detailed analysis of clinical parameters and cytogenetic findings was performed to facilitate risk stratification, according to the International Prognostic Scoring System (IPSS) and the revised IPSS (IPSS-R). The median follow-up of the cohort was 28.1 months (range, 1-187.4 months). During this follow-up 49% of patients died and 29% evolved to secondary acute myeloid leukemia (sAML) (**Supplementary Table S1**).

Written informed consent from all patients was obtained in accordance with the Declaration of Helsinki guidelines, and the studies were approved by the Local Ethics Committee ("Comité Ético de Investigación Clínica, Hospital Universitario de Salamanca").

2.2 | Analysis of cohesin mutations

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).

Targeted-deep sequencing

All gDNA samples underwent targeted-deep sequencing using an in-house custom capture-enrichment panel of 117 genes previously related to the pathogenesis of myeloid malignancies (**Supplementary Table S2**), according to a Nextera sequencing design using *Illumina DesignStudio*. Sequencing libraries were prepared according to the manufacturer´s instructions (*Nextera Rapid Capture Enrichment, Illumina*), using unique barcodes for each sample, multiplexed and sequenced on Illumina NextSeq 500 and MiSeq platforms. Sequencing data was analyzed by applying an in house informatics pipeline as previously described in our group (27). After this analysis, we performed a variant filtering and 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, variants suspected of being sequencing errors after visual inspection with the *IGV browser* and described in our internal artefacts

database 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.

The mean coverage of the sequencing was 537X (p10-p90: 305-929) where 98.4% of target regions were captured at >100X.

2.3 | Statistical analyses

Baseline characteristics were described as frequencies for categorical variables and as the medians and ranges for quantitative variables. Comparisons of proportions between patient subsets were performed by χ^2 test or Fisher´s exact test, as appropriate, while comparisons of quantitative values and ranks were performed by the Student *t*-test and by Mann-Whitney *U* test for means and medians of continuous variables, respectively. Binary logistic regressions were performed to predict disease phenotype based on mutational profile.

The Kaplan-Meier method was used to analyze survival outcomes using the log-rank test to compare the curves of each group. For the analysis of OS, patients who were still alive were censored at the date of last observation and survival time was counted from diagnosis. Leukemia-free survival (LFS) was defined from the time of diagnosis until evolution to sAML or death from any cause, whichever came first. Evolution to sAML was also analyzed with time to leukemic progression (TTL), which was defined as the time from diagnosis to the time of evolution to sAML only. Univariate and multivariate Cox proportional hazard regression analysis was used to identify independent prognostic factors for OS, LFS and TTL. Significant covariates were dichotomized at their medians, and the resulting 2-group log-rank test differences were visualized using Kaplan-Meier methods and Cox models. Two-sided values of $p<0.05$ were considered statistically significant. Analyses were performed with the statistical software package *IBM SPSS Statistics 22.0* (IBM Corp., Armonk, NY) and *GraphPad Prism 6.0* (GraphPad software, San Diego, CA).

3 | RESULTS

The main clinical characteristics of all enrolled patients are summarized in **Supplementary Table S1**. The median age at diagnosis was 74.9 years (range 29.4-92.2), with 58.9% males. According to 2017 WHO classification, most of patients belonged to MDS with multilineage dysplasia with or without ring sideroblasts (MDS-RS-MLD and MDS-MLD, 45.6%), followed by MDS with excess blasts type-2 (MDS-EB2, 16.3%), with excess blasts type-1 (MDS-EB1, 14.7%) and MDS with isolated del(5q) (10.1%). By IPSS-R classification, most patients had very low (26.8%) and low (44.4%) risk, 15.3% had intermediate risk while the remaining 13.6% had high-risk MDS (high and very high risk). Regarding cytogenetics, 84.5% of the series harbored normal karyotype or clonal alterations of very good or good risk.

3.1 | Frequency and distribution of cohesin mutations

Sequencing analysis of the global cohort identified 52 mutations in cohesin genes in 48 patients (11.5%) (**Supplementary Table S3**). The most frequently mutated gene was *STAG2* (n=34, 69% of all cohesin-mutated patients), followed by *SMC3* (n=7), *RAD21* (n=5), *SMC1A* (n=3) and *CTCF* (n=1). Cohesin mutations were almost always mutually exclusive, although >1 mutated gene was detected in 2 patients (**Figure 1A and 1B**).

Figure 1

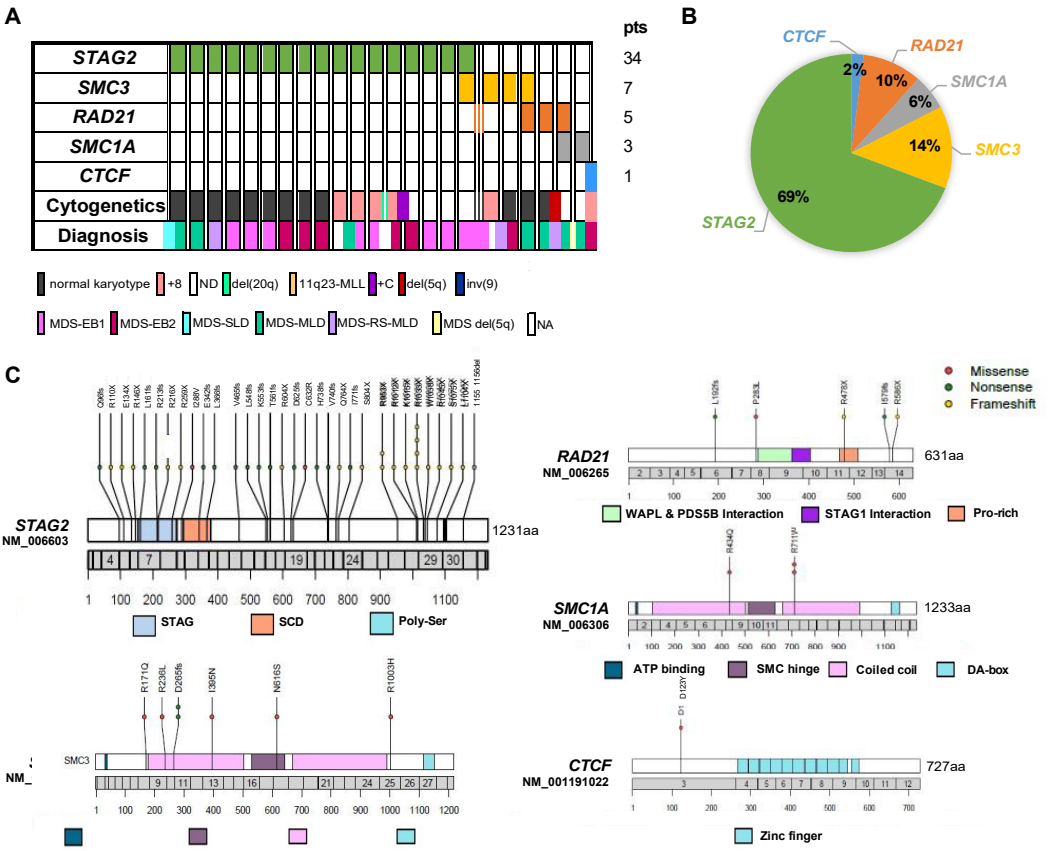


FIGURE 1: Genetic alterations of the cohesin complex in myelodysplastic syndromes. A) Landscape of mutations identified in a 418 MDS patients' cohort. Cytogenetic alterations of cohesin-mutated patients are represented by color: grey, normal karyotype; pink, trisomy 8, green, deletion of 20q; yellow, 11q23-MLL rearrangement; purple, gain of C chromosome; red, deletion of 5q; navy blue, inversion of chromosome 9; and white, not determined. Also, WHO 2017 diagnosis are represented by color: pink, MDS-EB; sky blue, MDS-SLD; green, MDS- MLD; lilac, MDS-RS-MLD; light yellow, MDS del(5q); and white, not available. **B) Distribution of mutations in cohesin family identified in 418 MDS patients.** The most commonly mutated gene was *STAG2*, representing the 69% of all cohesin-mutated patients. **C) Location and type of mutations in the cohesin genes (*STAG2*, *SMC3*, *SMC1A*, *RAD21*, *STAG1* and *CTCF*) in 418 MDS patients.** No mutations hotspots were identified. Different types of mutations are represented by color: red, missense mutations; green, nonsense mutations; and yellow, frameshift mutations. The amino acid changes are referred to using their three-letter abbreviations.
Abbreviations: pts, patients; MDS-EB1, myelodysplastic syndromes with excess blasts type 1; MDS-EB2, myelodysplastic syndromes with excess blasts type 2; MDS-SLD, myelodysplastic syndromes with single lineage dysplasia; MDS-MLD, myelodysplastic syndromes with multilineage dysplasia; MDS-RS-MLD, myelodysplastic syndromes with ring sideroblasts and multilineage dysplasia; MDS del(5q), syndrome myelodysplastic with isolated del(5q); ND, not determined; NA, nota available.

All cohesin mutations were heterozygous, except for mutations in *STAG2* and *SMC1A*, due to these genes are located on chromosome X and result in the absence of expression of a wild-type allele. Of note, *STAG2* mutations were more frequent in males (n=28, hemizygous) than females (n=8, heterozygous) while, in *SMC1A*, mutations in 1 female (heterozygous) and in 2 males (hemizygous) were observed. Interestingly, no hotspot mutation was identified in any of the genes and most of them were loss-of-function type mutations (39 out of 52, 75%), including nonsense (n=22) and frameshift (n=17) mutations. In particular, 92% of *STAG2*

mutations were nonsense (n=20) or frameshift (n=13) and 80% of *RAD21* mutations were loss-of-function type. On the other hand, missense mutations were more frequently found in *SMC3*, with only 2 out of 7 were frameshift (**Figure 1C**).

3.2 | Mutational profile of cohesin mutated patients

In order to characterize the genetic background and profile of cohesin mutated patients, we compared the median number of mutations and mutational incidence of other MDS-related genes in patients harboring cohesin mutations (from now on referred to as "cohesin^{-MUT}") versus patients with no mutations in cohesin genes (from now on referred to as "cohesin^{-WT}").

The cohesin^{-MUT} patients presented a median of four mutations (p10-p90: 2-8), whereas the cohesin^{-WT} patients had a median of two mutations (p10-p90: 0-4), showing a significantly higher number of mutations at cohesin^{-MUT} patients (p<0.0001) (**Figure 2A**). In

Figure 2

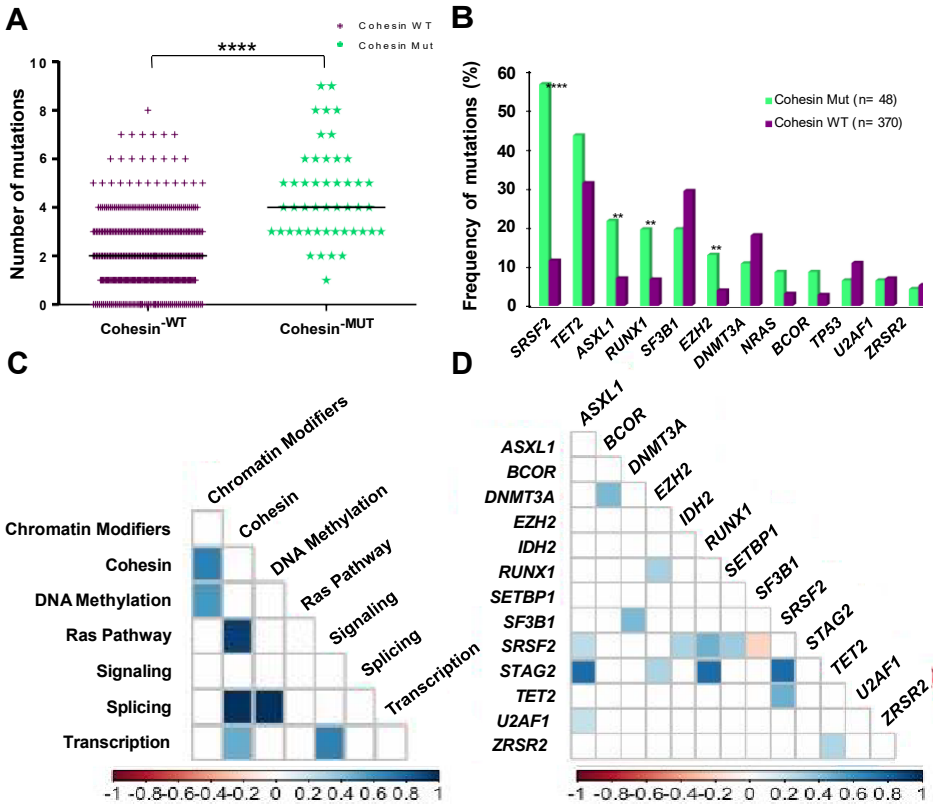


FIGURE 2: Co-occurring somatic mutations in cohesin mutated patients. A) Differences in the number of mutations between cohesin-mutated patients (cohesin^{-MUT}) and patients with no mutations in cohesin genes (cohesin^{-WT}). These graphs show a statistically significant increase in the median of the number of mutations in cohesin mutated patients (2 *vs*. 4 mutations, p<0.0001). Individual as well as median values are depicted. **B) Mutational frequencies and associations in the MDS cohort according to the presence of cohesin mutations.** Significant p-values are annotated with asterisks. Cohesin^{-MUT} patients showed higher mutation frequencies of *SRSF2* (54.2% *vs.* 11.1%, p<0.0001), *ASXL1* (20.8% *vs.* 6.8%, p=0.0031), *RUNX1* (18.8% *vs.* 6.5%, p=0.0074) and *EZH2* (12.5% *vs.* 3.8%, p=0.0185) genes. **C) Co-occurrence of cohesin mutations and other MDS-related pathways** and **D) the co-occurrence of the most frequent cohesin gene *STAG2* and other MDS-related genes in 418 MDS patients.** Strong significant co-occurrences between cohesin and splicing mutations (p<0.0001), Ras pathway alterations (p=0.001) and chromatin modifiers (p=0.007) were detected in our MDS cohort. In this sense, strong significant co-occurrences between *STAG2* and *SRSF2* and *ASXL1* were also detected (p<0.0001, both).
Abbreviations: ****, p<0.0001; **, p<0.01.

addition, cohesin^{-MUT} patients showed higher mutation frequencies of *SRSF2* (p<0.0001), *ASXL1* (p=0.0031), *RUNX1* (p=0.0074) and *EZH2* (p=0.0185) genes (**Figure 2B**).

Furthermore, with this same aim, gene mutations were analyzed for co-occurrence regarding individual genes and the functional pathways they are implicated in. Interestingly, a strong significant co-occurrence was observed between mutations in cohesin and in splicing genes (p<0.0001), Ras pathway alterations (p=0.001) and chromatin modifiers (p=0.007) (**Figure 2C**). In this sense, a robust significant co-occurrence was found between mutations in *STAG2* and *SRSF2* splicing gene (p<0.0001), as well as with *ASXL1* chromatin modifier (p<0.0001) (**Figure 2D**).

Figure 3

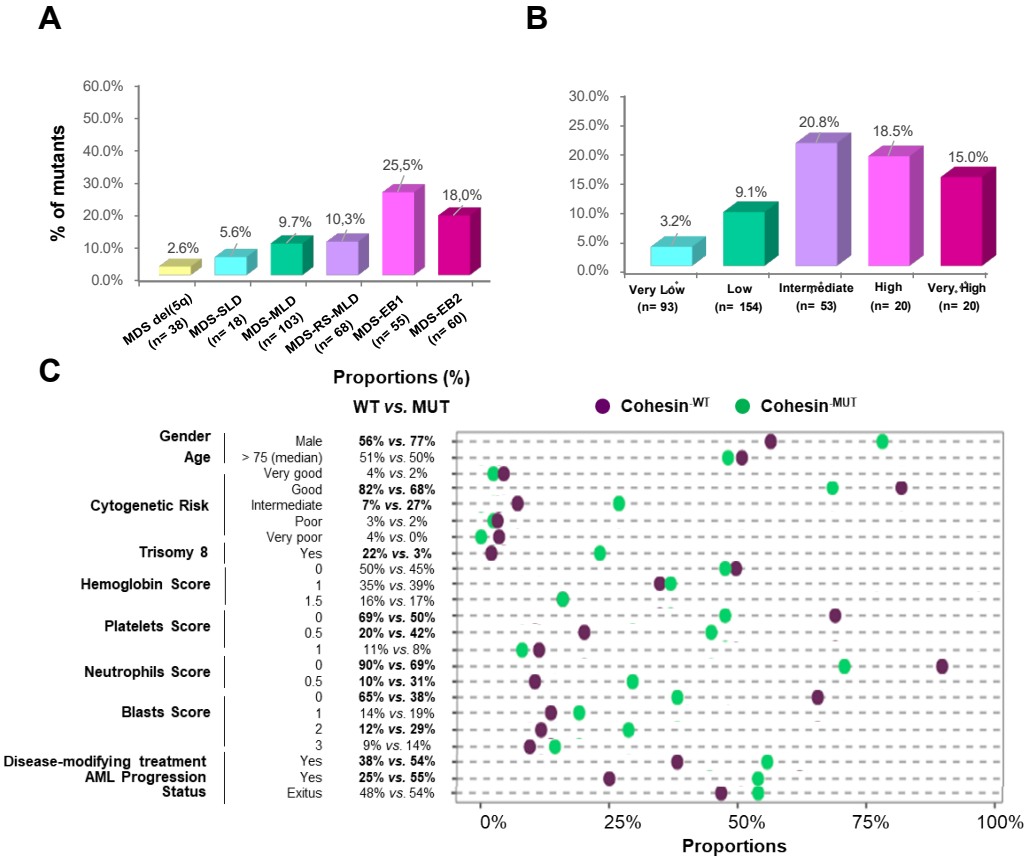


FIGURE 3: Disease phenotype of patients with cohesin family gene mutations. A) Frequency of cohesin gene mutations within each category of the WHO 2017 classification. A higher frequency of cohesin mutations was found in excess blasts subtypes (p=0.004), where somatic alterations were detected in 25.5% of MDS-EB1 and 18.0% of MDS-EB2. **B) Frequency of cohesin family gene mutations within each category of the IPSS-R.** Cohesin^{-MUT} patients were more likely to be associated with intermediate and high-risk IPSS-R categories (p=0.001), where mutations were detected in 20.8% of intermediate-risk and 18.5% of high-risk IPSS-R patients. **C) Clinical and biological characteristics of patients with mutations in the cohesin genes within the entire cohort (n=418).** This graph shows the comparison of the proportions of cohesin^{-WT} versus cohesin^{-MUT} patients regarding gender, age, cytogenetic risk, presence of trisomy 8, hemoglobin levels, platelets and neutrophils cell counts, blasts in bone marrow, need for disease-modifying treatment, AML progression and vital status. For the establish of the cytogenetic, hemoglobin, platelets, neutrophils and blast categories, the scores of IPSS-R calculation were used. The statistically significant proportions are represented in bold (gender, p=0.006; cytogenetic risk, p=0.001; trisomy 8, p<0.0001; platelets, p=0.009; neutrophils, p=0.0008; blasts, p=0.0014; need for disease-modifying treatment, p=0.049; and AML progression, p<0.0001).

3.3 | Disease phenotype of patients harboring cohesin mutations

We analyzed the clinical and biological characteristics of patients with mutations in the cohesin genes within the entire cohort (n = 418) (**Table 1**). Regarding WHO 2017 classification, a higher frequency of cohesin mutations was found in excess of blasts subtypes (p=0.004), where somatic alterations were detected in 25.5% of MDS-EB1 and 18.0% of MDS-EB2 (**Figure 3A)**. In addition, cohesin^{-MUT} patients were more likely to be associated with intermediate and high-risk IPSS-R categories (p=0.001) (**Figure 3B**). Normal cytogenetics was present in 52% of cohesin^{-MUT} patients, a proportion comparable to cohesin^{-WT} patients. However, among patients with an abnormal cytogenetics, cohesin mutations were associated with intermediate risk alterations (cytogenetic-risk score 2, p=0.001), especially with the presence of trisomy 8 (22% cohesin^{-MUT} *vs.* 3% cohesin^{-WT}), showing a significant co-occurrence of these two events (p<0.0001) (**Figure 3C**).

Additionally, cohesin^{-MUT} patients showed a higher incidence of traditional poor prognostic markers, such as lower platelets, absolute neutrophils and white blood cells counts in PB (p=0.009, p=0.0008 and p=0.016, respectively), and a higher number of BM blasts (4% vs. 1.4%, p=0.0014) (**Supplementary Figure S1**). Moreover, a higher proportion of patients displaying cohesin gene mutations received disease-modifying treatment (54% *vs.* 38%, p=0.049) and showed a higher rate of progression to sAML (55% *vs.* 25%, p<0.0001) (**Figure 3C**) (**Table 1**).

Taking into account that *SRSF2* mutations were found to be strongly enriched in cohesin^{-MUT} patients, all aforementioned clinical characteristics (blasts, hemoglobin, platelets, neutrophils, cytogenetics, AML progression and status) were assessed by a logistic regression in order to identify if any of them could be directly associated with the co-occurrence of *SRSF2* and cohesin mutations. Thereby, this co-occurrence was associated solely with AML progression (p=0.017; odds ratio [OR] 6.8, 95% confidence interval [CI] 1.4-32.6) (**Supplementary Table S4**).

3.4 | Clinical outcome of MDS patients harboring cohesin mutations

Median follow-up time for all patients was 28.1 months (range, 1-187.4 months). First, we analyzed the impact on clinical outcome in the entire cohort considering all cohesin gene mutations as a whole. In this sense, cohesin^{-MUT} patients showed a trend towards shorter OS (median: 3.1 *vs.* 5.2 years; Hazard ratio [HR] 1.5, 95% CI 1.0-2.3, p=0.069), a significantly shorter LFS (median: 1.3 *vs.* 3.5 years; HR 2.1, 95% CI 1.4-3.1, p<0.0001) and a shorter TTL (median: 1.4 *vs.* 8.3 years; HR 3.3, 95% CI 2.0-5.5, p<0.0001) (**Figure 4A**). Next, we evaluated the prognostic impact of each cohesin gene separately in the global cohort. Due to these analyses were limited by the small number of mutated patients per gene, it was only performed for *STAG2* and *SMC3*, which were mutated in more than five patients. Kaplan-Meier curve analysis showed that *STAG2* had a negative influence on both OS (p=0.022), LFS (p<0.0001) and TTL (p<0.0001), whereas *SMC3* had no influence on neither OS (p=0.574) nor LFS (p=0.919), although a trend on TTL was observed (p=0.072) (**Supplementary Figure S2**).

Furthermore, we studied the prognostic impact of the co-occurrence of cohesin mutations and *SRSF2* mutations on OS, LFS and TTL. Therefore, we considered the following subgroups of patients: (i) cohesin-*SRSF2* double mutated patients (double

TABLE 1: Clinical and biological characteristics of cohesin^{-MUT} vs. cohesin^{-WT} patients

Variables	Cohesin^{-MUT} (n = 48)	Cohesin^{-WT} (n= 370)	*p*-value
Age at diagnosis, median (range, years)	73 (34-88)	75 (29-92)	0.757
Gender, n (%)			**0.006**
Male	37 (77.1%)	209 (56.5%)	
Female	11 (22.9%)	161 (43.5%)	
WHO 2017 Classification at diagnosis, n (%)			**0.004**
MDS-SLD	1 (2.3%)	17 (5.1%)	
MDS-MLD	10 (22.7%)	93 (28.1%)	
MDS-RS-SLD	0	21 (6.3%)	
MDS-RS-MLD	7 (15.9%)	61 (18.4%)	
MDS-EB1	14 (31.8%)	41 (12.4%)	
MDS-EB2	11 (25.0%)	50 (15.1%)	
MDS with isolated del(5q)	1 (2.3%)	37 (11.2%)	
MDS unclassifiable	0	11 (3.3%)	
IPSS stratification, n (%)			**<0.0001**
Low	6 (15.0%)	164 (50.5%)	
Intermediate 1	25 (62.5%)	103 (31.7%)	
Intermediate 2	7 (17.5%)	44 (13.5%)	
High	2 (5.0%)	14 (4.3%)	
IPSS-R stratification, n (%)			**0.001**
Very low	3 (8.3%)	90 (28.9%)	
Low	14 (38.9%)	140 (45.0%)	
Intermediate	11 (30.6%)	42 (13.5%)	
High	5 (13.9%)	22 (7.1%)	
Very high	3 (8.3%)	17 (5.5%)	
Cytogenetic risk, n (%)			**0.001**
Very good	1 (2.4%)	15 (4.4%)	
Good	28 (68.3%)	277 (81.7%)	
Intermediate	11 (26.8%)	24 (7.1%)	
Poor	1 (2.4%)	11 (3.2%)	
Very poor	0	12 (3.5%)	
Blood count at diagnosis			
Hemoglobin level, g/dL, median (range)	9.7 (5.5-14.8)	9.9 (3.8-15.4)	0.686
Platelet count, x10^9/L, median (range)	100.5 (2.0-314.0)	161.0 (4-1067)	**0.009**
ANC, x10^9/L, median (range)	1.2 (0.1-56.0)	2.0 (0.1-11.8)	**0.0008**
WBC, x10^9/L, median (range)	3.6 (1.6-8.4)	4.1 (1.2-16.0)	**0.016**
Bone marrow study at diagnosis			
% blasts, median (range)	4.0 (0.0-19.0)	1.4 (0.0-19.9)	**0.0014**
% ring sideroblasts, median (range)	0.0 (0.0-48.0)	0.0 (0.0-100.0)	0.583
Disease-modifying treatment, yes, n (%)	23 (53.5%)	127 (37.9%)	**0.049**
sAML progression, yes, n (%)	21 (55.3%)	65 (24.9%)	**<0.0001**
Status, deceased, n (%)	26 (54.2%)	172 (47.9%)	0.415

Note. Significant values are shown in bold. Percentage are based on the number of non-missing values.
Abbreviations: MDS-SLD, syndrome myelodysplastic with single lineage dysplasia; MDS-RS-SLD, syndrome myelodysplastic with ring sideroblasts and single lineage dysplasia; MDS-MLD, syndrome myelodysplastic with multi-lineage dysplasia; MDS-RS-MLD, syndrome myelodysplastic with ring sideroblasts and multi-lineage dysplasia; MDS-EB1, myelodysplastic syndrome with excess blasts type-1; MDS-EB2, myelodysplastic syndrome with excess blasts type-2; IPSS, International Prognostic Scoring System; IPSS-R, International Prognostic Scoring System Revised; WBC, white blood cell; ANC, absolute neutrophil count; sAML, secondary acute myeloid leukemia.

mutants); (ii) *SRSF2* single mutated patients (*SRSF2*$^{-MUT}$); (iii) cohesin single mutated patients (cohesin^{-MUT}); and (iv) cohesin-*SRSF2* wild-type (cohesin^{-WT} + SRSF2^{-WT}). Kaplan-Meier survival analysis showed that the median OS of the double mutants was shorter than *SRSF2*$^{-MUT}$, cohesin^{-MUT} and cohesin^{-WT} + SRSF2^{-WT} patients (18.3 *vs.* 45.1 *vs.* 40.1 *vs.* 64.6 months, respectively, p=0.031) (**Supplementary Figure S3**). Moreover, the LFS as well as the TTL of the double mutants was shorter than *SRSF2*$^{-MUT}$ and cohesin^{-WT} + SRSF2^{-WT} patients, but similar to cohesin^{-MUT} patients (median LFS: 13.8 *vs.* 27.4 *vs.* 45.4 *vs.* 17.7 months, respectively, p<0.0001; median TTL: 14.5 *vs.* 56.6 *vs.* 99.6 *vs.* 18.6 months, p<0.0001) (**Figure 4B**).

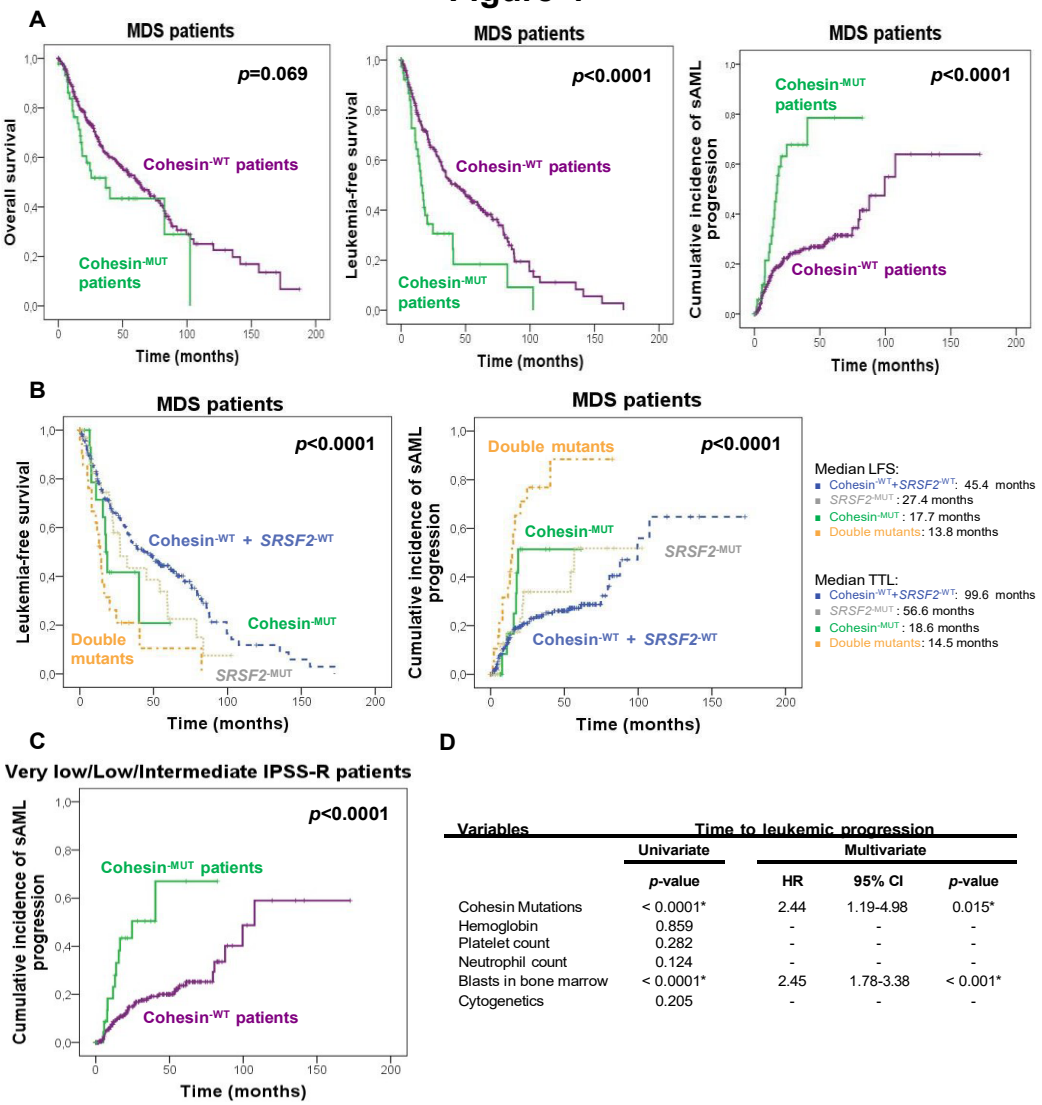
To further study the prognostic value of cohesin mutations, OS, LFS and TTL analyses were carried out for each IPSS-R group, separately. Interestingly, IPSS-R very low/low/intermediate-risk patients harboring cohesin mutations displayed a significant worse outcome, showing a shorter LFS (median: 24.7 *vs.* 53.8 months, p=0.017) (**Supplementary Figure S4**) and an earlier sAML progression than cohesin^{-WT} patients (median TTL: 24.7 *vs.* 107.7 months, p<0.0001) (**Figure 4C**). However, no significant differences in OS were observed (p=0.218) (**Supplementary Figure S4**). Within this subset of patients, additional univariate and multivariate *Cox* regression analyses were performed on TTL, by assessing the presence of cohesin mutations and the IPSS-R prognostic variables BM blasts, hemoglobin, platelets, neutrophils and cytogenetic. Thereby, we identified that cohesin mutations, together with BM blasts, were significantly associated with a shorter time to sAML progression (HR 2.44, 95% CI 1.19-4.98; p=0.015) (**Figure 4D**).

Finally, in view of these results, we further assessed the prognostic value of the cohesin mutations on the IPSS-R stratification. Due to limited number of cohesin^{-MUT} patients in the very low-risk and in the very high-risk groups, the very low and low-risk patients were considered as one category, as well as the very high and high-risk patients. Of note, the presence of these mutations stratified very low/low-risk IPSS-R patients into two groups with completely distinct outcome regarding OS, LFS and TTL (p<0.0001, both). In fact, very low/low-risk patients who harbored cohesin mutations displayed shorter OS, LFS and TTL than cohesin^{-WT} patients, but a nearly identical outcome to that of those of the next higher IPSS-R risk category, intermediate-risk (**Figure 5A**). In contrast, cohesin mutations had no impact on intermediate-risk patients when compared to those of high/very high-risk (**Figure 5B**). Similarly, we investigated the impact of cohesin mutations on clinical outcomes in MDS patients according to the presence of excess of blasts by WHO classification, MDS-EB versus the remaining categories, except MDS unclassifiable which were excluded from the analyses. It is noteworthy that, in the cohort of patients without excess of blasts (MDS-no-EB), cohesin mutations stratified these patients in two different groups regarding OS, LFS and TTL (p=0.003, p<0.0001 and p<0.0001, respectively); cohesin^{-MUT} patients had shorter OS, LFS and TTL than cohesin^{-WT}, but comparable to those with excess of blasts (**Supplementary Figure S5**).

4 | DISCUSSION

Herein, we have analyzed a well-characterized cohort of 418 patients with MDS demonstrating that the presence of cohesin mutations defined a subset of patients displaying a specific mutational profile and a poor prognosis disease phenotype. Moreover, we have proved that cohesin mutations had a negative impact on OS, LFS and TTL in MDS

Figure 4



Variables	Time to leukemic progression			
	Univariate	Multivariate		
	p-value	HR	95% CI	*p*-value
Cohesin Mutations	< 0.0001*	2.44	1.19-4.98	0.015*
Hemoglobin	0.859	-	-	-
Platelet count	0.282	-	-	-
Neutrophil count	0.124	-	-	-
Blasts in bone marrow	< 0.0001*	2.45	1.78-3.38	< 0.001*
Cytogenetics	0.205	-	-	-

FIGURE 4: Prognostic impact of the cohesin mutations on overall survival (OS), leukemia-free survival (LFS) and time to leukemic progression (TTL). A) Kaplan-Meier curves for OS, LFS and TTL in MDS patients according to the presence/absence of cohesin gene mutation. Cohesin^{-MUT} patients showed a trend towards shorter OS (median: 3.1 *vs.* 5.2 years; Hazard ratio [HR] 1.5, 95% CI 1.0-2.3, *p*=0.069), a significantly shorter LFS (median: 1.3 *vs.* 3.5 years; HR 2.1, 95% CI 1.4-3.1, *p*<0.0001) and a shorter TTL (median: 1.4 *vs.* 8.3 years; HR 3.3, 95% CI 2.0-5.5, *p*<0.0001). **B) Kaplan-Meier curves for LFS and TTL in the entire MDS cohort according to the presence of: cohesin-*SRSF2* double mutations, *SRSF2* single mutations (*SRSF2*$^{-MUT}$), cohesin single mutations (cohesin^{-MUT}), and cohesin-*SRSF2* wild type (cohesin^{-WT} + *SRSF2*$^{-WT}$).** Double mutants showed a shorter LFS and a shorter TTL than *SRSF2*$^{-MUT}$ patients, but similar to cohesin^{-MUT} patients (median LFS: 13.8 *vs.* 27.4 *vs.* 17.7 months, respectively, *p*<0.0001; median TTL: 14.5 *vs.* 56.6 *vs.* 18.6 months, *p*<0.0001). **C) Kaplan-Meier curves for TTL in very low/low/intermediate-risk IPSS-R patients according to the presence/absence of cohesin gene mutations.** IPSS-R very low/low/intermediate-risk patients harboring cohesin mutations displayed a significant earlier progression to sAML than cohesin^{-WT} patients (median: 24.7 *vs.* 107.7 months, *p*<0.0001). **D) Univariate and multivariate analysis (*Cox* regression) for the TTL in very low/low/intermediate-risk IPSS-R patients.** Hemoglobin, platelets, neutrophils, blasts in bone marrow and cytogenetics were included as categorical variable using the already established IPSS-R thresholds for these prognostic variables. **p*-value <0.05 was considered significant and HR>1 or <1 indicate an increased or decreased risk, respectively, of an event for the first category listed. Cohesin mutations were the sole factor significantly associated with an earlier progression to sAML (HR 2.44, 95% CI 1.19-4.98; *p*=0.015), together with BM blasts.

Abbreviations: OS, overall survival; LFS, leukemic free survival, LFS; HR, hazard ratio; CI, confidence interval; n.s, not significant.

patients, mainly, in very low, low and intermediate-risk IPSS-R patients.

Previous studies have reported that cohesin mutations occur in 10-15% of MDS patients, reaching to 20% of high-risk MDS or sAML patients, being most of these mutations nonsense and frameshift types (12-14, 28). According to this fact, we have demonstrated that somatic mutations in the cohesin genes occurred in 11.9% of MDS patients, being *STAG2* the most frequently mutated gene, that no mutation hotspot was identified, and most mutations were loss-of-function type. Therefore, these results evidenced that the loss of function of cohesin complex is a common event in MDS patients, as occurred in other myeloid neoplasms, such as AML.

Additionally, this study has expanded the current knowledge of the mutational profile of cohesin^{-MUT} MDS patients. We have identified that cohesin^{-MUT} patients showed a significantly higher number of mutations and higher mutation frequencies in other additional genes such as *SRSF2, ASXL1, RUNX1*, *EZH2* and *NRAS*, usually associated with worse outcome(5-7). Furthermore, we found a strong significant co-occurrence between cohesin and splicing mutations, specifically *SRSF2*, Ras pathway alterations and chromatin modifiers mutations. Previous studies have described some associations between cohesin and other pathways, such as RNA splicing and chromatin regulation, and genes, like *SRSF2*, *RUNX1* and *ASXL1*, in myeloid malignancies (8, 12, 29, 30), but our

Figure 5

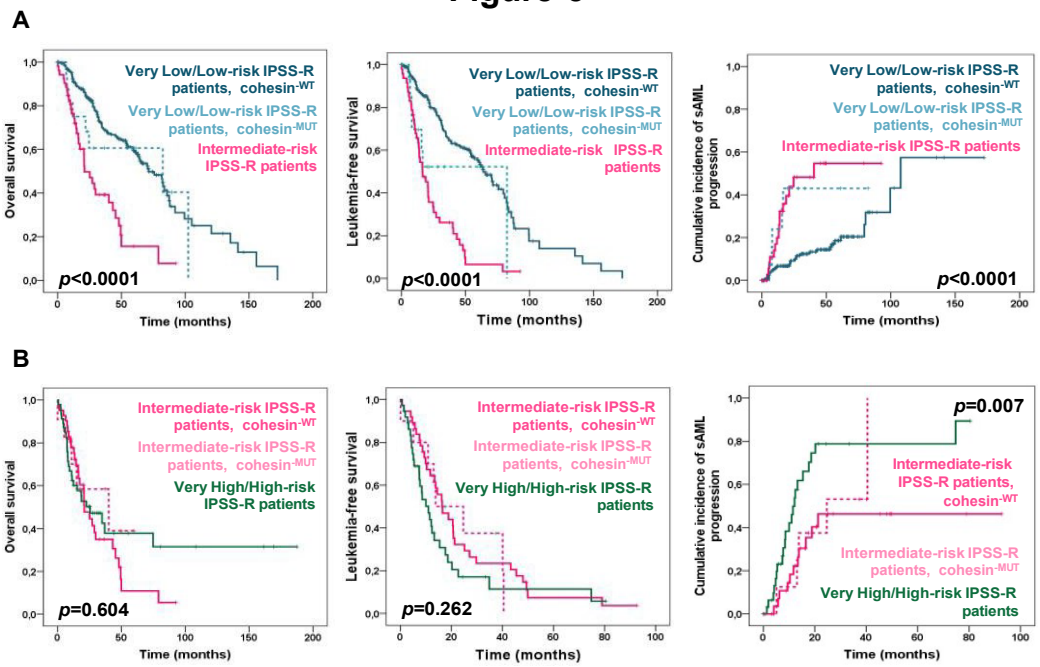


FIGURE 5: Overall survival, leukemia-free survival and time to sAML progression according to the IPSS-R risk categories and cohesin mutational status. A) Kaplan-Meier curves for overall survival (OS), leukemic free survival (LFS) and time to sAML progression (TTL) in very low/low-risk MDS patients according to the presence or absence of cohesin mutations, versus all intermediate-risk patients and B) in intermediate-risk MDS patients according to the cohesin mutational status versus, all high/very high-risk patients. The presence of cohesin mutations stratified very low/low-risk IPSS-R patients into two groups with completely distinct outcome, displaying very low/low-risk patients who harbored cohesin mutations a shorter OS (p<0.0001), LFS (p<0.0001) and TTL (p<0.0001) than cohesin^{-WT} patients, but a nearly identical outcome to that of those of the next higher IPSS-R risk category, intermediate-risk. In contrast, cohesin mutations had no impact on intermediate-risk patients when compared to those of high/very high-risk.

results confirm these events in a cohort of specifically MDS patients. Todisco *et al.* (30) have also reported that the co-occurrence between *STAG2* and *SRSF2* is associated with blast phenotype and, in this line, we have demonstrated that this combination is clearly impacting on the clinical outcome of *SRSF2* mutated patients, with double mutants displaying a shorter OS, LFS and TTL than *SRSF2* single mutated patients.

Moreover, we have identified that the presence of cohesin mutations defined a subset of MDS patients showing a disease phenotype of poor prognosis, characterized by lower platelet, absolute neutrophil and white blood counts in PB, a higher number of blasts in BM and a higher rate of progression to sAML. Similar to other studies(12, 13, 30), cohesin^{-MUT} MDS patients were more likely to be associated with the diagnostic subtypes of excess of blasts, according to WHO 2017, and intermediate and high-risk IPSS-R categories.

The canonical function of cohesin complex is the stabilization of sister chromatids during the metaphase(16, 18). However, recent studies have described that cohesin genes in myeloid malignancies are involved in additional functions such as double-strand DNA repair, chromatin accessibility and transcriptional activation, contributing to self-renewal and leukemogenesis (17, 20, 21). In agreement with this, the majority of patients showing mutations in cohesin genes in our cohort displayed a normal karyotype, supporting that cohesin genes are not involved in destabilization of chromosomal integrity in MDS, but rather alternative mechanisms such as transcriptional control and DNA repair. Interestingly, among patients with an abnormal karyotype, cohesin mutations were associated with trisomy 8 (12, 31). However, due to the small number of patients with this combination, analyses of clinical impact were limited and larger studies are required to study this co-occurrence in deep.

To our knowledge, there have been few studies integrating cohesin mutations data into IPSS-R and studying the effect of cohesin mutations and their co-occurrence with other mutations on the clinical outcome of MDS patients (12, 28, 30, 32). In this work, we comprehensively analyzed the impact of cohesin mutations on OS, LFS and TTL not only in the entire MDS cohort, but also in SRSF2-mutated patients' subgroup, demonstrating that cohesin mutated patients had a shorter overall survival and an earlier progression to sAML. Likewise, we further studied the negative impact of these mutations on OS, LFS and TTL for each IPSS-R group and we confirmed this outcome specially on time to leukemic progression in very low, low and intermediate-risk patients, where cohesin mutations were associated in the multivariate analysis with an earlier progression to sAML, together with BM blasts, the morphology-based metrics that defines sAML. Furthermore, we assessed the impact of these mutations on the IPSS-R stratification and we have proved that the presence of cohesin mutations could stratify the IPSS-R very low/low-risk patients into two groups with distinct outcome, being the OS, LFS and TTL of very low/low-risk patients who harbored cohesin mutations similar to the next higher IPSS-R risk category, namely intermediate-risk patients. However, this study could not identify this same negative impact in very high and high-risk patients, probably because other poor prognosis clinical characteristics mask the effect of cohesin mutations.

In summary, cohesin mutations, which likely result in loss of function, are recurrent genetic anomalies in MDS patients. Patients harboring these mutations display a specific mutational profile, a poor prognosis disease phenotype and a worse clinical outcome, characterized by

a higher rate of AML evolution and an earlier progression to sAML. These mutations also have impact on the IPSS-R stratification of very low and low-risk patients. Thus, as this study has demonstrated, mutational data could be useful in the monitoring of patients and the prediction of disease evolution and the incorporation of mutational profiles into current IPSS-R classification may improve the prognostic stratification of MDS patients. Consequently, an integrative study of mutations, alone or in combination, and clinical characteristics in a large series of MDS patients is necessary to define a new system of prognostic classification, the molecular IPSS-R.

ACKNOWLEDGMENTS

The authors would like to thank to Sara González, Irene Rodríguez, Teresa Prieto, Mª Ángeles Ramos, Filomena Corral, Mª Almudena Martín, Ana Díaz, Ana Simón, María del Pozo, Isabel M Isidro, Vanesa Gutiérrez, Sandra Pujante and Mª Ángeles Hernández from the Cancer Research Center of Salamanca, Spain, for their technical support.

COMPETING INTERESTS

The authors declare no competing interests regarding the publication of this article.

AUTHOR CONTRIBUTIONS

MMI designed the experiments, performed targeted-deep sequencing experiments, analyzed the data and wrote the paper. MDC designed the experiments, contributed to interpret the results and wrote the paper. JSR and AHS contributed to clinical analyses and interpret the results, JMHS and KJ performed NGS data analysis and JME contributed to statistical analyses. FLC, MT, MM, CO, AMO, JL, MS and MV provided patient samples and clinical information and SSM and CMG contributed to perform the NGS experiments. RB and JMHR contributed to data analysis, interpretation of the results and critically reviewed the manuscript and FR and MA conceived the study, designed the experiments and wrote the manuscript. All authors discussed the results and revised the manuscript.

DATA AVAILABILITY STATEMENT

Data available on request from the authors. The data that support the findings on this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Tefferi A, Vardiman JW. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. Leukemia. 2008;22(1):14-22.
2. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120(12):2454-2465.
3. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood. 2013;122(25):4021-4034.
4. Lindsley RC, Ebert BL. Molecular pathophysiology of myelodysplastic syndromes. Annu Rev Pathol. 2013;8:21-47.
5. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616-3627.
6. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-247.
7. Malcovati L, Papaemmanuil E, Ambaglio I, et al. Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia. Blood. 2014;124(9):1513-1521.
8. Hou HA, Tsai CH, Lin CC, et al. Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. Blood Cancer J. 2018;8(4):39.

9. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997;89(6):2079-2088.
10. Malcovati L, Germing U, Kuendgen A, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. J Clin Oncol. 2007;25(23):3503-3510.
11. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-2405.
12. Thota S, Viny AD, Makishima H, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. Blood. 2014;124(11):1790-1798.
13. Kon A, Shih LY, Minamino M, et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet. 2013;45(10):1232-1237.
14. Viny AD, Levine RL. Cohesin mutations in myeloid malignancies made simple. Curr Opin Hematol. 2018;25(2):61-66.
15. Gruber S, Haering CH, Nasmyth K. Chromosomal cohesin forms a ring. Cell. 2003;112(6):765-777.
16. Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. Annu Rev Genet. 2009;43:525-558.
17. Seitan VC, Merkenschlager M. Cohesin and chromatin organisation. Curr Opin Genet Dev. 2012;22(2):93-100.
18. Wassmann K. Sister chromatid segregation in meiosis II: deprotection through phosphorylation. Cell Cycle. 2013;12(9):1352-1359.
19. Mazumdar C, Shen Y, Xavy S, et al. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell Stem Cell. 2015;17(6):675-688.
20. Meisenberg C, Pinder SI, Hopkins SR, et al. Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability. Mol Cell. 2019;73(2):212-223.
21. Mondal G, Stevers M, Goode B, Ashworth A, Solomon DA. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. Nat Commun. 2019;10(1):1686.
22. Rao SSP, Huang SC, Glenn St Hilaire B, et al. Cohesin Loss Eliminates All Loop Domains. Cell. 2017;171(2):305-320.
23. Wendt KS, Yoshida K, Itoh T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008;451(7180):796-801.
24. Shaffer LG, McGowan-Jordan, J., Schmid, M. 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. 2013.
25. 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.
26. 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.
27. Martin-Izquierdo M, Abaigar M, Hernandez-Sanchez JM, et al. Co-occurrence of cohesin complex and Ras signaling mutations during progression from myelodysplastic syndromes to secondary acute myeloid leukemia. Haematologica. 2020.
28. Thol F, Bollin R, Gehlhaar M, et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. Blood. 2014;123(6):914-920.
29. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-2221.
30. Todisco G, Creignou M, Galli A, et al. Co-mutation pattern, clonal hierarchy, and clone size concur to determine disease phenotype of SRSF2(P95)-mutated neoplasms. Leukemia. 2020.
31. Drevon L, Marceau A, Maarek O, et al. Myelodysplastic syndrome (MDS) with isolated trisomy 8: a type of MDS frequently associated with myeloproliferative features? A report by the Groupe Francophone des Myelodysplasies. Br J Haematol. 2018;182(6):843-850.
32. Gu S, Xia J, Tian Y, Zi J, Ge Z. A novel scoring system integrating molecular abnormalities with IPSS-R can improve the risk stratification in patients with MDS. BMC Cancer. 2021;21(1):134.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

Chapter 3

RAS-pathway mutations are associated with a worse clinical outcome in myelodysplastic syndromes patients

Marta Martín-Izquierdo[1], Fernando Ramos[2], María Abáigar[1], Jesús M Hernández-Sánchez[1], Teresa González[1,3], Félix López-Cadenas[3], Miguel Quijada-Álamo[1], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Eva Lumbreras[1], Kamila Janusz[4], Mónica del Rey[1], Sofía Toribio-Castelló[1], Jesús M Hernández-Rivas[1,3], María Díez-Campelo[3*] and Rocío Benito[1*]

1. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
2. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
3. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
4. IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.

*MDC and RB contributed equally to this work and both are co-senior authors.

Cancers. Submitted.

Article

RAS-pathway Mutations Are Associated with a Worse Clinical Outcome in Myelodysplastic Syndromes Patients

Marta Martín-Izquierdo [1], Fernando Ramos [2], María Abáigar [1], Jesús M Hernández-Sánchez [1], Teresa González [1,3], Félix López-Cadenas [3], Quijada-Álamo M [1], Sandra Santos-Mínguez [1], Cristina Miguel-García [1], Eva Lumbreras [1], Kamila Janusz [4], Mónica del Rey [1], Sofía Toribio-Castelló [1], Jesús M Hernández-Rivas [1,3], María Díez-Campelo [3*] and Rocío Benito [1*]

1. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
2. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
3. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
4. IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.

*MDC and RB contributed equally to this work and both are co-senior authors.

Mutations in genes of RAS signaling pathway have been described in 5-10% of myelodysplastic syndromes (MDS) patients and their activation plays an important role in the MDS pathogenesis. Nevertheless, the clinical impact of *RAS*-pathway mutations is controverted. We used NGS to characterize the *RAS*-pathway mutations (*NRAS*, *KRAS*, *HRAS*, *FLT3*, *KIT*, *CBL*, *EGFR*, *NF1*, *PTPN11*, *PTPN1*, *BRAF* and *G3BP1*) in 418 MDS patients and we analyzed correlations between these mutations and the clinical characteristics of MDS patients with respect to genetic profile, disease phenotype and clinical outcome in order to assess the prognostic impact of these mutations. The study identified *RAS*-pathway mutations in 11.2% of MDS patients, being the most commonly mutated gene *NRAS* (n=15), followed by *NF1* (n=11), *KRAS* and *CBL* (n=6), and most of them were subclonal events. Interestingly, patients harboring *RAS*-pathway mutations displayed a specific mutational profile, characterized by the co-occurrence between *RAS* and cohesin mutations, and a poor prognosis phenotype, with presence of cytopenias and excess of blasts and a higher rate of death, despite disease-modifying treatment. Moreover, patients with *RAS*-pathway mutations showed a worse clinical outcome, characterized by a shorter LFS and TTL and, above all, *RAS*-pathway mutations were significantly associated in both the univariate and multivariate analysis with a shorter OS in the entire MDS cohort (HR 1.71, 95% CI 1.04-2.81; *p*=0.033) and in IPSS-R very low/low-risk patients (HR 1.86, 95% CI 1.01-3.43, *p*=0.045). Therefore, the incorporation of mutational data into current classifications may improve the prognostic stratification of MDS patients.

Keywords: RAS-pathway mutations, myelodysplastic syndromes, prognostic value, clinical impact

Correspondence:

Jesús María Hernández-Rivas
IBMCC, CIC USAL-CSIC, Hospital Universitario de Salamanca
Paseo de San Vicente s/n,
37007, Salamanca, Spain
Tel: +34923291316
E-mail: jmhr@usal.es

1. Introduction

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic neoplasms characterized by ineffective hematopoiesis, bone marrow (BM) dysplasia in at least one myeloid lineage, peripheral blood (PB) cytopenia and a high risk of progression to secondary acute myeloid leukemia (sAML) (1-5). Over the last decade, our knowledge about the MDS mutational landscape has been continuously improved due to the implementation of next generation sequencing (NGS) methodologies. In fact, 78 to 90% of MDS patients carry at least one genomic abnormality and mutations were detected in several functional pathways: RNA splicing, epigenetic regulation, transcription, cohesin and signal transduction, the later one mainly represented by *RAS* signaling activating mutations (6-9). In addition, some of these mutations has been associated with distinct clinical features and shown to be of prognostic value (6, 10-18). Altogether has expanded our understanding of the MDS pathogenesis. Nevertheless, lack of clinical validation has restrained the diagnostic and prognostic categorization of patients regarding sequencing data.

In recent years, mutations in *RAS* oncogenes (*NRAS*, *KRAS* and *HRAS* genes) have been commonly found in a diverse range of human hematological malignancies, including myelofibrosis, chronic myelomonocytic leukemia (CMML), acute myeloid leukemia (AML) and MDS (19-23). These mutations has been described to lead to increased signaling through the RAS/RAF/MEK and RAS/PI3K pathways, which has been associated with increased proliferation of hematopoietic progenitor cells (24, 25). Besides direct activating mutations in *RAS* oncogenes, RAS-dependent signaling can be increased in response to mutations in other genes that are upstream or downstream of RAS in the signaling cascade (mutations in *FLT3*, *KIT*, *CBL*, *EGFR* and mutations in *BRAF*, respectively) or due to mutations in genes that are involved in RAS signaling regulation (mutations in *NF1*, *PTPN11*, *PTPN1*) (26, 27). In the case of MDS, *RAS*-pathway mutations have been reported in 5-10% of cases, being *NRAS* the most frequently mutated gene, and patients harboring these mutations were most frequently diagnosed with MDS with excess of blasts and belonging to high-risk IPSS-R categories (28-30). The prognostic impact of *RAS*-pathway mutations has been extensively studied in patients with myeloproliferative neoplasms and AML, but limited information is available regarding the prognosis of MDS patients with these mutations. Moreover, their clinical impact is controverted because some studies have described that *RAS*-pathway mutations do not appear to be correlated with poor prognosis (30, 31), while other recent works have described that they may promote cell proliferation and be associated with a high-risk of sAML progression and worse prognosis (32-34).

Hence, the prognostic impact of *RAS*-pathway mutations is still undefined and further investigation using large patient cohorts with detailed clinical information and data from deep sequencing studies is required to define the relevance of these mutations to disease phenotype and prognostic impact. With this aim, in the present work we studied a cohort of 418 patients diagnosed with MDS and we searched correlations between *RAS*-pathway mutations and the clinical characteristics of MDS patients with respect to genetic profile, disease phenotype and clinical outcome in order to assess the prognostic impact of these mutations.

2. Materials and Methods

2.1 Patients

Diagnostic BM or PB samples of 418 patients diagnosed with MDS from different Spanish institutions were analyzed at the time of diagnosis. Written informed consent from all patients was obtained in accordance with the Declaration of Helsinki guidelines and the studies were approved by the Local Ethics Committee ("Comité Ético de Investigación Clínica, Hospital Universitario de Salamanca").

In all 418 cases detailed clinical data, including age, gender, diagnosis, BM morphology, blood counts, time to sAML progression and overall survival (OS), was collected. Additionally, conventional cytogenetic and FISH analyses were performed in all samples, as previously described (35, 36). Diagnoses were established according to the 2008 and 2017 World Health Organization (WHO) criteria (5, 37). Regarding risk stratification, patients were classified according to the International Prognostic Scoring System (IPSS) and the revised IPSS (IPSS-R) (38, 39). The median follow-up of the cohort was 28.1 months (range, 1-187.4 months). During this follow-up 49% of patients died and 29% evolved to sAML (**Supplementary Table S1**).

2.2 Analysis of RAS-pathway mutations

Genomic DNA from PB and BM samples were extracted using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's standard protocol and quantity and quality for sequencing was tested by a Qubit® 2.0 Fluorometer system (Life Techonologies, Carlsbad, CA, USA), a nanodrop spectrophotometer (ND-1000, NanoDrop Technologies, Wilmington, DE, USA) and a TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA).

In all samples, targeted-deep sequencing was performed by a capture-enrichment approach using an in-house custom panel of 117 genes previously related to myeloid malignancies (**Supplementary Table S2**). Sequencing libraries were carried out according to the manufacturer's instructions (*Nextera Rapid Capture Enrichment, Illumina*) and were sequenced on Illumina NextSeq 500 and MiSeq platforms. Sequencing data was analyzed by applying an in house informatics pipeline and a subsequent stringent variant filtering as previously described by our group (40). The mean coverage of the sequencing was 540X (p10-p90: 305-929), where 98.8% of target regions were captured at >100X.

In this study we specifically focused on the study of the mutational status of Ras signaling pathway. Because of their functional significance, *RAS*-pathway mutations were considered as somatic mutations in the following genes included in the panel: *NRAS*, *KRAS*, *HRAS*, *FLT3*, *KIT*, *CBL*, *EGFR*, *NF1*, *PTPN11*, *PTPN1*, *BRAF* and *G3BP1*.

2.3 Statistical analyses

Baseline characteristics were described as frequencies for categorical variables and as medians and ranges for quantitative variables. χ^2 test or Fisher´s exact tests were used to compare categorical variables, while Student *t* and Mann-Whitney *U* tests were performed to compare means and medians of continuous variables, respectively. Regarding survival analyses, the *log-rank* test and the *Kaplan-Meier* method were used to compare the curves of each group. For the OS analysis, survival time was counted from diagnosis and patients who were still alive were censored at the date of last follow-up. For analysis of sAML evolution, date of progression replaced the date of death, as the uncensored variable, for estimating time to leukemic

progression (TTL), and leukemia-free survival (LFS) was defined as the time from diagnosis until sAML progression or death from any cause, whichever came first. Univariate and multivariate *Cox* proportional hazard regression analyses were carried out to identify independent prognostic factors for OS, LFS and TTL. Two-sided values of $p<0.05$ were considered statistically significant. Analyses were performed with the statistical software package *IBM SPSS Statistics 22.0* (IBM Corp., Armonk, NY) and *GraphPad Prism 6.0* (GraphPad software, San Diego, CA).

3. Results

3.1 Frequency and features of RAS-pathway mutations in MDS patients

Sequencing analyses were performed in the whole cohort of 418 well-characterized MDS patients, where 45.6% were patients with MDS with multilineage dysplasia with or without ring sideroblasts (MDS-RS-MLD and MDS-MLD), 16.3% with MDS with excess of blasts type-2 (MDS-EB2), 14.7% type-1 (MDS-EB1) and 10.1% with MDS with isolated del(5q) (**Supplementary Table S1**). Overall, 59 mutations in any of the studied *RAS*- pathway genes (**Supplementary Table S2**) were detected in 47 patients (incidence=11.2%) (**Supplementary Table S3**). The most frequently mutated gene was *NRAS*, being mutated in 15 patients, followed by *NF1* (n=11), *KRAS* and *CBL* (n=6, each), *PTPN11* and *KIT* (n=5, each), *BRAF* (n=3), *EGFR* (n=2) and *FLT3* (n=1) (**Figure 1A**). Of note, 3 patients harbored >1 mutations in the *NRAS* gene with different variant allele frequencies (VAF), suggesting that these mutations could be present in at least two different clones. Similarly, 1 patient carried two mutations in the *PTPN11* gene, but in this case both mutations were present in the same VAF, suggesting that both may be in the same clone. *RAS*-pathway mutations were predominantly missense (53/59, 89.8%), while frameshift mutations were mainly found in the *NF1* gene. No clear mutational hotspots were identified in the majority of genes, except for the *NRAS* gene, where common hotspots were the aminoacid positions G12 and G13, corresponding to the GTP binding domain (**Supplementary Table S3**).

Regarding the analyses of median VAF of all validated somatic mutations, we found that the median VAF of *RAS*-pathway mutations was 11.27% (p25-p75: 5.60%-29.55%), while the median VAF from the remaining detected mutations was 33.83% (p25-p75: 17.49%-43.83%), thereby *RAS*-pathway mutations showed a lower median VAF ($p<0.0001$) (**Figure 1B**). Moreover, to deepen the study of the signaling pathways, the remaining detected mutations were classified into their functional pathways. The median VAF of *RAS*-pathway mutations was significantly lower than the median VAF of mutations in splicing, epigenetic regulation (DNA methylation and chromatin modification), transcription, cohesins and other signaling pathways ($p<0.0001$, each). Thus, *RAS*-pathway mutations were present at lower mutational burden than mutations in other genes, suggesting they were subclonal events (**Figure 1C**).

3.2 Genetic profile of MDS patients harboring RAS-pathway mutations

In this study, all the genes included in the sequencing panel were analyzed, thereby we were able to characterize the genetic background of those patients harboring *RAS*-pathway mutations. Similarly, patterns of preferential co-occurrences and mutual exclusivity mechanisms of functional pathways were analyzed to look for potential cooperativity among *RAS*-pathway mutations and the other genes analyzed in the whole cohort.

Figure 1

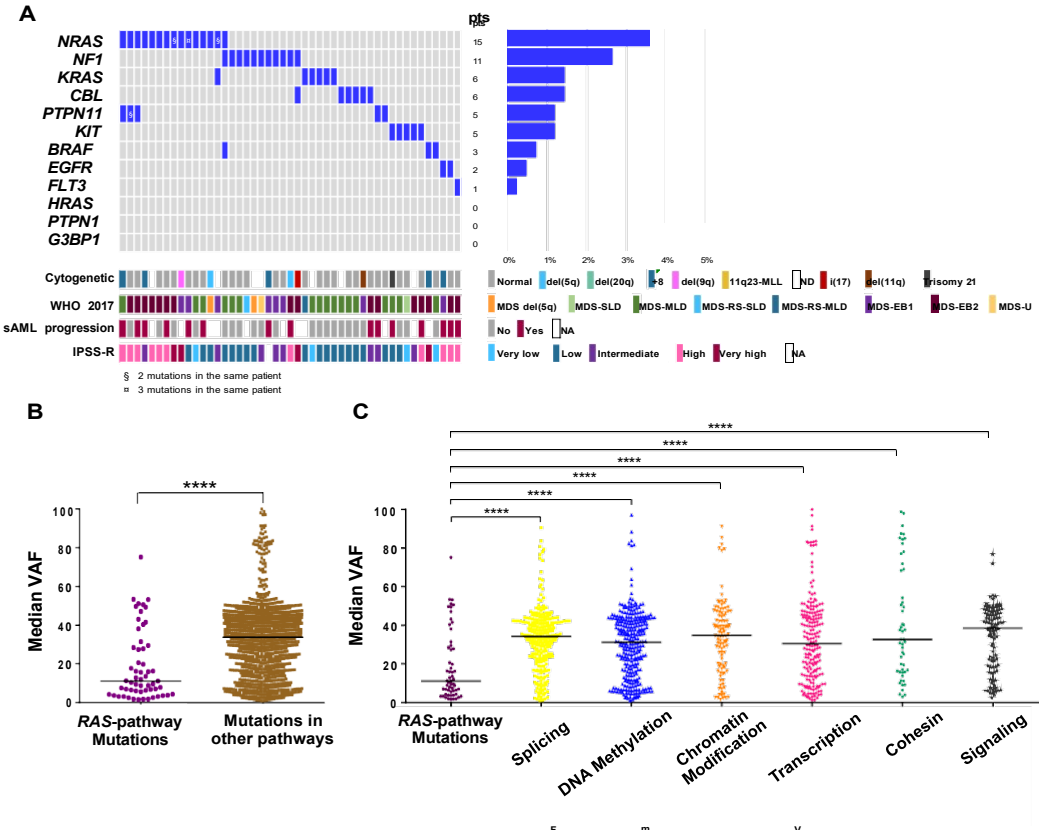


FIGURE 1: Genetic alterations of RAS signaling pathway. A) Landscape of mutations identified in 418 MDS patients. Frequencies of each gene of the RAS pathway are represented in right panel, being *NRAS* the most commonly mutated gene. Cytogenetic alterations of patients harboring *RAS* mutations are represented by color: grey, normal karyotype; blue, deletion of 5q; green, deletion of 20q; dark blue, trisomy 8; pink, deletion of 9q; yellow, 11q23-MLL rearrangement; red, isochromosome 17q; brown, deletion of 11q; dark grey, trisomy 21; and white, not determined. In addition, WHO 2017 diagnosis, sAML progression and IPSS-R categories are represented by color. WHO diagnosis: orange, MDS with isolated deletion of 5q (MDS del(5q)); light green, MDS with single-lineage dysplasia (MDS-SLD); dark green, MDS with multilineage dysplasia (MDS-MLD); sky blue, MDS with ring sideroblasts and single-lineage dysplasia (MDS-RS-SLD); navy blue, MDS with ring sideroblasts and multilineage dysplasia (MDS-RS-MLD); purple, MDS with excess of blasts type-1 (MDS-EB-1); mulberry, MDS with excess of blasts type-2 (MDS-EB-2); light yellow, MDS unclassifiable (MDS-U). sAML progression: grey, patient did not evolve to sAML; magenta, sAML progression; white, not available. IPSS-R categories: light blue, very low-risk; dark blue, low-risk; purple, intermediate-risk; pink, high-risk; magenta, very high-risk. **B-C) Median variant allele frequency (VAF) of somatic mutations in RAS pathway compared to other pathways in a cohort of 418 MDS patients.** This graph shows that the median VAF of *RAS*-pathway mutations was significantly lower than the median VAF of the remaining mutations **(B)** considered as one group (11.27% *vs.* 33.83%, $p<0.0001$), as well as classified into their functional pathways **(C)**: splicing (11.27% *vs.* 34.15%, $p<0.0001$), DNA methylation (11.27% *vs.* 31.33%, $p<0.0001$), chromatin modification (11.27% *vs.* 34.68%, $p<0.0001$), transcription (11.27% *vs.* 30.46%, $p<0.0001$), cohesins (11.27% *vs.* 32.62%, $p<0.0001$) and other signaling pathways (11.27% *vs.* 38.25%, $p<0.0001$). Abbreviations: pts, patients; WHO diagnosis, World Health Organization diagnosis; sAML, secondary Acute Myeloid Leukemia; ND, not determined; NA, not available; VAF, Variant Allele Frequency.

Classifying all detected mutations into the functional pathways they are implicated in, a significant co-occurrence between mutations in RAS pathway and cohesins was observed ($p=0.001$) (**Figure 2A**). Similarly, regarding the mutation frequencies of each analyzed gene in patients with *RAS*-pathway mutations, there was a positive association between *RAS*-pathway mutations and *STAG2* cohesin gene ($p=0.003$), as well as *SRSF2* splicing gene ($p=0.021$), *EZH2*

chromatin modifier (p=0.046) and *NPM1* and *GATA2* transcription factors (p=0.008, both). On the other hand, there was an exclusivity relationship between *RAS*-pathway mutations and *SF3B1* splicing gene (p=0.047) (**Figure 2B**).

3.3 Disease phenotype of MDS patients harboring RAS-pathway mutations

Evaluation of the clinical and biological characteristics of patients with mutations in the RAS pathway was performed (**Table 1**). Statistical analyses showed that MDS patients carrying *RAS*-pathway mutations (from now on referred to as "*RAS*$^{-MUT}$ patients") were more likely to be associated with excess of blasts subtypes (p=0.001), where mutations were detected in 26.2% of MDS-EB-2 (**Figure 3A**), and with high-risk IPSS-R categories (p=0.008), finding mutations in 33.3% of high-risk and 20.0% of very high-risk patients (**Figure 3B**). Regarding cytogenetic risk, it is noted that no *RAS*$^{-MUT}$ patients displayed poor-risk alterations, whereas a significant association between *RAS*-pathway mutations and intermediate-risk cytogenetic alterations was observed (p=0.028). In fact, *RAS*$^{-MUT}$ patients showed a higher proportion of cases harboring trisomy 8 (12.5% *vs.* 4.4%, p=0.031).

Moreover, *RAS*$^{-MUT}$ patients showed a higher incidence of poor prognostic markers, such as lower platelet and neutrophil cell counts in PB (p=0.004 and p=0.008, respectively), and a higher proportion of blasts in BM (p<0.0001). In this sense, they displayed a lower percentage of ring sideroblasts (p=0.006). In addition, a higher proportion of *RAS*$^{-MUT}$ patients received disease-modifying treatment (53.5% *vs.* 37.9%, p=0.049) and died during the follow-up time of the present study (63.8% *vs.* 46.7%, p=0.027), and a trend to a higher rate of progression to sAML was observed (41.7% *vs.* 27.0%, p=0.068) (**Table 1**).

Figure 2

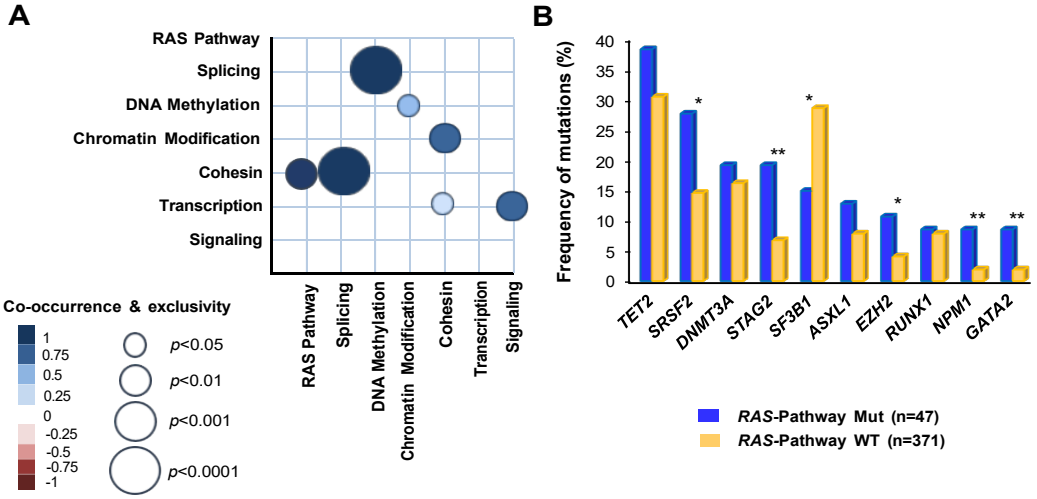


FIGURE 2: Genetic profile of MDS patients harboring *RAS*-pathway mutations. A) Bubble plot of mutual exclusivity and co-occurrence in mutations in MDS-related pathways. This graph shows a significant co-occurrence between *RAS*-pathway and cohesin mutations (p=0.001). **B) Mutational frequencies and associations in a MDS cohort according to the presence of *RAS*-pathway mutations.** Significant p-values are annotated with asterisks. MDS patients harboring *RAS*-pathway mutations showed higher mutation frequencies of *SRSF2* (27.7% *vs.* 14.6, p=0.021), *STAG2* (19.1% *vs.* 6.7%, p=0.003), *EZH2* (10.6 *vs.* 4.0, p=0.046), *NPM1* (8.5 *vs.* 1.9, p=0.008) and *GATA2* (8.5 *vs.* 1.9, p=0.008) genes, while they displayed lower mutation frequency of *SF3B1* gene (14.9 *vs.* 28.6, p=0.047).
Abbreviations: *, p<0.05; **, p<0.01.

TABLE 1: Clinical and biological characteristics of patients harboring *RAS*-pathway mutations compared to patients with wild-type *RAS* genes

Variables	*RAS*$^{-MUT}$ patients (n = 47)	*RAS*$^{-WT}$ patients (n= 371)	*p*-value
Age at diagnosis, median (range, years)	72 (36-88)	75 (29-92)	0.441
Gender, n (%)			0.673
Male	61.7%	58.5%	
Female	38.3%	41.5%	
WHO 2017, n (%)			**0.001**
MDS-SLD	0%	4.9%	
MDS-MLD	29.8%	24.0%	
MDS-RS-SLD	0%	5.7%	
MDS-RS-MLD	6.4%	17.5%	
MDS-EB1	17.0%	12.7%	
MDS-EB2	34.0%	12.1%	
MDS with isolated del(5q)	4.3%	9.7%	
MDS unclassifiable	2.1%	2.7%	
IPSS classification, n (%)			**<0.0001**
Low	21.3%	43.1%	
Intermediate 1	40.4%	29.4%	
Intermediate 2	27.7%	10.2%	
High	2.1%	4.0%	
IPSS-R classification, n (%)			**0.008**
Very low	8.5%	24.0%	
Low	38.3%	36.7%	
Intermediate	12.8%	12.7%	
High	19.1%	4.9%	
Very high	8.5%	4.3%	
Cytogenetic risk, n (%)			**0.028**
Very good	4.3%	3.8%	
Good	66.0%	73.9%	
Intermediate	19.1%	7.0%	
Poor	0%	3.2%	
Very poor	0%	3.2%	
Blood counts			
Hemoglobin level, g/dL, median (range)	9.9 (4.4-15.2)	9.9 (3.8-15.4)	0.877
Platelet count, x10^9/L, median (range)	97.0 (2.0-499.0)	161.0 (4-1067)	**0.004**
ANC, x10^9/L, median (range)	1.5 (0.1-5.1)	2.0 (0.1-56.0)	**0.008**
WBC, x10^9/L, median (range)	3.5 (1.3-16.0)	4.1 (1.2-14.4)	0.148
Bone marrow study			
% blasts, median (range)	4.0 (0.0-18.6)	1.2 (0.0-19.9)	**<0.0001**
% ring sideroblasts, median (range)	0.0 (0.0-30.0)	0.0 (0.0-100.0)	**0.006**
Disease-modifying treatment, yes (%)	53.5%	37.9%	**0.049**
sAML progression, yes (%)	41.7%	27.0%	0.068
Status, deceased (%)	63.8%	46.7%	**0.027**

Note. Significant values are shown in bold.

Abbreviations: WT, wild-type; MDS-SLD, syndrome myelodysplastic with single lineage dysplasia; MDS-RS-SLD, syndrome myelodysplastic with ring sideroblasts and single lineage dysplasia; MDS-MLD, syndrome myelodysplastic with multi-lineage dysplasia; MDS-RS-MLD, syndrome myelodysplastic with ring sideroblasts and multi-lineage dysplasia; MDS-EB1, myelodysplastic syndrome with excess blasts type-1; MDS-EB2, myelodysplastic syndrome with excess blasts type-2; IPSS, International Prognostic Scoring System; IPSS-R, International Prognostic Scoring System Revised; WBC, white blood cell; ANC, absolute neutrophil count; sAML, secondary acute myeloid leukemia.

3.4 Impact of RAS-pathway mutations on overall survival and time to sAML progression

Median follow-up time for all patients was 28.1 months (range, 1-187.4 months), with a median OS of 59.8 months and a median TTL of 87.7 months. Firstly, the impact on clinical outcome in the entire cohort considering all mutations in *RAS*-pathway genes was analyzed. In this regard, univariate analyses showed that *RAS*$^{-MUT}$ patients showed a shorter OS (median: 31.1 *vs.* 64.6 months, Hazard ratio [HR] 1.58, 95% Confidence Interval [CI] 1.05-2.37, *p*=0.028), a shorter LFS (median: 20.4 *vs.* 40.5 months, HR 1.77, 95% CI 1.19-2.63, *p*=0.004) and a shorter TTL (median: 54.2 *vs.* 87.7 months, HR 1.86, 95% CI 1.06-3.25, *p*=0.028) (**Figure 4**). Moreover, in order to evaluate the impact of *RAS*-pathway mutations among all conventional variables with known prognostic value, such as BM blasts, hemoglobin, platelets, neutrophils and cytogenetic (IPSS-R prognostic variables), multivariate *Cox* regression analyses were carried out for OS, LFS and TTL and *RAS*-pathway mutations were independently associated with a shorter OS (HR 1.71, 95% CI 1.04-2.81; *p*=0.033) (**Table 2**).

TABLE 2: Results of Multivariate *Cox* Proportional Hazard Model for overall survival in 418 MDS patients

Overall survival			
Covariate	**Hazard Ratio**	**95% CI**	***p*-value**
***RAS*-pathway mutations**	1.71	1.04-2.81	**0.033**
BM blasts			
Score= 0		Reference	
Score= 1	1.70	1.02-2.81	**0.040**
Score= 2	1.29	0.72-2.30	0.391
Score= 3	1.80	0.97-3.34	0.064
Hemoglobin			
Score= 0		Reference	
Score= 1	1.44	0.95-2.19	0.082
Score = 1.5	1.72	1.05-2.82	**0.032**
Platelets			
Score = 0		Reference	
Score= 0.5	1.66	1.07-2.58	**0.024**
Score= 1	1.29	0.74-2.25	0.370
Neutrophils	0.88	0.48-1.60	0.667
Cytogenetic			
Score= 0		Reference	
Score= 1	1.74	0.69-4.41	0.243
Score= 2	0.81	0.27-2.37	0.694
Score= 3	4.35	1.12-16.87	**0.034**
Score= 4	5.23	1.33-20.54	**0.018**

Hemoglobin, platelets, neutrophils, blasts in bone marrow and cytogenetics were included as categorical variable using the already established IPSS-R thresholds for these prognostic variables. **p*-value <0.05 was considered significant and HR>1 or <1 indicate an increased or decreased risk, respectively, of an event for the first category listed. *RAS*-pathway mutations were significantly associated with a shorter OS (HR 1.71, 95% CI 1.04-2.81; p=0.033).

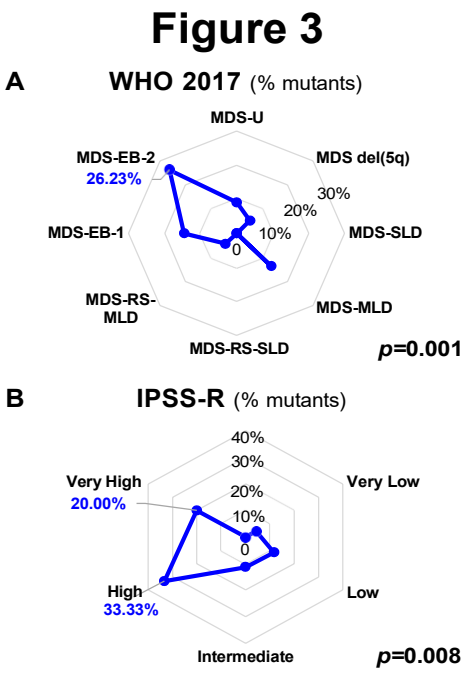


FIGURE 3: Disease phenotype of MDS patients carrying *RAS*-pathway mutations. Frequency of *RAS*-pathway mutations within each category of A) the WHO 2017 classification and B) the IPSS-R stratification. A higher frequency of *RAS*-pathway mutations was found in excess of blasts subtypes (*p*=0.001), where somatic alterations were detected in 26.2% of MDS-EB-2, and high-risk IPSS-R categories (*p*=0.008), finding mutations in 33.3% of high-risk and 20.0% of very high-risk patients.

Next, the prognostic impact of individual *RAS*-pathway genes was assessed in the global cohort. It was only performed for genes that were mutated in more than five patients: *NRAS*, *NF1*, *KRAS* and *CBL*. Survival analyses showed that *NF1* had a negative influence on OS (20.4 *vs.* 62.0 months, $p<0.0001$) and LFS (6.7 *vs.* 36.9 months, $p=0.003$), but no influence on TTL were observed ($p=0.572$), whereas *NRAS* had no influence on OS ($p=0.514$), but a shorter LFS (11.0 *vs.* 36.9 months, $p=0.043$) and a shorter TTL (11.0 *vs.* 99.6 months, $p=0.010$) were observed among patients with *NRAS* mutations (**Supplementary Figure 1**). On the hand, *KRAS* and *CBL* had no influence on neither OS, LFS nor TTL.

Finally, to further study the prognostic value of *RAS*-pathway mutations, the OS, LFS and TTL were analyzed for each IPSS-R group, separately. Due to limited number of RAS^{-MUT} patients in the very low-risk and the very high-risk groups, the very low and low-risk patients were considered as one category, as well as the very high and high-risk patients. Of note, IPSS-R very low/low-risk RAS^{-MUT} patients showed a significant shorter OS than patients without these mutations (32.1 *vs.* 81.6 months, $p=0.009$) (**Figure 5A**), whereas no significant differences in LFS ($p=0.183$) and TTL ($p=0.644$) were observed. Furthermore, this impact on IPSS-R very low/low-risk RAS^{-MUT} patients was confirmed by multivariate *Cox* proportional hazard analyses, being *RAS*-pathway mutations the only factor associated with a shorter OS (HR 1.86, 95% CI 1.01-3.43, $p=0.045$) (**Figure 5B**). By contrast, the presence of *RAS*-pathway mutations had no impact on OS, LFS and TTL of intermediate, high and very high-risk IPSS-R patients.

4. Discussion

RAS signaling pathway plays an important role in proliferation, differentiation and survival of the cell, thereby alterations in its function promote tumoral proliferation. In the case of MDS, the lack of integration of molecular and clinical data has limited analysis of prognostic significance of *RAS*-pathway mutations in patients with MDS. In the present study, we have analyzed a clinically well-characterized cohort of 418 MDS patients by NGS, identifying that the presence of *RAS*-pathway mutations was associated with a specific genetic profile and a disease phenotype of poor prognosis. Moreover, this study has demonstrated that the presence of *RAS*-pathway mutations at diagnosis could define a subset of patients with a poor outcome, displaying a shorter OS.

Figure 4

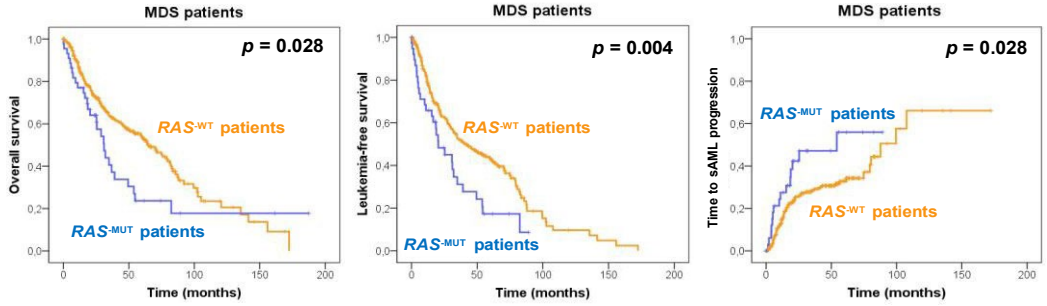


FIGURE 4: Prognostic impact of the *RAS*-pathway mutations on overall survival (OS), leukemia-free survival (LFS) and time to sAML progression (TTL). Kaplan-Meier curves for OS, LFS and TTL in MDS patients according to the presence/absence of *RAS*-pathway mutations. Patients carrying *RAS*-pathway mutations showed a shorter OS (median: 31.1 *vs.* 64.6 months, HR 1.58, 95% CI 1.05-2.37, $p=0.028$), a shorter LFS (median: 20.4 *vs.* 40.5 months, HR 1.77, 95% CI 1.19-2.63, $p=0.004$) and a shorter TTL (median: 54.2 *vs.* 87.7 months, HR 1.86, 95% CI 1.06-3.25, $p=0.028$).

Previous studies have analyzed the mutational status of the *RAS* oncogenes *NRAS* and *KRAS* in MDS patients, describing that these mutations occur in 5-10% MDS cases and *NRAS* was the most frequently mutated gene (28-31). Besides direct activating Ras signaling mutations in *RAS* oncogenes (*NRAS*, *KRAS* and *HRAS*), mutations in other Ras pathway components that are upstream or downstream of signaling cascade or involved in their regulation can increase the Ras-dependent signaling (26, 27). Thus, we have studied the mutational status of not only the *RAS* oncogenes *NRAS*, *KRAS* and *HRAS*, but also other 9 genes involved in RAS pathway (*FLT3*, *KIT*, *CBL*, *EGFR*, *NF1*, *PTPN11*, *PTPN1*, *BRAF* and *G3BP1*).

Somatic mutations in RAS signaling pathway occurred in 11.2% of MDS patients, being *NRAS* the most frequently mutated gene (15/418, 3.6%) and most mutations were missense. In the case of *NRAS* gene, mutation hotspots were identified in the aminoacid positions G12 and G13, which are located in GTP binding domain and they play an important role in protein function (41). Moreover, the median VAF of *RAS*-pathway mutations was lower than median VAF from mutations in other pathways, suggesting these represent sub-clonal events that could be acquired late in the disease course, as previously described (8). Therefore, these results proved that *RAS*-pathway mutations are common events in MDS patients which are acquired late in the disease course and they could lead to the activation of RAS pathway.

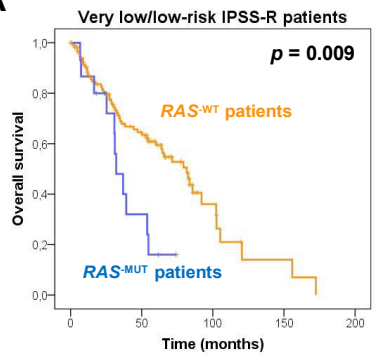
To our knowledge, this study is the first work that assessed the molecular status of the RAS pathway in MDS patients by means of an NGS approach. The abnormal cells in the BM of MDS patients evolve over time giving rise to sub-clonal populations, often coexisting different sub-clones that can retain important clinical significance (42). Previous studies have assessed the mutational status of RAS pathway in MDS patients by PCR (29-31), so mutations at 15-20% level or below might be present, but were not detected by this technique, masking their prognostic significance. In addition, these studies were only focused on the mutational state of *RAS*-pathway genes, hindering the characterization of the mutational profile of MDS patients harboring *RAS*-pathway mutations. Herein, the presence of *RAS*-pathway mutations has been associated with mutations in cohesin genes and, specifically with *STAG2* gene, as well as with other genes, such as *SRSF2*, *EZH2*, *NPM1* and *GATA2*. By contrast, in our study *RAS*-pathway and *SF3B1* mutations were mutually exclusive. Thus, we have demonstrated that *RAS*$^{\text{-MUT}}$ patients display a specific genetic profile characterized mainly by the co-occurrence of *RAS*-pathway and cohesin mutations.

Similar to other studies (28, 30), *RAS*$^{\text{-MUT}}$ patients were most frequently diagnosed with MDS with excess of blasts and with high-risk IPSS-R categories. Additionally, the presence of *RAS*-pathway mutations was associated with cytopenias, such as lower platelet and neutrophil counts in PB, and with a higher number of blasts in BM. Similarly, *RAS*$^{\text{-MUT}}$ patients showed a higher incidence of poor prognostic marker, since a higher proportion of *RAS*$^{\text{-MUT}}$ patients received disease-modifying treatment and deceased. Accordingly, the presence of *RAS*-pathway mutations defined a subset of MDS patients that showed a disease phenotype of poor prognosis.

Furthermore, we comprehensively analyzed the impact of *RAS*-pathway mutations on OS, LFS and TTL. Regarding the prognosis of *RAS*$^{\text{-MUT}}$ patients, limited information is available now and the clinical impact of these mutations is controverted, because some studies have described that *RAS*-pathway mutations could be associated with poor outcomes (32-34), while in other works these mutations do not appear to correlate with a poor prognosis (30, 31). In the entire

Figure 5

A



B

Variables	Overall survival			
	Univariate	Multivariate		
	p-value	HR	95% CI	*p*-value
RAS-pathway Mutations	0.009*	1.86	1.01-3.43	0.045*
Hemoglobin	0.014*	1.35	0.96-1.90	0.087
Platelet count	0.431	-	-	-
Neutrophil count	0.927	-	-	-
BM Blasts	0.705	-	-	-
Cytogenetics	0.967	-	-	-

FIGURE 5: Prognostic impact of the *RAS*-pathway mutations on overall survival (OS) in very low/low-risk IPSS-R patients. A) Kaplan-Meier curve according to the presence/absence of *RAS*-pathway mutations and B) univariate and multivariate analysis (*Cox* regression) for OS in very low/low-risk IPSS-R patients. IPSS-R very low/low-risk patients carrying *RAS*-pathway mutations showed a significant shorter OS than patients without these mutations (32.1 *vs.* 81.6 months, *p*=0.009). For multivariate analysis, hemoglobin, platelets, neutrophils, blasts in bone marrow and cytogenetics were included as categorical variable using the already established IPSS-R thresholds for these prognostic variables. **p*-value <0.05 was considered significant and HR>1 or <1 indicate an increased or decreased risk, respectively, of an event for the first category listed. The negative impact on OS was confirmed by multivariate *Cox* proportional hazard analyses, being *RAS*-pathway mutations the only factor associated with a shorter OS (HR 1.86, 95% CI 1.01-3.43, *p*=0.045).

Abbreviations: HR, hazard ratio; CI, confidence interval.

MDS cohort, we have identified that *RAS*$^{-MUT}$ patients showed a shorter LFS and TTL and, above all, *RAS*-pathway mutations were significantly associated in both the univariate and multivariate analysis with a shorter OS. In addition, the prognostic impact of each *RAS* pathway gene was analyzed, confirming the negative impact of the most frequently mutated genes of RAS pathway, namely *NRAS* and *NF1*. Likewise, we further studied the clinical impact of *RAS*-pathway mutations for each IPSS-R group, identifying that IPSS-R very low/low-risk patients carrying *RAS*-pathway mutations showed a significant shorter OS than patients without these mutations. Hence, the incorporation of *RAS* mutational data could be useful in the monitoring of these patients and in the prediction of disease evolution.

5. Conclusions

In summary, *RAS*-pathway mutations are recurrent genetic events in MDS patients, being acquired late in the disease course. Patients carrying these mutations display a specific mutational profile, highlighting the co-occurrence between *RAS*-pathway and cohesin mutations, and a poor prognosis disease phenotype, with presence of cytopenias and excess of blasts and with a higher rate of death, despite disease-modifying treatment. Moreover, patients harboring *RAS*-pathway mutations show a worse clinical outcome, characterized, above all, by a shorter OS. Therefore, the incorporation of mutational data into current classifications may improve the prognostic stratification of MDS patients.

Supplementary Materials: **Supplementary Table S1:** Main clinical and biological characteristics of the entire cohort included in the study (n=418); **Supplementary Table S2:** Panel of 117 myeloid-related genes used for targeted-deep sequencing; **Supplementary Table S3:** List of all cohesin mutations found by targeted-deep sequencing in 418 MDS patients; **Supplementary Figure S1:** Prognostic influence of mutations in the RAS pathway genes *NRAS* and *NF1*.

Author Contributions: MMI designed the experiments, performed targeted-deep sequencing experiments, analyzed the data and wrote the paper. FR and MA designed the experiments, contributed to interpret the results and wrote the paper. JMHS performed NGS data analysis and TG and FLC provided patient samples and clinical information. SSM and CMG contributed to perform the NGS experiments and MQA, EL, KJ, MR and STC contributed to clinical analyses and interpret the results. JMHR contributed to data analysis, interpretation of the results and critically reviewed the manuscript and MDC and RB conceived the study, designed the experiments and wrote the manuscript. All authors discussed the results and revised the manuscript.

Funding: This work was supported by grants from the Spanish Fondo de Investigaciones Sanitarias FIS PI18/01500, PI17/01741, Instituto de Salud Carlos III (ISCIII), Fondo de Investigación Sanitaria (Instituto de Salud Carlos III – Contratos Río Hortega (CM17/0017), European Regional Development Fund (ERDF), Una manera de hacer Europa, European Union Seventh Framework Programme [FP7/2007-2013] under Grant Agreement nº306242-NGS-PTL, SYNtherapy: Synthetic Lethality for Personalized Therapy-based Stratification in Acute Leukemia (ERAPERMED2018-275); ISCIII (AC18/00093), Proyectos de Investigación del SACYL, Gerencia Regional de Salud de Castilla y León: GRS2155/A/2020, GRS1850/A18, GRS1653/A17, and Centro de Investigación Biomédica en Red de Cáncer (CIBERONC CB16/12/00233).
MMI is supported by a predoctoral grant from the Junta de Castilla y León, and by the Fondo Social Europeo (JCYL-EDU/556/2019 PhD scholarship).

Institutional Review Board Statement: Written informed consent from all patients was obtained in accordance with the Declaration of Helsinki guidelines and the studies were approved by the Local Ethics Committee ("Comité Ético de Investigación Clínica, Hospital Universitario de Salamanca").

Data Availability Statement: Data available on request from the authors. The data that support the findings on this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank to Sara González, Irene Rodríguez, Teresa Prieto, Mª Ángeles Ramos, Filomena Corral, Mª Almudena Martín, Ana Díaz, Ana Simón, María del Pozo, Isabel M Isidro, Vanesa Gutiérrez, Sandra Pujante and Mª Ángeles Hernández from the Cancer Research Center of Salamanca, Spain, for their technical support. We also thank to Javier Sánchez-Real, Mar Tormo, Marta Megido, Carmen Olivier, Andrés Madinaveitia-Ochoa, Julio Dávila, Jorge Labrador, Carlos Aguilar, Juan N Rodríguez, Guillermo Martín-Núñez, Magdalena Sierra and Manuel Vargas for providing patients samples and clinical information.

Conflict of Interest: The authors declare no competing interests regarding the publication of this article.

References

1. Mufti GJ, Bennett JM, Goasguen J, et al. Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. Haematologica. 2008;93(11):1712-1717.
2. Lindsley RC, Ebert BL. Molecular pathophysiology of myelodysplastic syndromes. Annu Rev Pathol. 2013;8:21-47.
3. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood. 2013;122(25):4021-4034.
4. Tefferi A, Vardiman JW. Myelodysplastic syndromes. N Engl J Med. 2009;361(19):1872-1885.
5. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-2405.
6. 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.
7. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-247.
8. Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet. 2017;49(2):204-212.
9. Nagata Y, Maciejewski JP. The functional mechanisms of mutations in myelodysplastic syndrome. Leukemia. 2019;33(12):2779-2794.
10. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011;365(15):1384-1395.
11. Kulasekararaj AG, Smith AE, Mian SA, et al. TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. Br J Haematol. 2013;160(5):660-672.
12. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. Nat Med. 2020;26(10):1549-1556.
13. Chen CY, Lin LI, Tang JL, 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):405-414.
14. Liang S, Zhou X, Pan H, Yang Y, Shi L, Wang L. Prognostic value of DNMT3A mutations in myelodysplastic syndromes: a meta-analysis. Hematology. 2019;24(1):613-622.
15. Thol F, Friesen I, Damm F, et al. Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. J Clin Oncol. 2011;29(18):2499-2506.
16. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood. 2012;119(15):3578-3584.
17. Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011;364(26):2496-2506.
18. Malcovati L, Karimi M, Papaemmanuil E, et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. Blood. 2015;126(2):233-241.
19. Greenberger JS. ras mutations in human leukemia and related disorders. Int J Cell Cloning. 1989;7(6):343-359.
20. Nakagawa T, Saitoh S, Imoto S, et al. Multiple point mutation of N-ras and K-ras oncogenes in myelodysplastic syndrome and acute myelogenous leukemia. Oncology. 1992;49(2):114-122.
21. Ricci C, Fermo E, Corti S, et al. RAS mutations contribute to evolution of chronic myelomonocytic leukemia to the proliferative variant. Clin Cancer Res. 2010;16(8):2246-2256.
22. Tyner JW, Erickson H, Deininger MW, et al. High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients. Blood. 2009;113(8):1749-1755.

23. Buradkar A, Bezerra E, Coltro G, et al. Landscape of RAS pathway mutations in patients with myelodysplastic syndrome/myeloproliferative neoplasm overlap syndromes: a study of 461 molecularly annotated patients. Leukemia. 2021;35(2):644-649.
24. Khan AQ, Kuttikrishnan S, Siveen KS, et al. RAS-mediated oncogenic signaling pathways in human malignancies. Semin Cancer Biol. 2019;54:1-13.
25. Zhang J, Wang J, Liu Y, et al. Oncogenic Kras-induced leukemogeneis: hematopoietic stem cells as the initial target and lineage-specific progenitors as the potential targets for final leukemic transformation. Blood. 2009;113(6):1304-1314.
26. Lavoie H, Therrien M. Regulation of RAF protein kinases in ERK signalling. Nat Rev Mol Cell Biol. 2015;16(5):281-298.
27. Santos FPS, Getta B, Masarova L, et al. Prognostic impact of RAS-pathway mutations in patients with myelofibrosis. Leukemia. 2020;34(3):799-810.
28. Yunis JJ, Boot AJ, Mayer MG, Bos JL. Mechanisms of ras mutation in myelodysplastic syndrome. Oncogene. 1989;4(5):609-614.
29. Constantinidou M, Chalevelakis G, Economopoulos T, et al. Codon 12 ras mutations in patients with myelodysplastic syndrome: incidence and prognostic value. Ann Hematol. 1997;74(1):11-14.
30. Al-Kali A, Quintas-Cardama A, Luthra R, et al. Prognostic impact of RAS mutations in patients with myelodysplastic syndrome. Am J Hematol. 2013;88(5):365-369.
31. Neubauer A, Greenberg P, Negrin R, Ginzton N, Liu E. Mutations in the ras proto-oncogenes in patients with myelodysplastic syndromes. Leukemia. 1994;8(4):638-641.
32. Takahashi K, Jabbour E, Wang X, et al. Dynamic acquisition of FLT3 or RAS alterations drive a subset of patients with lower risk MDS to secondary AML. Leukemia. 2013;27(10):2081-2083.
33. Meggendorfer M, de Albuquerque A, Nadarajah N, et al. Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. Haematologica. 2015;100(12):e487-490.
34. Pellagatti A, Roy S, Di Genua C, et al. Targeted resequencing analysis of 31 genes commonly mutated in myeloid disorders in serial samples from myelodysplastic syndrome patients showing disease progression. Leukemia. 2016;30(1):247-250.
35. 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.
36. 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.
37. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-951.
38. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997;89(6):2079-2088.
39. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120(12):2454-2465.
40. Martin-Izquierdo M, Abaigar M, Hernandez-Sanchez JM, et al. Co-occurrence of cohesin complex and Ras signaling mutations during progression from myelodysplastic syndromes to secondary acute myeloid leukemia. Haematologica. 2020.
41. Nyiri K, Koppany G, Vertessy BG. Structure-based inhibitor design of mutant RAS proteins-a paradigm shift. Cancer Metastasis Rev. 2020;39(4):1091-1105.
42. Murphy DM, Bejar R, Stevenson K, et al. NRAS mutations with low allele burden have independent prognostic significance for patients with lower risk myelodysplastic syndromes. Leukemia. 2013;27(10):2077-2081.

General Discussion

Myelodysplastic syndromes (MDS) are a **highly heterogeneous disease** at genomic and clinical level. Genetically, it is characterized by the presence of a wide variety of genomic and epigenetic abnormalities which result in the deregulation of a large number of biological processes [4,73,186,187]. This biological complexity has shown to impact the development and the clinical evolution of MDS patients, which actually present a highly variable clinical course, ranging from indolent conditions with a near-normal life expectancy spanning years, to more aggressive forms rapidly progressing to sAML [5,6]. Thereby, diagnosis, prognostic and risk stratification of MDS patients and even clinical decision making may be difficult and challenging.

MDS patients present a **high risk of eventually progression to MDS-related acute myeloid leukemia** (sAML), occurring this evolution in approximately a third of cases with a well-known dismal prognosis [32,57]. In addition, one of the goals of the treatment of MDS is preventing the evolution towards an overt sAML, because most of patients who evolved to sAML are resistant to currently available treatments and their long-term survival rate is less than 10% after a couple of years [42,58,125].

Therefore, the understanding of the molecular alterations and mechanisms underlying MDS pathogenesis and the evolution of the disease to more advance stages has become essential to assess patient prognosis and to adopt the best therapeutic decisions, as well as to evaluate novel tailored therapeutic approaches based on the genetic profile of each patient. High-throughput genome-wide and sequencing studies, such as genomic and expression arrays (CGH, SNP and GEP arrays), DNA sequencing (DNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq) and transcriptome sequencing (RNA-seq), have provided great insights into our current understanding of MDS, in relation to their biology and pathogenesis. In particular, since the first human genome was sequenced in 2001 [188,189], the **molecular characterization of genetic disease, and specially MDS, has dramatically improved**, describing that 78 to 90% of MDS patients carry at least one genetic mutation and that more than 40 driver genes are involved in MDS pathogenesis [62,63,71]. In addition, it has recently described that some of these driver genes tend to appear together and they are mutated simultaneously, suggesting a complex structure of gene-gene interactions and extensive subclonal diversification [64,75].

Moreover, **patient's genomic profile critically impacts on clinical phenotype**, prognosis and response to therapy, foregrounding the MDS genetics for predicting clinical course as well as optimizing therapy and management of patients. Thus, *SF3B1* mutations are strongly associated with increased ring sideroblasts and generally predict a favorable prognosis, defining a subset of MDS patients [34,36]. Consequently, the mutational status of *SF3B1* has recently been included as a parameter in WHO 2017 classification [32]. On the other hand, the presence of some mutations, such as *SRSF2*, *U2AF1*, *TP53*, *DNMT3A*, *RUNX1*, *EZH2* and *ASXL1* is associated with a worse clinical outcome [62,82-89].

Nevertheless, nowadays the biological and, consequently, clinical significance of genetic mutations on the development of MDS and during the course of the disease is not completely established. The great inter-patient and intraclonal **heterogeneity** [62-64,71,73], together to the lack of efficient integration of genetics and clinical information, **has restrained the interpretation of the sequencing results** and, thus, the adoption of sequencing data into real-life clinical practice. Therefore, there is a need to carry out studies from bench to bedside and back again.

To date, new high-throughput genomic technologies, such as Next Generation Sequencing (NGS), have enabled to describe some of the molecular alterations and mechanisms underlying the disease progression, which has been defined as a complex process of clonal evolution together with a great clonal heterogeneity [70,135,141]. However, due to this complexity and the few sequential studies including paired samples from the same patients [70,135-137], the clonal evolution have been described individually, without finding common and preferential patterns, and **the mutational dynamics underlying the evolution to sAML remain partially unknown**. To overtake this issue, in the first study from this PhD research, we addressed this critical question by combining NGS tools with the use of a high number of paired samples from patients at diagnosis and after sAML progression in order to better understand how mutations, alone or in combination, could contribute to leukemic transformation.

In this PhD research, we performed a massive screening by whole-exome sequencing (WES) in a selection of paired samples from 20 patients. WES is a helpful tool to study MDS, particularly in unveiling the clonal evolution, due to the high number of assessed regions. However, this methodology required high amounts of DNA input and time-consuming-bioinformatics analysis, as well as some regions were poorly covered, missing some subclonal mutations with low variant allele frequency (VAF). Thereby, WES actually presented limitations to interpret sequencing results and to incorporate them into clinical practice [172-174]. For these reasons, a more **comprehensively study about disease progression was performed** by applying target-deep sequencing (TDS) in a larger cohort of serially collected samples. TDS is a powerful approach that can fulfil the best balance between the accurate identification of targeted events with great sensitivity and the overall cost and data burden for large-scale executions [166], and, thus, it allowed a better characterization of intraclonal heterogeneity, occurred in MDS evolution.

Previous studies described that cancer, in general, is a clonal disease that arises from the accumulation of acquired abnormalities in an individual cell and, specifically, MDS when evolving to advanced stages and progressing to sAML are associated with an increase of clonal heterogeneity [70,135,190]. In this line, our study identified that MDS patients gained more mutations during disease progression and higher VAF were found after sAML evolution, associating the leukemic transformation with a more

complex clonal architecture and a heavy mutational burden due to clonal expansion. Therefore, our results suggest that **pronounced genomic instability is occurring during progression to sAML**, triggering an increase of clonal heterogeneity. Moreover, it has previously described that the number of oncogenic mutations in MDS patients negatively correlates with leukemia-free survival and provides independent prognostic information further to the IPSS-R stratification [62]. Thereby, these clinical results may be explained by the biological results of our study, associating adverse outcome and, specifically earlier sAML progression, with the increase of clonal heterogeneity during disease evolution and, thus, the number of mutations, as well as the increase of their VAF, could be useful to monitor patients before sAML evolution occur.

The mechanisms of MDS progression to sAML show a great heterogeneity [70,135,137,191]; however, our work identified **four patterns of mutational dynamics** during the evolution. Some mutations remain **stable (i)** during the progression with a similar allelic burden at the time of diagnosis and at sAML stage. This type of mutations was found in RNA splicing and DNA methylation genes and they have high VAFs, suggesting that these pathways could be directly involved in early stages of MDS pathogenesis (driver role), such as Clonal Hematopoiesis of Indeterminate Potential (CHIP) [59,142,143], but not in sAML progression (passenger role). Moreover, some mutations whose VAF significantly **decreased at sAML stage (ii)** were identified randomly throughout all the genes, probably as result of clone sweeping, a previously described event [70]. On the other hand, mutations that were initially present at the time of MDS diagnosis and they were also detected at the sAML stage but, in this case, with higher clone size, **(iii) increasing mutations**, and mutations which were **newly acquired (iv)** at the sAML stage, are of particular note, because their dynamic patterns suggested that they were positively selected during disease evolution. In this PhD work, we have observed that increasing mutations were mainly found in genes of the cohesin complex, such as *STAG2*, suggesting that mutations in this pathway could play a significant role in the leukemic transformation, as previously described [70,77,104]. Moreover, this dynamic process suggests that these mutations are early events in sAML progression and they would create an environment more prone to progress to more aggressive stages. Thus, the identification of these mutations is key to anticipate the disease evolution. Likewise, newly acquired mutations were not detected at the time of diagnosis, even by an amplicon-based TDS approach that allows a much high coverage (around 5000X), confirming that they were present only at the sAML stage. These mutations were found in *RAS*-pathway genes, such as *NRAS*, indicating that they are late events which might drive leukemic transformation, in accordance with previous studies that have described that *RAS*-pathway alterations could lead sAML evolution [121-123].

Furthermore, under this context of complex clonal architecture, the combination of mutations is not totally random and the co-occurrence of some genetic lesions may trigger a cascade of events leading to malignant transformation. Thereby, the question about how the combination of mutations contributes to disease evolution is relevant not only to determine what type of co-occurrence orchestrates the initiation of sAML, but also to identify possible targets for therapeutic approaches. In our study, we identified a significant **co-occurrence between mutations in cohesin complex and in Ras signaling pathway** and, particularly, between *STAG2* and *NRAS* genes, finding this combination in 15-20% of all patients who evolved to sAML. Taken into account this result and the distinguished mutational dynamics of these pathways and genes, our study suggests that this progressive combination could play an important role in the progression to sAML, being cohesin mutations early events, whereas Ras mutations are acquired later and drive the clone expansion and the disease evolution. Unfortunately, these are not absolute rules and this co-occurrence does not fully explain the sAML progression, but it would be a **preferential mechanism underlying MDS progression to sAML**.

For more than a decade, DNA hypomethylating agents (HMA) have been considered standard of care for MDS and they were associated with longer overall survival in controlled clinical trials [155-158]. However, they do not eradicate neoplastic clones and, in practice in the real-life data, the proportion of MDS patients who respond and maintain responses for a substantial period comprises fewer than half of treated patients [159]. Thus, it is clear that the HMA effect is transient and, despite treatment, MDS patients eventually progress to sAML. Moreover, several studies have described that HMA therapy alters clonal distribution [192-194] and, even, could induce C>G transversion in genes implicated in MDS and AML pathogenesis [162,195]. In our sequential study, we investigated the mechanisms of progression to sAML in MDS patients who were treated with disease-modifying agents (HMA and lenalidomide) before their disease evolution and we researched whether these mechanisms could be slightly different than those of non-treated patients. A higher proportion of mutations displaying newly acquired or increasing dynamics in chromatin modifiers genes was identified in treated patients, while in untreated patients the majority of mutations in these genes were stable during the evolution. Therefore, this PhD work confirms that therapy could influence on clonal evolution by the selection of resistant clones that were already present at diagnosis and by the acquisition of new mutations during treatment, providing insights that **mutations in chromatin-modifier genes could be related to the evolution of treated patients**. Nevertheless, there are still limitations and further questions that should be acknowledged in this regard. In fact, our study is based on the analysis of a small cohort of paired samples including 22 treated patients, leaving unsolved questions about the mechanisms underlying disease progression of treated patients. Thus, more studies with larger numbers of treated patients should be carried out to

validate this result and to confirm whether HMA therapy might alter clonal evolution underlying sAML evolution. Moreover, functional studies with *in vitro* and *in vivo* models that replicate disease characteristics could be useful to study biological mechanisms and mutations involved on HMA resistance. Similarly, the application of genome-wide CRISPR/Cas9 library screening in these models could be helpful in order to determine the role of biological pathways in the response and the resistance of these drugs.

Overall, this PhD work have gained insight into the molecular mechanisms of MDS progression to sAML, identifying a progressive combination or co-occurrence between mutations in cohesin complex and *RAS*-pathway genes which is involved in this leukemic transformation. Meanwhile, the molecular mechanisms by which mutations in these pathways promote leukomogenesis are not well understood and the clinical significance of these, alone or in combination, is not clear in MDS patients. Hence, **studying these mutations at diagnosis is required to define their prognostic impact and to anticipate disease evolution** in real-life.

To gain insight in this purpose, we studied a cohort of more than 400 MDS patients at diagnosis combining genomic data and reliable clinical information in order to **define the relevance of mutations in cohesin complex (chapter 2) and Ras signaling pathway (chapter 3) to clinical phenotypes and outcomes, as well as to assess the prognostic impact** of the integration of these mutations with the MDS risk factors (IPSS and IPSS-R). Moreover, in these studies, we implemented NGS techniques that assessed the mutational status of more than 100 MDS-related genes, thereby we could expand the knowledge of mutational profile of patients harboring cohesin and *RAS*-pathway mutations, respectively, and identify associations between pathways and genes.

In the case of cohesin mutations, previous studies reported that they, mostly nonsense and frameshift, occur in 10-15% of MDS patients [77,103,104], but our study confirmed that the loss of function of cohesin complex is a common event in MDS patients (11.9% of cases) and most of cohesin mutations are clustered in *STAG2* gene, occurring in around 70% of all cohesin-mutated patients. In addition, we found that MDS patients carrying cohesin mutations display a specific profile characterized by a higher number of mutations and by strong co-occurrences with mutations in *RAS*-pathway, chromatin modifier and splicing genes, specifically *SRSF2* splicing factor, as previously reported [81]. Likewise, this co-occurrence of *SRSF2* and cohesin mutations was also found in our study of patients who evolved to sAML (chapter 1), but *SRSF2* mutations displayed a stable dynamic during disease evolution and, for this reason, we do not focus on this combination. Nevertheless, this study at diagnosis has enabled to demonstrate that the **co-occurrence between mutations in cohesin complex and *SRSF2* gene is clearly impacting on the clinical outcome of *SRSF2*-mutated patients**, with double-mutated patients displaying a shorter overall survival

and leukemia-free survival, as well as an earlier progression to sAML, than *SRSF2*-single-mutated patients.

Although the canonical function of cohesin complex is the stabilization of sister chromatids during the metaphase in cell division [92,196], the majority of patients carrying mutations in cohesin genes in our cohort displayed a normal karyotype, supporting that cohesin genes are not involved in destabilization of chromosomal integrity in MDS; but rather alternative mechanisms such as transcriptional control and DNA repair, as described in recent studies [197-199]. In addition, our study identified that the presence of **cohesin mutations defined a subset of patients with a disease phenotype of poor prognosis**, showing lower platelet, absolute neutrophil and white blood counts in PB, a higher number of blasts in BM and a higher rate of progression to sAML. Besides, MDS patients harboring cohesin mutations were more likely to be associated with the diagnostic subtypes of excess of blasts and intermediate and high-risk IPSS-R categories, like other previous studies [77,103].

Regarding the prognostic impact, we comprehensively analyzed the impact of cohesin mutations on overall and leukemia-free survival and time to sAML progression, as well as we assessed their impact on the IPSS-R stratification. To date, some studies have described that cohesin mutations trend towards a worse prognosis [63,77,94], though not independently prognostic value was identified. Thanks to the integration of cohesin mutations data into IPSS-R stratification in our study, we demonstrated that cohesin mutations influence on leukemic transformation, showing patients with these mutations an earlier progression to sAML, especially in very low, low and intermediate-risk IPSS-R patients and we proved that the incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of MDS patients. Therefore, **studying the mutational status of cohesin genes could be useful in the monitoring of patients and the prediction of disease evolution, potentially allowing for early intervention**.

In the case of *RAS*-pathway mutations, previous studies analyzed the mutational status of the *NRAS* and *KRAS* genes, describing that these mutations occur in 5-10% of MDS patients [112,119,120]. However, in our work we studied the mutational status of not only the *RAS* oncogenes *NRAS*, *KRAS* and *HRAS*, but also other 9 genes involved in RAS pathway (*FLT3*, *KIT*, *CBL*, *EGFR*, *NF1*, *PTPN11*, *PTPN1*, *BRAF* and *G3BP1*) and whose mutations could activate this signaling pathway. We identified that somatic mutations in *RAS* pathway occurred in 11.2% of MDS patients and the median VAF of these mutations was lower than median VAF from mutations in other pathways, suggesting that these are subclonal events that are acquired at late stages of disease course. Moreover, to our knowledge, our study is the first work that evaluated the molecular status of RAS-pathway in MDS patients by a NGS approach, allowing the identification of associations between pathways and genes. Herein, we found that **patients with *RAS*-pathway mutations display a specific genetic profile**

characterized by the co-occurrence of mutations in *RAS* pathway and cohesin complex, whereas *RAS*-pathway and *SF3B1* mutations were mutually exclusive.

In order to identify whether *RAS*-pathway mutations are associated with a specific disease phenotype, we evaluated the clinical and biological characteristics of patients harboring mutations in *RAS* pathway. Previous studies reported that *RAS*-pathway mutations were most frequently found in MDS with excess blasts and in high-risk IPSS-R categories [112,118]. This PhD work identified, by integrating the molecular and clinical data, that the **presence of *RAS*-pathway mutations defined a subset of MDS patients who showed a disease phenotype of poor prognosis**. We observed that patients with these mutations displayed more accentuated cytopenias, such as lower platelet and neutrophil counts in PB, and a higher number of blasts in BM, as well as a higher incidence of poor prognostic marker, since a higher proportion of patients carrying *RAS*-pathway mutations received disease-modifying treatment and deceased.

Regarding the prognosis of MDS patients carrying *RAS*-pathway mutations, some studies described that mutations in *RAS* pathway could be associated with poor outcomes [121-123], whereas other works did not find correlation with a poor prognosis [112,120]. Therefore, to date, information about the clinical impact of *RAS*-pathway mutations is limited and controverted. To overtake this issue, we comprehensively analyzed the significance of *RAS*-pathway mutations on overall and leukemia-free survival and time to progression towards sAML. Thereby, we demonstrated that **patients carrying these mutations showed, above all, a shorter overall survival,** further confirming this negative impact by adjusting for the IPSS-R prognostic variables and in the subset of very low/low-risk IPSS-R patients.

Hence, **the activation of RAS pathway is a late event in the disease course that could be involved in a more aggressive disease stage**, associated with a dismal prognosis of patients. Consequently, these results point out that Ras pathway could be a potential therapeutic target in MDS patients and target drugs, such as rigosertib, could be useful in the management of patients with Ras signaling mutations [200,201]. Nevertheless, more studies from a functional perspective, for example by the implementation of CRISPR/Cas9 system, should be carried out in order to explore these therapeutic options.

Collectively, on the basis of the results from this PhD research, we propose a model for MDS development and progression to sAML (**Figure 9**). Previous studies described that **mutations follow a typical order of acquisition during disease progression**, identifying mutations that are usually acquired in early stages of disease, even in older people without hematological neoplasms defined as CHIP, while other mutations are more common in late stages of disease course [57,59,70,136,142]. In this line, Gilliland *et al.* described a traditional simplistic model of progression to sAML, postulating that two classes of cancer associated genes are required to

disease evolution [146,147]. Firstly, loss-of-function mutations in transcription factors are acquired, impairing hematopoietic differentiation and leading to the accumulation of more primitive cells. However, these mutations are not sufficient to cause acute leukemia and second mutations in activated signaling genes are needed to promote proliferative and/or survival advantage to hematopoietic stem cells.

Thereby, a **new model of genetic evolution** could be suggested for MDS development and progression to sAML **consisting of cohesin mutations as an early event in the progression of the disease, followed by mutations in *RAS* pathway that are associated with a more aggressive disease stage and drive the evolution**. We have described that mutations in cohesin complex showed an increasing dynamic during sAML transformation, being present at MDS diagnosis, but their tumor burden was higher at sAML stage. In addition, when we analyzed the mutational status of this pathway in a large MDS cohort, we have identified that patients with cohesin mutations displayed a higher number of mutations, compatible to pronounced genomic instability, as identified during the progression towards sAML. Similarly, patients harboring cohesin mutations showed an earlier progression to sAML, impacting on the current prognostic stratification of MDS patients. Thus, mutations in cohesin complex could be an early event in the sAML evolution which promote the genomic instability and the impair hematopoietic differentiation, due to their role in transcription regulation and DNA repair. Meanwhile, in our sequential study with paired samples we have found that *RAS*-pathway mutations were acquired at sAML stage. Additionally, we analyzed the mutational status of RAS pathway in a MDS cohort at diagnosis and we identified that these mutations are subclonal events that associate with MDS with excess of blasts and high-risk IPSS-R categories, suggesting that *RAS*-pathway mutations are acquire late in the disease course. Moreover, patients carrying these mutations at diagnosis showed a disease phenotype of poor prognosis and a shorter overall survival. Hence, mutations in RAS pathway could be the 'second hit' in the disease progression that promote proliferative and/or survival advantage to hematopoietic stem cells, leading the evolution to a more aggressive disease stage, such as sAML **(Figure 9)**.

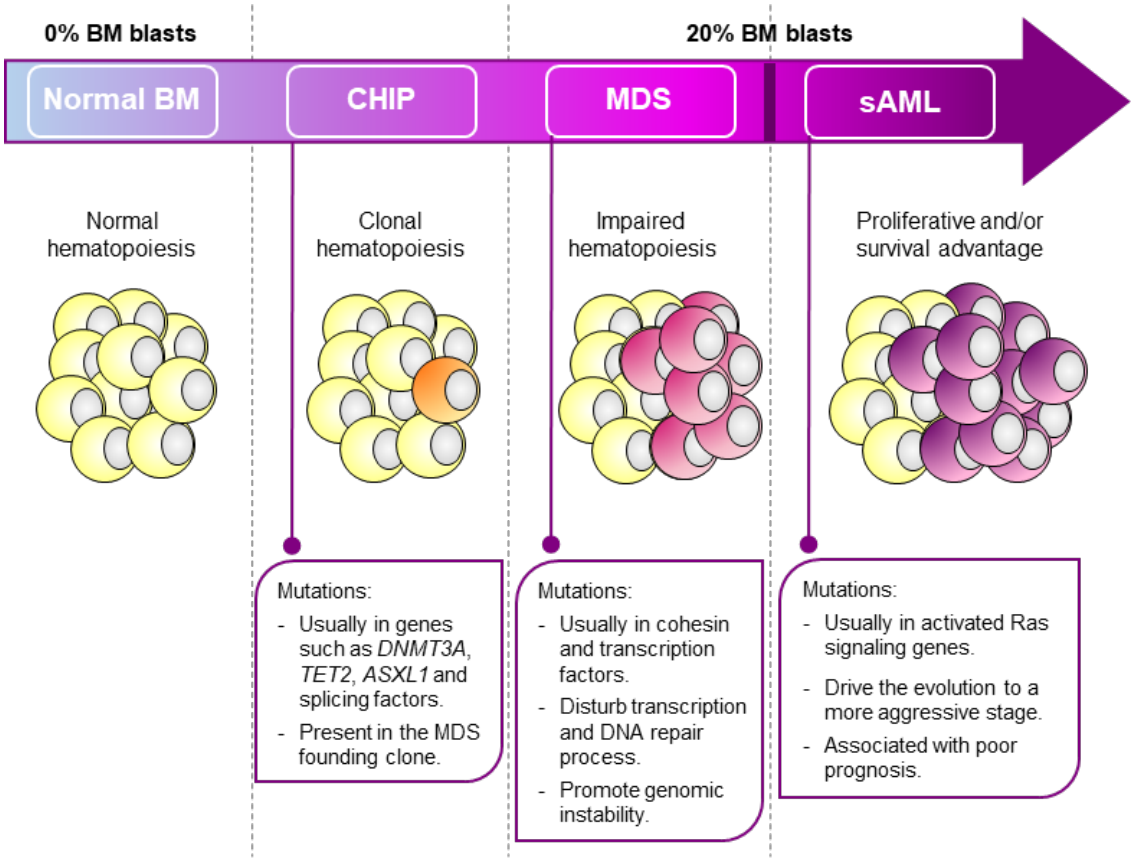


Figure 9. Proposed genetic model to MDS development and their progression to sAML. Mutations involved in each disease stages are shown: normal bone marrow (normal BM), Clonal Hematopoiesis of Indeterminate Potential (CHIP), myelodysplastic syndromes (MDS) and secondary acute myeloid leukemia (sAML).

In summary, the results provided in this PhD show that the implementation of high-throughput techniques helps to understand the mechanisms underlying the MDS pathogenesis and their progression to more aggressive stages, as leukemic transformation to sAML is. Furthermore, the integration of genomic data with clinical-biological parameters, such as those already included in current prognostic classifications, allows to the prediction and monitoring of disease evolution and, as a result, achieves a better management of patients, due to early intervention. Therefore, the use of new NGS techniques may be a used as a complementary tool to conventional tests in the MDS study in order to advance towards a more accurate integrated diagnosis, a more precise prognostic stratification and even a more adequate therapeutic orientation, due to the identification of potential therapeutic targets, opening new avenues in drug discovery.

1. Future perspectives

In the last decade, tremendous progress has been made in defining the genetic landscape of MDS patients, due to the implementation of large-scale sequencing techniques [57,62-64,70,71,136]. However, **models that replicate the identified mutations are now required** to study gene function, biological mechanisms and the role of mutations on disease pathogenesis [202-204].

Leukemia cell lines are universal model systems for investigating hematological malignancies, but **the number of MDS cell lines accurately reflecting the heterogeneity of the disease is very limited** [205,206]. In this regard, it is easier to establish cell lines in the leukemic AML/MDS phase than in the MDS phase, due to it appears that more aggressive cell lines are easier of establishment [207]. In addition, a large percentage of purposed MDS cell lines are cross-contaminated with established leukemia cell lines that grow faster than MDS malignant cells [207], even with B-lymphoblastoid cell lines [205]. Thus, nowadays, only a handful of immortalized MDS cell lines are available for research. Moreover, most of these MDS cell lines present very complex cytogenetic abnormalities, such as hyperdiploidy and hypodiploidy, and alterations that are not usual in MDS patients [208]. Then, these MDS *in vitro* models do not cover all the main genetic background of the MDS patients and they are not representatives of the common disease.

Beside the lack of cell lines harboring MDS alterations, added to the **difficulty of manipulating primary MDS cells**. Clonal cells from patients with MDS engraft poorly in xenotransplantation assays (*in vivo* models) and genetic and mechanistic studies in primary MDS cells (*ex vivo* models) are hampered by the difficulty of transfection or transduction of these cells [8,209]. Furthermore, there is a significant variability between patient samples and inter-patient genetic heterogeneity might not be well recapitulated if large cohort of primary MDS cells are not used [8].

Under this scenario, genome engineering is a helpful research tool to generate *in vitro* and *in vivo* models that reproduce disease characteristics by the targeted custom modification of the genome in a precise and efficient way. Thereby, genome engineering allows **to investigate the functional consequences of alterations with unknown biological significance**. In this PhD work, we have proved that the implementation of high-throughput techniques helps to understand the mechanisms underlying the MDS pathogenesis and their progression to more aggressive stages, as leukemic transformation to sAML is. Nevertheless, there are still limitations and further questions about the biological functions involved in disease evolution should be acknowledged.

The most recent and fastest growing method for genome editing is based on the Clustered Regions of Interspersed Palindromic Repeats (CRISPR) viral defense system found in bacteria and archaea [210]. The **CRISPR/Cas9 system** is easy to customize and optimize because the site selection for DNA cleavage is guided by a short sequence of RNA, called a guide RNA o gRNA. This guide RNA utilizes standard Watson-Crick binding to recognize the target sequence, termed the protospacer and position a CRISPR-associated nuclease (Cas9) to create a double strand break (DSB) in genomic DNA. When DSBs occur naturally, cells respond by activating DNA repair machinery and the damage is repaired by homology directed repair (HDR) or non-homologous end joining (NHEJ) mechanisms. Based on this principle, CRISPR/Cas9 system produce a DSB in a specific target region of DNA and the utilization of these repair processes allows for either NHEJ or HDR recombination resulting in the insertion, deletion or mutation of specific DNA sequences [203,211-213] (**Figure 10**).

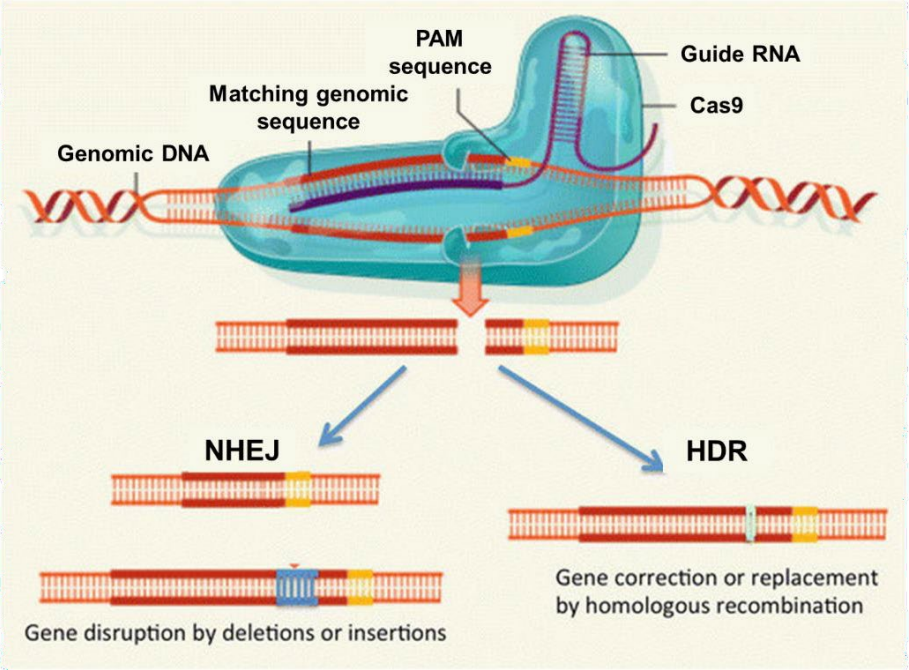


Figure 10. CRISPR/Cas9 targeting system (adapted from Tu, *Mol Neurodegener*, 2015) [214]. In the CRISPR/Cas9 system, a RNA guide hybridizes with the genomic DNA, specifically with a sequence immediately preceding an NGG DNA motif (PAM sequence), resulting in a double-strand break 3 bp upstream of the PAM sequence. This double-strand break is produced by the endonuclease Cas9 and, then, is repaired by endogenous cellular DNA repair machinery that catalyze non-homologous end joining (NHEJ) or homology-directed repair (HDR).

NGS studies have described which pathway are altered during sAML evolution and, specifically, in this PhD work we have described a preferential pattern of sequential acquisition of mutations in cohesin and Ras signaling pathway, but the accurate mechanisms through these alterations driving leukemic transformation are yet unsolved. Therefore, studies from a **functional perspective should be carried out in order to validate these genetic results and to widen the knowledge about biological mechanisms involved in MDS progression**. Moreover, nowadays MDS progression to sAML is associated with a well-known dismal prognosis, due to most of patients who evolved to sAML are resistant to currently available treatments, as well as the lack of efficient treatments that prevent evolution. Thus, functional studies in *in vitro* and *in vivo* models could be helpful to identify new therapeutic targets, as well as to test efficacy and efficiency of already existing drugs by drug screenings.

To overtake these issues, during the last year of this project we have edited myeloid cell lines by CRISPR/Cas9 technology in order to **generate cell models of MDS progression to sAML that recreate the mutations identified in this PhD study**. In point of fact, we propose a genetic model to MDS evolution to sAML that consists of the progressive combination of loss-of-function mutations in cohesin complex and activating mutations in Ras pathway as a 'second hit'. Thereby, we aim to reproduce this model of genetic evolution by an *in vitro* approach.

For this purpose, we chose to take forward for **CRISPR/Cas9 gene editing the SKM-1 cell line** which is an MDS/AML transition cell line, representative of MDS patients with loss of function of cohesin complex, since it harbors a stop gained mutation in *STAG2* (R259X). Moreover, we used the Alt-R™ CRISPR-Cas9 system from Integrated DNA technologies (IDT®) company to target *NRAS*, the most frequently mutated gene of RAS pathway. This system consists of a CRISPR/Cas9 ribonucleoprotein (RNP) complex, a guide RNA (sgRNA) targeting aminoacid 12 from *NRAS* and a DNA template with the change of interest in order to perform homology-directed insertion of a *NRAS* G12D mutation. Additionally, a sgRNA was designed not to target the human genome and it was used as a negative control. Then, SKM-1 cells were electroporated using a Neon Transfection System (Invitrogen) with the three components of the CRISPR/Cas9 complex. After 48 hours from electroporation, cells were sorted by fluorescence-activated cell sorting (FACS) using FACS-Aria (BD Bioscience), since the ATTO™ 550 fluorescent dye was added to the sgRNA, allowing the selection of cells with the CRISPR/Cas9 complex incorporated. Following successful nucleofection of the CRISPR complexes, single-cell were seeded in 96-well plates by serial dilutions. After cell expansion during 2 weeks, DNA was extracted and clones were screened by PCR and Sanger sequencing for the presence of the *NRAS* G12D mutation.

In total, we generated four different **clones of SKM-1 harboring *NRAS* G12D mutation** and four clones derived from SKM-1 cells electroporated with sgRNA designed not to target the human genome that were used as controls. Thereby, we have generated an *in vitro* model that reproduces the genetic changes in *STAG2* (loss-of-function) and *NRAS* (G12D activating mutation) underlying the MDS progression to sAML, where control clones are similar than cells observed at MDS time, while *NRAS*-mutated clones recreate the sAML stage where co-occur *STAG2* and *NRAS* mutations.

Hence, our next steps will be to explore the effect of these mutations on proliferation, cell cycle, cell death and apoptosis and differentiation. In addition, we are going to study using RNAseq methodology the changes on expression and regulation of other biological pathways produced as consequence of the acquisition of *NRAS* mutation. Moreover, in the future we would like to use this *in vitro* model to explore drug response for currently available drugs, such as HMA and venetoclax, as well as to identify new potential therapeutic options for treatment of patients who evolve to sAML, such as PARP inhibitors.

Concluding Remarks

1. MDS progression to sAML is a very complex and dynamic process of clonal evolution and clonal selection. The implementation of NGS tools in paired samples from patients who evolved to sAML allows the characterization of the mutational landscape underlying sAML progression, which is characterized by greater genomic instability, irrespective of the MDS subtypes at diagnosis.

    1.1. The mechanisms underlying this evolution are heterogeneous, however, four types of mutational dynamics could be identified, being increasing and newly acquired mutations of particular importance. Of note, the co-occurrence of cohesin mutations, mainly displaying an increasing dynamic, and mutations in Ras signaling pathway, mostly acquired at sAML stage, was identified as a preferential mechanism in 15-20% of MDS patients who progressed to sAML.

    1.2. The treatment with disease-modifying agents (DNA hypomethylating agents and lenalidomide) could impact the clonal evolution patterns in MDS patients. The mechanisms of sAML progression in treated patients are different from those of non-treated patients, characterized by the presence of mutations in chromatin-modifier genes that increased their clone size or were newly acquired at sAML stage.

2. The presence of cohesin loss-of-function mutations is a common event in MDS patients, occurring in 11.9% of cases at diagnosis, being *STAG2* the most frequently mutated gene.

    2.1. MDS patients harboring cohesin mutations display a specific profile characterized by a higher number of mutations and by strong co-occurrences with mutations in *RAS*-pathway, in the chromatin modifier genes and in the splicing genes, especially in the *SRSF2* splicing factor.

    2.2. The presence of mutations in cohesin genes define a subset of patients with a disease phenotype of poor prognosis, characterized by more accentuated cytopenias in peripheral blood and a higher number of blasts in bone marrow. Moreover, cohesin mutations are poor prognostic markers regarding the leukemic transformation, being associated with a higher rate and an earlier progression to sAML, especially in very low, low and intermediate-risk IPSS-R patients.

    2.3. The incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of MDS patients of very low and low-risk patients. Thereby, the study of the mutational status of cohesin genes could be useful in the monitoring of patients and the prediction of disease evolution, potentially allowing for early intervention.

3. Mutations in *RAS*-pathway are recurrent genetic events in MDS patients, occurring in 11.2% of cases and usually being subclonal and acquired late during the course of the disease.

   3.1. MDS patients carrying *RAS*-pathway mutations display a specific genetic profile highlighting the co-occurrence of mutations in RAS pathway and cohesin complex, whereas the presence of *RAS*-pathway and *SF3B1* mutations were mutually exclusive.

   3.2. The presence of *RAS*-pathway mutations defines a subset of MDS patients who showed a disease phenotype of poor prognosis, with presence of accentuated cytopenias and excess of blasts, as well as a higher rate of death, despite disease-modifying treatment. Moreover, patients harboring *RAS*-pathway mutations show a shorter overall survival, further confirming this negative impact by adjusting for the IPSS-R prognostic variables. Therefore, the activation of RAS pathway could be involved in the progression to an aggressive disease stage, associating with a dismal prognosis of MDS patients.

Departamento de **MEDICINA, HEMATOLOGÍA**

CENTRO DE INVESTIGACIÓN DEL CÁNCER-IBMCC (USAL-CSIC)

Resumen en castellano

Tesis Doctoral

Estudio de los mecanismos moleculares y de la evolución clonal implicados en la progresión de los Síndromes Mielodisplásicos a Leucemia Aguda Mieloblástica:

Caracterización genómica mediante Secuenciación Masiva

Supervisores:

Prof. Dr. Jesús María Hernández Rivas

Dra. M. del Rocío Benito Sánchez

Dra. María Abáigar Alvarado

Marta Martín Izquierdo

2021

Introducción

1. Síndromes Mielodisplásicos

1.1. Perspectiva general

Los Síndromes Mielodisplásicos (SMD) son un grupo heterogéneo de enfermedades clonales que afectan a las células madre hematopoyéticas. Estos desórdenes se caracterizan por presentar una hematopoyesis ineficaz en la médula ósea (MO), que da lugar a citopenias en sangre periférica (SP), a pesar de poseer una médula ósea normo o hipercelular, así como por presentar una morfología aberrante o displásica en las células hematopoyéticas de al menos un linaje mieloide. Además, los SMD presentan un riesgo elevado de progresión a Leucemia Aguda Mieloblástica secundaria (LAMs) [1-4].

La forma de presentación de la enfermedad y el curso clínico de los pacientes con SMD es muy variable, desde formas indolentes, con una esperanza de vida cercana a la normal durante años, a formas más agresivas que progresan rápidamente a LAMs [5-7].

El proceso de desarrollo de las células sanguíneas (hematopoyesis) y cómo se ve afectado en los SMD está aún en debate. En los SMD, el clon celular mutado puede suprimir la hematopoyesis normal y alterar el desarrollo de todos los linajes celulares, dando lugar a displasias y deficiencias (citopenias sanguíneas) de múltiples linajes celulares (**Figura 1**).

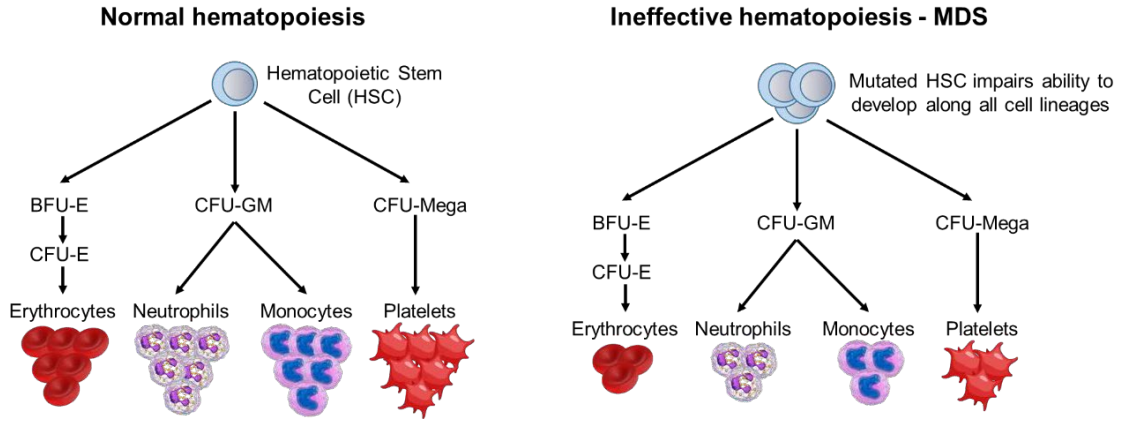


Figura 1. Los Síndromes Mielodisplásicos alteran la hematopoyesis (adaptada de Corey *et al.*, *Nat Rev Cancer*, 2007) [8]. En los pacientes con SMD, las células madre hematopoyéticas (HSC) se encuentran alteradas, por lo que el desarrollo de los distintos linajes celulares se ve afectado, anulando la hematopoyesis normal. De este modo, los progenitores de los distintos linajes celulares, como el progenitor eritroide (CFU-E), el progenitor de granulocito-monocito (CFU-GM) y el progenitor de megacariocito (CFU-Mega), pueden verse alterados, dando lugar a displasia y deficiencias o citopenias a lo largo de todos los linajes celulares.

1.2. Epidemiología y etiología

Los SMD son una de las enfermedades hematológicas más prevalentes, siendo descrita más habitualmente en la población de edad avanzada, con una mediana de 65-70 años al diagnóstico y <10% de los pacientes son menores de 50 años [8-10].

La incidencia anual de los SMD oscila entre los 2-12 casos por 100.000 habitantes por año. Sin embargo, esta incidencia aumenta con la edad llegando a 40-50 por 100.000 habitantes entre la población mayor de 70 años [8,9,11]. Aunque no existen diferencias étnicas en la incidencia, algunos estudios sugieren que en Asia la edad de aparición de la enfermedad es menor con una mediana de 53 años al diagnóstico [12,13]. Con respecto al sexo de los pacientes, los SMD afectan más frecuentemente a hombre que a mujeres, excepto para los casos con deleción aislada 5q, los cuales predominan en el sexo femenino [14,15].

La causa de los SMS solo es conocida en un 15% de los casos, sin embargo, su asociación con el envejecimiento sugiere que una de las causas podía ser el daño genético causado por exposición peligrosa o por susceptibilidad heredada. Algunos de los factores medioambientales de riesgo que pueden estar relacionados con este daño genético son la quimioterapia, la radiación, el benceno y sus derivados, el petróleo, la ingesta abusiva de alcohol, el tabaco, las infecciones virales o los agentes inmunosupresores [8,16-20]. De hecho, la incidencia de los SMD relacionados con la terapia ha aumentado en los últimos años, probablemente por el uso de quimioterapia y radioterapia en el tratamiento del cáncer [10]. En relación con la susceptibilidad heredada, aunque los casos de SMD de edad más avanzada suelen presentan mutaciones somáticas adquiridas con la edad que conducen la evolución clonal, en el caso de los niños y adultos jóvenes es más habitual que porten condiciones genéticas de origen germinal las cuales conducen al desarrollo de la enfermedad. Esta susceptibilidad genética ocurre en un tercio de los casos pediátricos, incluyendo niños con Síndrome de Down, anemia de Falconi, Síndrome de Schwachman-Diamond y neurofibromatosis [9]. Por el contrario, en los adultos la predisposición hereditaria es menos común, aunque en los últimos años gracias a la implantación de la secuenciación masiva en el ámbito clínico se han descrito un mayor número de alteraciones hereditarias que predisponen al desarrollo de un SMD [21].

1.3. Diagnóstico

El diagnóstico de los SMD a menudo supone un reto debido a que son un grupo de enfermedades muy heterogéneo. Principalmente se basa en la presencia de citopenias en un análisis rutinario de SP y los síntomas más comunes son fatiga, infecciones y sangrado como resultado de las citopenias en SP [11,16]. De acuerdo con la Organización Mundial de la Salud (OMS) y con el Sistema Internacional de puntuación pronóstica IPSS, las citopenias se definen como: anemia, cuando los

niveles de hemoglobina son menores de 10 g/dl; trombocitopenia, cuando el recuento de plaquetas es inferior a 100 x10^9/L; y neutropenia, cuando el recuento absoluto de neutrófilos es menor de 1,6 o 1,8 x10^9/L, según la OMS y el IPSS, respectivamente [22].

Para el diagnóstico de los SMD es necesario un examen morfológico de un frotis de sangre periférica o de un aspirado de médula ósea el cual confirme que las citopenias observadas son consecuencia de un fallo en la médula ósea [23]. La médula ósea de los pacientes con SMD normalmente es hiper o normocelular y muestra displasia en uno o en varios linajes hematopoyéticos mieloides (eritroide, granulocítico o megacariocítico), considerándose como displasia cuando más del 10% de las células de al menos un linaje mieloide presenta cambios morfológicos inequívocos [16,24,25]. Además, la proporción de blastos en la MO debe ser valorada para proporcionar una correcta clasificación, puesto que las formas con <2% de blastos en MO presentan un mejor pronóstico que aquellas con >2% de blastos en MO, así como la presencia de más de un 20% de blastos en MO es el límite que define la OMS para el diagnóstico de LAM. Por el contrario, una biopsia de la MO no es obligatoria en el diagnóstico de los SMD, aunque es importante para la identificación de los SMD fibróticos o hipocelulares [9].

Del mismo modo, durante el diagnóstico de un SMD es recomendable realizar un estudio cuidadoso que permita identificar pacientes con citopenias, pero no displasia, los cuales suelen tener características similares a los pacientes con SMD. Además, la característica clave de los SMD es la naturaleza clonal de la displasia. Por tanto, la detección de una anormalidad cromosómica (cariotipo y/o FISH), el análisis por citometría de flujo y un análisis mutacional deben ser considerados como estudios complementarios para definir mejor el subtipo diagnóstico del SMD, así como el pronóstico de los pacientes [9,11,26,27].

1.4. Clasificación

Debido a la heterogeneidad de los SMD, es necesario categorizar los pacientes en grupos por características morfológicas similares, etiología molecular común, pronósticos semejantes o por presentar un riesgo similar de progresión a LAMs o de respuesta a las terapias disponibles [28].

Clasificación morfológica

En 1982, el grupo cooperativo Francés-Americano-Británico (FAB) propuso el primer sistema de clasificación de acuerdo a la citomorfología y al porcentaje de blastos al diagnóstico [29]. A lo largo de los años, este sistema de clasificación ha sufrido varias revisiones para refinar la clasificación diagnóstica, incorporándose información referente a las citopenias sanguíneas, las características morfológicas, la citogenética y los datos genéticos. En este sentido, más recientemente han surgido las

clasificaciones de la OMS 2001, OMS 2008 y la actual OMS 2017 que han sido usadas con éxito para la clasificación de los pacientes con SMD [30-32].

El último sistema de clasificación utilizado en los pacientes con SMD es la clasificación de la OMS 2017. Esta clasificación divide a los MSD en varias entidades basadas en los hallazgos morfológicos encontrados en la SP y la MO, tales como el tipo y el grado de la displasia, el número de citopenias, la presencia de sideroblastos en anillo, la proporción de blastos en MO y en SP, la presencia de alteraciones cromosómicas específicas (deleción del cromosoma 5q) y, por primera vez, la información genética (presencia de mutaciones en *SF3B1*) [32-36].

La última edición de la clasificación de la OMS identifica 9 subtipos diferentes de SMD: SMD con displasia unilínea (SMD-DU), SMD con displasia multilínea (SMD-DM), SMD con sideroblastos en anillo y displasia unilínea (SMD-SA-DU), SMD con sideroblastos en anillo y displasia multilínea (SMD-SA-DM), SMD con del(5q) aislada, SMD con exceso de blastos tipo 1 y tipo 2 (SMD-EB1 y SMD-EB2), SMD inclasificable y anemia refractaria de la infancia (**Tabla 1**).

Sistemas de estratificación de riesgo

Los pacientes con SMD presentan un curso clínico muy variable, con grandes diferencias en la supervivencia global (SG) y en el riesgo de transformación a LAM, pudiendo darse formas indolentes de la enfermedad que permanecen estables durante años y formas que progresan rápidamente a leucemia [5,6,28]. Por esta razón, se han desarrollado sistemas de puntuación pronóstica que estratifican los pacientes según su riesgo, teniendo en cuenta su supervivencia esperada y su riesgo de evolución a LAM, con el fin de mejorar el manejo de los pacientes y la estimación y evaluación de su pronóstico.

Algunos de estos sistemas de puntuación pronóstica son el Sistema Internacional de Puntuación Pronóstica (International Prognostic Scoring System, IPSS), Sistema de Puntuación Pronóstica basado en la clasificación de la OMS (WHO classification-based Prognostic Scoring System, WPSS), el Sistema Puntuación Pronóstica de SMD de bajo riesgo (Lower-Risk MDS Prognostic Scoring System, LR-PSS), el Sistema de Puntuación Exhaustivo del MD Anderson (MD Anderson Comprehensive Scoring System, MDA-CSS) y el Sistema Internacional de Puntuación Pronóstica Revisado (Revised International Prognostic Scoring System, IPSS-R) [6,28,37-41].

Tabla 1. Criterio de la OMS 2017 para la clasificación de los pacientes con SMD (adaptada de Arber *et al.*, *Blood*, 2016) [32].

Nombre	Linajes displásicos	Citopenias*	Sideroblastos en anillo en MO (%)	Blastos en MO y SP (%)	Citogenética
SMD con displasia unilínea (SMD-DU)	1	1 o 2	<15% / <5%†	MO <5%, SP <1%, no bastones de Auer	Ninguna, a menos que cumpla con todos los criterios para SMD con del(5q) aislada
SMD con displasia multilínea (SMD-DM)	2 o 3	1-3	<15% / <5%†	MO <5%, SP <1%, no bastones de Auer	Ninguna, a menos que cumpla con todos los criterios para SMD con del(5q) aislada
SMD con sideroblastos en anillo y displasia unilínea (SMD-SA-DU)	1	1 o 2	≥15% / ≥5%†	MO <5%, SP <1%, no bastones de Auer	Ninguna, a menos que cumpla con todos los criterios para SMD con del(5q) aislada
SMD con sideroblastos en anillo y displasia multilínea (SMD-SA-DM)	2 o 3	1-3	≥15% / ≥5%†	MO <5%, SP <1%, no bastones de Auer	Ninguna, a menos que cumpla con todos los criterios para SMD con del(5q) aislada
SMD con del(5q) aislada	1-3	1-2	No	MO <5%, SP <1%, no bastones de Auer	Del(5q) única o con 1 alteración adicional, excepto -7 o del(7q)
SMD con exceso de blastos tipo 1 (SMD-EB-1)	0-3	1-3	No	MO 5%-9% o SP 2%-4%, no bastones de Auer	Ninguna
SMD con exceso de blastos tipo 2 (SMD-EB-2)	0-3	1-3	No	MO 10%-19% o SP 5%-19% o bastones de Auer	Ninguna
SMD Inclasificable (MDS-U)					
Con 1% de blastos en sangre	1-3	1-3	No	MO <5%, SP = 1%, ‡ no bastones de Auer	Ninguna
Con displasia unilínea y pancitopenia	1	3	No	MO <5%, SP <1%, no bastones de Auer	Ninguna
Basado en alteraciones citogenéticas	0	1-3	<15%§	MO <5%, SP <1%, no bastones de Auer	Alteraciones típicas de SMD
Anemia Refractaria de la Infancia	1-3	1-3	No	MO <5%, SP <2%	Ninguna

***** Citopenias definidas como: hemoglobina, <10 g/dL; plaquetas, <100 x 10^9/L; y recuento absoluto de neutrófilos <1.8 x 10^9/L. Con poca frecuencia, los SMD pueden presentarse con anemia leve o trombocitopenia por encima de estos niveles. El recuento de monocitos en SP debe ser <1 x 10^9/L.

† Mutación de *SF3B1* presente.

‡ El porcentaje del 1% de blastos en SP debe registrarse al menos en 2 ocasiones distintas.

§ Casos con ≥15% de sideroblastos en anillo por definición presentan displasia eritroide significativa y son clasificados como SMD-SA-DU.

El predictor de pronóstico más utilizado para pacientes con SMD es el IPSS, publicado por primera vez en 1997 por Greenberg et al [38]. Sin embargo, a lo largo del tiempo se ha descrito que el IPSS no predecía de manera precisa el pronóstico en los SMD de riesgo bajo [42]. De este modo, el IPSS ha sido revisado (IPSS-R) para refinar la estratificación de los pacientes incorporando subgrupos citogenéticos más amplios y más diferenciados, una descripción de los blastos más estricta y una información más detallada de las citopenias[41]. A continuación, el Sistema de Puntuación Pronóstica Revisado IPSS-R es detallado (**Tabla 2**).

El IPSS-R incluye 5 grupos de riesgo que presentan una supervivencia global y un riesgo de progresión a LAM significativamente diferentes. Los dos grupos de riesgo inferiores son conocidos como "SMD de bajo riesgo", mientras que los pacientes de los dos niveles superiores se conocen como "SMD de alto riesgo". Sin embargo, el grupo intermedio, el cual representa casi el 20% de todos los casos de SMD, es una categoría heterogénea que engloba pacientes indolentes semejantes a los SMD de bajo riesgo y pacientes con una enfermedad más agresiva [43,44].

Una de las limitaciones más importantes del IPSS-R es la falta de información genética en su sistema de puntuación. Recientemente, se ha generado una gran cantidad de datos moleculares de pacientes con SMD, lo que ha permitido aumentar el conocimiento sobre algunas mutaciones genéticas y sobre su influencia en el pronóstico de los pacientes. En concreto, se ha descrito que las mutaciones en *SF3B1* se asocian con un resultado clínico favorable, mientras que las mutaciones en *TP53*, *SRSF2*, *DNMT3A*, *IDH2*, *RUNX1* and *ASXL1* se asocian con una pobre evolución clínica [36,45-51]. Por tanto, en la actualidad se está llevando a cabo una propuesta para incluir este conocimiento molecular en el sistema de estratificación IPSS-R.

Tabla 2. Sistema Internacional de Puntuación Pronóstica Revisado (IPSS-R) (adaptado de Greenberg *et al.*, *Blood*, 2012) [41].

	Valores de Puntuación						
	0	0.5	1	1.5	2	3	4
Citogenética	-Y, del(11q)	-	Normal, del(5q), del(12p), del(20q), doble incluyendo del(5q)	-	Del(7q), +8, +19, i(17q), cualquier otro clon independiente único o doble	-7, inv(3)/t(3q)/del(3q), doble incluyendo -7/del(7q), Complejo: 3 alteraciones	Complejo: >3 alteraciones
% Blastos en MO	≤2	-	>2% to <5%	-	5% a 10%	>10%	
Hemoglobina (g/dL)	≥10	-	8 to 10	<8			
Plaquetas ($x10^9$/L)	≥100	50 a <100	<50				
Neutrófilos (ANC) ($x10^9$/L)	≥0,8	<0,8					

	Categorías y Resultado Clínico				
	Muy Bajo	**Bajo**	**Intermedio**	**Alto**	**Muy Alto**
Valor de riesgo	≤1,5	>1,5 a 3	>3 a 4,5	>4,5 a 6	>6
Pacientes (%)	19%	38%	20%	13%	10%
Supervivencia (mediana, años)	8,8	5,3	3,0	1,6	0,8
Tiempo hasta evolución a LAM (15% de casos) (mediana, años)	No alcanzado	10,8	3,2	1,4	0,7

1.5. Resultado clínico

Los pacientes con SMD presentan un gran rango de evoluciones clínicas, desde condiciones indolentes durante años y cercanas a una esperanza de vida normal hasta formas que progresan rápidamente a leucemia aguda. De este modo, la heterogeneidad clínica complica el manejo de los pacientes, pues dificulta la elección del tratamiento, así como el momento de la intervención [5]. Además de las variables específicas de la enfermedad, existen una serie de factores relacionados con el propio paciente, como son la edad o la presencia de otras patologías, que son esenciales para la estimación del riesgo de los pacientes. En concreto, debido a que la mediana de edad es cercana a los 70 años, los pacientes con SMD suelen presentar comorbilidades que influyen en la evolución clínica y en el espectro de terapias disponibles [43].

Tratamiento de los SMD

Las opciones terapéuticas para los pacientes con SMD varían desde un tratamiento de soporte y sintomático de las citopenias sanguíneas, basado principalmente en transfusiones, hasta un trasplante alogénico de progenitores hematopoyéticos (alo-TPH), dependiendo de los factores de riesgo propios de la enfermedad y relacionados con el paciente [7]. Además, los SMD principalmente afectan a población de edad avanzada, lo que significa que la mayoría de los pacientes no puede tolerar abordajes terapéuticos intensivos, como la quimioterapia intensiva o el trasplante alogénico [52].

Hasta el momento, tres fármacos han sido aprobados por la Administración de Alimentos y Medicamentos de los Estados Unidos (Food and Drug Administration, FDA) y por la Agencia Europea del Medicamente (European Medicines Agency, EMA) para su uso en indicaciones relacionado con los SMD: lenalidomida (aprobada en 2005), un fármaco inmunomodulador que se administra oralmente; y azacitidina y decitabina (aprobadas en 2004 y 2006, respectivamente), dos análogos de los nucleótidos que son agentes hipometilantes del ADN (DNA hypomethylating agents, HMA). Además de estos medicamentos, existe un extenso uso de agentes estimuladores de la eritropoyesis (rythropoiesis-stimulating agents, ESA) fuera de lo indicado, como son la eritropoyetina y la darbepoyetina [43]. Sin embargo, desde la aprobación de la decitabina en 2006, han pasado más de 15 años sin un nuevo fármaco para los pacientes con SMD, a excepción del luspatercept, aprobado para el tratamiento de la anemia en pacientes con SMD en 2020 [53].

Por tanto, el tratamiento de los SMD es elegido en base a la clasificación pronóstica y los objetivos propuestos a alcanzar con la terapia son diferentes en pacientes de bajo riesgo a los propuestos en pacientes de alto riesgo e, incluso, a los propuesto en aquellos pacientes en los que el tratamiento con agentes hipometilantes ha fallado.

En pacientes con SMD de bajo riesgo, la meta es reducir la necesidad de transfusiones gracias al tratamiento con agentes estimuladores de la eritropoyesis y luspatercept, así como prevenir la transformación a un subtipo de alto riesgo o a LAMs y mejorar la supervivencia. En el caso de los pacientes de bajo riesgo con del(5q), el tratamiento con lenalidomida puede alcanzar altas tasas de respuesta, consiguiendo en estos pacientes independencia transfusional e, incluso, remisión citogenética [52,54,55]. Por el contrario, en el caso de los SMD de alto riesgo, el objetivo principal es prolongar la supervivencia y las terapias actuales disponibles incluyen soporte de factores de crecimiento, lenalidomida, agentes hipometilantes, quimioterapia intensiva y alo-TPH. Antes de iniciar el tratamiento de cualquier paciente con un SMD de alto riesgo, se debe evaluar la opción de realizar un alo-TPH, puesto que, actualmente, es el único tratamiento curativo disponible y, por tanto, el único que puede tener un efecto importante en la supervivencia del paciente. Por el contrario, en aquellos pacientes que no son candidatos a trasplante, la opción terapéutica más apropiada sería el tratamiento con agentes hipometilantes [54,55].

Aunque aproximadamente la mitad de los pacientes con SMD son sometidos a tratamiento con agentes hipometilantes, la mayoría no responde o termina perdiendo la respuesta inicial. Para estos pacientes con enfermedad refractaria o en vías de progresión tras la terapia con agentes hipometilantes, en este momento no existe una intervención indicada y las opciones terapéuticas incluyen la participación en un ensayo clínico, tratamiento con citarabina o un alo-TPH [52,54,55].

A pesar de estas opciones, existe un gran número de pacientes cuya enfermedad progresa y evoluciona a LAMs. Por tanto, existe una enorme necesidad clínica de encontrar nuevas terapias para los pacientes con SMD.

Progresión de los SMD a LAMs

Los pacientes con SMD presentan un elevado riesgo de, finalmente, progresar a una leucemia aguda mieloide (LAM secundaria, LAMs), ocurriendo dicha progresión en aproximadamente un tercio de los pacientes. Por definición, la progresión a LAM sucede cuando se observa en un evaluación morfológica de la MO un aumento en el número de blastos superior al 20% [32,56].

La progresión de enfermedad está asociada a un pronóstico lúgubre, puesto que la mayoría de los pacientes que progresan son resistentes a los tratamientos actuales, tienen tasas inferiores de remisión completa y la tasa de supervivencia a largo plazo de los pacientes tratados es menor del 10% [42,57]. Además, como sucede en otras patologías premalignas de alto riesgo, se debe tener en cuenta el riesgo de progresión a LAM a la hora de la toma de decisiones clínicas, como puede ser si el paciente se somete a un alo-TPH y cuándo [57].

Más información sobre la progresión de los SMD a LAMs está incluida en el apartado 3 de la sección de Introducción.

2. Alteraciones biológicas y moleculares de los SMD

Es bien sabido que en los SMD, como otros tipos de cáncer, suceden rondas sucesivas de selecciones positivas, donde las mutaciones y otras alteraciones genéticas juegan un papel muy importante [58-60]. Además, es poco común encontrar una única alteración genética aislada, siendo mucho más habitual la combinación de mutaciones y/o alteraciones citogenéticas [61,62]. Por tanto, para descifrar qué alteraciones o qué mecanismos están involucrados en esta selección positiva (alteraciones "drivers") es crucial conocer la patogénesis de los SMD y los distintos fenotipos de la enfermedad [63].

Los primeros estudios genéticos en pacientes con SMD se centraron en las alteraciones citogenéticas detectadas por citogenética convencional (cariotipo). Aproximadamente un 50% de los casos diagnosticados de SMD presentan alteraciones citogenéticas, siendo la citogenética uno de los factores más importantes que influye en el resultado clínico de los pacientes [33,38,41,64]. La mayoría de las alteraciones observadas en los SMD son cambios no balanceados, los cuales resultan en ganancias y pérdidas de material cromosómico. Las alteraciones más frecuentes son: -5/del(5q) y -7/del(7q), seguidas de la trisomía 8, del(20q),-18/del(18q), -Y, ganancia del cromosoma 21, del(17p)/iso(17q) y inv/t(3q). Además, la mayoría de estas lesiones a menudo coocurren siendo parte de un cariotipo complejo [33,64]. En la **Figura 2** se muestran las alteraciones cromosómicas más recurrentes en pacientes con SMD y su incidencia.

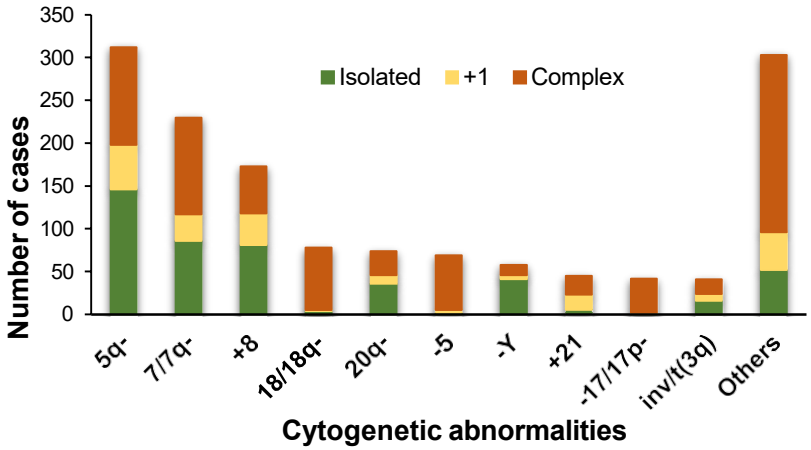


Figura 2. Frecuencias de las alteraciones citogenéticas más frecuentes en pacientes con SMD, divididas en aisladas, presentes con una anomalía adicional o presentes en un contexto de cariotipo complejo (adaptada de Haase *et al.*, *Blood*, 2007) [33]. Este gráfico muestra las alteraciones citogenéticas más frecuentes en pacientes con SMD. La presencia de estas alteraciones junto con otras alteraciones o aisladas es representada en color: verde para las alteraciones citogenéticas aisladas, amarillo para los casos con una anomalía adicional, y naranja para los casos con cariotipo complejo.

Sin embargo, más del 50% de los pacientes poseen una citogenética normal mientras que varios estudios han demostrado que casi la totalidad de los SMD poseen mutaciones genéticas [61-63,65].

2.1 Perfil mutacional de los SMD

El perfil genético de los pacientes con SMD influye de manera crítica en el fenotipo clínico, en el pronóstico y en la respuesta al tratamiento, por lo que es vital ampliar y actualizar nuestro conocimiento sobre la genética de los SMD para, de este modo, predecir la evolución clínica de los pacientes y seleccionar la terapia más adecuada para cada paciente [63].

Hasta el año 2010, el estudio de la genética de los SMD se reducía al estudio de un pequeño número de genes, como son *TP53*, *EZH2*, *RUNX1*, *TET2*, *NRAS*, *KRAS*, *FLT3*, *ASXL1*, por cariotipo mediante SNP arrays y secuenciación Sanger [66-68]. Sin embargo, durante la última década, nuestro conocimiento acerca de los genomas de pacientes con SMD ha aumentado exponencialmente gracias a la detección exhaustiva de mutaciones somáticas mediante las nuevas técnicas avanzadas de secuenciación de alto rendimiento [61,62,69].

En los últimos años, se ha descrito que de un 78 hasta un 90% de los casos con SMD portan al menos una mutación genética [61,62,70], encontrándose éstas principalmente en pacientes con un cariotipo normal puesto que aproximadamente dos tercios de los pacientes con cariotipo normal poseen alguna mutación genética [3,65]. Asimismo, el número de mutaciones varía según el subtipo de SMD. De hecho, mientras que un paciente con un SMD de bajo riesgo, como son los SMD-DU con o sin sideroblastos en anillo y los SMD con del(5q) aislada, presentan una mediana de 2-3 mutaciones drivers detectadas por secuenciación dirigida y 6 por secuenciación de exoma completo, un SMD de alto riesgo (SMD-EB y SMD-DM) y pacientes con LAMs tienen un número mayor de mutaciones que oscila entre 9 y 13 alteraciones[63].

Hasta el momento, se han descrito más de 40 genes drivers implicados en la patogénesis de los SMD, los cuales pueden ser categorizados según la ruta funcional en la que participan: splicing del ARN (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*, entre otros), metilación del ADN (*TET2*, *DNMT3A*, *IDH1*/*IDH2*), transcripción (*RUNX1*, *ETV6*, *TP53*, *BCOR*, etc.), modificación de la cromatina (*EZH2*, *ASXL1*, *ATRX*, …), complejo de las cohesinas (*STAG2*, *STAG1*, *RAD21*, *SMC1A*, *SMC3*, *CTCF*) y señalización (*NRAS*, *KRAS*, *FLT3*, *CBL*, *JAK2*, etc.). Los genes más frecuentemente mutados son los implicados en las vías de la metilación del ADN y modificación de la cromatina, así como los implicados en el splicing. Sin embargo, la mayoría de estos genes se encuentran mutados en una frecuencia baja, con solo 6 de ellos mutados en >10% de los pacientes [61,62,70-74] (**Figura 3**).

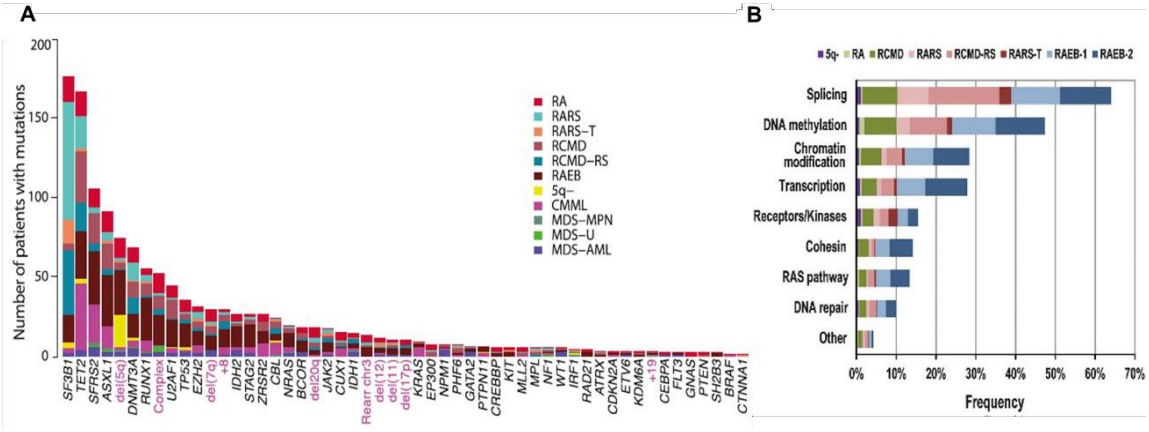


Figura 3. Frecuencias de las alteraciones drivers identificadas por secuenciación o por citogenética, así como de las vías más frecuentemente alteradas, considerando los distintos subtipos de SMD (adaptada de Papaemmanuil *et al.*, *Blood*, 2013 and de Haferlach *et al.*, *Leukemia*, 2014) [61,62]. a) Las alteraciones más frecuentes son mutaciones en *SF3B1*, *TET2*, *SRSF2*, *ASXL1*, la deleción del cromosoma 5q, mutaciones en *DNMT3A*, *RUNX1*, el cariotipo complejo y mutaciones en *U2AF1* y *TP53*. **b)** Con respecto a las vías funcionales que pueden verse alteradas en pacientes con SMD, las más frecuentes son el splicing del ARN y la metilación del ADN.

Teniendo en cuenta este contexto donde múltiples genes drivers se encuentran mutados, la combinación de mutaciones no es totalmente al azar y algunas tienden a aparecer juntas (coocurrencia) más frecuentemente que lo esperado por casualidad, mientras que otras son mutuamente excluyentes, puesto que rara vez son observadas juntas. Este hecho sugiere que las interacciones gen-gen, mutacionalmente hablando, están detrás de las selecciones clonales positivas y negativas durante todo el curso de la enfermedad [63,74]. Por ejemplo, las mutaciones en los componentes del splicing, así como las mutaciones en los genes del complejo cohesinas son consideradas prácticamente excluyentes, probablemente debido a la letalidad sintética que se produce cuando dos componentes de una misma vía celular se encuentran mutados [75,76]. Por el contrario, las mutaciones implicadas en el splicing y en la metilación del ADN parece que coocurren en estados tempranos de la enfermedad, siendo descritas como mutaciones "fundadoras" en más del 50% de los pacientes con SMD [77], del mismo modo que las mutaciones en *RUNX1* se asocian con alteraciones en *ASXL1* [78] o mutaciones en *STAG2* coocurren con otras en genes como *RUNX1*, *SRSF2*, *ASXL1* o *EZH2* [76,79,80].

Por tanto, los SMD se caracterizan por presentar mutaciones en múltiples genes drivers, una compleja estructura de interacciones gen-gen y una diversificación subclonal extensa, lo que demuestra una gran heterogeneidad tanto a nivel interpaciente como intratumoral. No obstante, a pesar de las bajas frecuencias mutacionales y la variabilidad interpaciente, algunas de estas mutaciones han demostrado relacionarse con el fenotipo de la enfermedad y poseer un valor pronóstico.

Algunos ejemplos de este valor pronóstico son las mutaciones en *TP53* que se asocian con un cariotipo complejo [45,81] o las mutaciones en *SF3B1* que se relacionan fuertemente con un mayor número de sideroblastos en anillo en la MO y normalmente predicen un pronóstico favorable [34,36,51]. Asimismo, la presencia de algunas mutaciones se ha asociado con un peor desenlace clínico y el propio número de mutaciones oncogénicas se relaciona negativamente con la supervivencia libre de leucemia (SLL), proporcionando información pronóstica independiente más allá de la estratificación del IPSS-R [61]. A diferencia del pronóstico favorable que suelen presentar los casos mutados en *SF3B1*, los pacientes con mutaciones en otros genes del splicing, como pueden ser *SRSF2* o *U2AF1*, presenta una peor supervivencia global [82-84]. Del mismo modo, mutaciones en *TP53*, *DNMT3A*, *RUNX1*, *EZH2* y *ASXL1* se han descrito como predictores de una peor supervivencia global [61,81,85-88]. Además, recientemente algunos estudios han reportado el impacto en el fenotipo de la enfermedad y en la evolución de los SMD de algunas concurrencias de mutaciones. En esta línea, la concurrencia entre mutaciones en *SRSF2* y en genes de la vía de RAS se ha asociado con leucocitosis, mientras que la concurrencia de mutaciones en *SRSF2* y *STAG2*, *RUNX1* y *IDH1/2* es más frecuente en fenotipos con presencia de blastos [80] o la concurrencia de mutaciones en *SF3B1* con otros genes específicos, como son *SRSF2*, *IDH2*, *BCOR*, *NUP98* y *STAG2*, modifica el pronóstico favorable de *SF3B1* y se asocia con un pronóstico adverso [89].

En la **Figura 4** se muestra un resumen de las mutaciones más frecuentes encontradas en pacientes con SMD y su impacto pronóstico asociado.

Por consiguiente, la identificación de mutaciones en diversas vías de señalización celular ha mejorado nuestro conocimiento acerca del impacto clínica de estas mutaciones, sugiriendo que las clasificaciones diagnósticas y pronósticas de los pacientes pueden ser refinadas por la inclusión de los datos genéticos. No obstante, a día de hoy, siguen siendo necesarias grandes bases de datos para el entendimiento de los mecanismos implicados en la patogénesis de los SMD y para adaptar las opciones terapéuticas de acuerdo a las mutaciones específicas. Además, la falta de validación clínica limita la implantación de los resultados de la secuenciación de estos genes en la práctica clínica habitual.

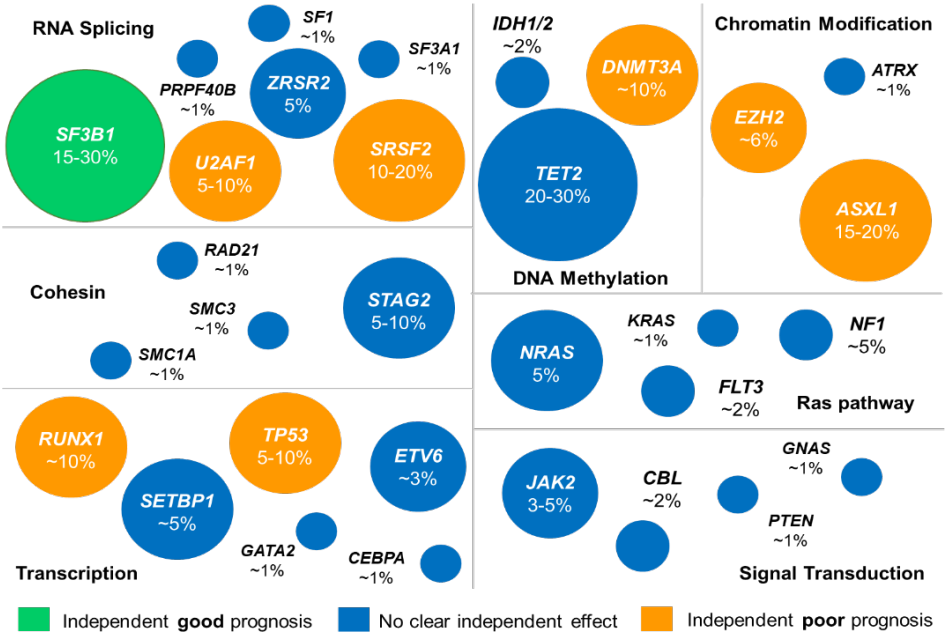


Figura 4. Resumen de las mutaciones más frecuentes en pacientes con SMD y su impacto pronóstico (adaptada de Steensma, *Mayo Clin Proc*, 2015)[23]. En esta figura se muestran las frecuencias mutacionales de los genes alterados en pacientes con SMD, así como su valor pronóstico, representado por colores: verde, genes cuyas mutaciones se relacionan con pronóstico favorables (como *SF3B1*); naranja, mutaciones que se asocian con un pronóstico adverso (como mutaciones en *SRSF2*, *ASXL1*, *DNMT3A*, *RUNX1*, *TP53*, *U2AF1* y *EZH2*); y azul, para aquellos genes cuyo impacto clínico no está definido.

2.2 El complejo cohesina

El complejo cohesina es un complejo proteico multimérico conservado a lo largo de las especies, el cual está compuesto por dos subunidades de mantenimiento estructural de los cromosomas (Structural Maintenance of Chromosomes, SMC), la subunidad SMC1A y la subunidad SMC3, junto con una subunidad α-kleisina llamada RAD21, que forma un anillo uniéndose a una de las subunidades de antígeno estromal (STAG1/2) [90,91] (**Figura 5**).

El complejo cohesina es conocido por su función canónica de alinear y estabilizar las cromátidas hermanas durante la metafase en la división celular [91]. Sin embargo, se han descrito en varias neoplasias mutaciones en las cohesinas que no se relacionan con aneuplodía o con alteraciones citogenéticas complejas [92,93], de modo que estos resultados sugieren que las cohesinas participan en un mecanismos alternativo que cuando se ve alterado permite la expansión clonal y la transformación maligna. En esta línea, recientemente se ha descrito el papel del complejo cohesina en la reparación del ADN tras la replicación [94-96] y su implicación en la activación transcripcional a través de las interacciones *cis* de largo alcance, en concreto, en la formación de bucles en el ADN entre enhancers y promotores en la que también participa el facto de unión CCCTC (CTCF) [91,97-101].

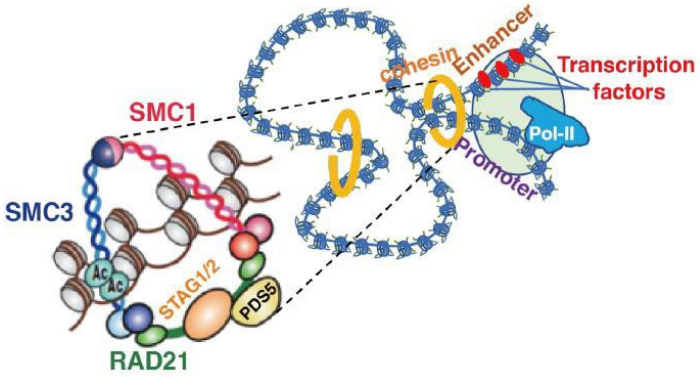


Figura 5. Representación del complejo cohesina (adaptada de Ogawa, *Blood*, 2019) [63]. El complejo cohesinas está compuesto por dos subunidades de mantenimiento de los cromosomas SMC1A and SMC3 junto con una subunidad α-kleisina llamada RAD21 que forma un anillo uniéndose a una de las subunidades de antígeno estromal (STAG1/2).

En el caso de las neoplasias mieloides, las mutaciones en los genes que codifican para las subunidades del complejo cohesina (*STAG2*, *STAG1*, *SMC3*, *SMC1A*, *RAD21* and *CTCF*) se han descrito en un 10-15% de los pacientes, siendo más comunes en los SMD de alto riesgo o en pacientes con LAM, donde la incidencia puede llegar hasta un 20% de los casos [76]. Además, gracias a análisis clonales de estos pacientes se ha descrito que son eventos tempranos [102] y gracias al estudio de sus frecuencias alélicas (Variant Allele Frequency, VAF) parece que estas mutaciones se encuentran en el clon dominante [93]. Del mismo modo, tras el estudio retrospectivo de la evolución clínica de pacientes mutados en estos genes se ha encontrado que las mutaciones en las cohesinas tienden a asociarse, aunque sin valor independiente, con un peor pronóstico en pacientes con neoplasias mieloides [62,76,93]. Con respecto a su papel en el desarrollo del SMD, como ha sucedido con otras neoplasias, algunos estudios han descrito que las mutaciones en las cohesinas no contribuyen a la transformación hematopoyética a través de la inestabilidad cromosómica [93,102], sino que pueden alterar la regulación transcripcional que sucede durante el proceso de diferenciación y proliferación y la reparación del daño en el ADN y, por consiguiente, participar en la evolución clonal y en la tumorogénesis [76,96]. Además, algunos autores sugieren que las mutaciones de las cohesinas pueden tener un impacto negativo en la hematopoyesis, generando un clon ancestral susceptible a un segundo "hit" que provoca la transformación leucémica [103].

Con respecto a las concurrencias de las mutaciones de las cohesinas en pacientes mieloides, se ha descrito que concurren con mutaciones en otros genes drivers en la evolución clonal, como son las mutaciones en el splicing (*SRSF2*) [79,104], en modificadores de la cromatina (*BCOR* and *ASXL1*), reguladores transcripcionales (*RUNX1*, *NPM1*) [76,93] y genes implicados en la vía de RAS [76]. Por tanto, estos datos sugieren que la desregulación epigenética puede ser sinérgica con respecto al mecanismo de acción afectado por la haploinsuficiencia de las cohesinas. Igualmente,

la vía de señalización de RAS cuando se encuentra activada puede alterar la expresión de genes claves, que están aún más desregulados por la inactividad de las cohesinas[76].

Sin embargo, los mecanismos moleculares por los que las mutaciones en las cohesinas promueven la leucomogénesis no están completamente establecidos y la significancia clínica de estas mutaciones aún no está clara.

2.3 Vía de señalización de Ras

RAS es un proto-oncogén que pertenece a la familia de las GTPasas pequeñas y está implicado en la transducción celular de señales [105,106]. Esta familia de genes está compuesta por los genes *NRAS*, *KRAS* y *HRAS*, los cuales codifican para proteínas homólogas que tienen la propiedad bioquímica de unirse a los nucleótidos de Guanina y poseer actividad GTPasa intrínseca. Por tanto, estas proteínas están involucradas en la transducción de las señales externas a la célula [107].

Las mutaciones activantes de la vía de Ras, es decir, las alteraciones que permiten que RAS esté en un estado activo y, por consiguiente, prevenga la hidrólisis de GTP, se encuentran muy habitualmente en un diverso rango de cánceres, incluyendo los SMD, la LAM, la leucemia mielomonocítica crónica (LMMC) y la leucemia mielomonocítica juvenil [108-112]. La activación de RAS permite un aumento de la señalización a través de las vías *RAS/RAF/MEK* y *RAS/PI3K* [113] y, por tanto, las mutaciones activantes de *RAS* dirigen un aumento de la proliferación de las células progenitoras hematopoyéticas [114,115]. Asimismo, además del efecto directo de las mutaciones activantes de *RAS*, la cascada de señalización dependiente de RAS puede verse inducida como respuesta a mutaciones en otros genes que codifican factores de niveles superiores de la vía de Ras (*FLT3*, *KIT*, *CBL*), genes involucrados en la regulación de dicha vía (*NF1*, *PTPN11*, *PTPN1*) o efectores de niveles inferiores de la vía de Ras (*BRAF*) [115,116].

En el caso de las neoplasias mieloides, las mutaciones de la vía de Ras se han descrito en un 10-15% de los pacientes diagnosticados de LAM, sin embargo, son menos frecuentes en pacientes con SMD, encontrándose en un 5-10% de los casos y siendo *NRAS* el gen más común [111,117,118]. Aunque se ha estudiado ampliamente el impacto pronóstico de las mutaciones en estos genes en pacientes con neoplasias mieloproliferativas (NMP) y con LAM, la información disponible respecto al pronóstico de los pacientes con SMD que portan mutaciones en los genes de la vía de Ras es limitada y contradictora [111].

Algunos trabajos han descrito que las mutaciones de la vía de Ras no tienden a relacionarse con un pronóstico adverso [111,119], mientras que estudios más recientes han reportado que estas mutaciones pueden promover la proliferación celular y

asociarse con un mayor riesgo de progresión a LAM y un peor pronóstico [120-122]. Del mismo modo, los pacientes con mutaciones en la vía de Ras son más frecuentemente diagnosticados con un SMD con exceso de blastos y con subtipos pertenecientes a las categorías de alto riesgo del IPSS-R, así como suelen presentar un mayor recuento de leucocitos y un mayor número de blastos en MO [111,123]. Con respecto a su función y papel en la patofisiología de los SMD, las mutaciones en la vía de Ras pueden provocar una activación de la vía de señalización constitutiva siendo hipersensible a los estímulos endógenos y, por consiguiente, estas mutaciones pueden aumentar la proliferación, reducir la apoptosis y alterar el crecimiento celular y la supervivencia [4]. Sin embargo, actualmente se necesitan estudios funcionales en células de SMD y modelos in vivo para confirmar estos efectos.

En pacientes diagnosticados con SMD, las mutaciones de la vía de Ras se han descrito como eventos tardíos en el curso de la enfermedad, siendo alteraciones subclonales que concurren con otras mutaciones fundadoras [69]. En concreto, los pacientes con mutaciones en la vía de Ras presentan más comúnmente mutaciones en *U2AF1*, *RUNX1*, *WT1*, *PTPN11*, *ETV6* y *NPM1* [123], mientras que no suelen presentar mutaciones en *FLT3* (mutaciones de tipo ITD o TKD) [111].

Por tanto, la información disponible sobre el impacto clínico de las mutaciones de la vía de Ras en pacientes con SMD es limitada y contradictoria y los mecanismos moleculares por lo que la vía de Ras está implicada en la patogénesis de los SMD y en su progresión aún son desconocidos.

3. Progresión de los SMD a Leucemia Aguda Mieloblástica secundaria

El curso clínico de los SMD es altamente variable, pero aproximadamente un tercio de los pacientes termina progresando a LAM, conocida como LAM secundaria o LAMs, lo que conlleva un pronóstico lúgubre [42,57,124].

La LAMs representa hasta un 25-35% del total de casos con LAM, proviniendo la mayoría de ellos (60-80%) de un SMD previo [125-127], y estos pacientes presentan características clinicopatológicas distintas que la LAM *de novo*. Por este motivo, en la clasificación de la OMS de 2017 de neoplasias hematológicas, la LAM proveniente de un SMD previo es clasificada como una entidad diferente, llamada LAM con cambios displásicos (SMD-LAM) [32], la cual presenta tasas de remisión completa más bajas, así como una supervivencia libre de recaída y una supervivencia global menor a la observada en la LAM *de novo* [127-129].

Aunque los SMD y la LAM secundaria son consideras como entidades diferentes, en realidad, representan un proceso continuo de la enfermedad sometido a una evolución clonal genética. De hecho, como se define en el IPSS-R, la supervivencia global de los SMD de alto riesgo y muy alto riesgo [41] es similar a la observada en los pacientes con SMD-LAM [130], indicando que los SMD de riesgo más alto y la LAMs son entidades clínicamente similares y ambas podrían representar dos estadios cancerígenos de una enfermedad continua. Sin embargo, estas dos entidades muestran algunas diferencias celulares. La LAM se caracteriza por un completo bloqueo en la maduración celular hematopoyética, permitiendo la acumulación de mieloblastos en MO (≥20% de las células nucleadas), mientras que los pacientes con SMD producen células sanguíneas maduras, pero deficientes, debido a que la maduración y la morfología son aberrantes y, consecuentemente, se observa un número de células maduras bajo. Además, los SMD normalmente se caracterizan por presentar un aumento de la apoptosis y de la muerte celular inflamatoria, mientras que la LAMs, y la LAM en general, se asocia con un aumento de la supervivencia celular y la proliferación, sugiriendo que el aumento del fenotipo de muerte celular que sucede en el SMD deja paso a un fenotipo pro-supervivencia durante la progresión a LAMs [56].

Por definición, el recuento de blastos en la MO es el parámetro usado para diferenciar entre SMD y LAMs. Sin embargo, es una medida subjetiva, basada en una valoración morfológica que es sensible a la variabilidad del propio observador [131,132], además de que posee ciertas limitaciones debido a que la progresión de la enfermedad es un proceso continuo y definir un límite es bastante complejo. Por este motivo, la definición de transformación a LAMs ha evolucionado durante el tiempo y, aunque, actualmente el límite se encuentra en ≥20% de blastos en MO por la OMS en la anterior clasificación de la FAB el límite era de ≥30% [133]. Además, estudios genéticos

recientes han identificado clonalidad en los SMD, puesto que casi la totalidad de los pacientes tienen mutaciones clonales lo que supone una medida de carga tumoral que a menudo excede al porcentaje de blastos [69,134-136]. Por estas razones, la incorporación de datos de secuenciación en muestras secuenciales podría ser útil para la monitorización de los cambios dinámicos en la carga tumoral que suceden durante el SMD, y en combinación con el recuento de blastos y otros parámetros clínicos podría ayudar a la anticipación e identificación de progresión de manera más temprana.

3.1. La genética de la progresión de los SMD a LAMs: Evolución clonal

Los SMD y la LAMs muestran alteraciones citogenéticas similares, como son las deleción del cromosoma 5q, 7q, 20q o el cariotipo complejo, las cuales son definidas como alteraciones citogenéticas "asociadas al SMD". De hecho, la presencia de estas alteraciones define el diagnóstico de la OMS de LAM-SMD, incluso en ausencia de un diagnóstico previo de SMD o sin displasia morfológica. Por el contrario, los reordenamientos balanceados, como son las translocaciones t(15;17) y t(8;21), la inversión del cromosoma 16 o los reordenamientos en *KMT2A* son más comunes en la LAM *de novo* [32,64].

Del mismo, los estudios de secuenciación masiva a gran escala han identificado mutaciones en genes comunes entre los SMD y la LAMs, tales como son las mutaciones en genes del spliceosoma (*SRSF2*, *SF3B1*, *U2AF1*), en *EZH2*, *BCOR* y *STAG2*, cuya presencia sugiere que la LAM ha evolucionado desde un SMD, incluso sin un antecedente previo de diagnóstico de SMD [79,137-139]. Por tanto, estas mutaciones ocurren en un momento temprano de la patogénesis del SMD anterior a la progresión a LAMs.

Estos estudios de secuenciación masiva también han mostrado que la MO de pacientes con SMD puede ser clonalmente compleja con múltiples clones únicos, pero relacionados genéticamente, incluyendo un clon fundador y subclones derivados de este clon fundador [69,134,140]. A pesar de su complejidad, la progresión a LAMs ha sido definida como un proceso dinámico, donde el clon fundador del SMD persiste, aunque nuevos subclones con alteraciones genéticas diferentes crecen o aparecen, desencadenando un aumento de la heterogeneidad intraclonal [69]. Por tanto, la evolución a LAMs es un proceso de expansión clonal, que incluye diversificación genética y epigenética, y de selección clonal por la adaptabilidad dinámica, lo que resulta en una arquitectura clonal compleja, es decir, en una MO con gran heterogeneidad intraclonal.

En este contexto de heterogeneidad intraclonal, la concurrencia de mutaciones genéticas, así como el orden de adquisición son cruciales para la patogénesis de la enfermedad, aunque establecer el orden exacto de la adquisición de mutaciones en un paciente puede ser complicado.

Recientemente, algunos estudios han descrito que alrededor del 10% de las personas mayores de 70 años presentan mutaciones somáticas en sus células sanguíneas, lo que se conoce como Hematopoyesis Clonal de Potencial Incierto (Clonal Hematopoiesis of Indeterminate Potential, CHIP) [58,141,142]. Muchas de estas mutaciones (*DNMT3A*, *TET2*, *ASXL1* y genes del spliceosoma) se encuentran comúnmente mutadas en altos VAFs tanto en pacientes con SMD y LAMs, lo que podría indicar que las mutaciones en estos genes ocurren en un desarrollo temprano de la enfermedad y están presente en la mayoría de las células, es decir, están presentes en el clon fundador [69,134,136]. En cambio, otras mutaciones son más comunes en la LAMs, como las mutaciones en los factores de transcripción (*RUNX1*, *GATA2*, *CEBPA*) o en genes relacionados con la señalización (*RAS* family genes, *FLT3*), sugiriendo que estas mutaciones se adquieren más tarde durante la progresión de la enfermedad en un subconjunto de células en expansión [120,136,143,144].

En esta línea, Makishima *et al.* identificaron 2 grupos de mutaciones asociadas con la evolución de un SMD a LAMs: mutaciones en genes como *FLT3*, *PTPN11*, *WT1*, *IDH1*, *NPM1*, *IDH2* y *NRAS* que se encontraron enriquecidas en pacientes con LAMs comparados con SMD de alto riesgo, conocidas como mutaciones tipo-1; y las mutaciones tipo-2 que se encontraron enriquecidas en pacientes de alto riesgo en comparación con SMD de bajo riesgo (mutaciones en *TP53*, *GATA2*, *KRAS*, *RUNX1*, *STAG2*, *ASXL1*, *ZRSR2* y *TET2*)[69].

Considerando todos estos resultados juntos, las mutaciones parecen seguir un orden típico de adquisición durante la progresión. Sin embargo, nuestro conocimiento acerca del orden de adquisición de mutaciones permanece incompleto, puesto que la mayoría de los datos están basados en la frecuencia de las mutaciones en series de SMD y LAMs que no incluyeron muestras pareadas en lugar de en series que incluyan muestras seriadas de un mismo paciente.

Con el fin de superar esta limitación, estudios recientes han incluido algunas muestras pareadas de SMD/LAMs de un mismo paciente y han confirmado que existe un orden típico de adquisición de mutaciones durante la progresión de los SMD a LAM. Las mutaciones en los genes del spliceosoma y en modificadores epigenéticos están presentes con altos VAFs ya en el momento de la muestra del SMD , indicando que son adquiridas de manera temprana en el curso de la enfermedad, mientras que las mutaciones en genes de señalización raramente fueron detectadas en el SMD, indicado que son adquiridas más tarde en la progresión de la enfermedad [69,134-136]. Estos resultados concuerdan con el modelo simplista tradicional de evolución a LAMs que postula que las mutaciones pueden suceden en dos clases de genes [145,146]. Este

modelo propone que las mutaciones de pérdida de función en factores de transcripción tales como *RUNX1*, *CEBPA* y *RARA* (denominadas como mutaciones de "Clase II"), alteran la diferenciación hematopoyética y permiten la acumulación de más células primitivas. Sin embargo, estas variaciones solas no son suficientes para causar leucemia aguda y necesitan de una segunda mutación (mutaciones de "Clase I") en genes de señalización, tales como *FLT3*, *RAS* y *KIT*, para promover la ventaja proliferativa y de supervivencia de las células madre hematopoyéticas [147-149]. No obstante, estos estudios seriados también han descrito que el espectro mutacional es más complicado que lo que puede explicar un modelo de dos clases y que la progresión a LAMs es un proceso muy complejo y heterogéneo que incluye otros genes frecuentemente mutados que no encajan en ninguna de las categorías de Clase I y Clase II [133].

Además, los enfoques de secuenciación que se han utilizado en estos estudios han limitado la capacidad de detectar consistentemente mutaciones en bajos niveles y definir clonalidad. Por tanto, cuando las mutaciones están presentes en bajos VAFs la definición de la arquitectura clonal se ve limitada, siendo todo un reto determinar si las mutaciones coocurren en la misma célula o si existen en subclones paralelos. En estos casos, se necesitaría la metodología de secuenciación de célula única, sin embargo, esta técnica puede verse afectada por el fenómeno de los alelos nulos, o lo que es lo mismo, la pérdida del alelo mutado durante la amplificación del ADN durante la PCR debido a variaciones en la secuencia o mutaciones [56].

Teniendo en cuenta todos estos resultados, la progresión de los SMD a LAMs es claramente un proceso de evolución clonal, que puede ocurrir desde un periodo relativamente corto de tiempo (por ejemplo, semanas o meses). El análisis de la evolución clonal ha mostrado que la evolución puede seguir dos patrones: lineal, donde cada subclon sucesivo ocurre procedente de un clon parental previo, o ramificado, donde un clon fundador puede generar varios subclones paralelos [69,135,136]. Además, se ha identificado un proceso denominado "barrido clonal", el cual sucede cuando un subclon se expande hasta ocupar el tumor entero y fuerza hasta el colapso al resto de subclones [69] (**Figura 6**).

Por otra parte, recientemente Chen *et al.* han realizado experimentos de secuenciación masiva en un número limitado de poblaciones de células madre definidas fenotípicamente en pacientes con SMD y LAMs. Estos autores han descrito que la diversidad subclonal puede ser mayor en las células madre que en los blastos y que el clon dominante en las células madre de los SMD no siempre es el mismo como sucede con el clon dominante en la LAM, lo que indica una posible ruta no lineal de evolución clonal a nivel de las células madre [140]. Por tanto, estos resultados sugieren que existe un mayor nivel de diversidad clonal que el apreciada en muestra de MO total, planteando preguntas interesantes acerca de la diversidad clonal de las células madre en SMD y LAMs.

Por tanto, en este escenario de "juego de clones" que ocurre durante la progresión de los SMD a LAMs, son necesarios un mayor número de estudios que incluyan muestras pareadas de SMD y LAMs para proporcionar más información sobre los patrones genéticos de expansión clonal que suceden durante la progresión de la enfermedad. Además, estos estudios secuenciales también podrían proporcionar información sobre un orden crítico de adquisición de mutaciones que conduce la evolución de la enfermedad y predice la progresión a LAMs, lo que potencialmente permite una intervención temprana.

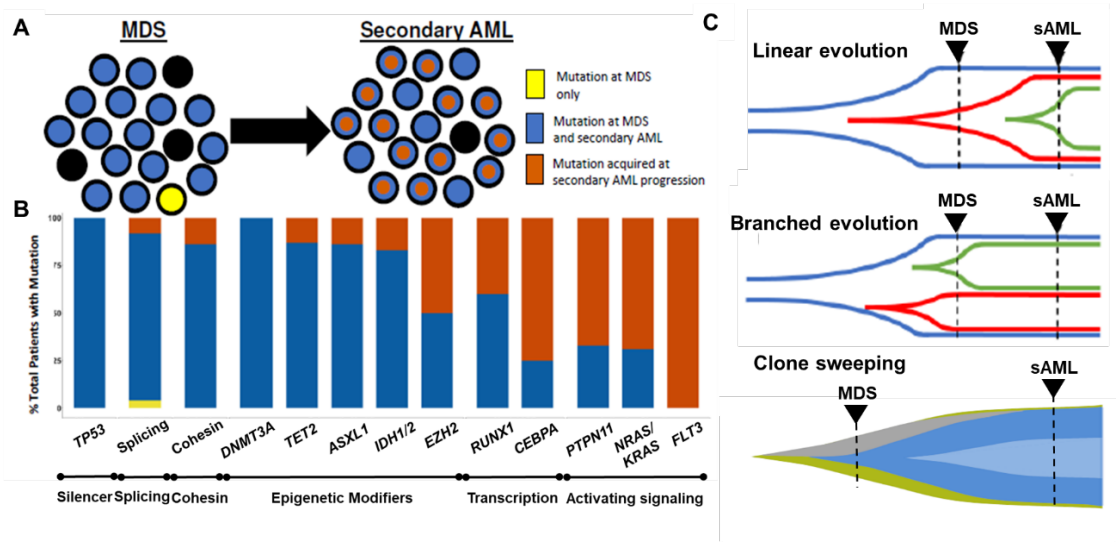


Figura 6. Evolución clonal durante la progresión de los SMD a LAMs (adaptada de Menssen, *Blood*, 2020) [56]. A) Modelo para la acumulación secuencial de mutaciones durante la progresión de los SMD a LAMs. Las células negras son normales tanto en el estado de SMD y LAMs. Las mutaciones en azul y Amarillo son adquiridas de manera temprana y están presentes en el clon fundador. Las mutaciones azules se expanden en la médula ósea durante el momento de la LAM, mientras que las mutaciones amarillas en el momento de la LAMs por el fenómeno del barrido clonal. Algunas células adquieren mutaciones naranjas, forman un subclone y se expande en el momento de la progresión a LAMs. **B) Porcentaje de pacientes con una mutación detectada en los momentos del SMD y la LAMs**. Las mutaciones detectables solo en el momento del SMD se representan en amarillo, las detectadas solo en el momento de la LAMs en naranja y las mutaciones detectadas en el momento del SMD que persisten durante la progresión de la enfermedad se representan en azul. **C) Patrones de evolución clonal durante la progresión de los SMD a LAMs.** Durante la progresión de los SMD a LAMs se observan tres patrones de expansión subclonal posibles: evolución lineal, evolución ramificada y barrido clonal.

3.2. Tratamiento y evolución clonal de los SMD

La evolución clonal que sucede durante la progresión a LAMs se caracteriza por la persistencia de un clon fundador mutado junto con la expansión de subclones que portan otras mutaciones. Aunque este patrón es similar para los pacientes que reciben tratamiento de soporte o con fármacos como la eritropoyetina, agonistas del receptor de trombopoyetina o con el factor estimulante G-CSF, otros fármacos pueden influir en los patrones de evolución clonal [56].

Una de las alteraciones citogenéticas más comunes en pacientes con SMD y LAMs es la deleción del cromosoma 5q, del(5q). La lenalidomida se relaciona con una disminución de las células clonales de la MO que portan la del(5q) y los pacientes que portan esta alteración y son tratados con este fármaco pueden alcanzar la remisión completa [150,151]. Sin embargo, los pacientes tratados con lenalidomida que finalmente recaen normalmente presentan una expansión de las células con la alteración del(5q) y otras alteraciones adicionales, como mutaciones en *TP53* [135,152]. Estas mutaciones en *TP53* se han encontrado en niveles bajos en los pacientes meses antes de la progresión de la enfermedad [153], lo que sugiere que el tratamiento con lenalidomida puede ejercer una presión selectiva que impacta en la evolución clonal que tiene lugar durante la progresión a LAMs.

Los agentes hipometilantes (HMAs) son uno de los fármacos más comunes en el tratamiento de los SMD y son considerados el tratamiento estándar desde hace más de una década. Aunque su uso se ha asociado con una supervivencia global más prolongadas en los ensayos clínicos controlados que se han realizado [154-157], un número alto de pacientes no responden al tratamiento o pierden la respuesta inicial [52,54,55,158], por lo que predecir la respuesta al tratamiento de los pacientes en base a sus mutaciones sigue siendo un reto y, en la actualidad, los resultados de los estudios realizados son muy variables. De manera interesante, algunos estudios han mostrado que el tratamiento con HMA puede inducir la transversión C>G en genes implicados en la patogénesis del SMD o la LAM [159,160]. Además, estas mutaciones C>G pueden ocurrir en un subclon que emerge en la recaída [161], lo que indica que los agentes hipometilantes HMA pueden influir en la evolución clonal, planteando preguntas sobre si los HMA pueden inducir mutaciones patogénicas.

Con respecto al trasplante de progenitores hematopoyéticos, éste es el único tratamiento curativo para los pacientes con SMD y LAMs, mejorando significativa su evolución clínica. Sin embargo, la recaída sigue siendo común [52,162]. Un estudio reciente sobre pacientes que progresaron después del trasplante ha descrito que, en todos los casos, durante la recaída son detectables las mutaciones del clon fundador y que los subclones que emergen en la progresión a menudo portan una nueva variante estructural [161], lo que sugiere que existe un patrón de evolución clonal y progresión de la enfermedad tras el trasplante, el cual podría ser útil en la monitorización molecular de enfermedad residual de los pacientes.

Tomando todos estos hallazgos en conjunto, las opciones terapéuticas disponibles actualmente para los pacientes con SMD no previenen que un número sustancial de pacientes presente progresión de la enfermedad y evolucione a LAMs. Además, algunos tratamientos como la lenalidomida, los agentes hipometilantes y el alo-TPH pueden influir en los patrones de evolución clonal que tiene lugar durante la progresión a LAMs.

4. Secuenciación masiva en el estudio de los SMD

La secuenciación del ADN es una herramienta muy relevante en la investigación biológica. La técnica más utilizada por más de tres décadas ha sido la secuenciación tradicional Sanger desde su descubrimiento en 1977 [163]. Esta metodología se basa en el uso de nucleótidos marcados fluorescentemente o radiactivamente que carecen del grupo hidroxilo del carbono 3' (dideoxinucleótidos) [163-165]. Sin embargo, esta técnica solo permite el descubrimiento de sustituciones y de pequeñas inserciones y deleciones (indels), de modo que han surgido nuevas metodologías más prometedoras que permiten una secuenciación más rápida y más precisa al realiza múltiples reacciones en paralelos [166].

La secuenciación masiva (Next Generation Sequencing, NGS) es la tecnología de secuenciación del ADN que ha revolucionado la investigación genómica [164,167]. La mayoría de los métodos conocidos como "secuenciación en paralelo" o "secuenciación masiva" se basan en el mismo principio enzimológico que la secuenciación Sanger, pero en una perspectiva orquestada y escalonada, que permite la secuenciación de decenas de miles e, incluso, billones de moldes simultáneamente [168].

Existen 3 tipos de secuenciación masiva del ADN para la identificación de mutaciones genómicas: secuenciación de genoma completo (whole-genome sequencing, WGS), secuenciación de exoma completo (whole-exome sequencing, WES) y secuenciación dirigida (targeted-deep sequencing, TDS) [169]. La secuenciación WGS proporciona información sobre el genoma completo, pero como aproximadamente el 98% del genoma son regiones no codificantes y se conoce muy poco acerca del papel de estas regiones en la patogénesis del SMD, no es un método ampliamente utilizado en el manejo clínico de los SMD [170]. En cambio, la secuenciación WES ha jugado un papel importante en el estudio de las neoplasias mieloides, particularmente en el estudio de la evolución clonal de los SMD y LAMs [134,135]. Aunque la secuenciación WGS y WES pueden ser consideradas *a priori* las mejores opciones por el gran número de regiones evaluadas, en realidad poseen limitaciones para incorporarse a la práctica clínica habitual. Algunas de estas limitaciones son los elevados costes, los análisis bioinformáticos trabajosos y la laboriosa interpretación clínica que requieren los resultados [171-173]. Además, las metodologías de WGS y WES pueden omitir algunas mutaciones subclonales de bajo VAF que son relevantes clínicamente, como las mutaciones en *TP53* [174]. Por estas razones, la secuenciación dirigida TDS, que permite la secuenciación solo de unas regiones de interés, es un enfoque poderoso que puede cumplir el mejor balance entre la identificación precisa y con gran sensibilidad de los eventos de interés y el coste total y la carga de datos que requieren estudios a gran escala. Adicionalmente, la profundidad de lectura de la TDS es mayor permitiendo la detección de alteraciones en bajos VAF, lo que es clave en el estudio de los SMD puesto que presentan una alta heterogeneidad intratumoral

y la inmensa mayoría de las variantes patogénicas conocidas se encuentran en un número relativamente bajo de genes [61,62,165].

Dentro de la secuenciación dirigida, existen dos tipos de estrategias, la secuenciación de amplicones o de captura. La secuenciación de amplicones utiliza primers diseñados específicamente para amplificar únicamente las regiones de interés antes de la preparación de la librería. Por el contrario, en la secuenciación por captura el ADN es fragmentado y, posteriormente, las regiones de interés son capturadas o enriquecidas vía hibridación con oligonucleótidos unidos a sondas biotiniladas, permitiendo el aislamiento de las regiones de interés [175-177]. El enriquecimiento basado en amplicones es el más barato de ambos enfoques, obtiene un mayor número de lecturas y requiere de mucho menos material de partida que la captura, por lo que es la opción ideal cuando el ADN de partida no suficiente para la secuenciación por captura. Sin embargo, la cobertura de las regiones de interés es más uniforme con un abordaje de captura [169]. Además, se ha demostrado que con la captura se obtiene un menor número de duplicados de PCR que con la secuenciación por amplicones (<40% *vs.* ~80%, respectivamente) y éstos pueden ser eliminados computacionalmente más fácilmente, puesto que la fragmentación aleatoria del ADN que se produce en la captura reduce la probabilidad de que dos fragmentos únicos alineen en las mismas coordenadas genómicas. Del mismo modo, las secuencias utilizadas como sondas en la captura son más largas lo que permite la secuenciación de regiones difíciles, como las regiones repetitivas, y un mayor nivel de especificidad [169]. Por tanto, las técnicas basadas en la captura proporcionan una selección más precisa y uniforme de las regiones de interés, mientras que la secuenciación por amplicones suele ser más útil en experimentos de pequeña escala donde la cantidad de muestra y el precio son factores limitantes.

Un resumen de los principales tipos de secuenciación masiva o NGS del ADN y de los dos tipos de estrategias de secuenciación dirigida para la identificación de mutaciones genómicas es mostrado en la **Figura 7**.

Todos los abordajes de secuenciación masiva implican una serie de pasos generales comunes: 1) preparación de la muestra y elección de la estrategia de secuenciación del ADN; 2) preparación de la librería; 3) secuenciación; y 4) análisis de los datos.

Preparación de la muestra y elección de la estrategia de secuenciación del ADN

El primer paso en la secuenciación masiva es la extracción del material genético, el ADN o el ARN. Aunque este paso puede considerarse trivial, la cantidad y la calidad del material de partida es un aspecto crucial, ya que la generación de microgramos de ADN desde un material de partida por PCR es un proceso sensible sujeto a múltiples variaciones.

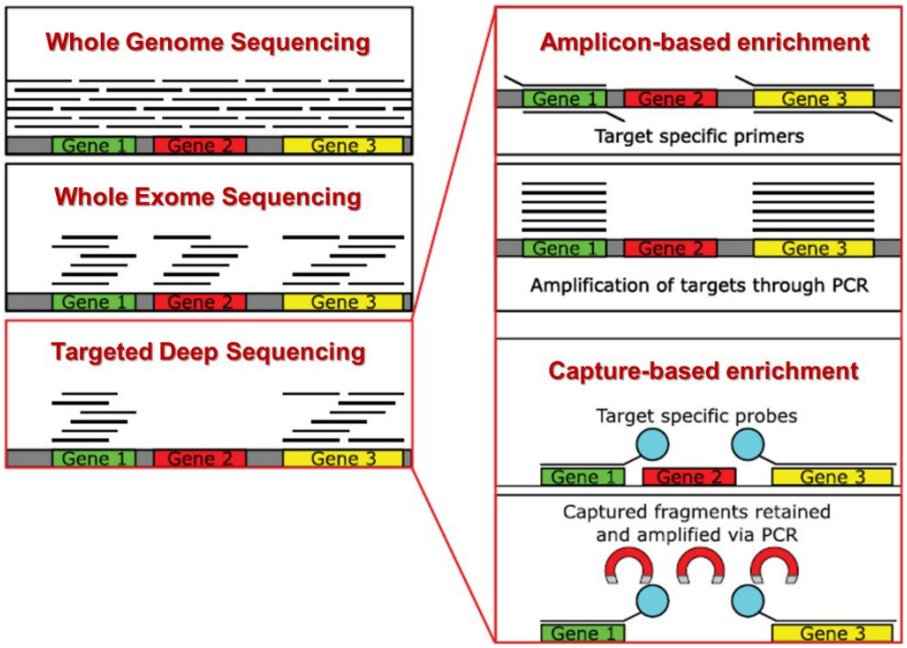


Figura 7. Los principales tipos de secuenciación masiva del ADN (adaptada de Bewicke-Copley, *Comput Struct Biotechnol J*, 2019) [169]. Se muestran la secuenciación de genoma y exoma completo y la secuenciación dirigida, así como las dos estrategias de secuenciación dirigida (enriquecimiento con amplicones o por captura) que suelen utilizarse para la identificación de las mutaciones genómicas.

De manera ideal, una librería de secuenciación debe representar a la perfección el ADN presente en la muestra biológica de partida y no puede crear más información que la que estaba presente en el molde original. Las librerías pueden generarse partiendo desde pequeñas cantidades de ADN, pero reducir la cantidad de partida puede comprometer la sensibilidad del experimento debido a la pérdida de heterogeneidad por una sobreamplificación de determinadas moléculas de ADN durante el proceso de generación de la librería (lecturas duplicadas). En concreto, estas lecturas duplicadas pueden provocar un fallo en la detección de variantes que estaban presentes en la muestra original, sobre o infra valorar determinadas variantes e identificar falsos positivos resultantes de errores de PCR que se han propagado a lo largo de la preparación de la librería y la secuenciación [178,179].

Del mismo modo, la elección de la estrategia de secuenciación es una de las primeras cosas a tener en cuenta. Como se ha descrito previamente, existen varios tipos de enfoques de secuenciación que difieren en la capacidad total de secuencia, la longitud de las lecturas, el tiempo de la carrera de secuenciación y la calidad y precisión de los datos obtenidos. Por tanto, estas características deben valorarse a la hora de considerar qué estrategia se adapta mejor a cada aplicación clínica específica[173].

Preparación de la librería

La preparación de la librería es un proceso que consiste en la fragmentación del ADN genómico en pequeños fragmentos de un tamaño particular, normalmente entre 100 y 500 pares de bases, los cuales tienen unidos en ambos extremos unos adaptadores específicos para cada plataforma.

Como se ha especificado previamente, existen dos estrategias para el enriquecimiento de las regiones de interés durante la preparación de la librería: enriquecimiento por amplicones o por captura. En el caso de los ensayos de captura, la preparación de la librería comienza con la fragmentación aleatoria del ADN genómica, por medios mecánicos o enzimáticos, seguida de la captura de las regiones de interés usando sondas de oligonucleótidos biotinilados complementarios a estas regiones y, por último, los fragmentos de interés se aíslan usando bolas magnéticas de estreptavidina. Por el contrario, la estrategia por amplicones conlleva la amplificación por PCR de las regiones del genoma de interés usando primers, por lo que las lecturas generadas tendrán las mismas coordenadas de inicio y fin marcadas por el diseño de los primers [165,168,173].

En ambos casos, los adaptadores específicos de la plataforma se unen a cada uno de los extremos de los fragmentos moldes. Estos adaptadores son complementarios a los oligonucleótidos utilizados, los cuales se unen a unas bolas o a una superficie sólida, llamada "flow cell", permitiendo la generación de la librería por la hibridación en puente que sucede en el secuenciador [165]. Por tanto, los fragmentos moldes de ADN son inmovilizados en la superficie de la flow cell que está diseñada para que el ADN se presente accesible a las enzimas a la vez que garantiza una alta estabilidad del molde de ADN unido a la superficie y una baja unión no específica de los nucleótidos marcados con fluorescencia.

Secuenciación

La mayoría de las plataformas comerciales usadas en la secuenciación masiva se basan en el concepto de la secuenciación por síntesis. En esencia, esta metodología permite la incorporación de nucleótidos usando una variedad de enzimas y esquemas de detección que permiten a la plataforma correspondiente recopilar los datos al mismo tiempo que sucede la síntesis enzimática del molde [168].

La secuenciación por síntesis consiste en la amplificación en paralelo de fragmentos de ADN usando la hibridación en puente y la PCR en presencia de nucleótidos marcados con fluorescencia. En detalle, después de la hibridación en la flow cell de las librerías amplificadas, los fragmentos de ADN actúan como molde de una nueva hebra complementaria la cual es sintetizada por PCR. Tras la síntesis de esta hebra complementaria, los moldes son eliminados y a medida que las hebras complementarias se hibridan en puente con los sitios de unión de los adaptadores vecinos, los ciclos consecutivos de PCR conducen a la amplificación por grupos o

clusters [165,168,180]. Por último, el secuenciador detecta la fluorescencia generada por la incorporación de los nucleótidos marcados en la hebra creciente de ADN [180,181]. Durante cada ciclo de secuenciación, se añade un único nucleótido marcado a la cadena de ADN y éste sirve como terminador de la polimerización, por lo que después de la incorporación de cada nucleótido la fluorescencia se visualiza para identificar la base y, posteriormente, éste es eliminado enzimáticamente para permitir la incorporación del siguiente nucleótido. El resultado es una secuenciación base a base muy precisa que elimina los errores específicos de secuencia y permite una robusta llamada de bases a lo largo del genoma, incluyendo las regiones repetitivas y con homopolímeros.

Sin embargo, la secuenciación por síntesis presenta algunas limitaciones. Por ejemplo, su longitud de secuencia (longitud de lectura) es limitada porque a medida que avanza la incorporación de nucleótidos y los ciclos de imagen aumenta el ruido y, por este motivo, aún su longitud de lectura es menor que la de la secuenciación Sanger. Por tanto, estas limitaciones deben tenerse en cuenta puesto que han impactado en cómo se analizan los datos de secuenciación [168].

Los principales pasos en la preparación de la librería, en la hibridación en puente, la amplificación por grupos o clusters y la secuenciación por síntesis son mostrados en la **Figura 8**.

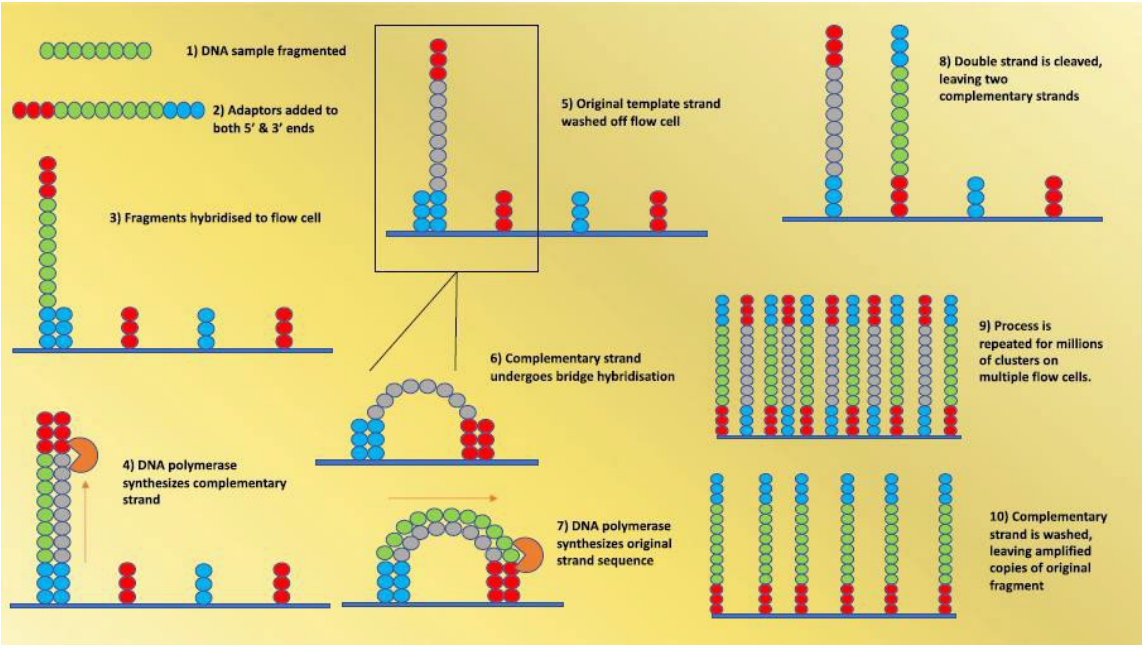


Figura 8. Pasos de la secuenciación masiva (adaptada de Spaulding, *Br J Haemat*, 2020) [165]. Se muestran los pasos de la preparación de la librería, la hibridación en puente, la amplificación por grupos o clusters y la secuenciación por síntesis.

Análisis de los datos

La secuenciación masiva genera grandes cantidades de datos que requieren una importante estructura computacional para el almacenaje, análisis e interpretación. Por tanto, el análisis de los datos de secuenciación masiva puede ser un proceso extremadamente complejo, dilatado en el tiempo y laborioso.

El flujo de trabajo del análisis de los datos de secuenciación incluye cuatro fases: 1) la llamada de bases, la asignación de la precisión y calidad de cada base y el demultiplexado; 2) el alineamiento de las lecturas frente a una referencia y la llamada de variantes, 3) la anotación de variantes; y 4) la interpretación de las variantes relevantes clínicamente [182]. Actualmente, existen un gran número de herramientas bioinformáticas y softwares para analizar los datos de secuenciación. La mayoría de ellos son algoritmos especializados en una única fase del análisis, sin embargo, existen múltiples *pipelines* que agrupan varios algoritmos y analizan los datos de manera seriada.

La primera fase del análisis ocurre en el mismo secuenciador, donde se generan las lecturas. Esta fase consiste en: i) la **llamada de bases**, que es la identificación de los nucleótidos específicos presentes en cada posición de una única lectura; ii) la **asignación de la precisión de cada base** medida por el score de calidad Phred (Q score), que indica la probabilidad de que una base dada haya sido anotada incorrectamente por el secuenciador; y iii) el **demultiplexado**, el cual consiste en la asociación computacional de las lecturas con un paciente cuando varias muestras se han agrupado o multiplexado en una carrera de secuenciación [182,183].

La segunda fase del análisis de secuenciación es el **mapeo y alineamiento** de cada lectura frente a un genoma humano de referencia, el cual está disponible en el Consorcio de Genomas de Referencia (la última versión puede encontrarse en http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/). Las herramientas de alineamiento usados en los laboratorios clínicos incluyen múltiples algoritmos que a veces cambian la velocidad por la precisión y pueden incorporar diferentes niveles de sensibilidad y especificidad. Sin embargo, se recomienda que el alineamiento incluya un procesamiento adicional para el realineamiento local, la eliminación de los duplicados de PCR y la recalibración de los scores de calidad de cada base. Como consecuencia del alineamiento, el siguiente paso es la llamada de variante, que es el proceso que compara las lecturas alineadas con un genoma de referencia para identificar las variaciones de nucleótidos. Hay una amplia variedad de softwares y estrategias disponibles para la llamada de variantes, pero ninguna de ellas actualmente es capaz de identificar todas las clases de variantes con la misma precisión. Por tanto, deben valorarse varios programas de llamada de variantes y/o la configuración de los parámetros para optimizar la detección de diferentes tipos de variantes [169,182,184].

A continuación de la llamada de variantes, la siguiente fase es la **anotación de variantes** en relación a los genes donde se encuentran, codón y posición del aminoácido y clasificar los tipos de variantes, como son las variantes sin sentido (nonsense), con cambio de sentido (missense), deleciones exónicas y variantes sinónimas, lo que permite comprender mejor las consecuencias funcionales en los genes identificados. En muchos estudios, solo las mutaciones del splicing y las exónicas no sinónimas son seleccionadas para su posterior análisis, centrándose solo en las mutaciones o en las variantes codificantes que tienen un efecto funcional. Sin embargo, estos criterios pueden cambiar dependiendo el propósito del estudio, por ejemplo, las variantes en las regiones UTRs o en los promotores pueden ser interesantes en determinados estudios [169].

La última fase del análisis de secuenciación es la **evaluación clínica** para determinar qué variantes deben ser reportadas y cómo en el informe de laboratorio. Existen algunos algoritmos que optimizan los ajustes para predecir las variantes deletéreas localizadas en genes clínicamente relevantes y para filtrar determinadas variantes como aquellas con una alta frecuencia en la población. Del mismo modo, existen algunas bases de datos, como ClinVar, COSMIC, ClinGen, algunos predictores proteicos o Human Variome Project, que pueden ser útiles para la evaluación de la relevancia de las variantes y para categorizarlas de acuerdo a su posible impacto funcional (benigna, potencialmente benigna, de significado incierto, potencialmente patogénica o patogénica). Sin embargo, la interpretación de las variantes es un gran reto porque no existe un formato estándar y común entre laboratorios. Por lo tanto, se necesitan colaboraciones y discusiones continuas entre los investigadores de los laboratorios, los clínicos, los fabricantes, los bioinformáticos, los desarrolladores de software y las organizaciones profesionales para garantizar la calidad de los ensayos y para guiar la toma de decisiones clínicas en las neoplasias mieloides [182,184].

A pesar de su potencial para informar y guiar la práctica clínica, la secuenciación masiva tiene algunas limitaciones. Aunque hoy en día los costes de la secuenciación masiva se han hecho más asequibles, éstos siguen siendo altos. Además, el informe con los datos de secuenciación puede tardar entre una y tres semanas, lo que puede ser demasiado tiempo en el caso de pacientes con una enfermedad agresiva donde la primera decisión clínica respecto al tratamiento debe ser rápida. Desde el punto de vista práctico, la secuenciación masiva no está estandarizada y todavía no existe un pipeline de análisis óptimo, lo que puede dar lugar a una escasa reproducibilidad entre laboratorios, incluso para los mismos conjuntos de datos. Por tanto, es necesario colaborar para formar conjuntos de datos más amplios con el fin de mejorar la precisión y eficacia en la llamada de variantes para las firmas pronósticas y terapéuticas y para inferir conclusiones significativas [165,169]. Además, aunque la secuenciación masiva tiene un límite de detección cercano al 1% y esto es una mejora con respecto a la secuenciación Sanger (límite de detección de >20%), la interpretación del VAF de manera significativa requiere un contexto, puesto que no

todas las mutaciones somáticas detectadas por secuenciación masiva son necesariamente patogénicas y contribuyen al desarrollo de las neoplasias mieloides. Por tanto, se necesitan más estudios para establecer un umbral significativo de sensibilidad de VAF [165].

En resumen, la secuenciación masiva ha revolucionado la investigación genómica y es una de las técnicas más prometedoras que cambiará el manejo de los pacientes. Sin embargo, aún es importante estandarizar los métodos de secuenciación y de llamada de variantes y continuar avanzando en el conocimiento del papel patogénico y pronóstico de las variantes para mejorar su utilidad clínica.

Hipótesis

Los SMD son un grupo heterogéneo de enfermedades clonales que afectan a las células madre hematopoyéticas los cuales presentan un curso clínico muy variable, por lo que su diagnóstico, pronóstico y toma de decisiones terapéuticas son todo un reto. En los últimos años, el uso de técnicas de alto rendimiento, como la secuenciación masiva, ha revolucionado la investigación genómica, aumentando nuestro conocimiento sobre las alteraciones moleculares subyacentes a la patogénesis del SMD y mejorando nuestra compresión acerca de los mecanismos implicados en la progresión de la enfermedad. Sin embargo, la ausencia de una combinación eficiente entre genética e información clínica ha limitado la interpretación de los datos de secuenciación y, por consiguiente, la implantación de la secuenciación masiva en la práctica clínica habitual.

Aunque la evolución clínica de los SMD es variable, aproximadamente un tercio de los pacientes finalmente progresa a LAMs, lo que supone un pronóstico muy adverso. El uso de técnicas de alto rendimiento ha permitido definir que esta evolución de la enfermedad es claramente un proceso de evolución clonal y algunas alteraciones moleculares subyacentes a esta progresión han sido identificadas. Sin embargo, la mayoría de los datos están basados en estudios de muestras no pareadas de SMD-LAMs en lugar de en muestras seriadas de un mismo paciente, de modo que el orden de adquisición de las mutaciones y las dinámicas mutacionales aún permanecen incompletos y, por consiguiente, los mecanismos implicados en la transformación leucémica no están establecidos. Por tanto, se necesitan estudios genómicos de secuenciación secuencial para determinar qué mutaciones, solas o en combinación, están implicadas en la progresión de los SMD a LAMs, así como para definir los patrones de evolución clonal subyacentes a esta progresión.

Durante los últimos años, nuestro conocimiento acerca del perfil mutacional de los SMD ha mejorado notablemente gracias a la aplicación de las técnicas de secuenciación masiva. Aproximadamente un 90% de los pacientes con SMD portan al menos una mutación y se ha identificado que algunas de ellas se relacionan con características clínicas y tienen valor pronóstico. A su vez, un gran número de genes drivers pueden verse alterados en pacientes con SMD y la combinación de mutaciones no es totalmente aleatoria, lo que sugiere que existen interacciones gen-gen involucradas en la evolución clonal. Asimismo, la falta de validación clínica ha limitado la categorización diagnóstica y pronóstica teniendo en cuenta los resultados de secuenciación. Por estas razones, se necesitan grandes estudios que combinen información genómica y clínica para definir el impacto pronóstico de las mutaciones y trasladar estos resultados a la práctica clínica.

Se han encontrado mutaciones de las cohesinas en un 10-15% de los pacientes diagnosticados con SMD y se ha descrito que estas mutaciones alteran la regulación transcripcional que sucede en los procesos de diferenciación/proliferación, lo que sugiere que el complejo cohesina juega un papel importante en la patogénesis del SMD. No obstante, debido a la heterogeneidad molecular de los SMD, aún no se ha determinado el impacto clínico de las mutaciones de las cohesinas. Por este motivo, es necesario conocer la relevancia de estas mutaciones en los fenotipos clínicos, así como en la evolución clínica, de manera que son necesarios más estudios de grandes series de pacientes que incluyan información clínica detallada.

Del mismo modo, se han reportado mutaciones en la vía de Ras en un 5-10% de los casos diagnosticados de SMD, siendo más prevalentes en pacientes de alto riesgo o con LAMs. La vía de señalización de Ras es una de las rutas funcionales que pueden verse implicadas en la patogénesis del SMD, produciendo las mutaciones activantes de esta vía un aumento de la proliferación en los progenitores hematopoyéticos. Pese a que se ha estudiado ampliamente el impacto pronóstico de las mutaciones de la vía de Ras en pacientes con NMP y LAM, la información acerca de su relevancia en pacientes con SMD es limitada, lo que implica que hay una falta de consenso acerca de cuál es el pronóstico de los pacientes que portan estas mutaciones. Por tanto, y similar a lo sucedido con las mutaciones de las cohesinas, se necesitan más estudios de investigación que incluyan grandes series de pacientes.

En resumen, el uso de las técnicas de alto rendimiento nos ayudará a esclarecer los mecanismos implicados tanto en la patogénesis del SMD como en su progresión leucémica. Además, la integración de los datos mutacionales con los parámetros clínico-biológicos ya incluidos en las actuales clasificaciones pronósticas, nos permitirá avanzar hacia un diagnóstico integrado más fiable, una estratificación pronóstica más precisa e, incluso, una orientación terapéutica más adecuada, puesto que se podrán desarrollar nuevos fármacos diana contra las mutaciones somáticas detectadas.

Objetivos

Objetivo general:

Caracterizar los mecanismos moleculares y la evolución clonal que están implicados en la patogénesis del SMD y en la progresión de los SMD a LAMs, tan bien como utilizar los datos moleculares para alcanzar un diagnóstico integrado más fiable y una estratificación pronóstica más precisa.

Objetivos específicos:

1. Caracterizar el perfil mutacional de los pacientes con SMD que evolucionan a LAMs.

    1.1. Evaluar las dinámicas mutacionales y los patrones de evolución clonal subyacentes a la progresión de los SMD a LAMs.

    1.2. Explorar los efectos del tratamiento modificador de la enfermedad sobre los perfiles de evolución clonal durante la transformación leucémica de los pacientes con SMD.

2. Definir la relevancia de las mutaciones de las cohesinas en pacientes con SMD:

    2.1. Analizar la incidencia de las mutaciones somáticas en genes de las cohesinas por secuenciación masiva en una serie grande de pacientes con SMD, así como de las concurrencias entre estas mutaciones y mutaciones en otros genes mieloides.

    2.2. Correlacionar la presencia de mutaciones en genes de las cohesinas, solas o en concurrencia con otras mutaciones adicionales, con las características clínicas y la evolución de los pacientes con SMD.

    2.3. Analizar el efecto de la integración de los datos moleculares de las cohesinas en las actuales clasificaciones diagnósticas y pronósticas.

3. Estudiar la significancia del perfil mutacional de la vía de señalización de Ras en la patogénesis del SMD:

    3.1. Establecer la incidencia de las mutaciones somáticas de la vía de Ras en pacientes con SMD mediante técnicas de secuenciación masiva, además de evaluar el perfil mutacional de los pacientes con estas mutaciones.

    3.2. Definir la correlación entre las mutaciones activantes de Ras y el fenotipo de la enfermedad y su impacto clínico en la supervivencia global y en el tiempo hasta la progresión a LAMs.

Resultados - Resúmenes

Capítulo 1 - Resumen

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*].

14. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
15. Research Program on Biomedical Informatics, Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain.
16. Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden.
17. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
18. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
19. Hematology, Hospital Universitario Miguel Servet, Zaragoza, Spain.
20. Hematology, Hospital del Bierzo, Ponferrada, León, Spain.
21. Hematology, Hospital Universitario de Burgos, Burgos, Spain.
22. Hematology, Hospital General de Segovia, Segovia, Spain.
23. Hematology, Hospital Nuestra Señora de Sónsoles, Ávila, Spain.
24. Hematology, Hospital Santa Bárbara, Soria, Spain.
25. Hematology, Hospital Juan Ramón Jiménez, Huelva, Spain.
26. Hematology, Hospital Virgen del Puerto, Plasencia, Spain.

*MA and MMI contributed equally to this work.
*MDC and JMHR contributed equally to this work and both are co-senior authors.

Haematologica (2020). doi: 10.3324/haematol.2020.248807. PMID: 32675227.

INTRODUCCIÓN Y OBJETIVOS

Los síndromes mielodisplásicos (SMD) presentan un elevado riesgo de progresión a Leucemia Aguda Mieloblástica (LAMs). Aproximadamente esta evolución sucede en un tercio de los pacientes y supone un pronóstico desalentador para el paciente, puesto que la mayoría de los pacientes que progresan son resistentes a los tratamientos actuales, presentan tasas de remisión completa más bajas y la tasa de supervivencia a largo plazo de los pacientes tratados es menor del 10%.

La realización durante las últimas décadas de grandes estudios genómicos ha permitido definir que esta progresión de la enfermedad es claramente un proceso complejo de evolución clonal, así como identificar algunas alteraciones moleculares subyacentes a esta progresión. Sin embargo, debido a la complejidad del proceso y a la falta de grandes series de muestras pareadas, las dinámicas mutacionales y los mecanismos implicados en la transformación leucémica aún no están establecidos. De este modo, en este estudio nos propusimos analizar por secuenciación masiva una amplia serie de muestras pareadas de pacientes con SMD que progresaron a LAMs, con el objetivo de determinar qué mutaciones, solas o en combinación, están implicadas en la progresión de los SMD.

PACIENTES Y MÉTODOS

En el estudio, se analizaron un total de 486 muestras procedentes de 437 pacientes, divididas en tres cohortes de estudio:

i) **Serie de descubrimiento**: consistió en 84 muestras pareadas de médula ósea (MO) procedentes de 42 pacientes que progresaron a LAMs, las cuales fueron analizadas por una combinación de secuenciación de exoma completo (WES) (una selección de 16 pacientes) y un panel de captura de 117 genes mieloides (todos los pacientes) en dos momentos diferentes: al diagnóstico (etapa SMD) y tras la progresión (etapa LAMs). Además, se contó con información referente al tratamiento recibido en los 42 pacientes: 16 recibieron azacitidina, 4 lenalidomida y 22 únicamente tratamiento de soporte.

ii) **Serie control**: incluyó 14 muestras pareadas procedentes de 7 pacientes que no progresaron a LAMs, analizadas por secuenciación masiva con el mismo panel de captura al momento del diagnóstico y tras un periodo mínimo de 3 años, en el caso de los SMD de bajo riesgo, o de 1 año, en los casos de alto riesgo, mostrando enfermedad estable.

iii) **Serie de validación**: consistió en 388 muestras de MO o sangre periférica procedentes de pacientes diagnosticados de SMD, las cuales fueron analizadas únicamente al momento del diagnóstico con el panel de captura.

RESULTADOS

Con el objetivo de caracterizar la serie principal del estudio, la serie de descubrimiento, 16 pacientes fueron inicialmente analizados por WES y, tras un estricto análisis, un total de 40 variantes drivers y 21 variantes passengers fueron identificadas (**Tabla Suplementaria S4**). Posteriormente, se analizaron estas mismas muestras por secuenciación masiva con un panel de captura (**Tabla Suplementaria S3**), identificando una alta correlación entre ambas metodologías (tasa de verdaderos positivos >89% y coeficiente *r* de Pearson = 0,90) (**Figura Suplementaria S2**).

Sin embargo, la aplicación de la secuenciación dirigida reveló algunas mutaciones drivers que no fueron detectadas por WES debido a la baja cobertura alcanzada con esta metodología, de modo que decidimos realizar un análisis más exhaustivo por secuenciación dirigida en una serie más amplia de muestras pareadas de 26 pacientes adicionales. Estos análisis de secuenciación por captura revelaron un total de 210 mutaciones en la etapa de la LAMs, de las cuales 159 ya estaban presentes en el momento del diagnóstico (**Tabla Suplementaria S5**). Además, cabe destacar que estas mutaciones se encontraron en 54 genes distintos, encontrándose el 59% de ellos mutados en menos de 3 pacientes, lo que muestra la gran heterogeneidad existente en los mecanismos de evolución de la enfermedad.

Con el fin de analizar los cambios en el tamaño clonal y su distribución durante la evolución a LAMs, se comparó el número de mutaciones, así como sus frecuencias alélicas (VAF), identificadas al diagnóstico *versus* la segunda muestra en la serie de descubrimiento y en la serie control (**Figura 1**). De este modo, se observó que aquellos pacientes que evolucionan a LAMs presentaron un aumento significativo en el número de mutaciones en la segunda muestra (4 *vs.* 5 mutaciones, $p<0,0001$), mientras que los pacientes controles no ($p=0,449$). Similarmente, los pacientes que evolucionaron a LAMs presentaron una mayor mediana de frecuencia alélica en la segunda muestra que los pacientes controles (29.11% *vs.* 36.76%, $p<0,0001$). Por tanto, estos resultados sugieren que existe una mayor inestabilidad genómica durante la evolución de los SMD a LAMs.

Con el objetivo de identificar las dinámicas mutacionales y qué mutaciones puede estar implicadas en la evolución clonal que sucede durante la transformación leucémica, se compararon los VAF de las mutaciones detectadas en ambos momentos de cada paciente (momento LAM *vs.* SMD) de la serie control y de descubrimiento. De esta manera, cuatro tipos de dinámicas mutacionales fueron identificadas: mutaciones de tipo 1, las cuales son mutaciones ya presentes en la etapa del SMD, pero cuyos VAF aumentan en el momento de la LAMs; mutaciones de tipo 2, que son aquellas cuyos VAF disminuyen durante la progresión; mutaciones tipo 3, que son las mutaciones que fueron adquiridas *de novo* en la fase de la LAMs; y las mutaciones tipo 4, que son las que persisten con una carga alélica similar en ambos

momentos (**Figura 2A**). Respecto a las mutaciones estables, éstas fueron detectadas principalmente en genes del splicing, como *SRSF2*, y de la metilación del ADN, como *DNMT3A*, y se observaron en VAF elevados, lo que sugiere que las alteraciones en estas vías podrían estar involucradas en la patogénesis del SMD (papel driver) más que en la progresión a LAMs (papel passenger). En el caso de las mutaciones cuyo VAF disminuyó durante la evolución, éstas fueron detectadas en minoría y distribuidas a lo largo de varios genes, sin mostrar ningún patrón, de manera que este fenómeno podría ser explicado como una consecuencia del "barrido clonal" que ejercen otros clones durante la progresión. Por el contrario, las mutaciones de mayor interés fueron las mutaciones adquiridas en la LAMs o las que cuyo VAF aumentó en la evolución, puesto que sus dinámicas indican que son seleccionadas positivamente durante la transformación leucémica. Estas mutaciones se localizaron principalmente en los genes de la vía de Ras y en los del complejo cohesinas, respectivamente.

A pesar del amplio espectro de mutaciones que fueron detectadas en el estudio, el 21,4% de los pacientes (9/42) incluidos en la serie de descubrimiento mostraron concurrencia de mutaciones en la vía del complejo cohesinas y en la vía de Ras (p=0,023). Además, también se observó esta asociación entre los genes más frecuentemente mutados de ambas vías, *STAG2* con mutaciones mayoritariamente de tipo 1, y *NRAS*, con mutaciones de tipo 3 (p=0,002) (**Figura 2B**). Para confirmar si esta combinación podría desempeñar un papel importante en la progresión, se estudió esta concurrencia de mutaciones en la serie de validación, observándose que estaba presente en 11 pacientes más, de los cuales 9 finalmente evolucionaron a LAMs (**Figura 2C**). Además, como la serie de validación solo incluyó pacientes que progresaron a LAMs, lo que involucra un pobre desenlace, se estudiaron en la serie de validación las consecuencias clínicas de esta combinación en el curso de la enfermedad, de modo que se identificó que los pacientes con mutaciones en las cohesinas y en la vía de Ras mostraban una menor supervivencia global y un menor tiempo hasta la progresión (**Figura 2D**).

Por último, como el 48% de los pacientes de la serie de descubrimiento fueron tratados con tratamiento modificador de la enfermedad, es decir, con azacitidina (n=16) o lenalidomida (n=4) y progresaron a LAMs después de la terapia, investigamos si los mecanismos de progresión pueden ser diferentes entre aquellos pacientes que reciben tratamiento modificador y aquellos que solo reciben tratamiento de soporte. En nuestro estudio en el caso de los pacientes tratados identificamos una mayor proporción de mutaciones cuya dinámica fue de tipo 1 (aumentan el VAF) o tipo 3 (adquiridas) en los genes modificadores de la cromatina, mientras que en los pacientes tratados las mutaciones en estos genes principalmente permanecían estables (p=0,031). Por el contrario, respecto al tratamiento, no se observaron diferencias en las dinámicas de las mutaciones en los genes de las cohesinas (p=1,00) o de la vía de Ras (p=0,329) (**Figura Suplementaria S5**). Por tanto, nuestros resultados sugieren que las mutaciones en los modificadores de la

cromatina pueden estar relacionadas con la evolución de pacientes con SMD los cuales son tratados con fármacos modificadores de la enfermedad, como son azacitidina y lenalidomida, antes de la progresión.

CONCLUSIONES

En resumen, la progresión de los SMD a LAMs se caracteriza por una gran inestabilidad genómica, independientemente del subtipo diagnóstico de SMD, y por la presencia de cuatro dinámicas mutacionales durante la evolución, destacando entre ellas las mutaciones cuyo VAF aumenta con la transformación leucémica y las adquiridas en la etapa de la LAMs (tipo 1 y tipo 3, respectivamente). Además, la concurrencia entre las mutaciones en las cohesinas y en la vía de Ras puede jugar un papel importante en el 15-20% de los pacientes que evolucionan a LAMs. Con respecto al tratamiento, las mutaciones en los genes modificadores de la cromatina pueden estar relacionadas con la evolución a LAMs de los pacientes que reciben tratamiento modificador de la enfermedad previo a la progresión.

Capítulo 2 - Resumen

Incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of very low/low-risk myelodysplastic syndromes

Marta Martín-Izquierdo[1], María Díez-Campelo[2], Javier Sánchez-Real[3], Alberto Hernández-Sánchez[2], Jesús M Hernández-Sánchez[1], Kamila Janusz[4], Félix López-Cadenas[2], Javier Martínez-Elicegui[1], Mar Tormo[5], Marta Megido[6], Carmen Olivier[7], Andrés Madinaveitia-Ochoa[8], Julio Dávila[9], Magdalena Sierra[10], Manuel Vargas[11], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Rocío Benito[1], Jesús M Hernández-Rivas[1,2], Fernando Ramos[3*] and María Abáigar[1*]

1. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
2. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
3. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
4. IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.
5. Hematology, Hospital Clínico Universitario de Valencia, INCLIVA Research Institute, Valencia, Spain.
6. Hematology, Hospital del Bierzo, Ponferrada, León, Spain.
7. Hematology, Hospital General de Segovia, Segovia, Spain.
8. Hematology, Hospital Universitario Miguel Servet, Zaragoza, Spain.
9. Hematology, Hospital Nuestra Señora de Sonsoles, Ávila, Spain.
10. Hematology, Hospital Virgen de la Concha, Zamora, Spain.
11. Hematology, Hospital de Jarrio, Asturias, Spain

*FR and MA contributed equally to this work and both are co-senior authors.

American Journal of Hematology. Second review.

INTRODUCCIÓN

Durante los últimos años el uso de la secuenciación masiva ha permitido avanzar en la caracterización molecular de los Síndromes Mielodisplásicos (SMD), identificando algunos mecanismos involucrados en su patogénesis. Sin embargo, debido a la gran heterogeneidad genética y clínica que presentan, los datos moleculares no se han incluido en las clasificaciones pronósticas actuales.

Una de las vías que pueden verse alteradas en los SMD es el complejo cohesina, encontrándose mutado en un 10-15% de los pacientes. Se ha descrito que estas mutaciones pueden alterar la regulación transcripcional que tiene lugar durante los procesos de diferenciación/proliferación, lo que sugiere que el complejo cohesina tiene un papel importante en la patogénesis del SMD. No obstante, debido a la heterogeneidad molecular de los SMD, aún no se ha determinado el impacto clínico de las mutaciones de las cohesinas. Por este motivo, en este estudio nos propusimos estudiar mediante secuenciación masiva estas mutaciones en una amplia serie de pacientes diagnosticados con SMD para identificar las características clínicas de los pacientes que portan estas mutaciones en cuanto a su perfil genético, su fenotipo y su evolución clínica, así como para evaluar el impacto pronóstico de integrar los datos moleculares de estos genes con el resto de factores de riesgo de los SMD (estratificación IPSS e IPSS-R).

PACIENTES Y MÉTODOS

En el estudio se incluyeron muestras de médula ósea o sangre periférica de un total de 418 pacientes diagnosticados con SMD y con una buena caracterización clínica y seguimiento. Todas las muestras fueron analizadas por secuenciación masiva mediante una estrategia de captura de secuencia (*Nextera Rapid Capture Custom Enrichment*) usando un panel de 117 genes previamente relacionados con patología mieloide, entre los que se incluían los genes del complejo cohesina: *STAG1*, *STAG2*, *SMC1A*, *SMC3* y *RAD21* (**Tabla Suplementaria S2**).

RESULTADOS

La mediana de edad fue de 74,9 años (rango 29,4-92,2), con un 48,9% de hombres. De acuerdo a la clasificación OMS 2017, la mayoría de los pacientes pertenecieron a las categorías de SMD con displasia multilínea, con y sin sideroblastos en anillo (SMD-DM y SMD-SA-DM, 45,6%), seguido de los SMD con exceso de blastos (16,3% SMD-EB2 y 14,7% SMD-EB1) y los SMD con deleción 5q (10,1%). Respecto a la estratificación del IPSS-R, la mayoría de los pacientes fueron de riesgo muy bajo (26,8%) o bajo (44,4%), 15,3% fueron de riesgo intermedio mientras que el restante 13,6% fueron SMD de alto riesgo (categorías de riesgo alto y muy alto) (**Tabla Suplementaria S1**).

El estudio de secuenciación masiva identificó 52 mutaciones en los genes del complejo cohesina en un total de 48 pacientes (11,5%) (**Tabla Suplementaria S3**). El gen más frecuentemente mutado fue *STAG2*, puesto que el 69% de todos los pacientes mutados en las cohesinas presentó mutaciones en este gen. Además, las mutaciones en las cohesinas se encontraron a lo largo de todo el gen, por lo que no se identificó ningún sitio hotspot. Asimismo, la mayoría de estas mutaciones fueron de tipo de pérdida de función, incluyendo mutaciones sin sentido (nonsense) y de cambio del marco de lectura (frameshift) (**Figura 1**).

Con el objetivo de caracterizar el perfil genético propio de los pacientes con mutaciones en las cohesinas, se comparó el número de mutaciones, así como la incidencia mutacional de otros genes mieloides, en los pacientes mutados versus los pacientes que no portaban mutaciones en las cohesinas. De este modo, se observó que los pacientes mutados en las cohesinas presentaron una mediana de 4 mutaciones frente a una mediana de 2 en el caso de los pacientes *wild-type*, lo que indicaba que los pacientes mutados tenían un número de mutaciones significativamente mayor (p<0,0001) (**Figura 2A**). Además, los pacientes con mutaciones en las cohesinas mostraban frecuencias mutacionales más altas en los genes *SRSF2* (p<0,0001), *ASXL1* (p=0,0031), *RUNX1* (p=0,0074) and *EZH2* (p=0,0185) (**Figura 2B**). De manera interesante, también se identificó una fuerte concurrencia entre las mutaciones en las cohesinas y genes del splicing (p<0,0001), alteraciones de la vía de Ras (p=0,001) y modificadores de la cromatina (p=0,007), y más especialmente entre mutaciones de las cohesinas y el factor del splicing *SRSF2* (p<0,0001) y el modificador de la cromatina *ASXL1* (p<0,0001) (**Figura 2C y 2D**).

Con respecto a las características clínicas y biológicas de los pacientes que portan mutaciones en las cohesinas, en este estudio se identificó que los pacientes que portaban estas mutaciones se asociaban con las categorías con exceso de blastos de la clasificación OMS 2017 (p=0,004) y con las categorías de riesgo intermedio y alto de la estratificación del IPSS-R (p=0,001) (**Figura 3A y 3B**). También cabe destacar que un 52% de los pacientes presentaron una citogenética normal, una proporción equiparable a la de los pacientes sin mutaciones en las cohesinas, sugiriendo que las cohesinas en los SMD no están implicadas en la desestabilización de la integridad cromosómica, sino más bien en otros mecanismos alternativos, como pueden ser el control transcripcional o la reparación del daño celular. Además, la presencia de las mutaciones de las cohesinas define un subconjunto de pacientes que muestran un fenotipo de mal pronóstico, caracterizado por un bajo recuento de plaquetas (p=0,009), neutrófilos (p=0,008) y leucocitos (p=0,016) en sangre periférica, un mayor número de blastos en médula (p=0,0014) y una mayor tasa de progresión a LAMs (p<0,0001) (**Tabla 1 y Figura 3C**).

Con el fin de determinar los efectos de las mutaciones de las cohesinas, solas o en combinación, en la evolución clínica de los SMD, estudiamos exhaustivamente el impacto de estas mutaciones en la supervivencia global (SG), en la supervivencia libre de progresión (SLP) y en el tiempo hasta la progresión. De manera interesante, se observó que los pacientes mutados en la serie global mostraron una tendencia a una menor supervivencia (mediana: 3,1 *vs.* 5,2 años, *p*=0,069), una menor supervivencia libre de progresión (mediana: 1,3 *vs.* 3,5 años, *p*<0,0001) y un menor tiempo hasta la progresión (mediana: 1,4 *vs.* 8,3 años, *p*<0,0001) (**Figura 4A**). Además, se demostró este mismo efecto en el subconjunto de pacientes con mutaciones en *SRSF2*, donde los pacientes doblemente mutados presentan una menor supervivencia libre de progresión (*p*<0,0001) y un menor tiempo hasta la progresión (*p*<0,0001) (**Figura 4B**), y, especialmente, en el subgrupo de pacientes de riesgo muy bajo, bajo e intermedio según el IPSS-R tanto en el análisis univariante (**Figura 4C**) como multivariante (HR 2,44, 95% CI 1,19-4,98; p=0,015) (**Figura 4D**).

Por último, y en vista a estos resultados, evaluamos el valor pronóstico de las mutaciones en las cohesinas en la estratificación del IPSS-R e identificamos que la presencia de mutaciones en las cohesinas puede estratificar los pacientes de riesgo muy bajo o bajo del IPSS-R en dos grupos diferentes, mostrando los pacientes mutados unas curvas de supervivencia global, libre de progresión y tiempo hasta la progresión muy similares a las de los pacientes de una categoría superior en el IPSS-R, es decir, los pacientes de riesgo intermedio (**Figura 5**).

CONCLUSIONES

Por tanto, las mutaciones en las cohesinas, que suelen resultar en una pérdida de su función, son alteraciones frecuentes en los SMD, que se asocian con un perfil mutacional específico, un fenotipo de mal pronóstico y una peor evolución clínica, caracterizada principalmente por una mayor tasa y una evolución a LAMs más temprana. Además, estas mutaciones tienen un efecto en la estratificación pronóstica de los pacientes de riesgo muy bajo y bajo según el IPSS-R, de manera que la incorporación de su estado mutacional puede mejorar la clasificación pronóstica de los pacientes con SMD y ser útil para la monitorización y predicción de la progresión de la enfermedad.

Capítulo 3 - Resumen

RAS-pathway mutations are associated with a worse clinical outcome in myelodysplastic syndromes patients

Marta Martín-Izquierdo[1], Fernando Ramos[2], María Abáigar[1], Jesús M Hernández-Sánchez[1], Teresa González[1,3], Félix López-Cadenas[3], Quijada-Álamo M[1], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Eva Lumbreras[1], Kamila Janusz[4], Mónica del Rey[1], Sofía Toribio-Castelló[1], Jesús M Hernández-Rivas[1,3], María Díez-Campelo[3*] and Rocío Benito[1*]

1. Institute of Biomedical Research of Salamanca (IBSAL), Cytogenetics-Molecular Genetics in Oncohematology, Cancer Research Center-University of Salamanca (IBMCC, USAL-CSIC).
2. Hematology, Hospital Universitario de León, Institute of Biomedicine (IBIOMED)-University of León, León, Spain.
3. University of Salamanca, IBSAL, Hematology, Hospital Clínico Universitario, Salamanca, Spain.
4. IMIBIC, Hematology, Hospital Universitario Reina Sofía, UCO, Córdoba, Spain.

*MDC and RB contributed equally to this work and both are co-senior authors.

Cancers. Submitted.

INTRODUCCIÓN

En los últimos años, gracias a la implementación de las nuevas técnicas de secuenciación masiva se han producido importantes avances en la caracterización molecular de los síndromes mielodisplásicos. Sin embargo, debido a su gran heterogeneidad y a la falta de validación clínica, el impacto clínico de las mutaciones no está completamente definido y no existe, a día de hoy, una clasificación pronóstica que tenga en cuenta los datos moleculares.

En concreto, las mutaciones de la vía de Ras se han descrito en un 5-10% de los pacientes con SMD, asociándose con un aumento en la proliferación de los progenitores hematopoyéticos. No obstante, pese a que se ha estudiado ampliamente el impacto pronóstico de las mutaciones de la vía de Ras en pacientes con neoplasias mieloproliferativas y LAM, la información acerca de su relevancia en pacientes con SMD es limitada y controvertida, de manera que se desconoce cuál es el pronóstico de los pacientes que portan este tipo de alteraciones. Por tanto, en este estudio nos propusimos como objetivo identificar las mutaciones de la vía de Ras en una serie amplia de pacientes con SMD mediante secuenciación masiva y analizar su relación con las características clínica de la enfermedad para determinar su valor clínico y pronóstico.

PACIENTES Y MÉTODOS

Un total de 418 pacientes diagnosticados con SMD y con información clínica detallada fueron incluidos en el estudio (**Tabla Suplementaria S1**). En todos los casos se analizó una muestra de médula ósea o sangre periférica al momento del diagnóstico mediante secuenciación masiva por una estrategia de captura de secuencia (*Nextera Rapid Capture Custom Enrichment*), para lo que se utilizó un panel de 117 genes mieloides, que incluía los siguientes genes implicados en la vía de Ras: *NRAS*, *KRAS*, *HRAS*, *FLT3*, *KIT*, *CBL*, *EGFR*, *NF1*, *PTPN11*, *PTPN1*, *BRAF* y *G3BP1* (**Tabla Suplementaria S2**).

RESULTADOS

El estudio de secuenciación masiva reveló 59 mutaciones en alguno de los genes estudiados de la vía de Ras, lo que supuso una incidencia del 11,2% (47 pacientes de un total de 418) (**Tabla Suplementaria S1**). El gen más frecuentemente mutado fue *NRAS*, encontrándose en 15 pacientes, seguido de *NF1* (n=11), *KRAS* y *CBL* (n=6, cada uno), *PTPN11* y *KIT* (n=5, cada uno), *BRAF* (n=3), *EGFR* (n=2) y *FLT3* (n=1) (**Figura 1A**). La mayoría de las mutaciones detectadas fueron de tipo missense (53/59, 89,8%) y para el gen *NRAS* se encontraron dos sitios hotspots, las posiciones de los aminoácidos G12 y G13 pertenecientes al dominio de unión de GTP. Además, se analizó la mediana de la frecuencia alélica (VAF) de todas las mutaciones somáticas identificadas,

identificando que las mutaciones en la vía de Ras aparecen en menores VAF que las mutaciones en el resto de vías, como son el splicing, la regulación epigenética, la transcripción, las cohesinas o el resto de rutas de señalización (p<0,0001, cada una), sugiriendo que estas mutaciones son eventos subclonales (**Figura 1B y 1C**).

Con el objetivo de identificar potenciales mecanismos de cooperatividad entre las mutaciones de la vía de Ras y otros genes, se analizaron los patrones de concurrencias y exclusividades entre las distintas vías funcionales analizadas en nuestro estudio. De este modo, se observó una fuerte concurrencia entre las mutaciones de la vía de Ras y en los genes de las cohesinas (*p*=0,001), así como con los genes *STAG2* (*p*=0,003), el factor de splicing *SRSF2* (*p*=0,021), el modificador de la cromatina *EZH2* (*p*=0,046) y los factores de transcripción NPM1 y GATA2 (*p*=0,008, ambos), mientras que se encontró una relación de exclusividad entre las mutaciones de la vía de Ras y el factor de splicing *SF3B1* (*p*=0,047) (**Figura 2**). Por tanto, los pacientes con mutaciones en la vía de Ras muestran un perfil genético específico, caracterizado principalmente por la concurrencia con las mutaciones de las cohesinas.

Con respecto a las características clínicas y biológicas de los pacientes con mutaciones en la vía de Ras, en nuestro estudio los pacientes mutados fueron más frecuentemente diagnosticados con un SMD con exceso de blastos (*p*=0,001) o con subtipos pertenecientes a las categorías de riesgo alto del IPSS-R (*p*=0,008) (**Figura 3**). Del mismo modo, no se encontraron pacientes con mutaciones en la vía de Ras que presentaran alteraciones citogenéticas de mal pronóstico, mientras que se asociaron significativamente con alteraciones de riesgo intermedio (*p*=0,028), en concreto, los pacientes mutados mostraron una proporción más alta de casos con la trisomía 8 (12,5% *vs.* 4,4%, *p*=0,031). Además, los pacientes que portaban alguna mutación en la vía de Ras presentaron una mayor incidencia de marcadores de mal pronóstico, como un menor recuento de plaquetas y neutrófilos en sangre periférica (*p*=0,004 y *p*=0,008, respectivamente) y una mayor proporción de blastos en médula ósea (*p*<0,0001), así como una mayor tasa de recibir un tratamiento modificador de la enfermedad (*p*=0,049) y de muerte (*p*=0,027) durante el seguimiento del estudio (**Tabla 1**).

Para determinar los efectos de las mutaciones de la vía de Ras en la evolución clínica de los SMD, estudiamos de manera exhaustiva su impacto en la supervivencia global (SG), en la supervivencia libre de progresión (SLP) y en el tiempo hasta la progresión. Estos análisis revelaron que los pacientes que albergaban mutaciones en la vía de Ras mostraban una menor supervivencia global (mediana: 31,1 *vs.* 64,6 meses, *p*=0,028), una menor supervivencia libre de progresión (mediana: 20,4 *vs.* 40,5 meses, *p*=0,004) y un menor tiempo hasta la progresión (mediana: 54,2 *vs.* 87,7 meses, *p*=0,028) en la serie global (**Figura 4**). Además, con el fin de evaluar el impacto de las mutaciones de Ras entre todas las variables convencionales con valor

pronóstico conocido, es decir, las variables incluidas en la estratificación del IPSS-R (blastos en médula ósea, hemoglobina, plaquetas, neutrófilos y citogenética), se realizó un análisis multivariante de regresión de *Cox* para la SG, SLP y el tiempo hasta la progresión, mostrando que los pacientes con mutaciones en la vía de Ras se asocian independientemente con una menor supervivencia global (SG: HR 1,71, 95% CI 1,04-2,81; *p*=0,033) (**Tabla 2**).

Finalmente, para confirmar el valor pronóstico de las mutaciones en la vía de Ras, se analizó la SG, SLP y el tiempo hasta la progresión para cada grupo del IPSS-R de manera individual. De manera interesante, los pacientes de las categorías de riesgo muy bajo y bajo con mutaciones en la vía de Ras mostraron una menor SG (32,1 *vs.* 81,6 meses, *p*=0,009) (**Figura 5A**), mientras que no se observaron diferencias en la SLP (*p*=0,183) y en el tiempo hasta la progresión (*p*=0,644). Además, este impacto fue confirmado por un análisis multivariante de regresión de *Cox*, donde las mutaciones de Ras fueron el único factor que se asoció con una menor supervivencia global (SG: HR 1,86, 95% CI 1,01-3,43, *p*=0,045) (**Figura 5B**).

CONCLUSIONES

En resumen, las mutaciones de la vía de Ras son eventos genéticos recurrentes en los pacientes con SMD (11,2% de los casos) y son adquiridas en etapas tardías del curso de la enfermedad. Los pacientes que portan estas mutaciones muestran un perfil mutacional específico, destacando la concurrencia entre las mutaciones de la vía de Ras y del complejo cohesinas, así como un fenotipo de mal pronóstico, con presencia de citopenias y exceso de blastos y una mayor tasa de muerte, a pesar de recibir tratamiento modificador de la enfermedad. Además, los pacientes con mutaciones en la vía de Ras muestran una peor evolución clínica, principalmente caracterizada por una menor supervivencia global. Por tanto, la incorporación de su estado mutacional puede mejorar la clasificación pronóstica de los pacientes con SMD.

Conclusiones

1. La progresión de los SMD a LAMs es un proceso muy complejo y dinámico de evolución y selección clonal. El uso de la secuenciación masiva en muestras pareadas de pacientes que evolucionaron a LAMs permite la caracterización del perfil mutacional subyacente a la progresión a LAMs, el cual es caracterizado por una gran inestabilidad genómica, independientemente del subtipo diagnóstico.

   1.1. Los mecanismos implicados en esta evolución son heterogéneos, sin embargo, se han podido identificar cuatro tipos de dinámicas mutacionales, siendo las más importantes aquellas mutaciones que aumentaron su carga tumoral o que se adquirieron durante la evolución. Además, cabe destacar que se identificó en un 15-20% de los pacientes con SMD que evolucionaron un mecanismo preferencial de progresión consistente en la concurrencia de mutaciones en las cohesinas, con una dinámica de aumento del tamaño clonal, y mutaciones en la vía de Ras, que fueron mayoritariamente adquiridas en la etapa de la LAMs.

   1.2. El tratamiento con agentes modificadores de la enfermedad (agentes hipometilantes y lenalidomida) puede impactar en los patrones de evolución clonal de los pacientes con SMD. Los mecanismos de progresión a LAMs en pacientes tratados pueden ser diferentes a los de los pacientes no tratados, caracterizándose por la presencia de mutaciones en genes modificadores de la cromatina que aumentan su tamaño clonal o son adquiridas durante la etapa de la LAMs.

2. La presencia de mutaciones de pérdida de función en las cohesinas es un evento común en pacientes con SMD, ocurriendo en un 11,9% de los casos al diagnóstico y siendo *STAG2* el gen más frecuentemente mutado.

   2.1. Los pacientes con SMD que portan mutaciones en las cohesinas presentan un perfil específico, caracterizado por tener un mayor número de mutaciones y por la fuerte concurrencia entre estas alteraciones y las mutaciones en la vía de Ras, en genes modificadores de la cromatina y genes de splicing, principalmente, con el factor de splicing *SRSF2*.

   2.2. La presencia de mutaciones en los genes de las cohesinas define un conjunto de pacientes con un fenotipo de mal pronóstico, que se caracteriza por marcadas citopenias en sangre periférica y un número elevado de blastos en médula ósea. Además, las mutaciones en las cohesinas son marcadores de mal pronóstico con respecto a la transformación leucémica, puesto que se asocian con una mayor tasa y una progresión a LAMs más temprana, especialmente en pacientes de riesgo muy bajo, bajo e intermedio según el IPSS-R.

2.3. La incorporación de los datos moleculares de las cohesinas en la clasificación actual del IPSS-R podría mejorar la estratificación pronóstica de los pacientes con SMD de muy bajo o bajo riesgo. Por tanto, el estudio del estado mutacional de las cohesinas podría ser útil en la monitorización de los pacientes y en la predicción de progresión de la enfermedad, permitiendo una intervención temprana.

3. Las mutaciones en la vía de Ras son eventos genéticos recurrentes en los pacientes diagnosticados de SMD, ocurriendo en un 11,2% de los casos y siendo normalmente subclonales y adquiridas de manera tardía en el curso de la enfermedad.

3.1. Los pacientes con SMD que portan mutaciones en la vía de Ras manifiestan un perfil genético específico, en el que destaca la concurrencia entre mutaciones en esta vía y en el complejo cohesina, mientras que son mutuamente excluyentes con las mutaciones en *SF3B1*.

3.2. La presencia de mutaciones en la vía de señalización de Ras define un conjunto de pacientes que exhiben un fenotipo de mal pronóstico, con marcadas citopenias y exceso de blastos, así como con una alta tasa de defunción, a pesar del tratamiento modificador de la enfermedad. Además, los pacientes con estas mutaciones presentan una supervivencia global más corta, independiente del resto de variables pronósticas incluidas en el IPSS-R. Por tanto, la activación de la vía de Ras puede estar implicada en la progresión a un estadio más agresivo de la enfermedad, provocando un pronóstico muy adverso en los pacientes con SMD.

Referencias

1 Mufti, G. J. *et al.* Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. *Haematologica* **93**, 1712-1717, doi:10.3324/haematol.13405 (2008).

2 Tefferi, A. & Vardiman, J. W. Myelodysplastic syndromes. *The New England journal of medicine* **361**, 1872-1885, doi:10.1056/NEJMra0902908 (2009).

3 Cazzola, M., Della Porta, M. G. & Malcovati, L. The genetic basis of myelodysplasia and its clinical relevance. *Blood* **122**, 4021-4034, doi:10.1182/blood-2013-09-381665 (2013).

4 Lindsley, R. C. & Ebert, B. L. Molecular pathophysiology of myelodysplastic syndromes. *Annual review of pathology* **8**, 21-47, doi:10.1146/annurev-pathol-011811-132436 (2013).

5 Malcovati, L. *et al.* Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: a basis for clinical decision making. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 7594-7603, doi:10.1200/JCO.2005.01.7038 (2005).

6 Malcovati, L. *et al.* Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS). *Haematologica* **96**, 1433-1440, doi:10.3324/haematol.2011.044602 (2011).

7 Hellstrom-Lindberg, E., Tobiasson, M. & Greenberg, P. Myelodysplastic syndromes: moving towards personalized management. *Haematologica* **105**, 1765-1779, doi:10.3324/haematol.2020.248955 (2020).

8 Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature reviews. Cancer* **7**, 118-129, doi:10.1038/nrc2047 (2007).

9 Ades, L., Itzykson, R. & Fenaux, P. Myelodysplastic syndromes. *Lancet* **383**, 2239-2252, doi:10.1016/S0140-6736(13)61901-7 (2014).

10 Deeg HJ, B. D., Gore SD, Haferlach T, Le Beau MM, Niemeyer C (eds). *Hematologic malignancies: Myelodysplastic syndromes*. (Springer, 2006).

11 Garcia-Manero, G. Myelodysplastic syndromes: 2014 update on diagnosis, risk-stratification, and management. *American journal of hematology* **89**, 97-108, doi:10.1002/ajh.23642 (2014).

12 Komrokji, R. Myelodysplastic syndromes: a view from where the sun rises and where the sun sets. *Leukemia research* **30**, 1067-1068, doi:10.1016/j.leukres.2006.04.004 (2006).

13 Matsuda, A. *et al.* Difference in clinical features between Japanese and German patients with refractory anemia in myelodysplastic syndromes. *Blood* **106**, 2633-2640, doi:10.1182/blood-2005-01-0040 (2005).

14 Neukirchen, J. *et al.* Incidence and prevalence of myelodysplastic syndromes: data from the Dusseldorf MDS-registry. *Leukemia research* **35**, 1591-1596, doi:10.1016/j.leukres.2011.06.001 (2011).

15 Rollison, D. E. *et al.* Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood* **112**, 45-52, doi:10.1182/blood-2008-01-134858 (2008).

16 Catenacci, D. V. & Schiller, G. J. Myelodysplasic syndromes: a comprehensive review. *Blood reviews* **19**, 301-319, doi:10.1016/j.blre.2005.01.004 (2005).

17 Pedersen-Bjergaard, J. *et al.* Increased risk of myelodysplasia and leukaemia after etoposide, cisplatin, and bleomycin for germ-cell tumours. *Lancet* **338**, 359-363, doi:10.1016/0140-6736(91)90490-g (1991).

18 Morrison, V. A. *et al.* Therapy-related myeloid leukemias are observed in patients with chronic lymphocytic leukemia after treatment with fludarabine and chlorambucil: results of an intergroup

study, cancer and leukemia group B 9011. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 3878-3884, doi:10.1200/JCO.2002.08.128 (2002).

19 Iwanaga, M. *et al.* Risk of myelodysplastic syndromes in people exposed to ionizing radiation: a retrospective cohort study of Nagasaki atomic bomb survivors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 428-434, doi:10.1200/JCO.2010.31.3080 (2011).

20 Cardis, E. *et al.* Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries. *Bmj* **331**, 77, doi:10.1136/bmj.38499.599861.E0 (2005).

21 Bannon, S. A. & DiNardo, C. D. Hereditary Predispositions to Myelodysplastic Syndrome. *International journal of molecular sciences* **17**, doi:10.3390/ijms17060838 (2016).

22 Greenberg, P. L. *et al.* Cytopenia levels for aiding establishment of the diagnosis of myelodysplastic syndromes. *Blood* **128**, 2096-2097, doi:10.1182/blood-2016-07-728766 (2016).

23 Steensma, D. P. Myelodysplastic Syndromes: Diagnosis and Treatment. *Mayo Clinic proceedings* **90**, 969-983, doi:10.1016/j.mayocp.2015.04.001 (2015).

24 Malcovati, L. *et al.* Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood* **122**, 2943-2964, doi:10.1182/blood-2013-03-492884 (2013).

25 Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951, doi:10.1182/blood-2009-03-209262 (2009).

26 Bejar, R. Myelodysplastic Syndromes Diagnosis: What Is the Role of Molecular Testing? *Current hematologic malignancy reports* **10**, 282-291, doi:10.1007/s11899-015-0270-5 (2015).

27 Duncavage, E. J. & Tandon, B. The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes. *International journal of laboratory hematology* **37 Suppl 1**, 115-121, doi:10.1111/ijlh.12361 (2015).

28 Bejar, R. Prognostic models in myelodysplastic syndromes. *Hematology. American Society of Hematology. Education Program* **2013**, 504-510, doi:10.1182/asheducation-2013.1.504 (2013).

29 Bennett, J. M. *et al.* Proposals for the classification of the myelodysplastic syndromes. *British journal of haematology* **51**, 189-199 (1982).

30 Vardiman, J. W., Harris, N. L. & Brunning, R. D. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* **100**, 2292-2302, doi:10.1182/blood-2002-04-1199 (2002).

31 Tefferi, A. & Vardiman, J. W. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia* **22**, 14-22, doi:10.1038/sj.leu.2404955 (2008).

32 Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405, doi:10.1182/blood-2016-03-643544 (2016).

33 Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110**, 4385-4395, doi:10.1182/blood-2007-03-082404 (2007).

34 Papaemmanuil, E. *et al.* Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England journal of medicine* **365**, 1384-1395, doi:10.1056/NEJMoa1103283 (2011).

35 Cazzola, M., Rossi, M., Malcovati, L. & Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie, M. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood* **121**, 260-269, doi:10.1182/blood-2012-09-399725 (2013).

36 Malcovati, L. *et al.* SF3B1 mutation identifies a distinct subset of

myelodysplastic syndrome with ring sideroblasts. *Blood* **126**, 233-241, doi:10.1182/blood-2015-03-633537 (2015).

37 Chen-Liang, T. H. Prognosis in Myelodysplastic Syndromes: The Clinical Challenge of Genomic Integration. *Journal of clinical medicine* **10**, doi:10.3390/jcm10102052 (2021).

38 Greenberg, P. *et al.* International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* **89**, 2079-2088 (1997).

39 Garcia-Manero, G. *et al.* A prognostic score for patients with lower risk myelodysplastic syndrome. *Leukemia* **22**, 538-543, doi:10.1038/sj.leu.2405070 (2008).

40 Kantarjian, H. *et al.* Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. *Cancer* **113**, 1351-1361, doi:10.1002/cncr.23697 (2008).

41 Greenberg, P. L. *et al.* Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* **120**, 2454-2465, doi:10.1182/blood-2012-03-420489 (2012).

42 Montalban-Bravo, G. & Garcia-Manero, G. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. *American journal of hematology* **93**, 129-147, doi:10.1002/ajh.24930 (2018).

43 Steensma, D. P. Myelodysplastic syndromes current treatment algorithm 2018. *Blood cancer journal* **8**, 47, doi:10.1038/s41408-018-0085-4 (2018).

44 Pfeilstocker, M. *et al.* Time-dependent changes in mortality and transformation risk in MDS. *Blood* **128**, 902-910, doi:10.1182/blood-2016-02-700054 (2016).

45 Haase, D. *et al.* TP53 mutation status divides myelodysplastic syndromes with complex karyotypes into distinct prognostic subgroups. *Leukemia* **33**, 1747-1758, doi:10.1038/s41375-018-0351-2 (2019).

46 Zheng, X. *et al.* Prognostic value of SRSF2 mutations in patients with de novo myelodysplastic syndromes: A meta-analysis. *PloS one* **12**, e0185053, doi:10.1371/journal.pone.0185053 (2017).

47 Lin, M. E. *et al.* Dynamics of DNMT3A mutation and prognostic relevance in patients with primary myelodysplastic syndrome. *Clinical epigenetics* **10**, 42, doi:10.1186/s13148-018-0476-1 (2018).

48 Hou, H. A. *et al.* Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. *Blood cancer journal* **8**, 39, doi:10.1038/s41408-018-0074-7 (2018).

49 Gangat, N. *et al.* Mutations and prognosis in myelodysplastic syndromes: karyotype-adjusted analysis of targeted sequencing in 300 consecutive cases and development of a genetic risk model. *American journal of hematology* **93**, 691-697, doi:10.1002/ajh.25064 (2018).

50 Mangaonkar, A. A. *et al.* Prognostic impact of ASXL1 mutations in patients with myelodysplastic syndromes and multilineage dysplasia with or without ring sideroblasts. *Leukemia research* **71**, 60-62, doi:10.1016/j.leukres.2018.07.010 (2018).

51 Malcovati, L. *et al.* SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS. *Blood* **136**, 157-170, doi:10.1182/blood.2020004850 (2020).

52 Platzbecker, U. Treatment of MDS. *Blood* **133**, 1096-1107, doi:10.1182/blood-2018-10-844696 (2019).

53 DeZern, A. E. Nine years without a new FDA-approved therapy for MDS: how can we break through the impasse? *Hematology. American Society of Hematology. Education Program* **2015**, 308-316, doi:10.1182/asheducation-2015.1.308 (2015).

54 Platzbecker, U. *et al.* Luspatercept for the treatment of anaemia in patients with lower-risk myelodysplastic syndromes (PACE-MDS): a multicentre, open-label phase 2 dose-finding study with long-term extension study. *The Lancet. Oncology* **18**, 1338-1347, doi:10.1016/S1470-2045(17)30615-0 (2017).

55 Garcia-Manero, G., Chien, K. S. & Montalban-Bravo, G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. *American journal of hematology* **95**, 1399-1420, doi:10.1002/ajh.25950 (2020).

56 Tobiasson, M. & Kittang, A. O. Treatment of myelodysplastic syndrome in the era of next-generation sequencing. *Journal of internal medicine* **286**, 41-62, doi:10.1111/joim.12893 (2019).

57 Menssen, A. J. & Walter, M. J. Genetics of progression from MDS to secondary leukemia. *Blood* **136**, 50-60, doi:10.1182/blood.2019000942 (2020).

58 Shukron, O., Vainstein, V., Kundgen, A., Germing, U. & Agur, Z. Analyzing transformation of myelodysplastic syndrome to secondary acute myeloid leukemia using a large patient database. *American journal of hematology* **87**, 853-860, doi:10.1002/ajh.23257 (2012).

59 Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine* **371**, 2477-2487, doi:10.1056/NEJMoa1409405 (2014).

60 McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613-628, doi:10.1016/j.cell.2017.01.018 (2017).

61 Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).

62 Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616-3627; quiz 3699, doi:10.1182/blood-2013-08-518886 (2013).

63 Haferlach, T. *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241-247, doi:10.1038/leu.2013.336 (2014).

64 Ogawa, S. Genetics of MDS. *Blood* **133**, 1049-1059, doi:10.1182/blood-2018-10-844621 (2019).

65 Haase, D. Cytogenetic features in myelodysplastic syndromes. *Annals of hematology* **87**, 515-526, doi:10.1007/s00277-008-0483-y (2008).

66 Bejar, R. Clinical and genetic predictors of prognosis in myelodysplastic syndromes. *Haematologica* **99**, 956-964, doi:10.3324/haematol.2013.085217 (2014).

67 Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *The New England journal of medicine* **360**, 2289-2301, doi:10.1056/NEJMoa0810069 (2009).

68 Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nature genetics* **42**, 665-667, doi:10.1038/ng.620 (2010).

69 Sanada, M. *et al.* Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904-908, doi:10.1038/nature08240 (2009).

70 Makishima, H. *et al.* Dynamics of clonal evolution in myelodysplastic syndromes. *Nature genetics* **49**, 204-212, doi:10.1038/ng.3742 (2017).

71 Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *The New England journal of medicine* **364**, 2496-2506, doi:10.1056/NEJMoa1013343 (2011).

72 Itzykson, R., Kosmider, O. & Fenaux, P. Somatic mutations and epigenetic abnormalities in myelodysplastic syndromes. *Best practice & research. Clinical haematology* **26**, 355-364, doi:10.1016/j.beha.2014.01.001 (2013).

73 Pellagatti, A. & Boultwood, J. The molecular pathogenesis of the myelodysplastic syndromes. *European journal of haematology* **95**, 3-15, doi:10.1111/ejh.12515 (2015).

74 Jhanwar, S. C. Genetic and epigenetic pathways in myelodysplastic syndromes: A brief overview. *Advances in biological regulation* **58**, 28-37, doi:10.1016/j.jbior.2014.11.002 (2015).

75 Nagata, Y. & Maciejewski, J. P. The functional mechanisms of mutations in myelodysplastic syndrome. *Leukemia* **33**, 2779-2794, doi:10.1038/s41375-019-0617-3 (2019).

76 Lee, S. C. *et al.* Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations.

Cancer cell **34**, 225-241 e228, doi:10.1016/j.ccell.2018.07.003 (2018).

77 Thota, S. *et al.* Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood* **124**, 1790-1798, doi:10.1182/blood-2014-04-567057 (2014).

78 Ganguly, B. B. & Kadam, N. N. Mutations of myelodysplastic syndromes (MDS): An update. *Mutation research. Reviews in mutation research* **769**, 47-62, doi:10.1016/j.mrrev.2016.04.009 (2016).

79 Wu, P. *et al.* Co-occurrence of RUNX1 and ASXL1 mutations underlie poor response and outcome for MDS patients treated with HMAs. *American journal of translational research* **11**, 3651-3658 (2019).

80 Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *The New England journal of medicine* **374**, 2209-2221, doi:10.1056/NEJMoa1516192 (2016).

81 Todisco, G. *et al.* Co-mutation pattern, clonal hierarchy, and clone size concur to determine disease phenotype of SRSF2(P95)-mutated neoplasms. *Leukemia*, doi:10.1038/s41375-020-01106-z (2020).

82 Bernard, E. *et al.* Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nature medicine* **26**, 1549-1556, doi:10.1038/s41591-020-1008-z (2020).

83 Makishima, H. *et al.* Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* **119**, 3203-3210, doi:10.1182/blood-2011-12-399774 (2012).

84 Thol, F. *et al.* Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* **119**, 3578-3584, doi:10.1182/blood-2011-12-399337 (2012).

85 Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature genetics* **44**, 53-57, doi:10.1038/ng.1031 (2011).

86 Kulasekararaj, A. G. *et al.* TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. *British journal of haematology* **160**, 660-672, doi:10.1111/bjh.12203 (2013).

87 Liang, S. *et al.* Prognostic value of DNMT3A mutations in myelodysplastic syndromes: a meta-analysis. *Hematology* **24**, 613-622, doi:10.1080/16078454.2019.1657613 (2019).

88 Chen, C. Y. *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. *British journal of haematology* **139**, 405-414, doi:10.1111/j.1365-2141.2007.06811.x (2007).

89 Thol, F. *et al.* Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 2499-2506, doi:10.1200/JCO.2010.33.4938 (2011).

90 Janusz, K. *et al.* Clinical, biological, and prognostic implications of SF3B1 co-occurrence mutations in very low/low- and intermediate-risk MDS patients. *Annals of hematology* **100**, 1995-2004, doi:10.1007/s00277-020-04360-4 (2021).

91 Haering, C. H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Molecular cell* **9**, 773-788, doi:10.1016/s1097-2765(02)00515-4 (2002).

92 Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annual review of genetics* **43**, 525-558, doi:10.1146/annurev-genet-102108-134233 (2009).

93 Balbas-Martinez, C. *et al.* Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nature genetics* **45**, 1464-1469, doi:10.1038/ng.2799 (2013).

94 Thol, F. *et al.* Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* **123**, 914-920, doi:10.1182/blood-2013-07-518746 (2014).

95 Strom, L. *et al.* Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317**, 242-245, doi:10.1126/science.1140649 (2007).

96 Watrin, E. & Peters, J. M. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *The EMBO journal* **28**, 2625-2635, doi:10.1038/emboj.2009.202 (2009).

97 Tothova, Z. *et al.* Cohesin mutations alter DNA damage repair and chromatin structure and create therapeutic vulnerabilities in MDS/AML. *JCI insight* **6**, doi:10.1172/jci.insight.142149 (2021).

98 Dorsett, D. *et al.* Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. *Development* **132**, 4743-4753, doi:10.1242/dev.02064 (2005).

99 Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801, doi:10.1038/nature06634 (2008).

100 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).

101 Schaaf, C. A. *et al.* Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. *PLoS genetics* **9**, e1003560, doi:10.1371/journal.pgen.1003560 (2013).

102 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324, doi:10.1016/j.cell.2017.09.026 (2017).

103 Kon, A. *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nature genetics* **45**, 1232-1237, doi:10.1038/ng.2731 (2013).

104 Viny, A. D. & Levine, R. L. Cohesin mutations in myeloid malignancies made simple. *Current opinion in hematology* **25**, 61-66, doi:10.1097/MOH.0000000000000405 (2018).

105 Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69, doi:10.1038/nature10496 (2011).

106 Varmus, H. E. The molecular genetics of cellular oncogenes. *Annual review of genetics* **18**, 553-612, doi:10.1146/annurev.ge.18.120184.003005 (1984).

107 Bos, J. L. ras oncogenes in human cancer: a review. *Cancer research* **49**, 4682-4689 (1989).

108 Janssen, J. W. *et al.* RAS gene mutations in acute and chronic myelocytic leukemias, chronic myeloproliferative disorders, and myelodysplastic syndromes. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 9228-9232, doi:10.1073/pnas.84.24.9228 (1987).

109 Greenberger, J. S. ras mutations in human leukemia and related disorders. *International journal of cell cloning* **7**, 343-359, doi:10.1002/stem.5530070603 (1989).

110 Nakagawa, T. *et al.* Multiple point mutation of N-ras and K-ras oncogenes in myelodysplastic syndrome and acute myelogenous leukemia. *Oncology* **49**, 114-122, doi:10.1159/000227023 (1992).

111 Ricci, C. *et al.* RAS mutations contribute to evolution of chronic myelomonocytic leukemia to the proliferative variant. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 2246-2256, doi:10.1158/1078-0432.CCR-09-2112 (2010).

112 Al-Kali, A. *et al.* Prognostic impact of RAS mutations in patients with myelodysplastic syndrome. *American journal of hematology* **88**, 365-369, doi:10.1002/ajh.23410 (2013).

113 Tyner, J. W. *et al.* High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients. *Blood* **113**, 1749-1755, doi:10.1182/blood-2008-04-152157 (2009).

114 Khan, A. Q. *et al.* RAS-mediated oncogenic signaling pathways in human malignancies. *Seminars in cancer biology* **54**, 1-13, doi:10.1016/j.semcancer.2018.03.001 (2019).

115 Zhang, J. *et al.* Oncogenic Kras-induced leukemogeneis: hematopoietic stem cells as the initial target and lineage-specific progenitors as the potential targets for

final leukemic transformation. *Blood* **113**, 1304-1314, doi:10.1182/blood-2008-01-134262 (2009).
116 Santos, F. P. S. *et al.* Prognostic impact of RAS-pathway mutations in patients with myelofibrosis. *Leukemia* **34**, 799-810, doi:10.1038/s41375-019-0603-9 (2020).
117 Lavoie, H. & Therrien, M. Regulation of RAF protein kinases in ERK signalling. *Nature reviews. Molecular cell biology* **16**, 281-298, doi:10.1038/nrm3979 (2015).
118 Yunis, J. J., Boot, A. J., Mayer, M. G. & Bos, J. L. Mechanisms of ras mutation in myelodysplastic syndrome. *Oncogene* **4**, 609-614 (1989).
119 Constantinidou, M. *et al.* Codon 12 ras mutations in patients with myelodysplastic syndrome: incidence and prognostic value. *Annals of hematology* **74**, 11-14, doi:10.1007/s002770050248 (1997).
120 Neubauer, A., Greenberg, P., Negrin, R., Ginzton, N. & Liu, E. Mutations in the ras proto-oncogenes in patients with myelodysplastic syndromes. *Leukemia* **8**, 638-641 (1994).
121 Takahashi, K. *et al.* Dynamic acquisition of FLT3 or RAS alterations drive a subset of patients with lower risk MDS to secondary AML. *Leukemia* **27**, 2081-2083, doi:10.1038/leu.2013.165 (2013).
122 Meggendorfer, M. *et al.* Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. *Haematologica* **100**, e487-490, doi:10.3324/haematol.2015.127985 (2015).
123 Pellagatti, A. *et al.* Targeted resequencing analysis of 31 genes commonly mutated in myeloid disorders in serial samples from myelodysplastic syndrome patients showing disease progression. *Leukemia* **30**, 247-250, doi:10.1038/leu.2015.129 (2016).
124 Huang, H. J. *et al.* [Molecular features and prognostic value of RAS mutations in patients with myelodysplastic syndromes]. *Zhonghua xue ye xue za zhi = Zhonghua xueyexue zazhi* **41**, 723-730, doi:10.3760/cma.j.issn.0253-2727.2020.09.004 (2020).
125 Steensma, D. P. & Bennett, J. M. The myelodysplastic syndromes: diagnosis and treatment. *Mayo Clinic proceedings* **81**, 104-130, doi:10.4065/81.1.104 (2006).
126 Leone, G., Mele, L., Pulsoni, A., Equitani, F. & Pagano, L. The incidence of secondary leukemias. *Haematologica* **84**, 937-945 (1999).
127 Granfeldt Ostgard, L. S. *et al.* Epidemiology and Clinical Significance of Secondary and Therapy-Related Acute Myeloid Leukemia: A National Population-Based Cohort Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **33**, 3641-3649, doi:10.1200/JCO.2014.60.0890 (2015).
128 Hulegardh, E. *et al.* Characterization and prognostic features of secondary acute myeloid leukemia in a population-based setting: a report from the Swedish Acute Leukemia Registry. *American journal of hematology* **90**, 208-214, doi:10.1002/ajh.23908 (2015).
129 Xu, X. Q. *et al.* Characteristics of acute myeloid leukemia with myelodysplasia-related changes: A retrospective analysis in a cohort of Chinese patients. *American journal of hematology* **89**, 874-881, doi:10.1002/ajh.23772 (2014).
130 Awada, H. *et al.* Machine Learning Integrates Genomic Signatures for Subclassification Beyond Primary and Secondary Acute Myeloid Leukemia. *Blood*, doi:10.1182/blood.2020010603 (2021).
131 Weinberg, O. K. *et al.* Clinical characterization of acute myeloid leukemia with myelodysplasia-related changes as defined by the 2008 WHO classification system. *Blood* **113**, 1906-1908, doi:10.1182/blood-2008-10-182782 (2009).
132 Ramos, F. *et al.* Myelodysplastic syndrome: a search for minimal diagnostic criteria. *Leukemia research* **23**, 283-290, doi:10.1016/s0145-2126(98)00166-0 (1999).
133 Senent, L. *et al.* Reproducibility of the World Health Organization 2008 criteria for myelodysplastic syndromes.

Haematologica **98**, 568-575, doi:10.3324/haematol.2012.071449 (2013).

134 Bejar, R. What biologic factors predict for transformation to AML? *Best practice & research. Clinical haematology* **31**, 341-345, doi:10.1016/j.beha.2018.10.002 (2018).

135 Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine* **366**, 1090-1098, doi:10.1056/NEJMoa1106968 (2012).

136 da Silva-Coelho, P. *et al.* Clonal evolution in myelodysplastic syndromes. *Nature communications* **8**, 15099, doi:10.1038/ncomms15099 (2017).

137 Kim, T. *et al.* The clonal origins of leukemic progression of myelodysplasia. *Leukemia* **31**, 1928-1935, doi:10.1038/leu.2017.17 (2017).

138 Lindsley, R. C. *et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* **125**, 1367-1376, doi:10.1182/blood-2014-11-610543 (2015).

139 Yokoyama, K. *et al.* Cell-lineage level-targeted sequencing to identify acute myeloid leukemia with myelodysplasia-related changes. *Blood advances* **2**, 2513-2521, doi:10.1182/bloodadvances.2017010744 (2018).

140 Sperling, A. S., Gibson, C. J. & Ebert, B. L. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nature reviews. Cancer* **17**, 5-19, doi:10.1038/nrc.2016.112 (2017).

141 Chen, J. *et al.* Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nature medicine* **25**, 103-110, doi:10.1038/s41591-018-0267-4 (2019).

142 Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine* **371**, 2488-2498, doi:10.1056/NEJMoa1408617 (2014).

143 Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine* **20**, 1472-1478, doi:10.1038/nm.3733 (2014).

144 Shiozawa, Y. *et al.* Gene expression and risk of leukemic transformation in myelodysplasia. *Blood* **130**, 2642-2653, doi:10.1182/blood-2017-05-783050 (2017).

145 Murphy, D. M. *et al.* NRAS mutations with low allele burden have independent prognostic significance for patients with lower risk myelodysplastic syndromes. *Leukemia* **27**, 2077-2081, doi:10.1038/leu.2013.160 (2013).

146 Kelly, L. M. & Gilliland, D. G. Genetics of myeloid leukemias. *Annual review of genomics and human genetics* **3**, 179-198, doi:10.1146/annurev.genom.3.032802.115046 (2002).

147 Gilliland, D. G. Molecular genetics of human leukemias: new insights into therapy. *Seminars in hematology* **39**, 6-11, doi:10.1053/shem.2002.36921 (2002).

148 Gaidzik, V. & Dohner, K. Prognostic implications of gene mutations in acute myeloid leukemia with normal cytogenetics. *Seminars in oncology* **35**, 346-355, doi:10.1053/j.seminoncol.2008.04.005 (2008).

149 Kitamura, T. *et al.* The molecular basis of myeloid malignancies. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **90**, 389-404, doi:10.2183/pjab.90.389 (2014).

150 Nolte, F. & Hofmann, W. K. Molecular mechanisms involved in the progression of myelodysplastic syndrome. *Future oncology* **6**, 445-455, doi:10.2217/fon.09.175 (2010).

151 List, A. *et al.* Efficacy of lenalidomide in myelodysplastic syndromes. *The New England journal of medicine* **352**, 549-557, doi:10.1056/NEJMoa041668 (2005).

152 List, A. *et al.* Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. *The New England journal of medicine* **355**, 1456-1465, doi:10.1056/NEJMoa061292 (2006).

153 Tehranchi, R. *et al.* Persistent malignant stem cells in del(5q) myelodysplasia in remission. *The New England journal of medicine* **363**, 1025-1037, doi:10.1056/NEJMoa0912228 (2010).

154 Jadersten, M. *et al.* TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 1971-1979, doi:10.1200/JCO.2010.31.8576 (2011).

155 Silverman, L. R. *et al.* Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 2429-2440, doi:10.1200/JCO.2002.04.117 (2002).

156 Dombret, H. *et al.* International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. *Blood* **126**, 291-299, doi:10.1182/blood-2015-01-621664 (2015).

157 Fenaux, P. *et al.* Azacitidine prolongs overall survival compared with conventional care regimens in elderly patients with low bone marrow blast count acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 562-569, doi:10.1200/JCO.2009.23.8329 (2010).

158 Fenaux, P. *et al.* Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *The Lancet. Oncology* **10**, 223-232, doi:10.1016/S1470-2045(09)70003-8 (2009).

159 Santini, V. How I treat MDS after hypomethylating agent failure. *Blood* **133**, 521-529, doi:10.1182/blood-2018-03-785915 (2019).

160 Uy, N., Singh, A., Gore, S. D. & Prebet, T. Hypomethylating agents (HMA) treatment for myelodysplastic syndromes: alternatives in the frontline and relapse settings. *Expert opinion on pharmacotherapy* **18**, 1213-1224, doi:10.1080/14656566.2017.1349100 (2017).

161 Jackson-Grusby, L., Laird, P. W., Magge, S. N., Moeller, B. J. & Jaenisch, R. Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4681-4685, doi:10.1073/pnas.94.9.4681 (1997).

162 Jacoby, M. A. *et al.* Subclones dominate at MDS progression following allogeneic hematopoietic cell transplant. *JCI insight* **3**, doi:10.1172/jci.insight.98962 (2018).

163 Tobiasson, M., Olsson, R., Hellstrom-Lindberg, E. & Mattsson, J. Early detection of relapse in patients with myelodysplastic syndrome after allo-SCT. *Bone marrow transplantation* **46**, 719-726, doi:10.1038/bmt.2010.179 (2011).

164 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467, doi:10.1073/pnas.74.12.5463 (1977).

165 Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology* **122**, e59, doi:10.1002/cpmb.59 (2018).

166 Spaulding, T. P., Stockton, S. S. & Savona, M. R. The evolving role of next generation sequencing in myelodysplastic syndromes. *British journal of haematology* **188**, 224-239, doi:10.1111/bjh.16212 (2020).

167 Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380, doi:10.1038/nature03959 (2005).

168 Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Archives of disease in childhood. Education and practice edition* **98**, 236-238, doi:10.1136/archdischild-2013-304340 (2013).

169 McCombie, W. R., McPherson, J. D. & Mardis, E. R. Next-Generation Sequencing Technologies. *Cold Spring Harbor perspectives in medicine* **9**, doi:10.1101/cshperspect.a036798 (2019).

170 Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K. & Wang, J.

Applications and analysis of targeted genomic sequencing in cancer studies. *Computational and structural biotechnology journal* **17**, 1348-1359, doi:10.1016/j.csbj.2019.10.004 (2019).

171 Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *Journal of human genetics* **59**, 5-15, doi:10.1038/jhg.2013.114 (2014).

172 Schwarze, K., Buchanan, J., Taylor, J. C. & Wordsworth, S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in medicine : official journal of the American College of Medical Genetics* **20**, 1122-1130, doi:10.1038/gim.2017.247 (2018).

173 Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D. & Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC genetics* **18**, 14, doi:10.1186/s12863-017-0479-5 (2017).

174 Bastida, J. M. *et al.* Molecular Diagnosis of Inherited Coagulation and Bleeding Disorders. *Seminars in thrombosis and hemostasis* **45**, 695-707, doi:10.1055/s-0039-1687889 (2019).

175 Wong, T. N. *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**, 552-555, doi:10.1038/nature13968 (2015).

176 Samorodnitsky, E. *et al.* Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human mutation* **36**, 903-914, doi:10.1002/humu.22825 (2015).

177 Mertes, F. *et al.* Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in functional genomics* **10**, 374-386, doi:10.1093/bfgp/elr033 (2011).

178 Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nature methods* **7**, 111-118, doi:10.1038/nmeth.1419 (2010).

179 Robin, J. D., Ludlow, A. T., LaRanger, R., Wright, W. E. & Shay, J. W. Comparison of DNA Quantification Methods for Next Generation Sequencing. *Scientific reports* **6**, 24067, doi:10.1038/srep24067 (2016).

180 McNulty, S. N., Mann, P. R., Robinson, J. A., Duncavage, E. J. & Pfeifer, J. D. Impact of Reducing DNA Input on Next-Generation Sequencing Library Complexity and Variant Detection. *The Journal of molecular diagnostics : JMD* **22**, 720-727, doi:10.1016/j.jmoldx.2020.02.003 (2020).

181 Grada, A. & Weinbrecht, K. Next-generation sequencing: methodology and application. *The Journal of investigative dermatology* **133**, e11, doi:10.1038/jid.2013.248 (2013).

182 Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **13**, 341, doi:10.1186/1471-2164-13-341 (2012).

183 Gargis, A. S. *et al.* Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nature biotechnology* **33**, 689-693, doi:10.1038/nbt.3237 (2015).

184 Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* **15**, 733-747, doi:10.1038/gim.2013.92 (2013).

185 McKerrell, T. *et al.* Development and validation of a comprehensive genomic diagnostic tool for myeloid malignancies. *Blood* **128**, e1-9, doi:10.1182/blood-2015-11-683334 (2016).

186 Shallis, R. M., Ahmad, R. & Zeidan, A. M. The genetic and molecular pathogenesis of myelodysplastic syndromes. *European journal of haematology* **101**, 260-271, doi:10.1111/ejh.13092 (2018).

187 Haferlach, T. The Molecular Pathology of Myelodysplastic Syndrome. *Pathobiology : journal of immunopathology, molecular and cellular biology* **86**, 24-29, doi:10.1159/000488712 (2019).

188 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

189 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).

190 Reiter, J. G. *et al.* An analysis of genetic heterogeneity in untreated cancers. *Nature reviews. Cancer* **19**, 639-650, doi:10.1038/s41568-019-0185-x (2019).

191 Stosch, J. M. *et al.* Gene mutations and clonal architecture in myelodysplastic syndromes and changes upon progression to acute myeloid leukaemia and under treatment. *British journal of haematology* **182**, 830-842, doi:10.1111/bjh.15461 (2018).

192 Polgarova, K. *et al.* Somatic mutation dynamics in MDS patients treated with azacitidine indicate clonal selection in patients-responders. *Oncotarget* **8**, 111966-111978, doi:10.18632/oncotarget.22957 (2017).

193 Cabezon, M. *et al.* Impact of mutational studies on the diagnosis and the outcome of high-risk myelodysplastic syndromes and secondary acute myeloid leukemia patients treated with 5-azacytidine. *Oncotarget* **9**, 19342-19355, doi:10.18632/oncotarget.25046 (2018).

194 Dohner, H. *et al.* Cytogenetics and gene mutations influence survival in older patients with acute myeloid leukemia treated with azacitidine or conventional care. *Leukemia* **32**, 2546-2557, doi:10.1038/s41375-018-0257-z (2018).

195 Uy, G. L. *et al.* Dynamic changes in the clonal structure of MDS and AML in response to epigenetic therapy. *Leukemia* **31**, 872-881, doi:10.1038/leu.2016.282 (2017).

196 Wassmann, K. Sister chromatid segregation in meiosis II: deprotection through phosphorylation. *Cell cycle* **12**, 1352-1359, doi:10.4161/cc.24600 (2013).

197 Seitan, V. C. & Merkenschlager, M. Cohesin and chromatin organisation. *Current opinion in genetics & development* **22**, 93-100, doi:10.1016/j.gde.2011.11.003 (2012).

198 Meisenberg, C. *et al.* Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability. *Molecular cell* **73**, 212-223 e217, doi:10.1016/j.molcel.2018.11.001 (2019).

199 Mondal, G., Stevers, M., Goode, B., Ashworth, A. & Solomon, D. A. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. *Nature communications* **10**, 1686, doi:10.1038/s41467-019-09659-z (2019).

200 Xu, F. *et al.* Rigosertib as a selective anti-tumor agent can ameliorate multiple dysregulated signaling transduction pathways in high-grade myelodysplastic syndrome. *Scientific reports* **4**, 7310, doi:10.1038/srep07310 (2014).

201 O'Bryan, J. P. Pharmacological targeting of RAS: Recent success with direct inhibitors. *Pharmacological research* **139**, 503-511, doi:10.1016/j.phrs.2018.10.021 (2019).

202 Pellagatti, A., Dolatshad, H., Valletta, S. & Boultwood, J. Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Archives of toxicology* **89**, 1023-1034, doi:10.1007/s00204-015-1504-y (2015).

203 Riordan, S. M., Heruth, D. P., Zhang, L. Q. & Ye, S. Q. Application of CRISPR/Cas9 for biomedical discoveries. *Cell & bioscience* **5**, 33, doi:10.1186/s13578-015-0027-9 (2015).

204 Tothova, Z. *et al.* Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell stem cell* **21**, 547-555 e548, doi:10.1016/j.stem.2017.07.015 (2017).

205 Drexler, H. G., MacLeod, R. A. & Uphoff, C. C. Leukemia cell lines: in vitro models for the study of Philadelphia chromosome-positive leukemia. *Leukemia research* **23**, 207-215, doi:10.1016/s0145-2126(98)00171-4 (1999).

206 Kida, J. I. *et al.* An MDS-derived cell line and a series of its sublines serve as an in vitro model for the leukemic evolution of MDS. *Leukemia* **32**, 1846-1850, doi:10.1038/s41375-018-0189-7 (2018).

207 Drexler, H. G., Dirks, W. G. & Macleod, R. A. Many are called MDS cell lines: one is chosen. *Leukemia research* **33**, 1011-1016, doi:10.1016/j.leukres.2009.03.005 (2009).

208 Palau, A. *et al.* Immunophenotypic, cytogenetic, and mutational characterization of cell lines derived from myelodysplastic syndrome patients after progression to acute myeloid leukemia. *Genes, chromosomes & cancer* **56**, 243-252, doi:10.1002/gcc.22430 (2017).

209 Rouault-Pierre, K. *et al.* Preclinical modeling of myelodysplastic syndromes. *Leukemia* **31**, 2702-2708, doi:10.1038/leu.2017.172 (2017).

210 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

211 Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244**, 1288-1292, doi:10.1126/science.2660260 (1989).

212 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

213 Sanchez-Rivera, F. J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nature reviews. Cancer* **15**, 387-395, doi:10.1038/nrc3950 (2015).

214 Tu, Z., Yang, W., Yan, S., Guo, X. & Li, X. J. CRISPR/Cas9: a powerful genetic engineering tool for establishing large animal models of neurodegenerative diseases. *Molecular neurodegeneration* **10**, 35, doi:10.1186/s13024-015-0031-x (2015).

List of Genes

ABL1	Abelson Murine Leukemia Viral Oncogene Homolog 1
AEBP2	Adipocyte Enhancer-Binding Protein 2
ARID2	AT-Rich Interaction Domain 2
ASXL1	Additional Sex Combs like Transcriptional Regulator 1
ATRX	Alpha Thalassemia/Mental Retardation Syndrome X-linked
BCAS1	Brain Enriched Myelin Associated Protein 1
BCOR	BCL6 Corepressor
BCORL1	BCL6 Corepressor-like 1
BCR	Breakpoint Cluster Region Protein
BMI1	BMI1 Proto-Oncogene, Polycomb Ring Finger
BRAF	B-Raf Proto-Oncogene, serine/threonine Kinase
CALR	Calreticulin
CBFb	Core-Binding Factor Subunit Beta
CBL	Cbl proto-oncogene
CBLB	Cbl proto-oncogene B
CBLC	Cbl proto-oncogene C
CD177	Human Neutrophil Alloantigen 2a
CDH13	Cadherin 13
CDH23	Cadherin 23
CDH3	Cadherin 3
CDK2	Cyclin Dependent Kinase 2
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CEBPA	CCAAT Enhancer Binding Protein Alpha
CREBBP	CREB Binding Protein
CSF3R	Colony Stimulating Factor 3 Receptor
CSNK1A1	Casein Kinase 1 Alpha 1
CTCF	CCCTC-Binding Factor
CTNNA1	Catenin Alpha 1
CUX1	Cut Like Homeobox 1
DNMT3A	DNA (cytosine-5)-Methyltransferase 3 Alpha
EED	Embryonic Ectoderm Development
EGFR	Epidermal Growth Factor Receptor
EIF2AK2	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 2
ENG	Endoglina
EP300	E1A Binding Protein P300
ETV6	Ets Variant 6
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FBXW7	F-Box And WD Repeat Domain Containing 7
FLT3	Fms-related tyrosine Kinase 3
G3BP1	Ras-GTPase-Activating Protein SH3-Domain-Binding Protein 2
GATA1	GATA Binding Protein 1
GATA2	GATA Binding Protein 2
GCAT	Glycine C-Acetyltransferase
GNAS	GNAS Complex Locus
GNB1	G Protein Subunit Beta 1
HRAS	HRas Proto-Oncogene, GTPase

IDH1	Isocitrate Dehydrogenase 1
IDH2	Isocitrate Dehydrogenase 2
IKZF1	IKAROS Family Zinc Finger 1
IL3	Interleukin 3
IRF1	Interferon Regulatory Factor 1
JAK1	Janus Kinase 1
JAK2	Janus Kinase 2
JAK3	Janus Kinase 3
JARID2	Jumonji And AT-Rich Interaction Domain Containing 2
JKAMP	JNK1/MAPK8 Associated Membrane Protein
KDM6A	Lysine (K)-specific Demethylase 6A
KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase
KMT2A	Lysine Methyltransferase 2A
KMT2D	Lysine Methyltransferase 2D
KRAS	KRAS Proto-Oncogene, GTPase
LUC7L2	LUC7 Like 2, Pre-MRNA Splicing Factor
MECOM	MDS1 And EVI1 Complex Locus
MPL	MPL Proto-Oncogene, Thrombopoietin Receptor
MTOR	Mechanistic Target Of Rapamycin Kinase
NF1	Neurofibromin 1
NOTCH1	Notch Receptor 1
NPM1	Nucleophosmin 1
NR2F6	Nuclear Receptor Subfamily 2 Group F Member 6
NRAS	NRAS Proto-Oncogene, GTPase
NTRK1	Neurotrophic Receptor Tyrosine Kinase 1
NUP98	Nucleoporin 98 and 96 Precursor
PARP	Poly adenosine diphosphate-ribose polymerase
PBRM1	Polybromo 1
PDGFRA	Platelet Derived Growth Factor Receptor Alpha
PDGFRB	Platelet Derived Growth Factor Receptor Beta
PHF19	PHD Finger Protein 19
PHF6	PHD Finger Protein 6
PHLPP1	PH Domain and Leucine Rich Repeat Protein Phosphatase 1
PTEN	Phosphatase And Tensin Homolog
PTPN1	Protein Tyrosine Phosphatase Non-Receptor Type 1
PTPN11	Protein Tyrosine Phosphatase Non-Receptor Type 11
RAD21	RAD21 Cohesin Complex Component
RARA	Retinoic Acid Receptor Alpha
RET	Ret Proto-Oncogene
RPS14	Ribosomal Protein S14
RUNX1	RUNX Family Transcription Factor 1
SALL4	Spalt Like Transcription Factor 4
SBDS	SBDS Ribosome Maturation Factor
SETBP1	SET Binding Protein 1
SETD2	SET Domain Containing 2, Histone Lysine Methyltransferase
SF1	Splicing Factor 1
SF3A1	Splicing Factor 3a Subunit 1
SF3B1	Splicing Factor 3b Subunit 1
SFPQ	Splicing Factor Proline And Glutamine Rich
SH2B3	SH2B Adaptor Protein 3
SMC1A	Structural Maintenance Of Chromosomes 1A
SMC3	Structural Maintenance Of Chromosomes 3

SPARC	Secreted Protein Acidic And Cysteine Rich
SRSF2	Serine and Arginine Rich Splicing Factor 2
STAG1	Stromal Antigen 1
STAG2	Stromal Antigen 2
SUZ12	SUZ12 Polycomb Repressive Complex 2 Subunit
TCL1B	TCL1 Family AKT Coactivator B
TERC	Telomerase RNA Component
TERT	Telomerase Reverse Transcriptase
TET2	Tet Methylcytosine Dioxygenase 2
TGM2	Transglutaminase 2
TIMM50	Translocase Of Inner Mitochondrial Membrane 50
TNFAIP3	TNF Alpha Induced Protein 3
TP53	Tumor Protein P53
TYK2	Tyrosine Kinase 2
U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1
UMODL1	Uromodulin Like 1
USB1	U6 SnRNA Biogenesis Phosphodiesterase 1
WASF3	WASP Family Member 3
WT1	WT1 Transcription Factor
ZRSR2	Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2

List of Tables and Figures

INTRODUCTION

RESULTS - CHAPTER 1

RESULTS - CHAPTER 2

RESULTS - CHAPTER 3

DISCUSSION

Supplementary Appendix

Supplementary Appendix - Chapter 1

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*].

Haematologica (2020). doi: 10.3324/haematol.2020.248807. PMID: 32675227.

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.

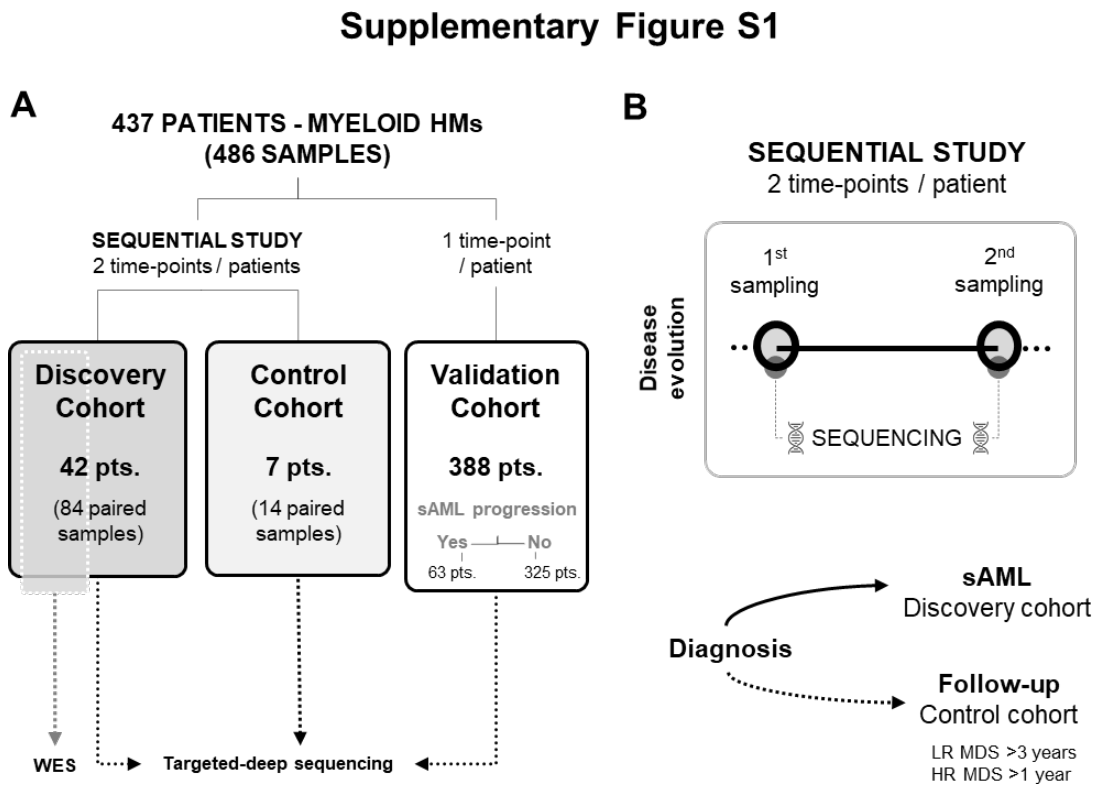


Figure S2: VAF comparison of the mutations detected by whole-exome sequencing vs. by targeted deed sequencing. The VAF correlation between these two platforms was high measured by Pearson coefficient (Pearson´s *r* = 0.90).

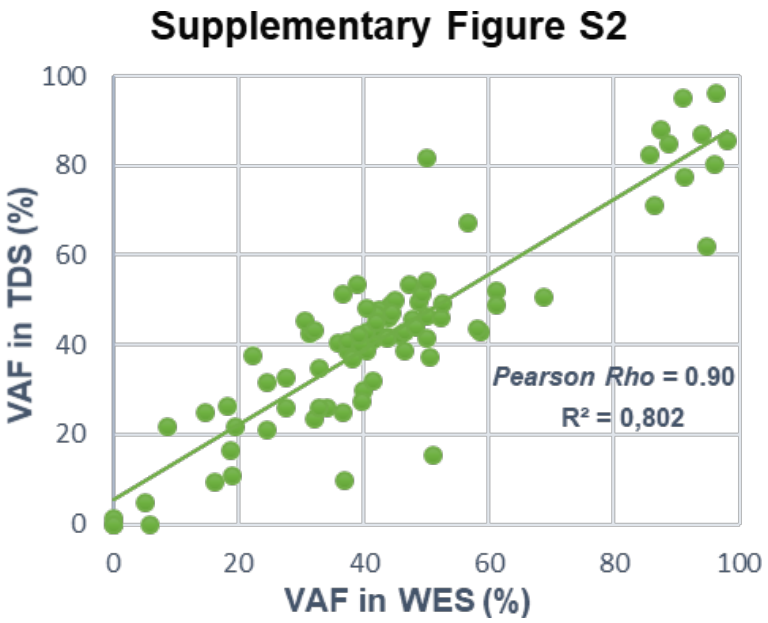


Figure S3: Representative examples of genes with different mutational dynamics in MDS patients who evolved to sAML. VAF at diagnosis and sAML of all mutations detected in *STAG2*, *NRAS*, *FLT3*, *SRSF2* and *DNMT3A* in our discovery cohort were represented showing type 1, type 3 and type 4 dynamics,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

Supplementary Figure S3

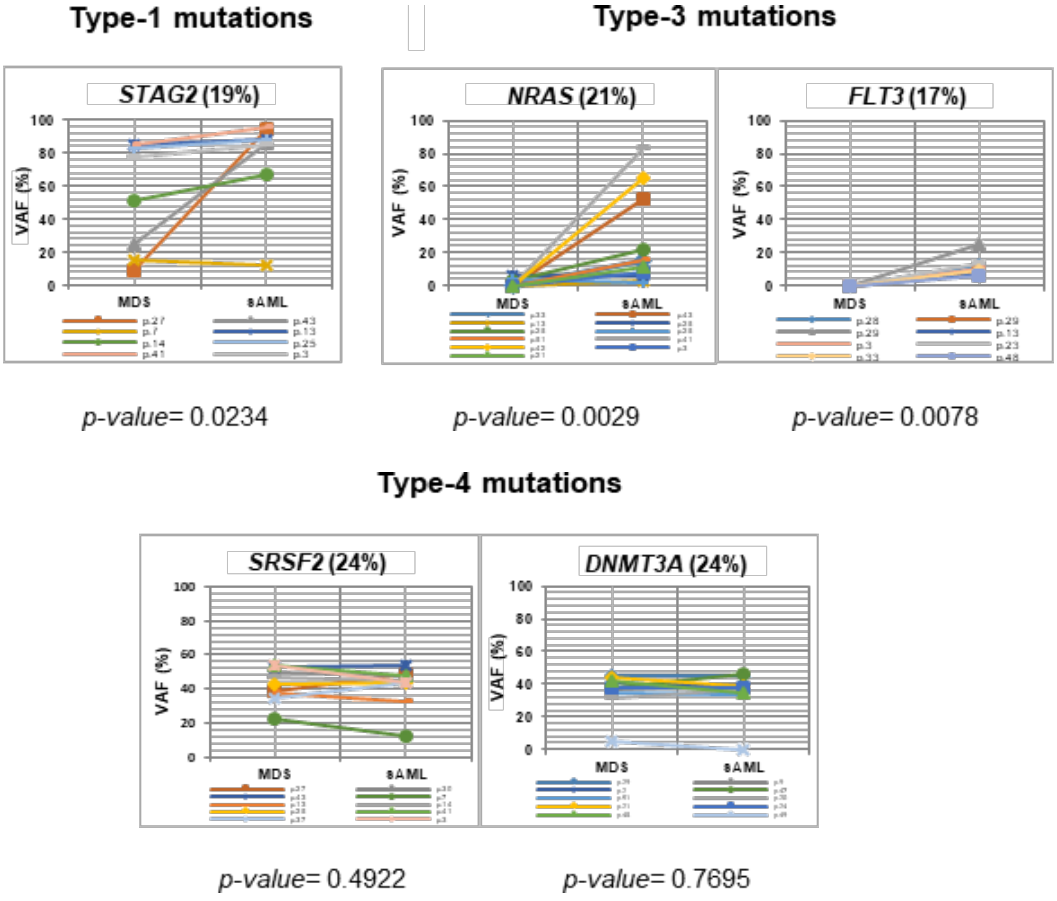


Figure S4: Prognostic impact of single mutations and co-occurring mutations in the cohesin complex and Ras pathway. Kaplan-Meier curves for overall survival in double-mutant and cohesin and Ras single mutant patients in the entire validation cohort.

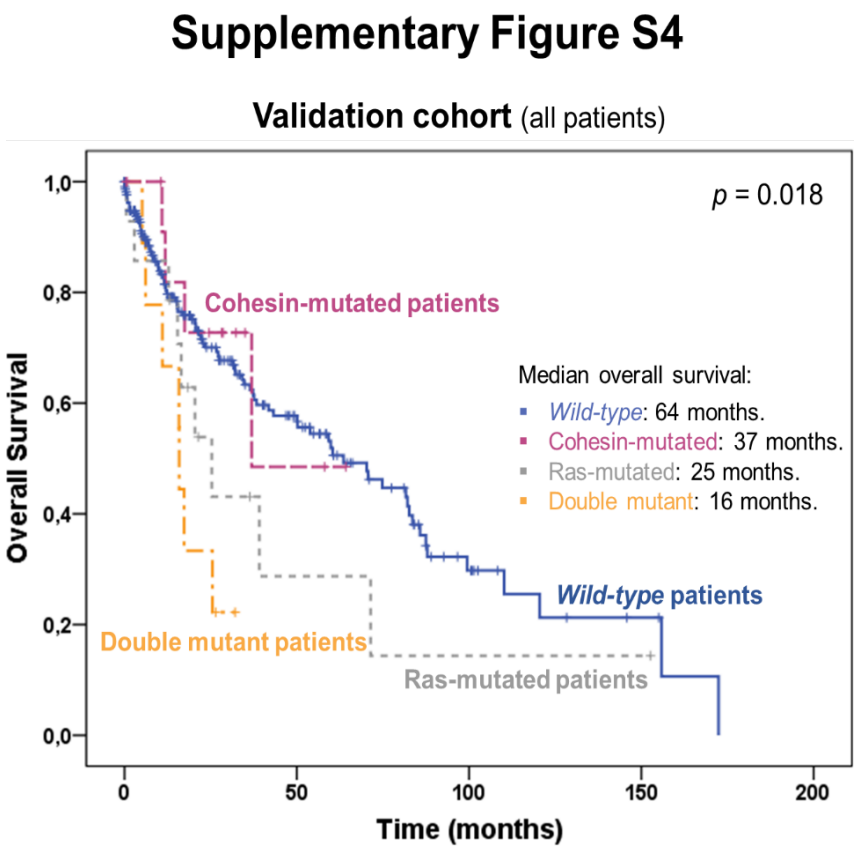
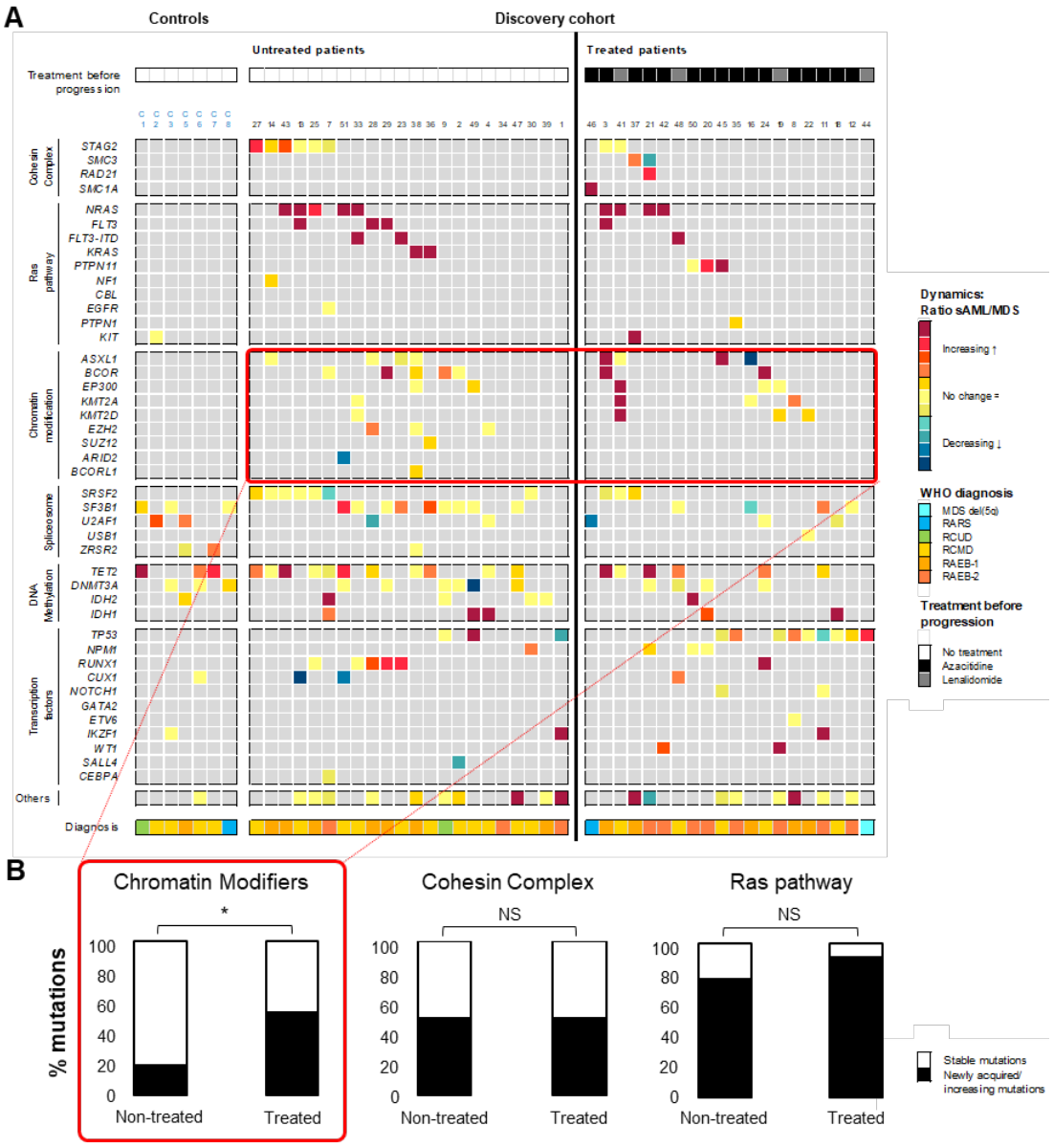


Figure S5: Landscape of mutational dynamics according to disease-modifying treatment on MDS patients who progressed to sAML. A) Patients in the discovery cohort were grouped according to whether they received treatment with a disease-modifying agent (38% 5-azacytidine or 10% lenalidomide) or supportive 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



Supplementary Appendix - Chapter 2

Incorporation of cohesin mutational data into current IPSS-R classification may improve the prognostic stratification of very low/low-risk myelodysplastic syndromes

Marta Martín-Izquierdo[1], María Díez-Campelo[2], Javier Sánchez-Real[3], Alberto Hernández-Sánchez[2], Jesús M Hernández-Sánchez[1], Kamila Janusz[4], Félix López-Cadenas[2], Javier Martínez-Elicegui[1], Mar Tormo[5], Marta Megido[6], Carmen Olivier[7], Andrés Madinaveitia-Ochoa[8], Julio Dávila[9], Magdalena Sierra[10], Manuel Vargas[11], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Rocío Benito[1], Jesús M Hernández-Rivas[1,2], Fernando Ramos[3*] and María Abáigar[1*]

American Journal of Hematology. Second review.

This section contains:

- 4 supplementary tables.
- 5 supplementary figures.

SUPPLEMENTARY TABLES

Table S1: Main clinical and biological characteristics of the entire cohort included in the study (n=418): demographics, WHO 2008 and 2017 subtypes, IPSS/IPSS-R risk classification, peripheral blood and bone marrow parameters and cytogenetics.

Table S2: Panel of 117 myeloid-related genes used for targeted-deep sequencing.

Table S3: List of all cohesin mutations found by targeted-deep sequencing in 418 MDS patients. The table includes information on gene, chromosome position (GRCh37/hg19), change at DNA level, Variant Allele Frequency (VAF) percentage, type of change or consequence at protein level, transcript by RefSeq nomenclature and change at protein level.

Table S4: Logistic Regression. The table includes information about the logistic regression analysis in patients harboring the co-occurrence of cohesin and *SRSF2* mutations.

Supplementary Table S1

Total cohort (n = 418)	
Gender, male, n (%)	246 (58.9%)
Age at diagnosis, median (range, years)	74.9 (29.4-92.2)
Vital status (dead), n (%)	48.6%
sAML Progression, (%)	28.8%
Disease-modifying Treatment, yes (%)	39.7%
WHO 2008 Classification at diagnosis, n (%)	418
RCUD	26 (6.2%)
RARS	40 (9.6%)
RCMD	186 (44.5%)
RAEB-1	56 (13.4%)
RAEB-2	60 (14.4%)
MDS del(5q)-	38 (9.1%)
MDS-U	12 (2.9%)
WHO 2017 Classification at diagnosis, n (%)	375
MDS-SLD	18 (4.8%)
MDS-RS-SLD	21 (5.6%)
MDS-MLD	103 (27.5%)
MDS-RS-MLD	68 (18.1%)
MDS-EB1	55 (14.7%)
MDS-EB2	61 (16.3%)
MDS del(5q)	38 (10.1%)
MDS-U	11 (2.9%)
IPSS classification, n (%)	365
Low	170 (46.6%)
Intermediate-1	128 (35.1%)
Intermediate-2	51 (14%)
High	16 (4.3%)
IPSS-R classification, n (%)	347
Very low	93 (26.8%)
Low	154 (44.4%)
Intermediate	53 (15.3%)
High	27 (7.8%)
Very high	20 (5.8%)
Blood count at diagnosis	361
Hemoglobin level, g/dL, median (range)	9.9 (3.8-15.4)
Platelet count, $x10^9$/L, median (range)	152 (2.0-1067.0)
ANC, $x10^9$/L, median (range)	1.9 (0.05-56.0)
WBC, $x10^9$/L, median (range)	4.0 (1.21-16.0)
Bone marrow study at diagnosis	362
% blasts, median (range)	1.6 (0.0-19.9)
%ring sideroblasts, median (range)	0.0 (0.0-100.0)
Cytogenetic risk at diagnosis (%)	380
Very good	16 (4.2%)
Good	305 (80.3%)
Intermediate	35 (9.2%)
Poor	12 (3.2%)
Very poor	12 (3.2%)

Abbreviations: sAML, secondary acute myeloid leukemia; RCUD, refractory cytopenia with unilineage dysplasia; RARS, refractory cytopenia with ring sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RAEB, refractory anemia with excess blasts; MDS del(5q), myelodysplastic syndrome with isolated del(5q); MDS-U, myelodysplastic syndrome unclassifiable; MDS-SLD, syndrome myelodysplastic with single lineage dysplasia; MDS-RS-SLD, syndrome myelodysplastic with ring sideroblasts and single lineage dysplasia; MDS-MLD, syndrome myelodysplastic with multi-lineage dysplasia; MDS-RS-MLD, syndrome myelodysplastic with ring sideroblasts and multi-lineage dysplasia; MDS-EB1, myelodysplastic syndrome with excess blasts type-1; MDS-EB2, myelodysplastic syndrome with excess blasts type-2; IPSS, International Prognostic Scoring System; IPSS-R, International Prognostic Scoring System Revised; WBC, white blood cell; ANC, absolute neutrophil count.

Supplementary Table S2

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 Table S3

Patient	Gene	Position (GRCh37/hg19)	Change (DNA)	VAF (%)	Type of change	Transcript (RefSeq)	Change (Protein)
#1	*STAG2*	chrX: 123215311	C>C/T	98,05	stopgain SNV	NM_006603	p.R953X
#2	*CTCF*	chr16: 67655488	G>G/T	19,85	nonsynonymous SNV	NM_001191022	p.D123Y
#3	*SMC3*	chr10: 112361839	G>G/A	40,09	nonsynonymous SNV	NM_005445	p.R1003H
#4	*STAG2*	chrX: 123171416	C>C/T	16,39	stopgain SNV	NM_006603	p.R110X
#5	*STAG2*	chrX: 123179197	C>C/T	17,14	stopgain SNV	NM_006603	p.R216X
#6	*RAD21*	chr8: 117859879	G>G/A	15,75	stopgain SNV	NM_006265	p.R586X
#6	*STAG2*	chrX: 123182897	A>A/G	98,76	nonsynonymous SNV	NM_006603	p.I288V
#7	*STAG2*	chrX: 123217380	C>C/T	50,56	stopgain SNV	NM_006603	p.R1012X
#8	*STAG2*	chrX: 123185051	insG	31,48	frameshift insertion	NM_006603	p.L366fs
#9	*STAG2*	chrX: 123196770	insAGAAGAC	91,51	frameshift insertion	NM_006603	p.K553fs
#10	*STAG2*	chrX: 123196795	insA	38,96	frameshift insertion	NM_006603	p.T561fs
#11	*STAG2*	chrX: 123224612	delTAAG	68,35	nonframeshift deletion	NM_006603	p.1155_1156del
#12	*SMC1A*	chrX: 53432009	G>G/A	87,56	nonsynonymous SNV	NM_006306	p.R711W
#13	*STAG2*	chrX: 123202438	C>C/T	85,29	stopgain SNV	NM_006603	p.Q764X
#14	*SMC3*	chr10: 112352865	A>A/G	47,32	nonsynonymous SNV	NM_005445	p.N616S
#15	*STAG2*	chrX: 123200239	insTAAT	25,00	frameshift insertion	NM_006603	p.V740fs
#16	*STAG2*	chrX: 123197044	C>C/T	84,52	stopgain SNV	NM_006603	p.R604X
#17	*STAG2*	chrX: 123220440	C>C/T	15,72	stopgain SNV	NM_006603	p.R1033X
#18	*STAG2*	chrX: 123197751	insA	37,68	frameshift insertion	NM_006603	p.D625fs
#19	*STAG2*	chrX: 123220476	C>C/T	9,21	stopgain SNV	NM_006603	p.R1045X
#20	*SMC3*	chr10: 112340744	G>G/A	36,86	nonsynonymous SNV	NM_005445	p.R171Q
#20	*RAD21*	chr8: 117869620	delG	2,93	frameshift deletion	NM_006265	p.L192fs
#21	*STAG2*	chrX: 123181311	C>C/T	85,00	stopgain SNV	NM_006603	p.R259X
#22	*SMC3*	chr10: 112342391	delTAAAATGGAGG	13,90	frameshift deletion	NM_005445	p.D265fs
#23	*STAG2*	chrX: 123217389	A>A/T	87,21	stopgain SNV	NM_006603	p.K1015X
#24	*STAG2*	chrX: 123205051	C>C/G	69,06	stopgain SNV	NM_006603	p.S804X

Supplementary Table S3

Patient	Gene	Position (GRCh37/hg19)	Change (DNA)	VAF (%)	Type of change	Transcript (RefSeq)	Change (Protein)
#25	*STAG2*	chrX: 123220440	C>C/T	49,40	stopgain SNV	NM_006603	p.R1033X
#26	*SMC3*	chr10: 112344033	T>T/A	32,60	nonsynonymous SNV	NM_005445	p.I395N
#27	*STAG2*	chrX: 123202460	insA	78,14	frameshift insertion	NM_006603	p.I771fs
#28	*STAG2*	chrX: 123200233	insA	10,00	frameshift insertion	NM_006603	p.H738fs
#29	*STAG2*	chrX: 123179034	insA	82,49	frameshift insertion	NM_006603	p.L161fs
#30	*RAD21*	chr8: 117868494	G>G/A	54,30	nonsynonymous SNV	NM_006265	p.P283L
#31	*RAD21*	chr8: 117864225	G>G/A	3,95	stopgain SNV	NM_006265	p.R478X
#32	*STAG2*	chrX: 123220440	C>C/T	32,64	stopgain SNV	NM_006603	p.R1033X
#33	*STAG2*	chrX: 123164975	delG	4,21	frameshift deletion	NM_006603	p.Q96fs
#34	*STAG2*	chrX: 123220456	G>G/A	15,56	stopgain SNV	NM_006603	p.W1038X
#35	*STAG2*	chrX: 123176433	G>G/T	11,28	stopgain SNV	NM_006603	p.E134X
#36	*STAG2*	chrX: 123197770	T>T/C	71,91	nonsynonymous SNV	NM_006603	p.C632R
#36	*STAG2*	chrX: 123179188	delAG	16,67	frameshift deletion	NM_006603	p.R213fs
#37	*STAG2*	chrX: 123220567	C>C/G	19,39	stopgain SNV	NM_006603	p.S1075X
#38	*SMC3*	chr10: 112342391	delTAAAATGGAGG	8,92	frameshift deletion	NM_005445	p.D265fs
#39	*SMC1A*	chrX: 53436388	C>C/T	40,54	nonsynonymous SNV	NM_006306	p.R434Q
#40	*STAG2*	chrX: 123220440	C>C/T	30,81	stopgain SNV	NM_006603	p.R1033X
#41	*RAD21*	chr8: 117859899	delAT	6,39	frameshift deletion	NM_006265	p.I579fs
#42	*STAG2*	chrX: 123176469	C>C/T	28,44	stopgain SNV	NM_006603	p.R146X
#42	*STAG2*	chrX: 123224457	delT	23,68	stopgain SNV	NM_006603	p.L1104X
#43	*STAG2*	chrX: 123196757	insA	77,12	frameshift insertion	NM_006603	p.L548fs
#44	*SMC1A*	chrX: 53432008	C>C/T	25,95	nonsynonymous SNV	NM_006306	p.R711Q
#45	*SMC3*	chr10: 112341840	G>G/T	51,88	nonsynonymous SNV	NM_005445	p.R236L
#46	*STAG2*	chrX: 123215311	C>C/T	11,11	stopgain SNV	NM_006603	p.R953X
#47	*STAG2*	chrX: 123191804	insT	37,50	frameshift insertion	NM_006603	p.V465fs
#48	*STAG2*	chrX: 123184977	insAA	30,23	frameshift insertion	NM_006603	p.E342fs

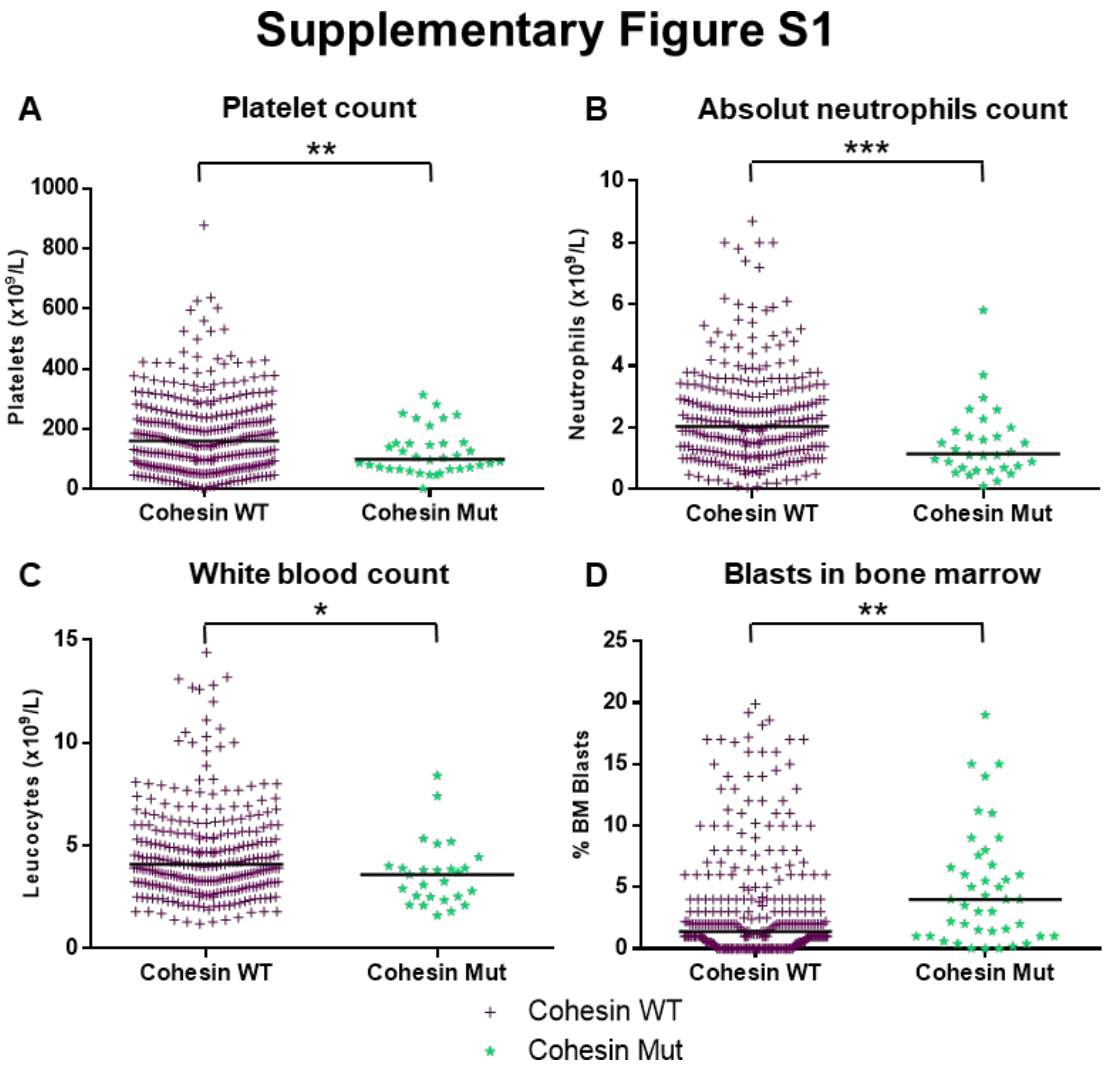
Supplementary Table S4

Logistic Regression

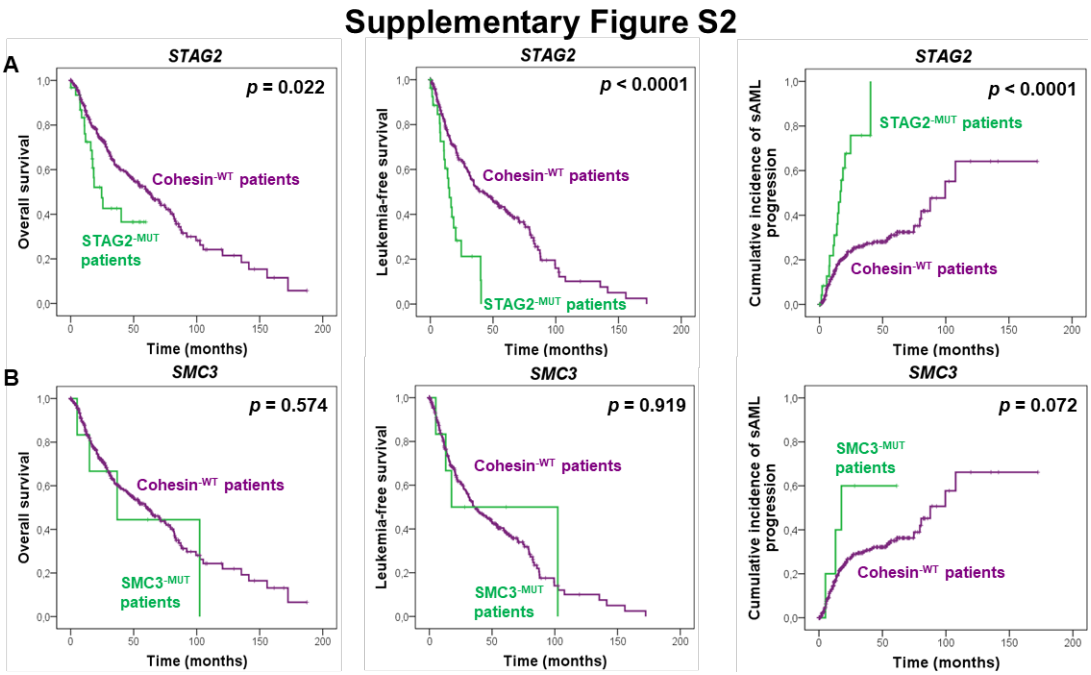
	p-value	Odds ratio	95% CI
BM Blasts	0.349	0.92	0.77-1.10
Hemoglobin	0.342	1.16	0.86-1.56
Platelets	0.094	0.99	0.98-1.00
Neutrophils	0.131	0.56	0.27-1.19
Cytogenetics	0.255	0.44	0.10-1.82
AML progression	**0.017***	**6.78**	**1.41-32.57**
Status	0.493	1.69	0.38-7.63

SUPPLEMENTARY FIGURES

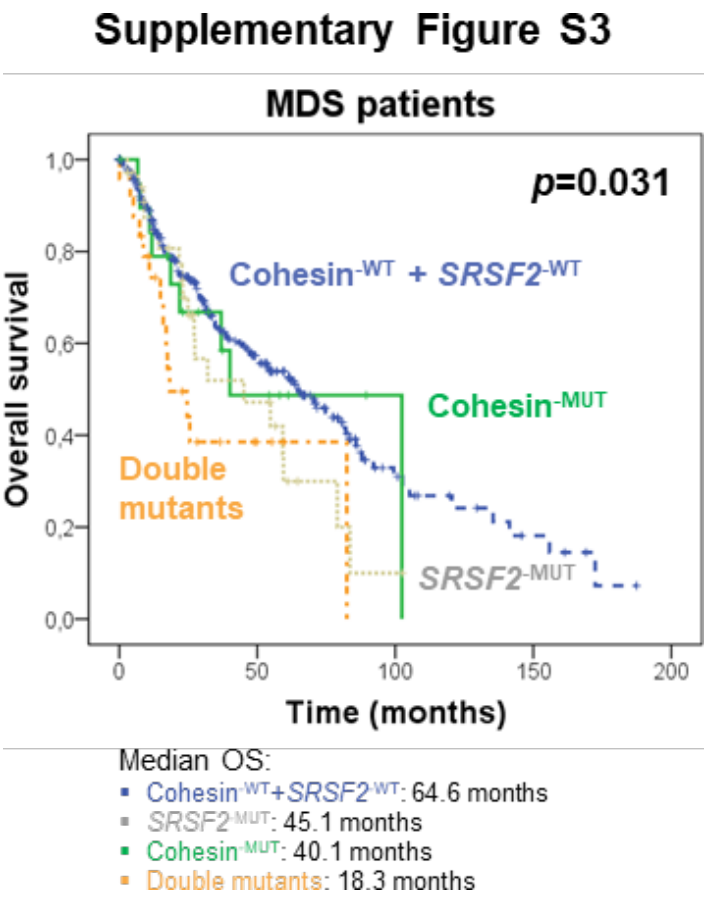
Supplementary Figure S1: Patients harboring cohesin mutations displayed a higher incidence of poor prognosis markers. A) Cohesin^{-MUT} patients showed a lower platelet count in peripheral blood (median 161 *vs.* 100.5 x10^9 platelets/L, *p*=0.009), **B)** a lower absolute neutrophils count (ANC) in peripheral blood (median 2.0 *vs.* 1.2 x10^9 neutrophils/L, *p*=0.0008), **C)** a lower white blood count in peripheral blood (median 4.1 *vs.* 3.6 x10^9 leucocytes/L, *p*=0.016) and **D)** a higher number of blasts in bone marrow (median 4% *vs.* 1.4% blasts, *p*=0.0014). Individual as well as median values are depicted.



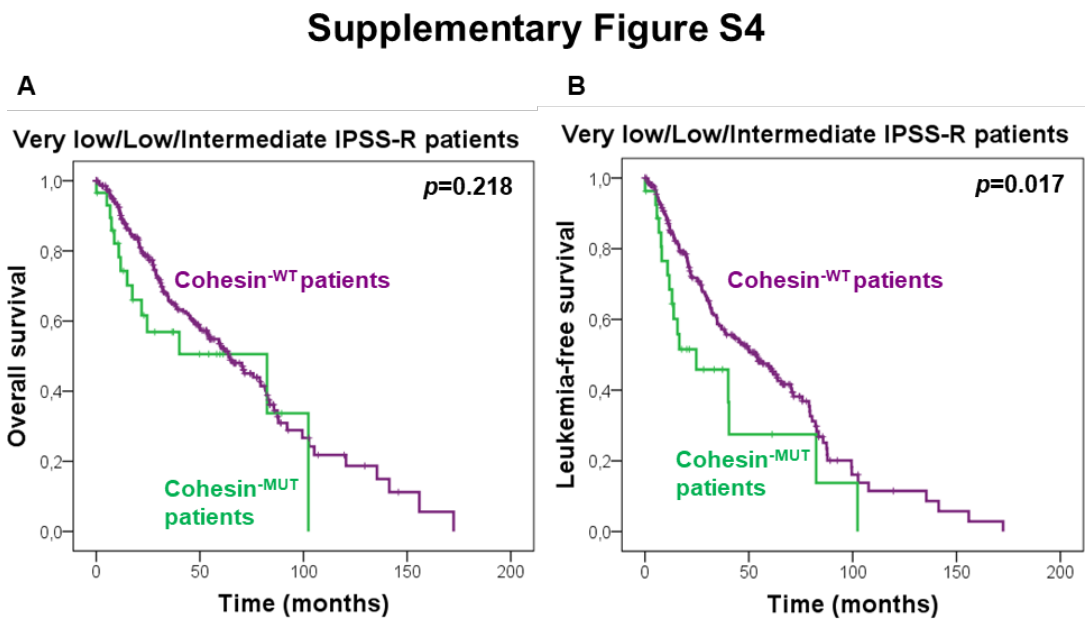
Supplementary Figure S2: Prognostic influence of mutations in the cohesin genes *STAG2* and *SMC3*. A) Kaplan-Meier curves of *STAG2* on overall survival (p=0.022), leukemia free survival (p<0.0001) and time to leukemic progression (p<0.0001). **B)** Kaplan-Meier curves of *SMC3* on overall survival (p=0.574), leukemia free survival (p=0.919) and time to leukemic progression (p=0.072).



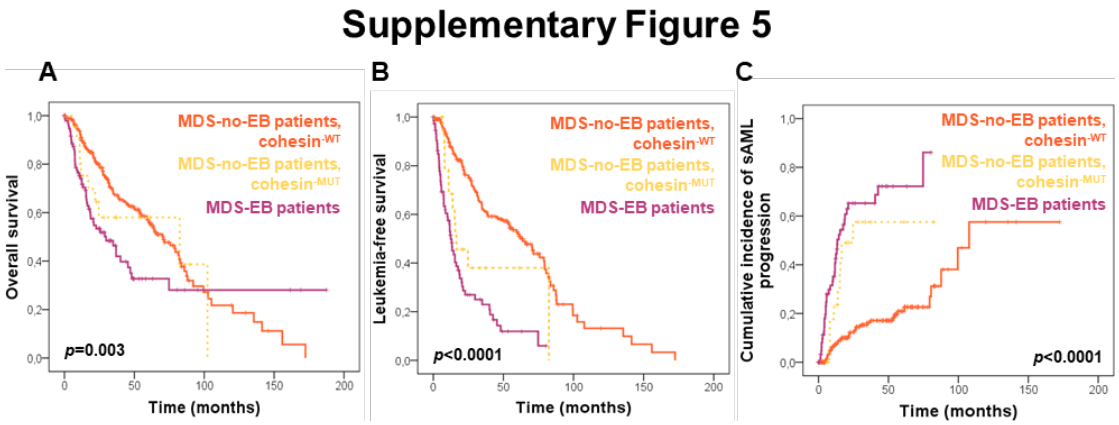
Supplementary Figure S3: Kaplan-Meier curves for OS in the entire MDS cohort according to the presence of: cohesin-*SRSF2* double mutations, *SRSF2* single mutations (*SRSF2*$^{-MUT}$), cohesin single mutations (cohesin^{-MUT}), and cohesin-*SRSF2* wild type (cohesin^{-WT} + *SRSF2*$^{-WT}$). Double mutants showed a shorter OS than *SRSF2*$^{-MUT}$ and cohesin^{-MUT} patients (18.3 *vs.* 45.1 *vs.* 40.1 months, respectively, *p*=0.031).



Supplementary Figure S4: Kaplan-Meier curves for overall survival and leukemia-free survival in very low/low/intermediate-risk IPSS-R patients according to the presence/absence of cohesin gene mutations. IPSS-R very low/low/intermediate-risk patients harboring cohesin mutations displayed a significant shorter leukemia-free survival than cohesin^{-WT} patients (median: 24.7 *vs.* 53.8 months, p=0.017), while no significant differences in OS were observed (p=0.218).



Supplementary Figure S5: Overall survival, leukemia free survival and time to sAML progression according to the presence of excess of blasts by WHO classification and cohesin mutational status. A) Kaplan-Meier curves for overall survival (OS), B) leukemic free survival (LFS) and C) time to sAML progression (TTL) in MDS patients without excess of blasts (MDS-no-EB) according to the presence or absence of cohesin mutations, versus all MDS patients with excess of blasts (MDS-EB). MDS with excess of blasts type 1 and 2 by WHO 2017 classification were considered as MDS-EB and the remaining categories were considered as MDS-no-EB, except MDS unclassifiable which were excluded in the analyses. The graphs show that cohesin mutations stratified MDS-no-EB patients in two different groups, displaying cohesin^{MUT} patients a shorter OS (p=0.003), LFS (p<0.0001) and TTL (p<0.0001) than cohesin^{WT}, but comparable to MDS-EB.



Supplementary Appendix - Chapter 3

RAS-pathway mutations are associated with a worse clinical outcome in myelodysplastic syndromes patients

Marta Martín-Izquierdo[1], Fernando Ramos[2], María Abáigar[1], Jesús M Hernández-Sánchez[1], Teresa González[1,3], Félix López-Cadenas[3], Miguel Quijada-Álamo[1], Sandra Santos-Mínguez[1], Cristina Miguel-García[1], Eva Lumbreras[1], Kamila Janusz[4], Mónica del Rey[1], Sofía Toribio-Castelló[1], Jesús M Hernández-Rivas[1,3], María Díez-Campelo[3*] and Rocío Benito[1*]

Cancers. Submitted.

This section contains:

- 3 supplementary tables.
- 1 supplementary figures.

SUPPLEMENTARY TABLES

Table S1: Main clinical and biological characteristics of the entire cohort included in the study (n=418): demographics, WHO 2008 and 2017 subtypes, IPSS/IPSS-R risk classification, peripheral blood and bone marrow parameters and cytogenetics.

Table S2: Panel of 117 myeloid-related genes used for targeted-deep sequencing. Genes were classified with respect to a list of seven biological pathways: Ras signaling pathway, splicing, DNA methylation, chromatin modification, cohesin complex, transcription and signaling.

Table S3: List of all cohesin mutations found by targeted-deep sequencing in 418 MDS patients. The table includes information on gene, chromosome position (GRCh37/hg19), transcript by RefSeq nomenclature, change at DNA level and at protein level, type of change or consequence at protein level and Variant Allele Frequency (VAF) percentage.

Supplementary Table S1

Total cohort (n = 418)	Patient numbers (ranges or percentages)
Gender, male	58.9%
Age at diagnosis, median (range, years)	74.9 (29.4-92.2)
WHO 2008 Classification (n=418)	
RCUD	6.2%
RARS	9.6%
RCMD	44.5%
RAEB-1	13.4%
RAEB-2	14.4%
MDS del(5q)-	9.1%
MDS-U	2.9%
WHO 2017 Classification (n=375)	
MDS-SLD	4.8%
MDS-RS-SLD	5.6%
MDS-MLD	27.5%
MDS-RS-MLD	18.1%
MDS-EB1	14.7%
MDS-EB2	16.3%
MDS del(5q)	10.1%
MDS-U	2.9%
IPSS classification (n=365)	
Low	46.6%
Intermediate-1	35.1%
Intermediate-2	14%
High	4.3%
IPSS-R classification (n=347)	
Very low	26.8%
Low	44.4%
Intermediate	15.3%
High	7.8%
Very high	5.8%
Cytogenetics risk at diagnosis (n=380)	
Very good	4.2%
Good	80.3%
Intermediate	9.2%
Poor	3.2%
Very poor	3.2%
Blood count at diagnosis (n=361)	
Hemoglobin level, g/dL, median (range)	9.9 (3.8-15.4)
Platelet count, $x10^9$/L, median (range)	152 (2.0-1067.0)
ANC, $x10^9$/L, median (range)	1.9 (0.05-56.0)
WBC, $x10^9$/L, median (range)	4.0 (1.21-16.0)
Bone marrow study at diagnosis (n=362)	
% blasts, median (range)	1.6 (0.0-19.9)
%ring sideroblasts, median (range)	0.0 (0.0-100.0)
Deceased	48.6%
sAML Progression	28.8%
Disease-modifying Treatment, yes	39.7%

Abbreviations: sAML, secondary acute myeloid leukemia; RCUD, refractory cytopenia with unilineage dysplasia; RARS, refractory cytopenia with ring sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RAEB, refractory anemia with excess blasts; MDS del(5q), myelodysplastic syndrome with isolated del(5q); MDS-U, myelodysplastic syndrome unclassifiable; MDS-SLD, syndrome myelodysplastic with single lineage dysplasia; MDS-RS-SLD, syndrome myelodysplastic with ring sideroblasts and single lineage dysplasia; MDS-MLD, syndrome myelodysplastic with multi-lineage dysplasia; MDS-RS-MLD, syndrome myelodysplastic with ring sideroblasts and multi-lineage dysplasia; MDS-EB1, myelodysplastic syndrome with excess blasts type-1; MDS-EB2, myelodysplastic syndrome with excess blasts type-2; IPSS, International Prognostic Scoring System; IPSS-R, International Prognostic Scoring System Revised; WBC, white blood cell; ANC, absolute neutrophil count.

Supplementary Table S2

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

Supplementary Table S3

Patient	Gene	Position (GRCh37/hg19)	Transcript (RefSeq)	Change (DNA)	Change (Protein)	Type of change	VAF (%)
#7	*BRAF*	chr7: 140434555	NM_004333	c.G2143T	p.E715X	stopgain	6,92
#14	*BRAF*	chr7: 140453155	NM_004333	c.G1780A	p.D594N	missense	2,58
#13	*BRAF*	chr7: 140482916	NM_004333	c.C1219A	p.P407T	missense	12,06
#25	*CBL*	chr11: 119148891	NM_005188	c.T1111C	p.Y371H	missense	3,1
#41	*CBL*	chr11: 119148921	NM_005188	c.T1141C	p.C381R	missense	16,18
#29	*CBL*	chr11: 119148967	NM_005188	c.G1187A	p.C396Y	missense	38,13
#39	*CBL*	chr11: 119149235	NM_005188	c.G1243A	p.G415S	missense	27,49
#27	*CBL*	**chr11: 119149251**	NM_005188	c.G1259C	p.R420P	missense	1,88
#33	*CBL*	**chr11: 119149251**	NM_005188	c.G1259A	p.R420Q	missense	13,5
#46	*EGFR*	chr7: 55229216	NM_005228	c.C1523A	p.A508D	missense	46,95
#47	*EGFR*	chr7: 55240785	NM_005228	c.C2029T	p.R677C	missense	14,97
#23	*FLT3*	chr13: 28592641	NM_004119	c.A2504T	p.D835V	missense	27,8
#20	*KIT*	**chr4: 55599321**	NM_000222	c.A2447T	p.D816V	missense	29,55
#40	*KIT*	**chr4: 55599321**	NM_000222	c.A2447T	p.D816V	missense	11,27
#43	*KIT*	chr4: 55602733	NM_000222	c.G2554A	p.V852I	missense	51,10
#24	*KIT*	chr4: 55602961	NM_000222	c.A2671G	p.S891G	missense	49,39
#30	*KIT*	chr4: 55603390	NM_000222	c.A2746G	p.T916A	missense	53,2
#11	*KRAS*	chr12: 25380279	NM_004985	c.G179T	p.G60V	missense	17,69
#38	*KRAS*	chr12: 25398211	NM_004985	c.A108G	p.I36M	missense	47,12
#44	*KRAS*	chr12: 25398262	NM_004985	c.G57T	p.L19F	missense	9,57
#35	*KRAS*	chr12: 25398281	NM_004985	c.G38A	p.G13D	missense	9,35
#45	*KRAS*	chr12: 25398282	NM_004985	c.G37T	p.G13C	missense	8,81
#36	*KRAS*	chr12: 25398284	NM_004985	c.G35A	p.G12D	missense	7,5
#14	*NF1*	chr17: 29527539	NM_000267	c.G988C	p.A330P	missense	10,38
#16	*NF1*	chr17: 29541542	NM_000267	c.A1466G	p.Y489C	missense	6,42
#25	*NF1*	chr17: 29546108	NM_000267	c.1614dupG	p.M538fs	frameshift insertion	50,60
#5	*NF1*	**chr17: 29553477**	NM_000267	c.2027dupC	p.T676fs	frameshift insertion	5,6
#21	*NF1*	**chr17: 29553477**	NM_000267	c.2027dupC	p.T676fs	frameshift insertion	12,87

Patient	Gene	Position (GRCh37/hg19)	Transcript (RefSeq)	Change (DNA)	Change (Protein)	Type of change	VAF (%)
#8	*NF1*	chr17: 29553493	NM_000267	c.G2042A	p.R681Q	missense	43,16
#37	*NF1*	chr17: 29553631	NM_000267	c.2180_2181insAGTGCAT	p.S727fs	frameshift insertion	15,65
#9	*NF1*	chr17: 29557868	NM_000267	c.T3122A	p.M1041K	missense	49,69
#3	*NF1*	chr17: 29585423	NM_000267	c.G4172C	p.R1391T	missense	6,82
#34	*NF1*	chr17: 29657445	NM_000267	c.C5678A	p.T1893K	missense	41,51
#15	*NF1*	chr17: 29683564	NM_000267	c.7639delA	p.K2547fs	frameshift deletion	75,2
#45	*NRAS*	chr1: 115256529	NM_002524	c.A182C	p.Q61P	missense	8,91
#31	*NRAS*	chr1: 115258703	NM_002524	c.C79T	p.H27Y	missense	53,35
#2	*NRAS*	**chr1: 115258744**	NM_002524	c.G38A	p.G13D	missense	11,09
#12	*NRAS*	**chr1: 115258744**	NM_002524	c.G38A	p.G13D	missense	1,88
#26	*NRAS*	**chr1: 115258744**	NM_002524	c.G38A	p.G13D	missense	4,29
#42	*NRAS*	**chr1: 115258744**	NM_002524	c.G38A	p.G13D	missense	1,64
#28	*NRAS*	chr1: 115258745	NM_002524	c.G37T	p.G13C	missense	6,29
#1	*NRAS*	**chr1: 115258747**	NM_002524	c.G35T	p.G12V	missense	2,64
#14	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	1,8
#18	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	31,48
#19	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	3,09
#22	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	28,44
#28	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	3,15
#45	*NRAS*	**chr1: 115258747**	NM_002524	c.G35A	p.G12D	missense	19,77
#2	*NRAS*	**chr1: 115258748**	NM_002524	c.G34A	p.G12S	missense	3,81
#4	*NRAS*	**chr1: 115258748**	NM_002524	c.G34A	p.G12S	missense	3,64
#6	*NRAS*	**chr1: 115258748**	NM_002524	c.G34T	p.G12C	missense	4,17
#28	*NRAS*	**chr1: 115258748**	NM_002524	c.G34A	p.G12S	missense	3,7
#32	*NRAS*	**chr1: 115258748**	NM_002524	c.G34C	p.G12R	missense	20,63
#10	*PTPN11*	chr12: 112888157	NM_002834	c.A173G	p.N58S	missense	40,39
#4	*PTPN11*	chr12: 112888195	NM_002834	c.T211C	p.F71L	missense	15,98
#26	*PTPN11*	chr12: 112888199	NM_002834	c.C215T	p.A72V	missense	7,06
#6	*PTPN11*	chr12: 112915523	NM_002834	c.A922G	p.N308D	missense	6,02
#17	*PTPN11*	chr12: 112919950	NM_002834	c.A1165C	p.K389Q	missense	7,84
#4	*PTPN11*	chr12: 112926885	NM_002834	c.C1505T	p.S502L	missense	16,3

SUPPLEMENTARY FIGURES

Supplementary Figure S1: Prognostic influence of mutations in the *RAS*-pathway genes *NRAS* and *NF1*. A) Kaplan-Meier curves of *NRAS* on overall survival (p=0.514), leukemia free survival (p=0.043) and time to leukemic progression (p=0.010). **B)** Kaplan-Meier curves of *NF1* on overall survival (p<0.0001), leukemia free survival (p=0.003) and time to leukemic progression (p=0.572).