Universitat Politècnica de València
School of Agricultural Engineering and Environment (ETSIAMN)
Bachelor's Degree in Biotechnology

MOLECULAR CHARACTERIZATION OF PATIENTS WITH HEREDITARY MYELOID NEOPLASMS BY NEXT-GENERATION SEQUENCING

Natalia Pardo Lorente

Final Degree Project
Academic year: 2017/2018
Valencia, June 2018

External Tutor: Dr. José Cervera Zamora
External Co-tutor: PhD. Mariam Ibáñez Company
Academic Tutor: Prof. José Gadea Vacas

MOLECULAR CHARACTERIZATION OF PATIENTS WITH HEREDITARY MYELOID NEOPLASMS BY NEXT-GENERATION SEQUENCING

Author: Natalia Pardo Lorente
Valencia, June 2018

Externat Tutor: Dr. José Cervera Zamora
External Co-tutor: PhD. Mariam Ibáñez Company
Academic Tutor: Prof. José Gadea Vacas

ABSTRACT

Hereditary myeloid malignancy syndromes (HMMSs) consist of a group of hematologic disorders with a germinal basis and with high levels of genetic and phenotypic heterogeneity. This group includes familial cases of myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). As a result of the technological progress of next-generation sequencing (NGS), germline alterations have been identified in a series of genes related to these hereditary myeloid neoplasms, and with a higher frequency than initially expected. In fact, in the 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia, these hereditary cases have been included as a new category. Consequently, current clinical guidelines strongly recommend studying every patient diagnosed with AML/MDS and suspicious of inherited predisposition, and for this purpose it is imperative to develop a NGS strategy that enables to identify new cases of familial myeloid neoplasms. An early detection of these familial cases is a key element when choosing an appropriate donor in case the patient is going to undergone allogeneic hematopoietic stem cell transplantation, due to the fact that, if the donor is related, both could carry the same mutation. Ultimately, it will allow for an improvement in patients' and carriers' management and clinical care, a better choice of treatment, specialized supervision and genetic counselling for both patients and family. This project aims to evaluate the frequency of germinal alterations in these patients. To this effect, a targeted multi-gene panel was designed in order to study simultaneously a group of several genes associated with HMMSs in a cohort of young patients (under the age of 60) diagnosed with AML/MDS. In each case, it is mandatory to perform a NGS analysis of the DNA at the moment of diagnosis as well as of a paired germinal sample. All the information contained in the personal and familial medical history must be also evaluated in order to find evidence that makes us suspicious of a HMMS. In this way, it will become easier to detect new cases and to evaluate the prevalence of HMMSs in adult population.

KEYWORDS: hereditary myeloid malignancy syndromes, germline predisposition, next-generation sequencing, targeted gene panel, myeloid neoplasm.

RESUMEN

Los síndromes hereditarios mieloides malignos (HMMSs) son un grupo de trastornos hematológicos de base germinal que presentan una gran heterogeneidad genética y fenotípica. Dentro de este grupo se incluye a los síndromes mielodisplásicos (SMD) y leucemias mieloides agudas (LMA) de carácter familiar. Como resultado de los avances en secuenciación masiva (NGS), se han podido identificar alteraciones germinales en una serie de genes relacionados con estas neoplasias mieloides hereditarias, y con una mayor frecuencia de lo que se esperaba inicialmente. De hecho, en la revisión de 2016 de la clasificación de la Organización Mundial de la Salud de neoplasias mieloides y leucemias agudas, se ha añadido una nueva categoría que incluye los casos asociados a mutaciones germinales. Las guías clínicas actuales recomiendan, por tanto, estudiar a todos los pacientes diagnosticados con LMA/SMD y con sospecha de predisposición hereditaria, para lo cual es imperativo desarrollar una estrategia de NGS que permita identificar nuevos casos de neoplasias mieloides familiares. La detección temprana de estos casos familiares es crucial para una adecuada selección del donante en caso de que el paciente se someta a un trasplante alogénico de células progenitoras hematopoyéticas, ya que, si el donante es emparentado, ambos podrían ser portadores de la misma mutación germinal. En última instancia, se consigue mejorar el manejo clínico y atención médica de pacientes y portadores, pudiendo seleccionar el tratamiento más adecuado, y proporcionar un seguimiento especializado y consejo genético tanto para los pacientes como para los familiares. El presente proyecto pretende evaluar la frecuencia de alteraciones germinales en estos pacientes. Para ello, se diseñó un panel de genes dirigido con el fin de estudiar de manera simultánea un conjunto de genes asociados a los HMMSs en una cohorte de pacientes jóvenes (menores de 60 años) diagnosticados con LMA/SMD. En cada caso, es imprescindible realizar un análisis de NGS en la muestra de DNA del momento de diagnóstico así como en una muestra pareada germinal. Toda la información contenida en la historia clínica personal y familiar del paciente también debe ser evaluada en busca de indicios que nos lleven a pensar que se trata de un HMMS. De este modo, se podrá facilitar la detección de nuevos casos y evaluar la prevalencia de HMMSs en la población adulta.

PALABRAS CLAVE: síndromes hereditarios mieloides malignos, predisposición germinal, secuenciación masiva, panel de genes dirigido, neoplasia mieloide.

ACKNOWLEDGMENTS

I had previously been at the IIS La Fe in an internship period during summer 2016, and it was such a gratifying and fruitful experience that I decided that I will come back two years later to do my final degree project there. Not only the IIS La Fe is an institution in the vanguard of biomedical research and with a national and international prestige, but also a pleasant working environment, in which I have been surrounded by expert scientists and students with different academic backgrounds, all of them working together with the same objective: the progression of human health.

I would like to express my very great appreciation specially to my external tutors, Mariam Ibáñez and José Cervera, for their expert advice and for the continuous supervision and encouragement through all the project. During an eight-month period of work, I have acquired a high knowledge about hematologic neoplasms, and I have experienced how the progression in different research areas is translated to a better management and clinical care of the patients, which is a really gratifying feeling. Thank you very much, not only for contributing to my training as a future researcher, but also for treating me with affection and respect. I hope that this personal and professional relationship will continue in the near future.

In addition, I am very grateful for the assistance given by other members of the research group of Hematology and Hemotherapy of the IIS La Fe and other members of the Hospital La Fe: Esperanza Such, Alessandro Liquori, Ángel Zúñiga, Juan Silvestre Oltra, Desiree Company, Eva Barragán, Claudia Sargas, Joaquín Panadero, Alejandra Pla, Mireia Boluda, Mireia Morote, Elisa González, Cristina Valiente and David Gómez. Each of them has contributed to my final project, either by helping me to cope with several difficulties or by providing me with useful information for my project. Thank you for making this stay in the IIS La Fe an exceptional experience of my life.

Finally, I would also like to offer my special thanks to my academic tutor, José Gadea. Because of his classes of the subject Genomics, I became interested in Omics technologies and this is why I decided to do this final project and to further continue my studies with a Master in Genomic Medicine. Thank you for accepting being my tutor from the very first moment, for supporting me whenever I had any problems or doubts, and guiding my project with your advice and suggestions.

.

INDEX

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

ACMG: American College of Medical Genetics and Genomics
AD: autosomal dominant
AML: acute myeloid leukemia
CBC: cell blood count
CLL: chronic lymphoid leukemia
CML: chronic myeloid leukemia
CMML: chronic myelomonocytic leukemia
CNV: copy number variation
DX: diagnosis
ELN: European Leukemia Net
ER: endoplasmic reticulum
Eur: European
FPDMM: familial platelet disorder with predisposition to myeloid malignancy
gDNA: genomic DNA
HMMSs: hereditary myeloid malignancy syndromes
HSCT: hematopoietic stem cell transplantation
HSF: Human Splicing Finder
IBMFSs: inherited bone marrow failure syndromes
InDels: small insertions and deletions
LOF: loss of function
MAF: minor allele frequency
MDS: myelodysplastic syndrome
MPN: myeloproliferative neoplasm
NA: not available
NCCN: National Comprehensive Cancer Network
NGS: next-generation sequencing
SNVs: single nucleotide variants
rear.: rearrangement
UTR: untranslated region
VAF: variant allele frequency
VUS: variant of uncertain significance
WES: whole-exome sequencing
WGS: whole-genome sequencing
WHO: World Health Organization

1. INTRODUCTION

1. INTRODUCTION

Hematologic neoplasms emerge from an uncontrolled proliferation of abnormal bone marrow cell populations carrying genetic alterations. Malignant hematologic disorders may affect myeloid or lymphoid lineages depending on the type of initiator cell, myeloblast or lymphoblast, respectively. This project focuses on myeloid hematologic neoplasms, specifically, acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS). MDS is a clonal bone marrow malignancy in which an altered hematopoiesis results in precursor cells with morphologic dysplasia and peripheral blood cytopenias (Arber *et al.,* 2016). MDS has usually a late onset and sometimes evolves to AML. AML is the most common type of *de novo* leukemia in adults. It is caused by malignant myeloblastic cells with acquired mutations that clonally expand and prevent downstream differentiation (De Kouchkovsky and Abdul-Hay, 2016). Recently, germline mutations have been reported in a small percentage of these neoplasms. Therefore, differentiating hereditary cases from acquired AML/MDS is imperative as patient management is totally different in both cases.

1.1 HEREDITARY MYELOID MALIGNANCY SYNDROMES

Hereditary myeloid malignancy syndromes (HMMSs) consist of myeloid neoplasms with a germline predisposition, including familial syndromes with predisposition to AML/MDS, inherited bone marrow failure syndromes (IBMFSs), familial myeloproliferative neoplasms (MPN) and traditional hereditary cancer syndromes (The University of Chicago Hematopoietic Malignancies Cancer Risk Team, 2016). Familial AML and MDS had been initially considered as rare neoplasms in adult population, as they were commonly related to childhood. However, it is increasingly frequent to diagnose hereditary hematologic syndromes with an increased risk of developing myeloid neoplasms in adulthood (Brown *et al*., 2017; Feurstein *et al*., 2016). The first hereditary myeloid neoplasm defined was Familial Platelet Disorder with predisposition to Myeloid Malignancy (FPDMM) due to germline *RUNX1* mutations, originally identified in 1999 (Song *et al*., 1999) and followed by AML with inherited *CEBPA* mutations, defined in 2004 (Smith *et al*., 2004). Since then, additional inherited hematologic syndromes, such as hereditary AML/MDS with mutated *DDX41* (Polprasert *et al*., 2015), familial thrombocytopenia-2 and thrombocytopenia-5 with altered *ANKRD26* and *ETV6,* respectively (Pippucci *et al*., 2011; Zhang *et al*., 2015), and familial AML/MDS with *GATA2* mutations, have been described (Hahn *et al.*, 2012).

Due to next-generation sequencing (NGS) improvements, the list of genes associated with predisposition to myeloid malignancies is continuously increasing. For instance, germinal *SRP72* mutations are related to familial aplastic anemia/MDS (Kirwan *et al*., 2012); *ATG2B* and *GSKIP* germline duplication is associated with familial MPN and AML (Saliba *et al.*, 2015); and germline mutations in cancer predisposition genes *BRCA1/BRCA2* and *TP53* also increase the risk of developing a leukemogenic process (Schulz *et al*., 2012).

To date, the list of genes associated with family cases of AML/MDS includes transcription factors such as *CEBPA, RUNX1, ETV6* and *GATA2,* helicases as *DDX41*, signalling molecules like *ANKRD26* and *GSKIP,* proteins involved in maintaining genomic stability

like *TP53, BRCA1* and *BRCA2,* in protein translation and transport such as *SRP72,* and in autophagy like *ATG2B* (Figure 1). The fact that these genes codify for proteins involved in a wide range of different molecular and cellular mechanisms reflects the heterogeneity of hematologic malignancies, which can arise from failures in diverse biological pathways (Brown *et al*., 2017; Porter, 2016). However, due to the enormous increase of sequencing projects, new information is becoming available on almost a daily basis and novel HMMSs-related genes are likely to be identified.

The increasing recognition of HMMSs is reflected by the new category of "myeloid neoplasms with germline predisposition" stated in the 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia (Table 1). This classification establishes three sub-groups: "myeloid neoplasms with germline predisposition without a pre-existing disorder or organ dysfunction", including AML with *CEBPA* mutations and AML/MDS with mutated *DDX41; "*myeloid neoplasms with germline predisposition and pre-existing platelet disorders", comprised of FPDMM due to *RUNX1* mutations, thrombocytopenia-2 with mutated *ANKRD26* and thrombocytopenia-5 with *ETV6* mutations; and "myeloid neoplasms with germline predisposition and other organ dysfunction", which includes myeloid neoplasms with *GATA2* mutations, myeloid neoplasms associated with IBMFSs, with telomere biology disorders, with neurofibromatosis, Noonan syndrome or Noonan syndrome-like disorders and with Down syndrome (Arber *et al*., 2016). Additionally, the National Comprehensive Cancer Network (NCCN) and the European Leukemia Net (ELN) have incorporated new guidelines to improve treatment and management of patients with hereditary AML/MDS (Dohner *et al*., 2017; Greenberg *et al.*, 2017). Accordingly, it is imperative to adopt a new paradigm for addressing hematologic neoplasms. These patients can be recognised by genetic testing of a large number of genes to detect either acquired or germinal mutations. In this regard, a multi-gene panel approach would enable the identification of familial cases, improving patients' management according to the risk and providing a better personalized treatment.

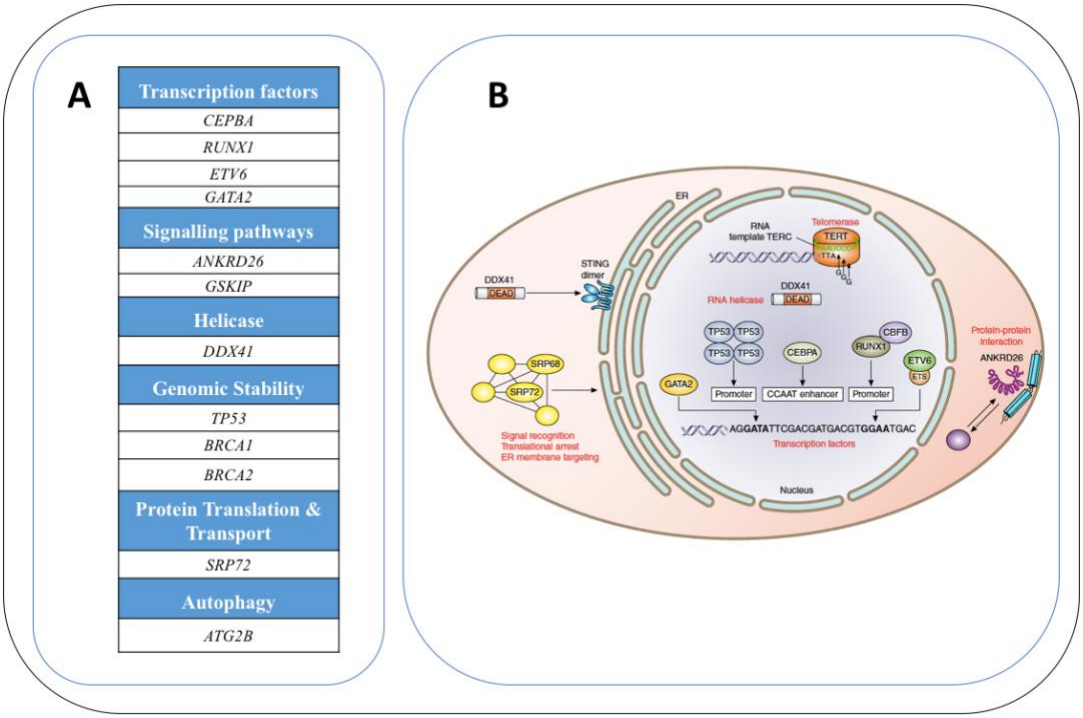

FIGURE 1 - **A) Biological functions of genes whose germinal mutations predispose to hematologic malignancies. B) Roles of proteins codified by HMMSs-related genes at the cellular level.** Reprinted from: The University of Chicago Hematopoietic Malignancies Cancer Risk Team, 2016.
ER, endoplasmic reticulum.

TABLE 1 – **WHO classification of myeloid neoplasms with germline predisposition** (Arber *et al*., 2016).

Myeloid neoplasms with germline predisposition
Myeloid neoplasms with germline predisposition without a pre-existing disorder or organ dysfunction
AML with germline *CEBPA* mutation
Myeloid neoplasms with germline *DDX41* mutation*
Myeloid neoplasms with germline predisposition and pre-existing platelet disorders
Myeloid neoplasms with germline *RUNX1* mutation*
Myeloid neoplasms with germline *ANKRD26* mutation*
Myeloid neoplasms with germline *ETV6* mutation*
Myeloid neoplasms with germline predisposition and other organs dysfunction
Myeloid neoplasms with germline *GATA2* mutation
Myeloid neoplasms associated with bone marrow failure syndrome
Myeloid neoplasms associated with telomere biology disorders
Myeloid neoplasms associated with neurofibromatosis, Noonan syndrome or Noonan syndrome-like disorders
Myeloid neoplasms associated with Down syndrome*

*Lymphoid malignancies with these germline mutations have also been reported.

1.1.1 WHO classification of hereditary myeloid neoplasms

1.1.1.1 Myeloid neoplasms with germline predisposition without a pre-existing disorder or organ dysfunction

1.1.1.1.1 AML with germline *CEBPA* mutation

CEBPA gene, located in the chromosomal region 19q13.1 and with one single exon, codifies for the protein CCAAT/enhancer-binding protein alpha (CEPBA), a transcription factor that recognizes the CCAAT motif located in the promoters of target genes. This transcription factor is involved in myeloid differentiation, as it activates promoters of myeloid-specific growth-factors receptors, namely the granulocyte colony-stimulating factor receptor and neutrophil granule proteins (Radomska *et al*., 1998). It contains a specific DNA sequence binding motif, a C-terminal basic leucine zipper domain for dimerization, and N-terminal transactivation domains (Smith *et al.*, 2004). This protein can work as an homodimer or as an heterodimer, together with CCAAT/enhancer-binding proteins beta and gamma (Pabst *et al*. , 2008).

Several germline mono-allelic mutations have been reported in families suffering from inherited hematologic malignancies (Figure 2). Frameshift mutations located in the 5' region result in an increased expression of an alternative shorter version of CEPBA protein. Mutant CEBPA loses its capacity to promote granulocytic differentiation, leading to a higher risk of developing AML (Pabst *et al.*, 2001). Besides the germinal mono-allelic mutation, a high percentage of individuals acquire a second mutation in the healthy allele, hindering the recognition of germline *CEBPA* alterations. In fact, approximately 10% of patients with *CEBPA* mutations turn out to have inherited mutations. Familial AML with germline *CEBPA* mutation behaves as an autosomal dominant (AD) disorder. Its incidence is ~1% from the total AML cases (Pabst *et al*., 2008), with an early onset, near-complete penetrance, favourable prognosis and a similar phenotype when compared to sporadic AML with somatic *CEPBA* mutations (Tawana *et al*., 2015). Proper discrimination between sporadic and hereditary cases is crucial for genetic counselling and patient monitoring.

1.1.1.1.2 Myeloid neoplasms with germline *DDX41* mutation

DDX41 gene is located in the chromosomal region 5q35.3 and includes 17 coding exons. It codifies for a DEAD box protein family member, characterized by the conserved DEAD/Asp-Glu-Ala-Asp motif. This protein is a putative RNA helicase that participates in the assembly of the spliceosome by interacting with several spliceosomal proteins. *DDX41* is expressed in precursor myeloid cells, suggesting a role in hematopoiesis. Nevertheless, its function in the development of the leukemogenic process remains unclear. Functional studies show that loss of expression of *DDX41* results in a higher proliferation and colony formation ability and impairs differentiation, providing transformed cells with a competitive advantage. This evidence suggests that this gene has a tumor suppressor function and that it is a relevant driver in the development of myeloid malignancies (Polprasert *et al*., 2015).

Hereditary AML/MDS with germline *DDX41* mutation, with an AD pattern, is presented with a late onset, notable penetrance, poor prognosis and an inferior overall survival, and its incidence is low (~0.75%). When these patients develop a myeloid malignancy, they are characterized by peripheral blood cytopenias, macrocytosis, a hypocellular bone marrow and erythroid dysplasia (Lewinsohn *et al*., 2016). The frameshift mutation p.D140fs*2 is present in most of the familial cases, however, there are other possibilities such as splice variants or missense mutations (Figure 2) (Cheah *et al*., 2017). Also, around 50% of individuals with germinal mutations then acquire secondary somatic mutations in the healthy allele of the gene (Polprasert *et al*., 2015).

1.1.1.2 Myeloid neoplasms with germline predisposition and pre-existing platelet disorders

1.1.1.2.1 Myeloid neoplasms with germline *RUNX1* mutation

RUNX1 (Runt-related transcription factor 1), located in 21q22.12 and with eight coding exons, codifies for a transcription factor involved in the regulation of hematopoiesis, particularly, in the maturation of hematopoietic stem cells. It contains a C-terminal transactivation domain and a N-terminal highly-conserved runt homology domain. RUNX1 protein, previously named core binding factor alpha 2, interacts with core binding factor beta *(*CBFβ*)*. This last protein facilitates the attachment of RUNX1 to DNA, and together, they form the core binding factor complex, an heterodimeric transcription factor (Schlegelberger and Heller, 2017). Functional studies revealed that mutated *RUNX1* alters the development of primitive erythroid cells and megakaryocytes, and granulocyte differentiation *(*Antony-Debré *et al., 2015;* Behrens *et al.*, 2016).

Patients with FPDMM with germline *RUNX1* mutations have a very variable phenotype. They may have mild to moderate thrombocytopenia, suffer from severe bleeding due to functional platelet defects and be at a high risk of developing a myeloid neoplasm. These malignancies have normally a childhood or early adulthood onset (Latger-Cannard *et al*., 2016; Schlegelberger and Heller, 2017). Despite its germinal incidence being unknown, it is estimated that 10-30% of the patients with AML carry mutations in *RUNX1* (Holme *et al.*, 2012; Mendler *et al.*, 2012). The mechanism by which these mutations result in hematologic neoplasms is thought to involve several biological pathways. Defective RUNX1 protein results in a higher clonogenic capacity and alters the differentiation process. This, together with the alteration of DNA repair pathways, a decrease in p53 protein levels (with the consequent down-regulation of apoptosis) and the fact that mutated *RUNX1* cells have a genotoxic stress-resistant phenotype, contribute to poor prognosis of *RUNX1* mutations (Bellissimo and Speck, 2017).

Germline mutations can occur in different positions of the gene and they can be point mutations, such as missense or nonsense, or small chromosomal alterations like insertions or deletions causing frameshift mutations (Figure 2). Dominant-negative mutations are more damaging that haploinsufficient mutations, as they are related to a higher risk of malignant transformation (Latger-Cannard *et al*., 2016). The acquisition of loss of function mutations in the healthy allele is common in these patients (Preudhomme *et al.*, 2009).

1.1.1.2.2 Myeloid neoplasms with germline *ANKRD26* mutation

Ankyrin repeat domain 26 (*ANKRD26*) gene is located in chromosome region 10p12.1 and contains 34 coding exons. It codifies for a protein localized in the inner part of the membrane, and with ankyrin repeats in the N-terminal region and spectrin-like coiled-coil domains, both important in protein-protein interactions with signalling molecules. It has a key role in megakaryopoiesis, as this gene is highly expressed in progenitor hematopoietic cells and its expression decreases with megakaryocytes maturation. RUNX1 and FLI1 transcription factors bind to regulatory regions of *ANKRD26* and downregulate this gene at the late stage of megakaryocyte differentiation. Hence, its expression is almost absent in platelets (Bluteau *et al*., 2014; Pippucci *et al*., 2011).

Germline mutations in *ANKRD26* gene are associated with thrombocytopenia-2, with an AD pattern. These patients are characterized by moderate thrombocytopenia and platelet dysfunction, a normal platelet size but typically with α-granule deficiency and significant dysmegakaryopoiesis in the bone marrow. Moreover, these individuals have an increased risk of developing MDS/AML and, on rare occasions, chronic lymphoid leukemia (CLL) or chronic myeloid leukemia (CML) (Noris *et al*., 2011).

These germinal mutations are normally mono-allelic point mutations that occur in the 5' untranslated region (UTR) of the gene, affecting its regulation (Figure 2) *(Pippucci et al.*, 2011). The incidence of germline *ANKRD26* mutations is ~11% in patients with an hereditary thrombocytopenia (Noris *et al*., 2011). In these patients, *ANKRD26* gene expression is preserved in megakaryocytes and platelets because 5'UTR mutations prevent the binding of RUNX1 and FLI1. The study of ERK pathway during megakaryocyte differentiation has revealed that its activation diminishes during the maturation process, so that, a reduction of MAPK signalling is required for proplatelet formation. Continuous *ANKRD26* expression has demonstrated to induce MAPK/ERK1/2 pathway. This permanent signalling in megakaryocytes leads to a defect in proplatelet formation that could explain the thrombocytopenia of these patients (Bluteau *et al*., 2014).

1.1.1.2.3 Myeloid neoplasms with germline *ETV6* mutation

ETS (E26 transformation-specific) variant 6 (*ETV6*), localized in chromosomal position 12p12.3 and with eight coding exons, encodes a transcriptional repression factor located in the nucleus. It consists of three functional domains: a N-terminal pointed domain which is involved in protein-protein interactions, a central regulatory domain and a C-terminal DNA-binding ETS domain (Zhang *et al.*, 2015). Homodimerization is required for its activity and it is achieved by the N-terminal pointed domain (Green *et al.*, 2010). Among other functions, *ETV6* has an important role in hematopoiesis, specifically in thrombopoiesis, as it regulates the activity of other transcription factors, such as FLI1, present in platelets and megakaryocytes. Therefore, mutated *ETV6* alters both megakaryocyte maturation and platelet formation (Kwiatkowski *et al*., 1998).

There is an association between AD thrombocytopenia-5 and germline *ETV6* mutations, as many of these patients have mutations that disrupt the activity of this protein. The incidence of germinal mutations is unknown. These individuals suffer from thrombocytopenia and severe bleeding, platelet size is normal, bone marrow shows

dysmegakaryopoiesis and they are prone to develop hematologic malignancies, both lymphoid and myeloid (Noetzli *et al.*, 2015).

Germline mutations in *ETV6* are normally missense mutations that affect the conserved DNA-binding region or the pointed domain, and they frequently act as dominant negative (Figure 2). Thus, they affect binding to DNA and might alter its dimerization. These mutations have an effect in its repression activity and alter intracellular localization, as mutant proteins are located in the cytoplasm instead of in the nucleus (Noetzli *et al.*, 2015).

1.1.1.3 Myeloid neoplasms with germline predisposition and other organs dysfunction

1.1.1.3.1 Myeloid neoplasms with germline *GATA2* mutation

GATA binding protein 2 (*GATA2*) is located in the chromosomal region 3q21.3 and has five coding exons. This gene codifies for a transcription factor that is a member of the GATA family, and it contains two zinc-finger domains involved in protein-protein interaction and DNA binding. This transcription factor has a key role in hematopoiesis regulation and it participates in hematopoietic stem cells' survival and serf-renewal (Crispino and Horwitz, 2017).

After familial MDS and AML were first described in 1999, *GATA2* deficiencies were identified as the third main entity of HMMSs in 2012 (Hahn *et al*., 2012). Its incidence in young patients diagnosed with MDS is 7%. The mechanism by which the leukemic process is initiated is unknown, but these mutations are associated with a high penetrance and an early onset (Wlodarski *et al*., 2017). *GATA2* deficiencies are also related to other syndromes such as AD and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome; dendritic cell, monocyte, B and NK lymphoid deficiency and primary lymphedema associated with predisposition to MDS/AML (Emberger syndrome) (Collin *et al.*, 2015).

Germinal mutations are typically mono-allelic and gathered in the conserved zinc-finger domains, often causing a loss-of-function effect (Figure 2) (Hahn *et al.,* 2012). As mutated *GATA2* alters hematopoiesis differentiation, carriers may develop cytopenias which may result in leukemogenic processes (Hirabayashi *et al.*, 2017). Additionally, MDS/AML patients with germline *GATA2* mutations commonly present abnormal karyotypes with monosomy 7 and trisomy 8, as well as somatic *ASXL1* mutations (Bödör *et al.*, 2012).

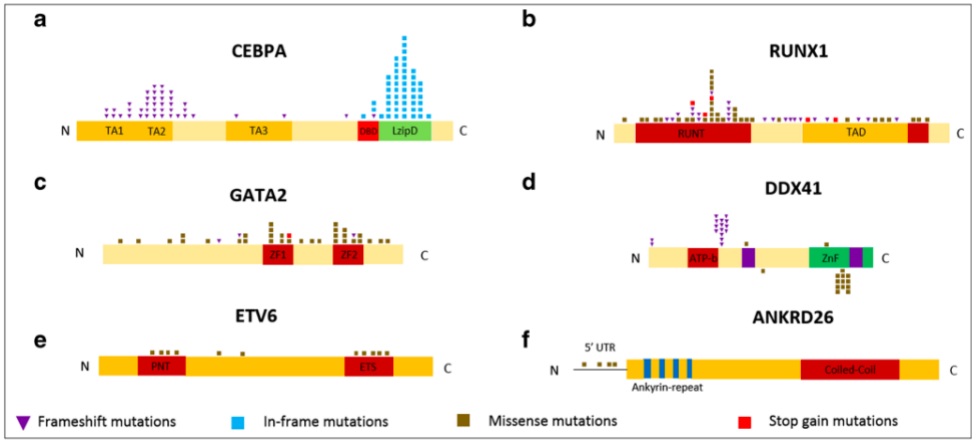

FIGURE 2 - **Most frequently mutated regions of familial AML/MDS predisposition genes.** Protein structure and domains of each gene are illustrated. Reprinted from: Király *et al.,* 2018.
UTR, untranslated region

1.1.2 Diagnosis and management of patients with hereditary myeloid malignancies

Diagnosis of HMMSs is complicated as many of the genes involved also take part in acquired myeloid neoplasms. To properly diagnose hematologic hereditary syndromes, there are several aspects to consider. Firstly, a detailed individual and familial medical history can be of great utility. Information about hematologic neoplasms and other types of cancer is a suspicious fact. Also, other non-cancer symptoms related to inherited myeloid syndromes must be taken into account, such as severe cytopenias, bleeding episodes, platelet dysfunction and thrombocytopenia (Brown *et al.*, 2017; The University of Chicago Hematopoietic Malignancies Cancer Risk Team, 2016). But HMMSs diagnosis should not only be based on relative precedents, as familial clinical history is not always available and this would omit a significant subgroup of patients. Besides, in some cases, HMMSs are diagnosed when analysing the hematopoietic stem cell donor, if this donor suffers from cytopenias or fails in hematopoietic precursors' mobilization (Churpek *et al.*, 2012).

Currently available commercial and custom NGS panels for AML/MDS diagnostic purposes usually include genes with recurrent somatic mutations and ignore known myeloid neoplasm predisposition genes. Hence, there is an urgent need to design new panels containing genes involved in HMMSs in order to identify these familial syndromes (Drazer *et al*., 2018). For this aim, apart from the sample of neoplastic tissue, a paired germinal sample should be tested in order to discern germline mutations (Brown *et al.*, 2017; Feurstein *et al.*, 2016).

In short, diagnosis of HMMSs should be based on personal and family clinical history, morphological and cytogenetic/FISH study of peripheral blood and bone marrow, and molecular analysis of a targeted NGS gene panel including predisposition genes that would allow for the detection of germinal mutations (Table 2, Figure 3) (Godley and Shimamura, 2017).

With reference to the management of patients with HMMSs, is of prime importance the optimal donor selection in case of bone marrow transplantation, as close relatives may

be also carriers. Family members must be thoroughly evaluated to discard any germinal mutation, although asymptomatic, to minimize the risk of choosing an affected donor. Hematopoietic stem cell transplantation (HSCT) from an unrelated donor is preferred and, if blood abnormalities are detected, these individuals must also undergo a genetic test to discard a germline mutation (Feurstein *et al*., 2016).

Early identification of familial myeloid syndromes is crucial for treatment choice and for patient supervision. These individuals, and their relatives, should be included in surveillance programmes and informed about the risk of developing myeloid neoplasms and the need of being subjected to continuous monitoring. This monitoring may include physical examination and blood cell counts to detect cytopenias or peripheral blasts in circulation every 3 to 6 months. And, in case the blood count is altered, a morphologic, cytogenetic/FISH and molecular analysis of the bone marrow is required (Figure 4) (Godley and Shimamura, 2017).

It is relevant to comment that there are several aspects that complicate the diagnosis and management of these patients, such as incomplete penetrance, variable phenotype and anticipation, and that the lifetime risk of developing a myeloid malignancy depends on the kind of syndrome. In addition, proper and meticulous anamnesis collection is crucial and must always include familial data to ease the identification of HMMSs.

TABLE 2 – **Guide for molecular genetic diagnosis of hereditary AML/MDS** (Dohner *et al*., 2017).

Molecular diagnosis of hereditary myeloid malignancies
Myelodysplastic predisposition / acute leukemia predisposition syndromes
CEBPA, DDX41, RUNX1, ANKRD26, ETV6, GATA2, SRP72, 14q32.2 genomic duplication *(ATG2B/GSKIP)*
Cancer predisposition syndromes
Li-Fraumeni syndrome *(TP53)*
Germline *BRCA1/BRCA2* mutations

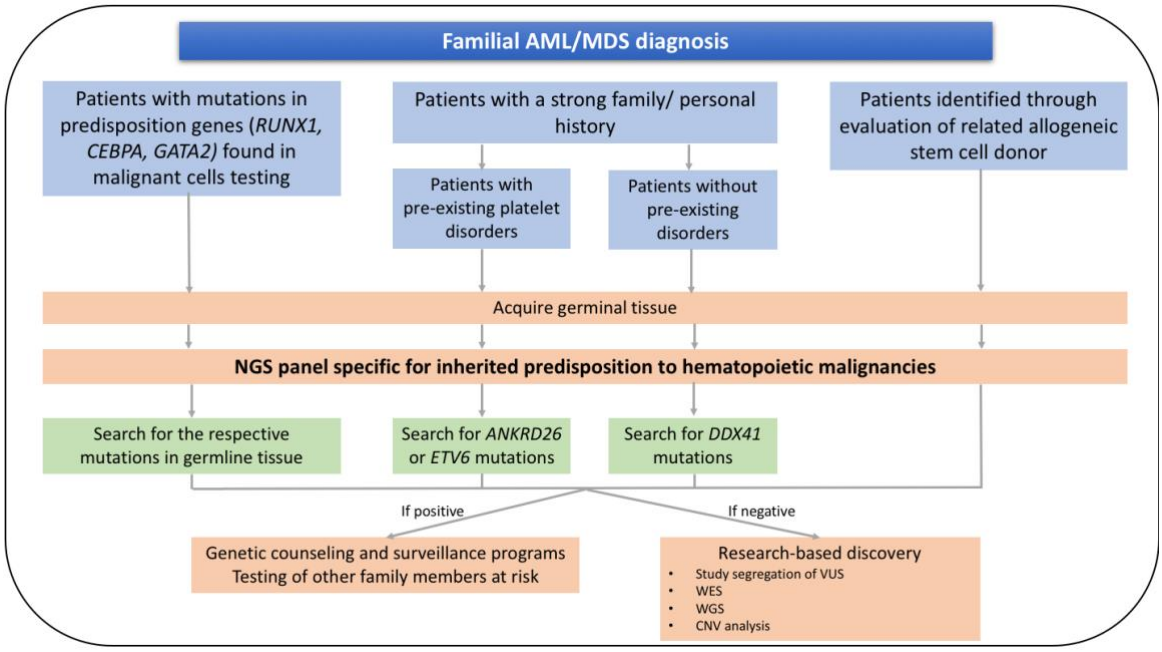

FIGURE 3 - **Diagnostic algorithm for familial cases of AML/MDS.** (Baptista *et al.*, 2017; Feurstein *et al.*, 2016; The University of Chicago Hematopoietic Malignancies Cancer Risk Team, 2016).
AML, acute myeloid leukemia. CNV, copy number variation. MDS, myelodysplastic syndrome. VUS, variant of uncertain significance. WES, whole-exome sequencing. WGS, whole-genome sequencing.

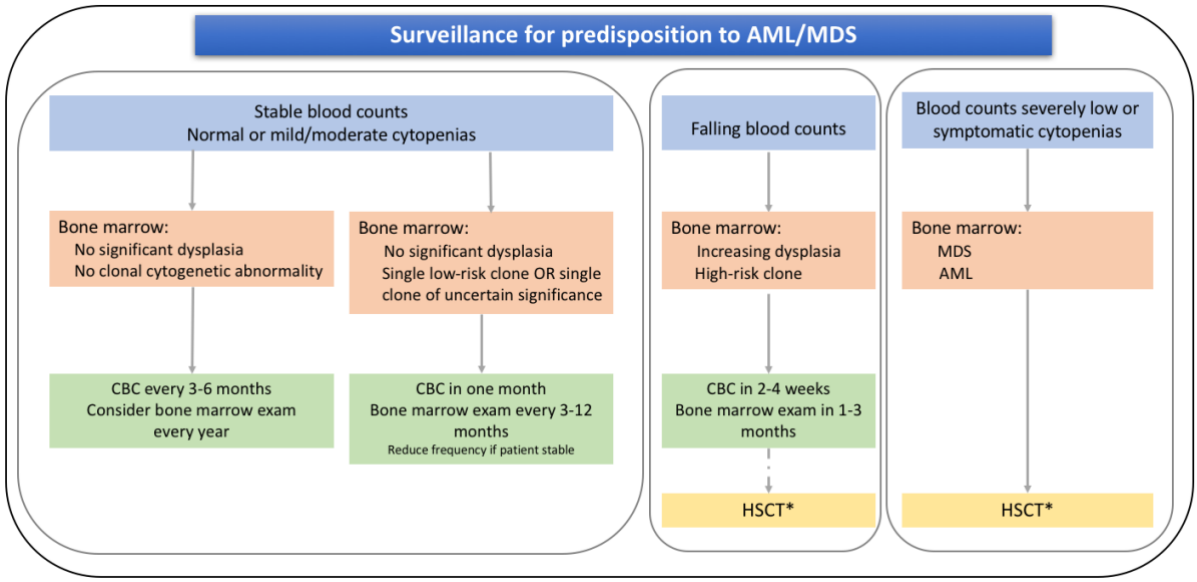

FIGURE 4 - **Surveillance for predisposition to AML/MDS.** General approach to manage patients with risk of developing a myeloid malignancy. Adapted from Godley and Shimamura, 2017.
AML, acute myeloid leukemia. CBC, cell blood count. HSCT, hematopoietic stem cell transplantation. MDS, myelodysplastic syndrome. *Chemotherapy may be considered for treatment.

1.2 USE OF MULTI-GENE PANELS TO DETECT GENETIC MUTATIONS UNDERLYING THE DEVELOPMENT OF CANCER

NGS is getting increasingly introduced into clinical practice. The notorious reduction of the cost of sequencing has allowed for the design of multi-gene panels, enabling simultaneous testing of multiple genes. These panels include several target genes of interest for specific neoplastic diseases based on previous evidence. Targeted sequencing is an efficient and sensitive tool to detect genetic alterations, both somatic and germline, and to provide the mutational spectrum of the patients. It can be useful to guide treatment selection, to provide information about the prognosis and tumor evolution, to avoid treatment resistance, to promote the development of new therapeutic drugs and to fully understand the molecular mechanisms underlying the progression of the tumorigenic process (Jensen *et al.*, 2018; Tsongalis *et a*l., 2014).

Additionally, genomic data can inform about the risk of developing cancer. Hereditary cancer predisposition was previously investigated only in a few well-known cases, such as *BRCA1* and *BRCA2* testing in patients with breast or ovarian cancer susceptibility. With the rise of multi-gene panels, individuals susceptible to develop cancer can be now detected. However, more efforts are needed to design strategies to effectively evaluate the risk, as gene mutation's implications may differ with the age, gender, genetic background or other characteristics of the patients, hindering an accurate risk stratification (Braun *et al*., 2018). On the other hand, tumor-only sequencing is effective when it comes to identify genetic variants, but it cannot distinguish between germinal and acquired mutations, so that germline tissue evaluation is needed for cancer predisposition diagnosis (Drazer *et al.*, 2018).

After multi-gene panel sequencing and the posterior bioinformatic analysis, a list of genetic variants detected in patients is obtained. These variants include small insertions and deletions (InDels) and single nucleotide variants (SNVs). Detected genetic alterations may include polymorphisms, synonymous, missense, nonsense, frameshift or splicing variants. In order to ascertain the clinical significance of these genetic changes, there are several public databases with information about human genetic variation, *in silico* predictors which estimate protein damage given an amino acidic change or assess potential splicing alterations, and laboratory-based functional assays. Taking into account all the information, variants are classified as pathogenic, likely pathogenic, variants of uncertain significance (VUS), likely benign or benign. But one may also find variants with conflicting interpretations, that is, variants which have been classified differently by distinct clinical institutions. Interpretation of the clinical repercussion of genetic variants remains a challenging procedure. A collaborative attitude between research and medical institutions, including data sharing, may allow the standardization of variant classification, therefore, reducing erroneous and uncertain interpretations (Balmaña *et al.*, 2016).

2. HYPOTHESIS
3. OBJECTIVES

2. HYPOTHESIS

HMMSs consist of an heterogeneous group of hematologic disorders with an inherited etiology. In particular, familial AML/MDS had been considered as rare neoplasms in the past, specially in adults. Recent technological developments have revealed that these hereditary cases are more frequent than previously expected, and have enabled the identification of a series of genes related to HMMSs, providing a major understanding of the altered molecular mechanisms in these patients. Due to the fact that some predisposition genes associated with hereditary myeloid malignancies are also frequently mutated in sporadic AML/MDS, a new methodological approach is needed to identify these familial cases. Moreover, individuals with HMMSs are often considered for hematopoietic stem cell transplantation. Hence, the correct diagnosis of these hereditary syndromes by means of germline mutation identification is crucial in order to ensure an appropriately selection of healthy donors and to offer these individuals genetic monitoring, cancer risk evaluation and family genetic counselling. Therefore, the purpose of this study is to perform an exhaustive genomic characterization of a subgroup of patients by means of a custom NGS panel with HMMS-related genes in order to integrate obtained data in current diagnostic and prognostic procedures and to improve hematopoietic stem cell donors' selection. In this way, the results of this study will be relevant in developing a NGS-based molecular diagnosis protocol to improve identification of these patients with the resulting improved management of these individuals and relatives.

3. OBJECTIVES

The present project aims to evaluate the frequency of germline mutations in a retrospective cohort of young patients (under the age of 60) with a diagnosis of sporadic MDS or AML. Recognition of these hereditary syndromes is of main importance to appropriately guide treatment selection and for the proper management and surveillance of these patients. In order to accomplish this overall aim, the following specific objectives were established: (1) to analyse the cohort of patients by a multi-gene NGS panel including genes associated with HMMSs in order to detect potentially germline genetic alterations, (2) to classify detected variants according to their pathogenicity, and (3) to validate detected mutations by Sanger.

4. MATERIALS AND METHODS

4. MATERIALS AND METHODS

4.1 DESCRIPTION OF COHORTS OF PATIENTS

The initial retrospective study population consisted of a total of 350 patients with *de novo* myeloid neoplasms, including 250 patients with AML and 100 with MDS, diagnosed at the Hospital La Fe (Valencia, Spain) between the years 2010 and 2018. For all cases, there was an exhaustive clinical and biological characterization, including cytomorphology, immunophenotyping, cytogenetic analysis, FISH and a molecular screening (in 17 cases by a NGS panel containing more than 30 genes with recurrent somatic mutations in hematologic malignancies). Among these 350 patients, the selected cohort consisted of 34 patients (23 AML, 11 MDS). Selection criteria were being under the age of 60 at diagnosis time and availability of paired germinal sample (Figure 5).

For all patients, an informed consent for undergoing molecular analysis of genetic alterations was obtained in accordance with the Declaration of Helsinki, the European Convention on Human Rights and Biomedicine, the Universal Declaration of the UNESCO on the Human Genome and Human Rights and the Spanish legislation in terms of biomedical research, personal data protection and bioethics.

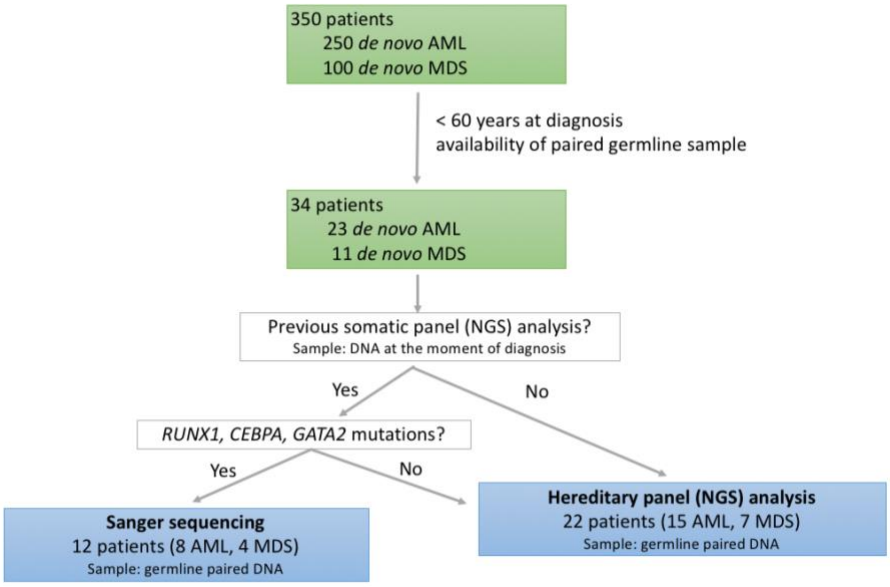

FIGURE 5 - **Description of the inclusion criteria to select the cohort of patients and posterior analysis.** Patients under the age of 60 and with a paired germline sample available were selected. Among the 34 selected patients, 17 of them had a previous somatic NGS panel analysis, which included three predisposition genes *(RUNX1, CEBPA* and *GATA2).* If they harboured mutations in one of these genes, these mutations were validated by Sanger in the paired germline sample. If they did not have mutations in one of the predisposition genes or did not have a previous NGS analysis, they were analysed by the hereditary NGS panel.
AML, acute myeloid leukemia. MDS, myelodysplastic syndrome. NGS, next-generation sequencing.

4.2 TISSUE SPECIMENS

For each patient, we obtained DNA samples from bone marrow aspiration at diagnosis time, and a paired germline DNA sample from bone marrow at complete molecular remission. All samples were supplied by La Fe Biobank. These samples were stored with the purpose of biomedical research in compliance with the current legislation (Law 14/2007, 3rd of July) and the addition of these samples to the Biobank collection was authorized by the Ethical Committee in Clinical Research of the University Hospital La Fe (Registration number 2014/0532). Automated DNA extraction from bone marrow samples was performed using robot QIAsymphony SP (QIAGEN), whose technology is based on silica-based DNA purification and the use of magnetic particles. This procedure includes four different steps: lysis, binding, washing and elution (Figure 6).

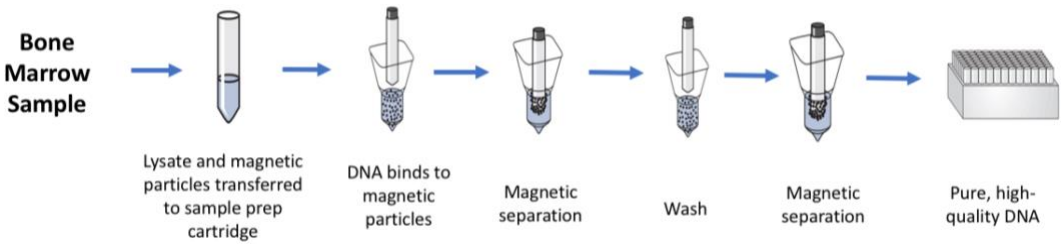

FIGURE 6 – **Automated DNA extraction**. QIAsymphony DNA Procedures. Image adapted from QIAsymphony® DNA Handbook. September 2010 (QIAGEN).

4.3 TARGETED GENE PANEL DESIGN (HEREDITARY PANEL)

The multi-gene panel was designed using the system SureDesign Custom Design Tool (Agilent Technologies Inc.). This panel included, among others, a series of genes related to hereditary malignant hematologic disorders: *CEBPA, DDX41, RUNX1, ANKRD26, ETV6, GATA2*, *SRP72, ATG2B, GSKIP, BRCA1, BRCA2* and *TP53.* This set of target genes was selected based on literature reviews and public databases. The multi-gene panel comprised the entire codifying region of each gene, the flanking 5'UTR and 3'UTR as well as promoter regions.

4.4 ENRICHED LIBRARY PREPARATION

Library preparation and target enrichment was conducted though "SureSelect^{QXT} Automated Target Enrichment for Illumina Multiplexed Sequencing. Featuring Transposase-Based Library Prep Technology. Automated using Agilent NGS Bravo. Version B1. December 2016" according to the manufacturer's instructions. The whole procedure is summarized in Figure 7.

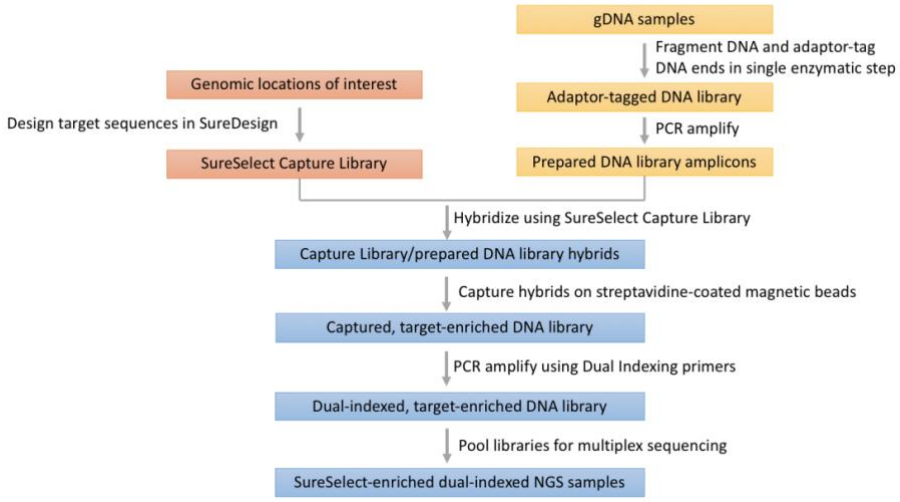

FIGURE 7 - **Sureselect^{QXT} NGS Target Enrichment Workflow.** Image adapted from "Sureselect^{QXT} Automated Target Enrichment for Illumina Multiplexed Sequencing. Featuring Transposase-Based Library Prep Technology."
gDNA, genomic DNA.

4.4.1 Genomic DNA library preparation

Genomic DNA (gDNA) was quantified and diluted to a final concentration of 10ng/µl to ensure optimal fragmentation. Quantification was performed by means of two serial fluorometric assays: Qubit dsDNA BR Assay and Qubit dsDNA HS Assay (Thermofisher Scientific). gDNA was then enzymatically fragmented with the transposase at the same time that adaptors were added to the ends of the fragments. Adaptor-tagged DNA samples were purified by using AMPure XP beads. These adaptor-ligated DNA samples were then amplified by PCR using the Herculase II Fusion DNA Polymerase, and further purified using AMPure XP beads. Library DNA quantity and quality was assessed by using the Agilent 4200 TapeStation and a D1000 ScreenTape (Agilent Technologies Inc.) to accurately determine DNA fragment size (245-325 bp).

4.4.2 Hybridization

The second step of the library preparation was to hybridize the gDNA library with the Capture Library in order to enrich targeted regions of the genome. To this effect, firstly, adaptor-ligated DNA libraries were normalized to 750ng. Then, the adaptor-tagged DNA library was hybridized to the capture library (probes labelled with biotin). After that, the hybridized library was captured using streptavidin-coated magnetic beads and the captured DNA-RNA hybrids were washed several times.

4.4.3 Indexing

Captured DNA libraries were amplified by PCR using the Herculase II Fusion DNA Polymerase to add dual indexing tags and purified using Agencourt AMPure XP beads. DNA quality and quantity was assessed by using the TapeStation (Agilent Technologies Inc.), being the average fragment length 331 bp.

4.5 PAIR-END MULTIPLEXED SEQUENCING

Samples were pooled and normalized to 4 nM for multiplexed sequencing. After checking its concentration using Qubit dsDNA HS Assay (Thermofisher Scientific), the pool was diluted to 10 pM and denatured by following the protocol "MiSeq System. Denature and Dilute Libraries Guide. April 2018." (Illumina Inc.). Additionally, SureSelect^{QXT} sequencing custom primers were combined with Illumina primers by carefully following the SureSelect^{QXT} manufacturer's instructions. 5% of PhiX Control v3 was used as a control library. Libraries were run on a MiSeq sequencer (Illumina Inc.), using the MiSeq Reagent Kit v3 and a read length of 2x150 bp. Our hereditary NGS panel provided median sequencing depth of 110X per sample and 99% of the target regions were covered. Adapter trimming was performed by MiSeq Illumina Reporter software. Afterwards, data files were de-multiplexed and converted into FASTQ data.

4.6 BIOINFORMATIC ANALYSIS

Sequencing read quality was evaluated with *fastQCv.0.11.2* (Andrew, 2010). Low quality reads were removed and adaptor remainders and low quality bases were trimmed by using *printseq lite v.0.20.4* (Schmieder and Edwards, 2011). After that, reads were mapped onto the reference human genome (GRCh37) using *bwa mem v.07.12* (Burrows-Wheeler Aligner) (Li, 2013) and visualized by *IGVv2.3* (Integrative Genomics Viewer) (Robinson *et al*., 2011). Variant calling was performed through *GATK* (Genome Analysis Toolkit), filtering by a minimum mapping quality score>5 (Mckenna *et al*., 2010). Functional annotation was performed using *Cartagenia* software (Agilent Technologies Inc.). This software uses as annotation sources: 1000Genomes, 1000GenomesPhase3, CIViC, COSMIC, ClinVar, ESP6500, ExAC, HGMDProfessional, OMIM, dbNSFP and dbSNP. Then, variants were filtered by discarding those with a coverage less than 20 and with a minor allele frequency (MAF) higher than 2% (polymorphisms). So, for each variant, there was information about its chromosomal position, type of variant (SNV, InDel), reference nucleotide, altered nucleotide, length, total coverage, minor allele coverage, MAF, gene, transcript, variant location (exonic, intronic, UTR), function (frameshift, non-frameshift, missense, nonsense, synonymous, splicing), variant nomenclature and information from several databases and biological predictors.

4.7 CLASSIFICATION OF DETECTED GENETIC VARIANTS

Final step was variant classification into five different categories according to their pathogenicity: benign, likely benign, VUS, likely pathogenic or pathogenic. All variants, including missense, nonsense, synonymous, splicing, frameshift or in-frame alterations, were evaluated in depth. For this purpose, "Standards and guidelines for the interpretation of sequence variants" from the American College of Medical Genetics and Genomics (ACMG) were taken into consideration (Figure 8) (Richards *et al.*, 2015). One important aspect to consider was if they had been described in databases such as ClinVar, a public collection of interpretation of the clinical significance of genetic variants (Landrum *et al.*, 2016); COSMIC (Catalogue of Somatic Mutations In Cancer); Varsome, a large data library with information about human genomics and variant annotation; or IARC TP53 Mutation Database, with data relative to reported *TP53* mutations (Table 3). Other helpful tool is the use of *in silico* biological predictors (Table 4). In order to analyse

the effect of missense mutations, predictors such as *SIFT* (Kumar *et al*., 2009), *PolyPhen-2* (Adzhubei *et al.,* 2010) and *MutationTaster2* (Schwarz *et al.*, 2014) were used to ascertain the degree of damage in the protein due to the amino acidic change. In the case of synonymous or splicing variants, *Human Splicing Finder (HSF)* (Desmet *et al*., 2009) makes a prediction whether an intronic or exonic mutation creates a splicing alteration or not. Another tool to analyse splicing missregulation due to genetic alterations is *SPANR (Splicing-based Analysis of Variants)* (Xiong *et al.*, 2015). In order to obtain substantial evidence for benignity/pathogenicity, predictors should give a common verdict. As for nonsense and frameshift mutations, they are generally considered as likely pathogenic despite not having been previously described, unless they are located in a terminal exon. With all this information, variants were characterized and finally, only pathogenic and likely pathogenic variants were taken into consideration.

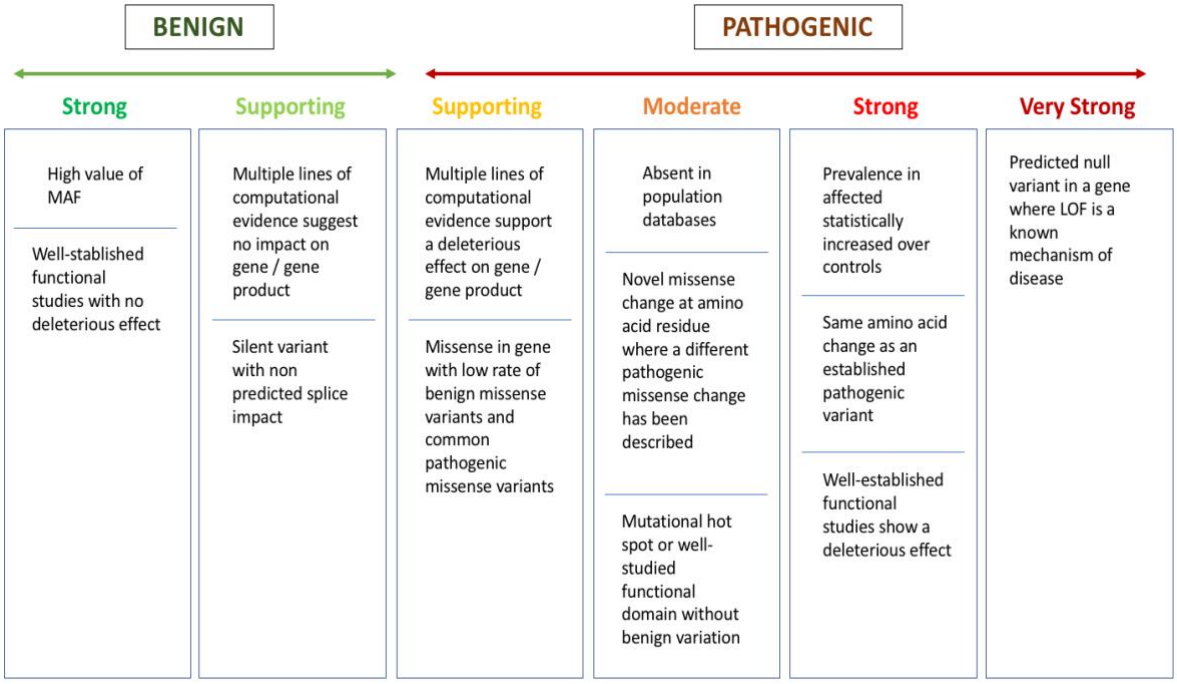

FIGURE 8 – **Criteria for interpreting the degree of pathogenicity of a genetic variant.** Adapted from Richards *et al*., 2015**.** LOF, loss of function.

TABLE 3 - **Databases consulted for variant classification.**

Biological Databases	
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
Varsome	https://varsome.com/
COSMIC	https://cancer.sanger.ac.uk/cosmic/
IARC TP53 Mutation Database	http://p53.iarc.fr/

TABLE 4 – ***In silico*** **predictors used for variant interpretation.**

In silico Predictors	
Protein effect prediction	
SIFT	http://sift.jcvi.org
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2
MutationTaster2	http://www.mutationtaster.org
Splicing prediction	
Human Splicing Finder	http://www.umd.be/HSF3/HSF.shtml
SPANR	http://tools.genes.toronto.edu/

4.8 VALIDATION OF DETECTED MUTATIONS BY SANGER SEQUENCING

Mutations detected by the somatic diagnosis NGS panel were validated by Sanger sequencing in order to assess if these mutations were germline or acquired by analysing a bone marrow DNA sample in complete remission. Additionally, pathogenic/likely pathogenic mutations and VUS variants located in conflictive genomic regions detected by the hereditary NGS panel were also validated. Primer sequences were designed by *Primer3Plus Version: 2.4.2*. The list of primers used is shown in Table 5.

The PCR reactions were performed using the kit AmpliTaq Gold™ DNA Polymerase with Buffer II and $MgCl_2$ (Applied Biosystems by Life Technologies, Carlsbad, CA). PCR amplification was as follows: one denaturation step at 95 ºC for 10 min, followed by 30 cycles with a denaturation step at 95 ºC for 15 s, an annealing step at the corresponding temperature and time (Table 5), and an extension step at 72 ºC for 30 s, ending with an additional extension step at 72 ºC for 5 min. After amplification, sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and samples were run in a 3500 Genetic Analyzer (Applied Biosystems). Data analysis was performed using software *MEGA7* (Kumar *et al.*, 2016) and *Chromas version 2.6* (Technelysium Pty Ltd, South Brisbane, Australia).

TABLE 5 – **List of primers used for validation of detected variants.** Annealing temperatures and times are displayed in the last column.

Primers				
Gene	**Exon**	**Forward (5'-3')**	**Reverse (5'-3')**	**Annealing**
RUNX1	Exon 4	GCTGTTTGCAGGGTCCTAAC	GGCCTCCGCCTGTCCTC	62 ºC – 45 s
	Exon 5	CATTGCTATTCCTCTGCAACC	GTTTGTTGCCATGAAACGTG	60 ºC – 1,5 min
	Exon 6	AAATTCCGGGAGTGTTGTCA	GAAAGGTTGAACCCAAGGAA	60 ºC – 1,5 min
	Exon 7	TGATCTCTTCCCTCCCTCCT	CAGTTGGTCTGGGAAGGTGT	58 ºC – 45 s
GATA2	Exon 5	ACTCCCTCCCGAGAACTTG	TAATTAACCGCCAGCTCCTG	56ºC – 30 s
ATG2B	Exon 42	TCCTACCTCCCTCCTCTACG	CTTCACACTGTCAGTTCCAAGC	56 ºC – 30 s
TP53	Exon 5-6	TGTTCACTTGTGCCCTGACT	TTAACCCCTCCTCCCAGAGA	58 ºC – 1 min
BRCA2	Exon 20	GGATTACAGATGTGAGCCA	GTCTCTAAGACTTTGTTCTCATA	56 ºC – 45 s

5. RESULTS

5. RESULTS

5.1 DESCRIPTION OF THE COHORT OF PATIENTS

Main biological and clinical characteristics of the 34 patients of the study cohort are summarized in Table 6. Among the cohort under study, two individuals were of special interest because they had a family history of thrombocytopenia and/or myeloid neoplasms (patients 21 and 22). And from this group of 34 patients, 12 of them, apart from the inclusive criteria, had mutations in *RUNX1* previously detected by the somatic diagnosis NGS pane (patients 23-34). So, as they were suspicious of a HMMS, they were selected to validate these mutations by Sanger.

TABLE 6 - **Clinical and biological characterization of the study cohort**.

Patient	Diagnosis	Age (at diagnosis)	Karyotype	Molecular Biology
1	AML	47	47,XX,+8[16]/46,XX[4]	*FLT3-ITD, NPM1, DNMT3A, WT1* (hiperexpression)
2	AML	43	45,XX,t(15;17)(q22;q21), -16[15]/46,XX[5]	*FLT3-ITD, PML-RARa, WT1* (hiperexpression)
3	AML	44	47,XY,+4[12]/ 47,idem,del(9)(p22)[7]/ 46,XY[1]	*FLT3-ITD, WT1* (low-moderate expression)
4	AML	38	46,XX,t(6;17;12;9)(p22;q12; q24.1;q34)[18]/46,XX[2]	*FLT3-ITD, DEK-CAN*. Oncomine Myeloid Research panel: rear. *DEK-CAN, FLT3* (c.1769_1770ins51)
5	AML	39	Normal karyotype	*FLT3-ITD, NPM1*
6	AML	44	46,XY,inv(16)(p13q22)[20]	*CBFb-MYH11, cKIT* (exon 17 D816V), *WT1* (low-moderate expression)
7	AML	54	Normal karyotype	Negative
8	AML	40	46,XY,inv(3)(q21q26)[20]	*FLT3-ITD, EVI1* (hiperexpression)
9	AML	42	Not available	*FLT3-ITD, NPM1*
10	AML	41	Normal karyotype	*FLT3-ITD, NPM1, WT1* (hiperexpression)
11	AML	19	Normal karyotype	*CFBFb-MYH11, WT1* (low-moderate expression)
12	AML	34	47,XY,+6[11]/46,XY[3]	*FLT3-ITD*
13	AML	46	47,XX,+8[11]	*FLT3-ITD*. AML Community panel: *FLT3* (c.1776_1777ins90)
14	MDS*	48	46,XX,del(5)(q13q31)[18]/ 46,XX[2]	Not available
15	MDS	30	46,XY,-7,+mar[20]	Not available
16	MDS*	50	Normal karyotype	*cKIT* (exon 17 D816V), *EVI1* (low-moderate expression). Oncomine Myeloid Research panel: *SF3B1* (c.2342A>G), *DNMT3A* (c.2711C>T)
17	MDS	35	Normal karyotype	Oncomine Myeloid Research panel: *DNMT3A* (c.2644C>T)

AML, Acute Myeloid Leukemia. MDS, myelodysplastic syndrome. *Later progression to AML. rear., rearrangement.

TABLE 6 - Clinical and biological characterization of the study cohort (continued).

Patient	Diagnosis	Age (at diagnosis)	Karyotype	Molecular Biology
18	AML	34	43-44,XY,-7,-10, add(13)(p10), add(16)(q22), del(17)(p11.2),-18,-19, -20,i(21)(q10)[8]/46,XY[9]	*EVI1* (low-moderate expression)
19	MDS	37	46,XX,del(5)(q13q31)[13]/ 46,XX[4]	Not available
20	MDS	34	Normal karyotype	AML Community panel: no relevant variants found
21	AML	39	45,XX,t(10;11)(p12;q23), -7[20]	*MLL-AF10* rear.
22	MDS	17	45,XY,-7[7]/46,XY[13]	*EVI1* (low-moderate expression)
23	MDS*	57	47,XY,+8[4]/46,XY[14]	AML Community panel: *RUNX1* (c.319C>T), *ASXL1* (C.1773C>A)
24	AML	32	46,XY,t(5;12)(q?13;p13)[5]/ 45,idem,-5[3]/46,XY[13]	*WT1* (low-moderate expression). AML Community panel: *RUNX1* (c.667G>T)
25	AML	52	Normal karyotype	*FLT3*. AML Community panel: *RUNX1* (c.497G>A), *FLT3* (c.2503G>C), *NRAS* (C.35G>A)
26	AML	36	47,XX,+8[18]/46,XX[2]	*IDH2* (R172K). AML Community panel: *IDH2* (c.515G>A), *RUNX1* (c.287dupA)
27	AML	42	Normal karyotype	*IDH2* (R140Q). AML Community panel: *IDH2* (c419G>A), *ASXL1* (c.1851_1852insA), *NRAS* (c.183A>C), *PTPN11* (c.182A>T), *RUNX1* (c.737C>T), *TET2* (c.470A>G)
28	MDS*	56	46,XY,del(6)(q15q23)[8]/ 46,XY[12]	*IDH2* (R140Q). AML Community panel: *IDH2* (c.419G>A), *JAK2* (c.1849G>T), *RUNX1* (c.374delC & c.610C>T)
29	MDS	32	45,XX,-7[5],idem+mar[5]/ 46,XX[10]	*RUNX1* (R166L). AML Community panel: *RUNX1* (c.497G>T)
30	AML	21	Normal karyotype	AML Community panel: *RUNX1* (c.496C>T)
31	MDS*	9	45,XX,-7[14]/46,XX[1]	AML Community panel: *DNMT3A* (c.89A>C), *RUNX1* (c.572_573insTG)
32	AML	15	48,XY,+3,+6,der(7)t(7;8) (q32;q22)[6]/46,XY[14]	AML Community panel: *RUNX1* (c.436A>G)
33	AML	30	Normal karyotype	AML Community panel: *RUNX1* (c.667G>T)
34	AML	35	47,XY,del(5)(q32q35),+8[19] /46,XY[1]	*FLT3-ITD*. AML Community panel: *FLT3* (c.1782_1783ins87), *RUNX1* (c.179C>T)

AML, Acute Myeloid Leukemia. MDS, myelodysplastic syndrome. *Later progression to AML. rear., rearrangement.

5.2 DETECTED VARIANTS

5.2.1 Exonic variants

The frequency of individuals with potentially germline exonic variants obtained in the selected cohort of patients with AML/MDS was of 44.1% (15 of 34 individuals studied) (Figure 9A). Among these 15 patients with exonic variants (12 AML, 3 MDS), 6 of them carried more than one variant. On one side, 18 distinct exonic variants were detected in 12 of the 22 patients evaluated by the hereditary NGS panel. These genetic alterations were considered as potentially germline because variant allele frequencies (VAF) were between 40-60% and bone marrow samples in remission were used for testing. In addition, from the variants previously identified in 12 patients analysed by the somatic NGS panel, 3 variants were detected in 3 individuals by Sanger sequencing in remission bone marrow samples, denoting a germline origin. So, in total, 21 distinct potentially germline exonic variants were identified in 15 patients.

Genetic variants were detected in *ANKRD26, ATG2B, BRCA1, BRCA2, ETV6, GATA2, RUNX1, SRP72* and *TP53* (Figure 9B). All variants are described in Table 7. The most recurrently mutated genes were *BRCA2* (n=5, 23.8%), *ATG2B* (n=5, 23.8%) and *RUNX1* (n=4, 19%). All variants were point mutations, specifically, missense (n=15) and synonymous (n=6). Genes with missense alterations were *BRCA2* (n=4), *RUNX1* (n=4), *ATG2B* (n=1), *SRP72* (n=1), *ANKRD26* (n=1), *BRCA1* (n=1), *ETV6* (n=1), *GATA2* (n=1) and *TP53* (n=1). Synonymous variants were located in *ATG2B* (n=4), *BRCA2* (n=1) and *SRP72* (n=1). Neither nonsense nor frameshift mutations were found.

5.2.2 Non-exonic variants

Among the individuals analysed by the hereditary gene panel, a significant number of genetic alterations in non-codifying regions with a VAF>40% was found. In particular, 26 distinct non-exonic germline variants were detected in 16 (12 AML, 4 MDS) of the 22 patients analysed (72.7%) (Figure 10A). The list of variants is detailed in Table 8. Of them, 18 were UTR variants and 8 were intronic. Non-exonic variants were localized in almost all genes associated with HMMSs: *ANKRD26* (n=2), *ATG2B* (n=9), *BRCA1* (n=1), *CEBPA* (n=1), *DDX41* (n=3), *ETV6* (n=4), *RUNX1* (n=5) and *TP53* (n=1) (Figure 10B). *ATG2B* gene had the largest number of non-exonic germline variants (34.6%), followed by *RUNX1* (19.2%), *ETV6* (15.4%) and *DDX41* (11.5%). The majority of individuals with non-exonic alterations (12 of 16) harboured more than one variant.

TABLE 7 - **Potentially germline exonic variants identified in the cohort under study**

Patient	DX	Age	Gene	RefSeq	Variant	VAF	Categorization	rs ID	MAF*
1	AML	47	*ETV6*	NM_001987	c.871A>G p.R291G	47%	VUS	rs970651209	4.078e-06
2	AML	43	*ATG2B*	NM_018036	c.4584C>T p.P1528=	44%	Likely benign	rs146028674	0.002
			ATG2B	NM_018036	c.1956T>C p.S652= [R]	47%	Likely benign	rs61736675	0.004 (Eur 0.014)
			BRCA1	NM_007300	c.3119G>A p.S1040N	42%	Likely benign	rs4986852	0.010 (Eur 0.029)
3	AML	44	*BRCA2*	NM_000059	c.2883G>A p.Q961=	44%	Likely benign	rs11571655	1.997e-04
5	AML	39	*SRP72*	NM_006947	c.58C>T p.R20W	50%	Likely benign	rs111673705	0.001
			SRP72	NM_006947	c.1803G>A p.G601=	48%	Likely benign	rs143643243	0.001
6	AML	44	*ATG2B*	NM_018036	c.1360G>A p.V454M	51%	VUS	rs117507139	0.002
7	AML	54	*ANKRD26*	NM_014915	c.3655G>A p.V1219I	43%	Likely benign	rs146819984	0.008 (Eur 0.013)
			ATG2B	NM_018036	c.1956T>C p.S652= [R]	47%	Likely benign	rs61736675	0.004 (Eur 0.014)
8	AML	40	*GATA2*	NM_001145661	c.920G>T p.R307L	49%	Likely pathogenic	NA	NA
			BRCA2	NM_000059	c.8629G>A p.E2877K	47%	VUS	NA	NA
15	MDS	30	*RUNX1*	NM_001754	c.499A>G p.S167G	52%	Likely pathogenic	NA	NA
			BRCA2	NM_000059	c.5744C>T p.T1915M	48%	Likely benign	rs4987117	0.009 (Eur 0.025)
16	MDS[+]	50	*ATG2B*	NM_018036	c.12G>A p.P4=	50%	Likely benign	rs72706804	0.008 (Eur 0.024)
17	MDS	35	*ATG2B*	NM_018036	c.6168C>T p.G2056=	53%	VUS	rs151219519	0.002
18	AML	34	*TP53*	NM_000546	c.404G>T p.C135F	57%	Pathogenic	NA	NA
21	AML	39	*BRCA2*	NM_000059	c.6100C>T p.R2034C	49%	Likely benign	rs1799954	0.001
			BRCA2	NM_000059	c.9371A>G p.N3124S	46%	Likely pathogenic	rs28897759	4.063e-06
25	AML	52	*RUNX1*	NM_001754	c.497G>A p.R166Q	44%	Pathogenic	rs1060499616	NA
27	AML	42	*RUNX1*	NM_001754	c.737C>T p.T246M	47%	VUS	rs555366994	1.997e-04
34	AML	35	*RUNX1*	NM_001754	c.179C>T p.A60V	54%	VUS	rs765314703	7.026e-05

*MAF values were obtained from 1000GenomesPhase3 data.
AML, acute myeloid leukemia. DX, diagnosis. Eur, European. MAF, minor allele frequency. MDS, myelodysplastic syndrome. NA, not available. VAF, variant allele frequency. VUS, variant of uncertain significance. [R], Variant found in two individuals. [+], patient with MDS who evolved to AML.

TABLE 8 - **Potentially germline non-exonic variants identified in the cohort under study.**

Patient	DX	Age	Gene	RefSeq	Variant	Type	VAF	Categorization	rs ID	MAF*
1	AML	47	*ATG2B*	NM_018036	c.3749+24T>C	intronic	55%	VUS	rs555213041	0.001
					c.3842+42G>A	intronic	52%	Likely benign	rs370805659	1.997e-04
			ANKRD26	NM_014915	c.531+41C>T [R]	intronic	63%	Likely benign	rs56325123	4.505e-06
2	AML	43	*RUNX1*	NM_001001890	c.-1544C>T [R]	5' UTR	50%	Likely benign	rs192533631	0.019
			ATG2B	NM_018036	c.*1803C>G	3' UTR	49%	Likely benign	rs148857991	0.002
					c.*5284G>C [R]	3' UTR	43%	Likely benign	rs147112707	0.004 (Eur 0.014)
4	AML	38	*ETV6*	NM_001987	c.*23C>T [R]	3' UTR	50%	Likely benign	rs72550782	0.008 (Eur 0.023)
5	AML	39	*DDX41*	NM_001321732	c.-257C>T	5' UTR	51%	VUS	rs374555260	1.225e-04
6	AML	44	*ATG2B*	NM_018036	c.*5335T>C	3' UTR	56%	VUS	NA	NA
					c.3038-25G>A	intronic	44%	Likely benign	rs148210333	0.007 (Eur 0.021)
7	AML	54	*ETV6*	NM_001987	c.*23C>T [R]	3' UTR	47%	Likely benign	rs72550782	0.008 (Eur 0.023)
					c.163+47C>T	intronic	41%	Likely benign	rs200112014	0.001
			ATG2B	NM_018036	c.*5284G>C [R]	3' UTR	48%	Likely benign	rs147112707	0.004 (Eur 0.014)
8	AML	40	*TP53*	NM_001126116	c.-107T>C [R]	5' UTR	64%	Likely benign	rs113530090	0.007 (Eur 0.017)
			CEBPA	NM_001285829	c.*361G>A	3' UTR	51%	Likely benign	rs41367646	0.005 (Eur 0.02)
			RUNX1	NM_001001890	c.*4238G>T [R]	3' UTR	58%	Likely benign	rs138870671	0.002
9	AML	42	*RUNX1*	NM_001001890	c.*2770C>G [R]	3' UTR	50%	Likely benign	rs75192893	0.016 (Eur 0.045)
			ATG2B	NM_018036	c.*5467C>T	3' UTR	46%	Likely benign	rs117053531	0.012 (Eur 0.022)
			DDX41	NM_001321732	c.193+13T>G	intronic	42%	Likely benign	rs182744895	0.001
10	AML	41	*DDX41*	NM_001321732	c.-486C>T	5' UTR	49%	Likely benign	rs139078584	0.006 (Eur 0.022)
			RUNX1	NM_001001890	c.*4238G>T [R]	3' UTR	51%	Likely benign	rs138870671	0.002
			ANKRD26	NM_014915	c.531+41C>T [R]	intronic	58%	Likely benign	rs56325123	4.505e-06
11	AML	19	*RUNX1*	NM_001001890	c.*2770C>G [R]	3' UTR	52%	Likely benign	rs75192893	0.016 (Eur 0.045)
			BRCA1	NM_007300	c.212+17T>C	intronic	59%	VUS	rs369461674	NA
12	AML	34	*ATG2B*	NM_018036	c.*2722A>T	3' UTR	50%	Likely benign	rs117614273	0.013 (Eur 0.036)
13	AML	46	*RUNX1*	NM_001001890	c.-1544C>T [R]	5' UTR	45%	Likely benign	rs192533631	0.019
			RUNX1	NM_001001890	c.*1138C>A [R]	3' UTR	100%	Likely benign	rs55744508	0.020 (Eur 0.064)
14	MDS+	48	*ATG2B*	NM_018036	c.*4581C>T	3' UTR	46%	VUS	rs571070929	6.869e-05
15	MDS	30	*TP53*	NM_001126116	c.-107T>C [R]	5' UTR	48%	Likely benign	rs113530090	0.007 (Eur 0.017
			ANKRD26	NM_014915	c.532-50C>T	intronic	42%	Likely benign	rs41280944	0.008 (Eur 0.013)
16	MDS+	50	*RUNX1*	NM_001001890	c.-435T>G	5' UTR	46%	VUS	NA	NA
			ETV6	NM_001987	c.*1812A>G	3' UTR	47%	VUS	rs1040537483	3.229e-05
					c.*514C>T [R]	3' UTR	48%	Likely benign	rs113420500	0.004
19	MDS	37	*RUNX1*	NM_001001890	c.*1138C>A [R]	3' UTR	48%	Likely benign	rs55744508	0.020 (Eur 0.064)
			ETV6	NM_001987	c.*514C>T [R]	3' UTR	45%	Likely benign	rs113420500	0.004

*MAF values were obtained from 1000GenomesPhase3 data.
AML, acute myeloid leukemia. DX, diagnosis. Eur, European. MAF, minor allele frequency. MDS, myelodysplastic syndrome. NA, not available. VAF, variant allele frequency. UTR, untranslated region. VUS, variant of uncertain significance. [R], Variant found in two individuals. [+], patient with MDS who evolved to AML.

5.3 CLASSIFICATION OF VARIANTS

After an exhaustive analysis of the potentially germline exonic variants detected, 10 were categorized as likely benign, 6 as VUS, 3 as likely pathogenic and 2 as pathogenic (Figure 9C). Thus, pathogenic and likely pathogenic alterations were 23.8% of the total variants (5/21) and were found in 14.7% patients of the cohort under study (5/34).

5.3.1 Pathogenic genetic alterations

Two pathogenic germline alterations, one in *TP53* (c.404G>T, p.C135F) and another one in *RUNX1* (c.497G>A, p.R166Q), were identified.

The *TP53* (NM_000546) mutation c.404G>T (p.C135F) was found in an individual who developed an AML at the age of 34. The variant was found with a VAF of 57% (50/88 reads). COSMIC's prediction was pathogenic (COSM10647) and ClinVar records showed that this variant had been described as either germline or somatic. Germline submission's categorization was 'likely pathogenic' and related to cancer predisposition syndrome. Besides, this variant was recorded in IARC TP53 Database as 'deleterious', always detected as somatic (64 counts). And it had 9 pathogenic *in silico* predictions *versus* no benign prediction.

The *RUNX1* (NM_001754) alteration c.497G>A (p.R166Q) was first identified in a patient by means of the somatic NGS panel. The variant was found with a VAF of 44%. COSMIC's prediction was pathogenic (COSM36055), described in hematopoietic tissue (15 samples) and associated with AML, MDS, chronic myelomonocytic leukemia (CMML) and MDS/MPN. ClinVar records classified this variant as 'likely pathogenic' and related to FPDMM. Also, this variant had 9 pathogenic predictions whereas no benign prediction, and UniProt also classified it as 'disease'.

5.3.2 Likely pathogenic genetic alterations

Three missense germline mutations were categorized as likely pathogenic: *BRCA2* c.9371A>G (p.N3124S), *RUNX1* c.499A>G (p.S167G) and *GATA2* c.920G>T (p.R307L).

BRCA2 (NM_000059) mutation c.9371A>G (p.N3124S) was identified in a patient suspicious of a HMMS. This variant was found with a VAF of 46% (61/132 reads). It had 8 pathogenic *in silico* predictions and only 1 benign prediction. Besides, another amino-acid missense variant at this position, p.N3124I, was classified as 'disease' by UniProt and as 'pathogenic' or 'likely pathogenic' by ClinVar.

RUNX1 (NM_001754) mutation c.499A>G (p.S167G) was detected by the hereditary gene panel with a VAF of 52% (158/306 reads). Although this variant was not described by Ensembl, ClinVar or COSMIC, *in silico* predictors supported a pathogenic verdict. Also, functional studies demonstrated that this mutation severely compromises DNA binding (Kwok *et al.*, 2009). And, as this variant was located at the end of exon 2, the effect on splicing was analysed using several predictors. HSF showed that this genetic alteration may break enhancer elements and create new silencers, and SPANR indicated that it would likely cause exon skipping, in which case, this would result in a change in the reading frame.

GATA2 (NM_001145661) variant c.920G>T (p.R307L) was identified with a VAF of 49% (49/101 reads). Although it had not been previously described neither in Ensembl nor in ClinVar, COSMIC (COSM255191) defined it as pathogenic, identified in hematopoietic tissue (5 samples) and associated to AML. *In silico* predictors were neither favourable.

5.3.3 Variants of uncertain significance

From the total cohort of patients, six VUS variants were identified: two alterations in *ATG2B,* one synonymous (c.6168C>T, p.G2056=) and the other one, missense (c.1360G>A, p.V454M); one missense variant in *ETV6* (c.871A>G, p.R291G); one missense change in *BRCA2* (c.8629G>A, p.E2877K); and two missense alterations in *RUNX1* (c.179C>T, p.A60V; c.737C>T, p.T246M).

The three missense variants located in *ATG2B* (NM_018036) c.1360G>A (p.V454M), *ETV6* (NM_001987) c.871A>G (p.R291G) and *BRCA2* (NM_000059) c.8629G>A (p.E2877K), were detected with a VAF of 51%, 47% and 47%, respectively. In all three cases, *in silico* predictors gave conflicting interpretations. In the case of *ETV6* c.871A>G, this variant was located outside any functional domain. *ATG2B* c.1360G>A variant was described by COSMIC as pathogenic (COSM3690272), but only detected in one sample (large intestine) and associated with colon cancer. Missense alterations *ATG2B* c.1360G>A and *BRCA2* c.8629G>A were located at the end of exon 9 and exon 20, respectively, so their possible effect on splicing was also evaluated. Regarding *ATG2B* c.1360G>A, predictors indicated that it probably did not cause an alteration in splicing. But in the case of *BRCA2* c.8629G>A, HSF indicated breaking of enhancer sites and creation of silencer motifs, and SPANR pointed to an exon skipping effect, although not significant.

Two missense variants in *RUNX1* (NM_001754) c.179C>T (p.A60V) and c.737C>T (p.T246M) had been previously identified in two different individuals by the somatic NGS panel with a VAF of 54% and 47%, respectively, and were confirmed to be germline by Sanger. In both cases, there were contradictory interpretations by *in silico* predictors. Also, COSMIC described both variants as pathogenic but only identified in one hematopoietic sample (COSM24762, COSM1030459). And in both alterations, clinical significance by ClinVar was ‘uncertain’ and related to FPDMM.

Synonymous *ATG2B* (NM_018036) variant c.6168C>T (p.G2056=) was detected with a VAF of 53% (186/351 reads). Its impact on splicing was evaluated and the predictor HSF showed that this variant may cause the creation of an alternative donor site in the last exon, so it would result in a shorter protein. But this potential shorter protein did not seem to alter any functional domain, although it could affect protein stability.

5.3.4 Likely benign or benign variants

From the total 21 different variants identified, most of them (10, 47.6%) were categorized as benign/likely benign. Half of them were synonymous and the other half, missense. One missense variant was identified in *BRCA1* (c.3119G>A, p.S1040N); three variants were detected in *BRCA2*, two missense (c.5744C>T, p.T1915M; c.6100C>T, p.R2034C) and one synonymous (c.2883G>A, p.Q961=); three alterations in *ATG2B*, all synonymous (c.12G>A, p.P4=; c.1956T>C, p.S652=; c.4584C>T, p.P1528=); two in *SRP72,* one missense (c.58C>T, p.R20W) and the other one, synonymous (c.1803G>A,

p.G601=) and one missense variant in *ANKRD26* (c.3655G>A, p.V1219I). These 10 benign germline alterations were found in 7 patients. In all cases, VAF was between 42-50%.

These alterations were classified as likely benign because of several reasons. As for missense mutations, most of them were considered 'benign/likely benign' by ClinVar and/or COSMIC, *in silico* predictors indicated a benign significance (except in one case: *SRP72* c.58C>T) and, in some of them, the MAF value in the European population was higher than 1%. Regarding synonymous variants, their effect on splicing was evaluated using different predictors and no potential alteration of splicing was detected. Besides, some of them were classified as 'likely benign' by ClinVar and/or had a European MAF higher than 1%.

5.3.5 UTR and intronic variants

From the total 26 distinct non-exonic variants detected, 18 were UTR variants and 8, intronic. Seven of the 26 variants (27%) were VUS and the rest of them were likely benign (Figure 10C). As for UTR variants, 13 of them were classified as likely benign because of having a European MAF higher than 1% or being present in several individuals, and the rest of them (n=5) were categorized as VUS, due to the lack of information in databases and predictors to analyse their effect. All intronic variants were analysed by the predictors HSF and SPANR. If both predictors agreed that the alteration had no effect on splicing, variants were considered as likely benign. Six of them were categorized as likely benign and 2 of them as VUS (*BRCA1,* c.212+17T>C; *ATG2B,* c.3749+24T>C). Both VUS intronic variants were predicted to create enhancer elements and break silencer sites by HSF and SPANR predicted an exon skipping effect.

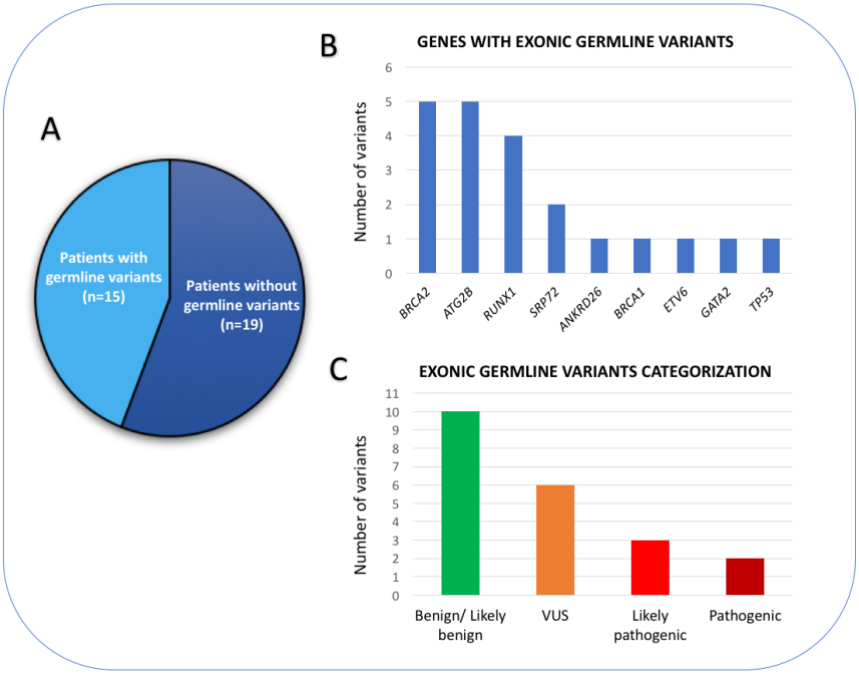

FIGURE 9 - **A) Pie chart showing the proportion of patients who carried potentially germline exonic variants. B) List of genes in which exonic germline variants were detected. C) Categorization of detected exonic germline variants.**
VUS, variants with uncertain significance.

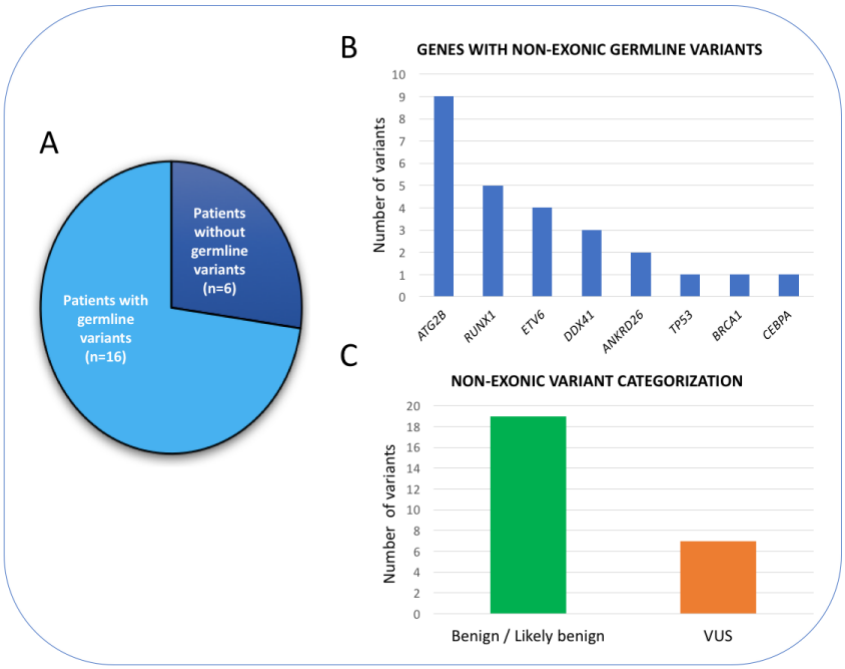

FIGURE 10 - **A) Pie chart showing the proportion of patients with potentially germline non-exonic variants. B) List of genes in which non-exonic germline variants were detected. C) Categorization of detected non-exonic germline variants.**
VUS, variants with uncertain significance.

5.4 VALIDATION OF DETECTED GENETIC ALTERATIONS BY SANGER

5.4.1 Validation of variants detected by the somatic NGS panel

Twelve *RUNX1* mutations previously detected in 12 individuals by the somatic NGS panel were validated by Sanger sequencing in the paired germline sample. From these 12 mutations, three of them were found in the sample in complete remission, confirming that these three patients carried a germline mutation in *RUNX1* (Figure 11). Mutations identified were c.497G>A (patient 25), c.737C>T (patient 27) and c.179C>T (patient 34), the first one categorized as pathogenic and the last two, as VUS.

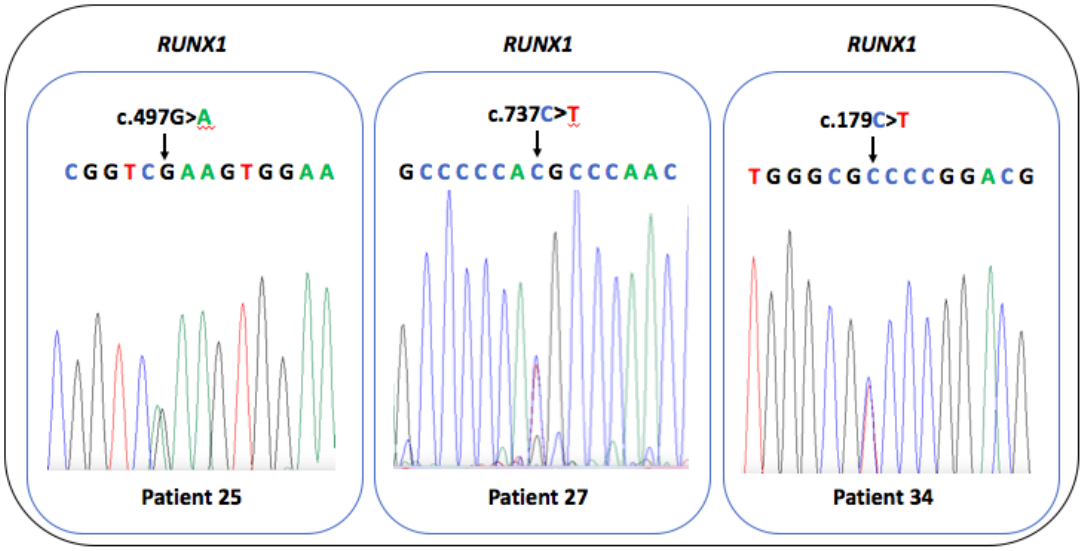

FIGURE 11 - **Sequencing electropherograms showing localization of *RUNX1* germline mutations identified in DNA from bone marrow samples in remission in patients 25, 27 and 34, by the somatic NGS panel.**

5.4.2 Validation of variants detected by the hereditary NGS panel

As paired germline DNA samples were directly used for the hereditary NGS panel analysis, we decided to only validate by Sanger pathogenic and likely pathogenic mutations as well as some VUS variants located in conflictive genomic regions. These variants were: *GATA2* c.920G>T and *BRCA2* c.8629G>A (patient 8) classified as likely pathogenic and VUS and detected with a VAF of 49% and 47%, respectively; *RUNX1* c.499A>G (patient 15) with a VAF of 52% and a likely pathogenic categorization; *ATG2B* c.6168C>T (patient 17) categorized as VUS and with a VAF of 53% and *TP53* c.404G>T (patient 18) with a pathogenic verdict and a VAF of 57%. Sanger electropherograms confirmed the existence of these variants and peak size was in accordance with VAF values detected by NGS (Figure 12).

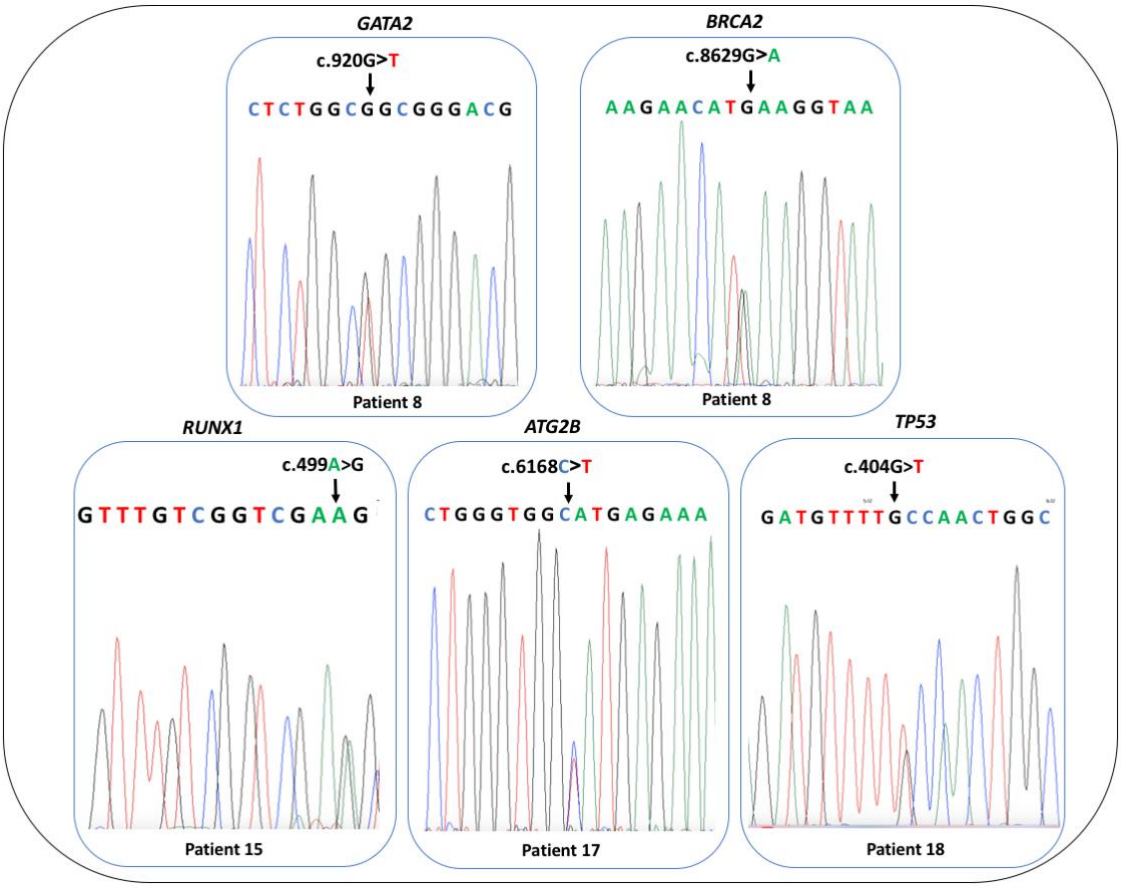

FIGURE 12 - **Sequencing electropherograms showing localization of *GATA2*, *BRCA2*, *RUNX1*, *ATG2B* and *TP53* germline mutations identified in DNA from bone marrow samples in remission in patients 8, 15, 17 and 18, respectively, by means of the hereditary NGS panel.**

6. DISCUSSION

6. DISCUSSION

The purpose of this study was to determine the frequency of germline mutations in patients diagnosed with AML/MDS by a targeted NGS panel including a series of genes related to HMMSs. The frequency of potentially germline exonic variants identified was higher than expected (44.1%), although the frequency of mutations with a pathogenic or likely pathogenic categorization was lower (14.7%). In total, 21 exonic variants were identified in the total cohort of 34 patients. Genes with more recurrent genetic alterations were *ATG2B* (23.8%*), BRCA2* (23.8%*)* and *RUNX1* (19%*).* Among the 21 variants, 10 were categorized as likely benign, 6 as VUS, 3 as likely pathogenic and 2 as pathogenic.

The fact that *ATG2B* and *BRCA2* were the most recurrently mutated genes could be explained by the large size of these genes. In addition, the high number of *RUNX1* variants could be related to the fact that our study population was enriched with patients with previously detected *RUNX1* mutations.

Potentially germline mutations identified with a categorization of pathogenic or likely pathogenic may more than likely have had a role in leukemogenesis and may have contributed to a competitive advantage of malignant cells *versus* healthy hematopoietic cells. The potential impact of these genetic variants in the oncogenic process is further discussed below.

It is estimated that 10-30% of the patients with AML carry a *RUNX1* mutation, although the incidence of germline mutations is unknown (Holme *et al.*, 2012; Mendler *et al.*, 2012). In our study, the frequency of individuals with a pathogenic germline *RUNX1* mutation was 5.9% (2 of 34 patients). Mutations detected were p.S167G and p.R166Q in patients 15 and 25, respectively. Germline *RUNX1* mutations are related to FPDMM. This gene codifies for a transcription factor with a key role in hematopoiesis regulation. As the two identified mutations affected the highly-conserved runt homology domain at the N-terminal part of the RUNX1 protein, which is important for binding DNA of target genes, these mutations may prevent affected cells from differentiation. Besides, there are studies which reveal that mutated *RUNX1* cells have a higher proliferative capacity, an altered DNA repair and a resistant phenotype against genotoxic stress (Bellissimo and Speck, 2017). *RUNX1* mutations are associated with a poor prognosis, so there is an urgent need to develop new targeted therapies in order to treat these patients and increase their survival rate.

GATA2 deficiencies constitute the third main entity of myeloid malignancy predisposition syndromes (Hahn *et al.*, 2012). In our cohort, only one patient (patient 8) with AML (2.9% of studied individuals) had a missense mutation in *GATA2* (p.R307L). *GATA2* is another transcription factor involved in regulating hematopoiesis and contributing to hematopoietic stem cells' survival. In this case, the described genetic alteration affected the conserved N-terminal zinc finger domain, relevant in DNA binding, so that it may contribute to the leukemogenic process in a similar way than *RUNX1* mutations, by preventing DNA binding and therefore, reducing its transcriptional activity. Some future therapeutic approaches may include gene editing procedures or increasing wild-type protein stability (Crispino and Horwitz, 2018).

In our study, a *TP53* mutation, p.C135F, was discovered in patient 18 (2.9% of the total cohort). Germline *TP53* mutations are related to Li-Fraumeni syndrome, a rare disorder

characterized by an increased risk of developing different neoplasms. *TP53* is a tumor suppressor gene, as its encoded p53 protein has a key function in cell cycle arrest, DNA repair and apoptosis. The detected mutation, also located in the DNA-binding region, may affect the activation of p53 target genes and the ability of cells to cope with stress and DNA damage, therefore promoting the acquisition of mutations. Several compounds have been identified which restore the function of wild-type p53 or that cause mutant p53 degradation, and there are also drugs which target signalling pathways necessary for growth and survival of p53 mutated cells. These drugs have been tested in cellular and/or animal models and constitute a promising future approach to treat these individuals (Zhao *et al.*, 2017).

The prevalence of hereditary breast cancer and ovarian cancer with germline mutations in *BRCA1/BRCA2* is 1/400 (Feurstein *et al*., 2016). Recently, it has been discovered that these mutations are not only associated with ovarian/breast cancer predisposition, but also with other malignancies, such as familial AML/MDS. In our case, a mutation in *BRCA2,* p.N3124S, was detected in patient 21. *BRCA2* is a tumor suppressor gene whose main role is repairing DNA damage. This mutation was located in the oligonucleotide/oligosaccharide-binding domain 3, involved in DNA binding. Therefore, by preventing binding of target genes, DNA repair may be altered, contributing to the progressive acquisition of driver mutations important in leukemia development.

The frequency of germline mutations detected in our study may be higher than previous published studies. This could be explained by the selection bias, as we have selected a cohort of patients under the age of 60 and, in addition, two of them were suspicious of having a HMMS because of a family history of hematologic neoplasms and/or thrombocytopenia.

Validation by Sanger was performed in all mutations detected by the somatic NGS panel in the paired germline sample, as they had been first identified in DNA samples at the moment of diagnosis. The hereditary NGS panel was directly analysed in the paired germline sample of each patient. For this reason, only mutations with a pathogenic/likely pathogenic classification and VUS localized in conflictive genomic regions were validated. Obtaining germline DNA can be a challenge in these patients, as blood, bone marrow or saliva samples may be contaminated with malignant cells. The most recommended sources of germline DNA are cultured fibroblasts from a skin biopsy, hair samples or buccal swab (Brown *et al.,* 2017). In our case, there was no availability of these recommended germline samples. So, this is why detected variants are referred as “potentially germline”. Due to the high VAF values (40-60%) and that only some of the samples had a small percentage of blasts, it is highly probable that these variants are of germline origin. However, it is true that some mutations may persist even in remission, and consequently, fibroblasts, hair or buccal swab samples would be needed to confirm their origin.

In our study, after an exhaustive analysis of the detected exonic variants, 10 were categorized as likely benign, 6 as VUS, 3 as likely pathogenic and 2 as pathogenic. But variant categorization remains a challenge. Due to the advance of NGS technologies and the increase of genomic data, there is an urgent need to stablish standardized criteria that facilitate variant classification and results interpretation in the clinical context. This categorization must be conducted taking into account clinical databases’ information, available literature, *in silico* predictions and gene actionability, that is, if the

target gene has diagnostic/prognostic/therapeutic implications in the disease under study. But, although there are guidelines for variant categorization, interpretation by different institutions may be different, resulting in discrepant reports and difficulties in medical management. Additionally, reporting VUS variants whose impact in disease is not clear remains controversial. Some institutions only report variants when there is a clear pathogenic effect, whereas others tend to report also VUS variants if they affect actionable genes. These uncertain variants should be reviewed from time to time with the new available literature trying to re-categorize them once their predicted effect is elucidated.

A further step would be to perform functional analysis to study the real effect of the detected alterations. As for genetic variants with a potential impact on splicing, splice reporter vectors, known as minigenes, could be used to verify whether the nucleotide change causes an aberrant splicing or not. For other types of genetic variants, it would be intriguing to study mutant transcription level, protein stability or localization using different approaches, such as quantitative real-time PCR, western blotting and immunoblotting or immunofluorescence.

Among all individuals with germline alterations, another further action would be to identify those cases with HSCT from a related donor. These donor samples could be analysed to test if these individuals harboured the same germline mutation as the patient. Moreover, family members of individuals with HMMSs should be also tested and included in surveillance programs. Also, segregation studies in these families would be useful to identify new variants that confer a predisposition to develop a myeloid neoplasm.

We acknowledge that the main limitation of our study is that, as it is a retrospective study, there were no paired germline samples obtained from patients' skin fibroblasts stored in La Fe Biobank, so bone marrow remission samples were used as germline DNA. In addition, the real incidence and prevalence of HMMSs in the population cannot be determined with this data, as the study cohort was small and there was a selection bias. However, this cohort is of great value as there is a small percentage of patients with familial AML/MDS.

7. CONCLUSIONS

7. CONCLUSIONS

1. The presence of potentially germline variants in HMMSs-related predisposition genes in patients with AML/MDS is more frequent than previously expected.

2. There is an urgent need to develop custom NGS panels including genes with predisposition to myeloid malignancies to be able to correctly diagnose HMMSs.

3. A meticulous anamnesis collection including familial medical history is crucial to ease the identification of HMMSs. And a paired germline sample (skin biopsy, hair samples or buccal swab, preferably) should be collected and stored in La Fe Biobank from these patients in suspicious cases.

4. Recognition of these familial cases of AML/MDS would result in a better management of these patients, specially in donor selection for HSCT and in genetic surveillance of patients and family members.

5. Further functional studies of the effect of germline mutations in HMMSs-related genes are necessary to elucidate their implication in the leukemogenic process, which could be of great utility for the future development of new therapeutic drugs.

8.REFERENCES

8. REFERENCES

ADZHUBEI, I.A.; SCHMIDT, S.; PESHKIN, L.; RAMENSKY, V.E.; GERASIMOVA, A.; BORK, P.; KONDRASHOV, A.S. AND SUNYAEV, S.R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4): 248–249.

ANDREW, S. (2010). FastQC: a quality control tool for high throughput sequence data. Cambridge, UK: Brabaham Institute.
Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

ANTONY-DEBRÉ, I.; MANCHEV, V.T.; BALAYN, N.; BLUTEAU, D.; TOMOWIAK, C.; LEGRAND, C.; LANGLOIS, T.; BAWA, O.; TOSCA, L.; TACHDJIAN, G.; LEHEUP, B.; DEBILI, N.; PLO, I.; MILLS, J.A.; FRENCH, D.L.; WEISS, M.J.; SOLARY, E.; FAVIER, R.; VAINCHENKER, W. AND RASLOVA, H. (2015). Level of RUNX1 activity is critical for leukemic predisposition but not for thrombocytopenia. *Blood*, 125(6): 930–940.

ARBER, D.A.; ORAZI, A.; HASSERJIAN, R.; THIELE, J.; BOROWITZ, M.J.; LE BEAU, M.M.; BLOOMFIELD, C.D.; CAZZOLA, M. AND VARDIMAN, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20): 2391–2405.

BALMAÑA, J.; DIGIOVANNI, L.; GADDAM, P.; WALSH, M.F.; JOSEPH, V.; STADLER, Z.K.; NATHANSON, K.L.; GARBER, J.E.; COUCH, F.J.; OFFIT, K.; ROBSON, M.E. AND DOMCHEK, S.M. (2016). Conflicting interpretation of genetic variants and cancer risk by commercial laboratories as assessed by the prospective registry of multiplex testing. *Journal of Clinical Oncology*, 34(34): 4071–4078.

BAPTISTA, R.L.R.; DOS SANTOS, A.C.E.; GUTIYAMA, L.M.; SOLZA, C. AND ZALCBERG, I.R. (2017). Familial Myelodysplastic/Acute Leukemia Syndromes-Myeloid Neoplasms with Germline Predisposition. *Frontiers in Oncology*, 7(206):1–8.

BEHRENS, K.; TRIVIAI, I.; SCHWIEGER, M.; TEKIN, N.; ALAWI, M.; SPOHN, M.; INDENBIRKEN, D.; ZIEGLER, M.; MÜLLER, U.; ALEXANDER, W.S. AND STOCKING, C. (2016). *RUNX1* downregulates stem cell and megakaryocytic transcription programs that support niche interactions. *Blood*, 127(26): 3369–3381.

BELLISSIMO, D.C. AND SPECK, N.A. (2017). *RUNX1* Mutations in Inherited and Sporadic Leukemia. *Frontiers in Cell and Developmental Biology*, 5(111): 1–11.

BLUTEAU, D.; BALDUINI, A.; BALAYN, N.; CURRAO, M.; NURDEN, P.; DESWARTE, C.; LEVERGER, G.; NORIS, P.; PERROTTA, S.; SOLARY, E.; VAINCHENKER, W.; DEBILI, N.; FAVIER, R. AND RASLOVA, H. (2014). Thrombocytopenia-associated mutations in the *ANKRD26* regulatory region induce MAPK hyperactivation. *Journal of Clinical Investigation,* 124(2): 580–591.

BÖDÖR, C.; RENNEVILLE, A.; SMITH, M.; CHARAZAC, A.; IQBAL, S.; ÉTANCELIN, P.; CAVENAGH, J.; BARNETT, M.J.; KRAMARZOVÁ, K.; KRISHNAN, B.; MATOLCSY, A.; PREUDHOMME, C.; FITZGIBBON, J. AND OWEN, C. (2012). Germ-line *GATA2* p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and *ASXL1* mutation demonstrating rapid onset and poor survival. *Haematologica*, 97(6): 890–894.

BRAUN, D.; YANG, J.; GRIFFIN, M.; PARMIGIANI, G. AND HUGHES, K.S. (2018). A Clinical Decision Support Tool to Predict Cancer Risk for Commonly Tested Cancer-Related Germline Mutations. *Journal of Genetic Counseling*, 2018:1-13.

BROWN, A.L.; CHURPEK, J.E.; MALCOVATI, L.; DÖHNER, H. AND GODLEY, L.A. (2017). Recognition of familial myeloid neoplasia in adults. *Seminars in Hematology*, 54: 60–68.

CHEAH, J.J.C.; HAHN, C.N.; HIWASE, D.K.; SCOTT, H.S. AND BROWN, A.L. (2017). Myeloid neoplasms with germline *DDX41* mutation. *International Journal of Hematology*, 106(2): 163–174.

CHURPEK, J.E.; NICKELS, E.; MARQUEZ, R.; ROJEK, K.; LIU, B.; LORENZ, R.; LEPORE, J.; MADZO, J.; WICKREMA, A.; ARTZ, A.S.; VAN BESIEN, K. AND GODLEY, L. A. (2012). Identifying familial myelodysplastic / acute leukemia predisposition syndromes

through hematopoietic stem cell transplantation donor with thrombocytopenia. *Blood*, 120(26): 5247–5249.

COLLIN, M.; DICKINSON, R. AND BIGLEY, V. (2015). Haematopoietic and immune defects associated with *GATA2* mutation. *British Journal of Haematology*, 169:173–187.

CRISPINO, J.D. AND HORWITZ, M.S. (2017). *GATA* factor mutations in hematologic disease. *Blood*, 129(15): 2103–2110.

DE KOUCHKOVSKY, I. AND ABDUL-HAY, M. (2016). Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood Cancer Journal*, 6(7): e441.

DOHNER, H.; ESTEY, E.; GRIMWADE, D.; AMADORI, S.; APPELBAUM, F.R.; BÜCHNER, T.; DOMBRET, H.; EBERT, B.L.; FENAUX, P.; LARSON, R.A.; LEVINE, R.L.; LO-COCO, F.; NAOE, T.; NIEDERWIESER, D.; OSSENKOPPELE, G.J.; SANZ, M.; SIERRA, J.; TALLMAN, M.S.; TIEN, H.F.; WEI, A.H.; LÖWENBERG, B. AND BLOOMFIELD, C.D. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, 129(4): 424–447.

DRAZER, M.W.; KADRI, S.; SUKHANOVA, M.; PATIL, S.A.; WEST, A.H.; FEURSTEIN, S.; CALDERON, D.A.; JONES, M.F.; WEIPERT, C.M.; DAUGHERTY, C.K.; CEBALLOS-LÓPEZ, A.A.; RACA, G.; LINGEN, M.W.; LI, Z.; SEGAL, J.P.; CHURPEK, J.E. AND GODLEY, L.A. (2018). Prognostic tumor sequencing panels frequently identify germ line variants associated with hereditary hematopoietic malignancies. *Blood Advances*, 2(2): 146–150.

FEURSTEIN, S.; DRAZER, M.W. AND GODLEY, L.A. (2016). Genetic predisposition to leukemia and other hematologic malignancies. *Seminars in Oncology*, 43(5): 598–608.

GODLEY, L.A. AND SHIMAMURA, A. (2017). Genetic predisposition to hematologic malignancies: management and surveillance. *Blood,* 130(4): 424–432.

GREEN, S.M.; COYNE, H.J.; MCINTOSH, L.P. AND GRAVES, B.J. (2010). DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *Journal of Biological Chemistry*, 285(24): 18496–18504.

GREENBERG, P.L.; STONE, R.M.; AL-KALI, A.; BARTA, S.K.; BEJAR, R.; BENNETT, J.M., CARRAWAY, H., DE CASTRO, C.M., DEEG, H.J., DEZERN, A.E., FATHI, A.T.; FRANKFURT, O.; GAENSLER, K.; GARCIA-MANERO, G.; GRIFFITHS, E.A.; HEAD, D.; HORSFALL, R.; JOHNSON, R.A.; JUCKETT, M.; KLIMEK, V.M.; KOMROKJI, R.; KUJAWSKI, L.A.; MANESS, L.J.; O'DONNELL, M.R.; POLLYEA, D.A.; SHAMI, P.J.; STEIN, B.L.; WALKER, A.R.; WESTERVELT, P.; ZEIDAN, A.; SHEAD, D.A. AND SMITH, C. (2017). Myelodysplastic Syndromes, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*, 15(1): 60–87.

HAHN, C.N.; CHONG, C.E.; CARMICHAEL, C.L.; WILKINS, E.J.; BRAUTIGAN, P.J.; LI, X.C.; BABIC, M.; LIN, M.; CARMAGNAC, A.; LEE, Y.K.; KOK, C.H.; GAGLIARDI, L.; FRIEND, K.L.; EKERT, P.G.; BUTCHER, C.M.; BROWN, A.L.; LEWIS, I.D.; TO, L.B.; TIMMS, A.E.; STOREK, J.; MOORE, S.; ALTREE, M.; ESCHER, R.; BARDY, P.G.; SUTHERS, G.K.; D'ANDREA, R.J.; HORWITZ, M.S. AND SCOTT, H.S. (2012). Heritable *GATA2* Mutations Associated with Familial Myelodysplastic Syndrome and Acute Myeloid Leukemia. *Nature Genetics*, 43(10): 1012–1017.

DESMET, F.O.; HAMROUN, D.; LALANDE, M.; COLLOD-BÉROUD, G.; CLAUSTRES, M. AND BÉROUD, C. (2009). Human Splicing Finder : an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, 37(9): 1–14.

HIRABAYASHI, S.; WLODARSKI, M.W.; KOZYRA, E. AND NIEMEYER, C. M. (2017). Heterogeneity of *GATA2*-related myeloid neoplasms. *International Journal of Hematology,* 106(2): 175–182.

HOLME, H.; HOSSAIN, U.; KIRWAN, M.; WALNE, A.; VULLIAMY, T. AND DOKAL, I. (2012). Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *British Journal of Haematology*, 158(2): 242–248.

JENSEN, K.H.; IZARZUGAZA, J.M.G.; JUNCKER, A.S.; HANSEN, R.B.; HANSEN, T.F.; TIMSHEL, P.; BLONDAL, T.; JENSEN, T.S.; RYGAARD-HJALSTED, E.; MOURITZEN,

P.; THORSEN, M.; WERNERSSON, R.; NIELSEN, H.B.; JAKOBSEN, A.; BRUNAK, S. AND SØRENSEN, F.B. (2018). Analysis of a gene panel for targeted sequencing of colorectal cancer samples. *Oncotarget*, 9(10): 9043–9060.

KIRÁLY, A.P.; KÁLLAY, K.; GÁNGÓ, A.; KELLNER, Á.; EGYED, M.; SZŐKE, A.; KISS, R.; VÁLYI-NAGY, I.; CSOMOR, J.; MATOLCSY, A. AND BÖDÖR, C. (2018). Familial Acute Myeloid Leukemia and Myelodysplasia in Hungary. *Pathology and Oncology Research,* 24(1): 83–88.

KIRWAN, M.; WALNE, A.J.; PLAGNOL, V.; VELANGI, M.; HO, A.; HOSSAIN, U.; VULLIAMY, T. AND DOKAL, I. (2012). Exome sequencing identifies autosomal-dominant *SRP72* mutations associated with familial aplasia and myelodysplasia. *American Journal of Human Genetics*, 90(5): 888–892.

KUMAR, P.; HENIKOFF, S. AND NG, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(8): 1073–1082.

KUMAR, S.; STECHER, G. AND TAMURA, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7): 1870–1874.

KWIATKOWSKI, B.A.; BASTIAN, L.S.; BAUER, T.R.; TSAI, S.; ZIELINSKA-KWIATKOWSKA, A.G. AND HICKSTEIN, D.D. (1998). The ets Family Member Tel Binds to the Fli-1 Oncoprotein and Inhibits Its Transcriptional Activity*. The Journal of Biological Chemistry*, 273(28): 17525–17530.

KWOK, C.; ZEISIG, B.B.; QIU, J.; DONG, S. AND SO, C.W.E. (2009). Transforming activity of AML1-ETO is independent of CBFbeta and ETO interaction but requires formation of homo-oligomeric complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8): 2853–2858.

LANDRUM, M.J.; LEE, J.M.; BENSON, M.; BROWN, G.; CHAO, C.; CHITIPIRALLA, S.; BAOSHAN, G.; HART, J.; HOFFMAN, D.; HOOVER, J.; JANG, W.; KATZ, K., OVETSKY, M.; RILEY, G.; SETHI, A.; TULLY, R.; VILLAMARIN-SALOMON, R.; RUBINSTEIN, W. AND MAGLOTT, D.R. (2016). ClinVar : public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, 44: 862–868.

LATGER-CANNARD, V.; PHILIPPE, C.; BOUQUET, A.; BACCINI, V.; ALESSI, M.C.; ANKRI, A.; BAUTERS, A.; BAYART, S.; CORNILLET-LEFEBVRE, P.; DALIPHARD, S.; MOZZICONACCI, M.J.; RENNEVILLE, A.; BALLERINI, P.; LEVERGER, G.; SOBOL, H.; JONVEAUX, P.; PREUDHOMME, C.; NURDEN, P.; LECOMPTE, T. AND FAVIER, R. (2016). Haematological spectrum and genotype-phenotype correlations in nine unrelated families with *RUNX1* mutations from the French network on inherited platelet disorders. *Orphanet Journal of Rare Diseases*, 11(49):1–15.

LEWINSOHN, M.; BROWN, A.L.; WEINEL, L.M.; PHUNG, C.; RAFIDI, G.; LEE, M.K.; SCHREIBER, A.W.; FENG, J.; BABIC, M.; CHONG, C.E.; LEE, Y.; YONG, A.; SUTHERS, G.K.; POPLAWSKI, N.; ALTREE, M.; PHILLIPS, K.; JAENSCH, L.; FINE, M.; D'ANDREA, R.J.; LEWIS, I.D.; MEDEIROS, B.C.; POLLYEA, D.A.; KING, M.C.; WALSH, T.; KEEL, S.; SHIMAMURA, A.; GODLEY, L.A.; HAHN, C.N.; CHURPEK, J.E. AND SCOTT, H. S. (2016). Novel germ line *DDX41* mutations define families with a lower age of MDS/AML onset and lymphoid malignancies. *Blood*, 127(8): 1017–1023.

LI, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*: 1303.3997.

MCKENNA, A.; HANNA, M.; BANKS, E.; SIVACHENKO, A.; CIBULSKIS, K.; KERNYTSKY, A.; GARIMELLA, K.; ALTSHULER, D.; GABRIEL, S.; DALY, M. AND DEPRISTO, M.A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9): 1297–1303.

MENDLER, J.H.; MAHARRY, K.; RADMACHER, M.D.; MRÓZEK, K.; BECKER, H.; METZELER, K.H.; SCHWIND, S.; WHITMAN, S.P.; KHALIFE, J.; KOHLSCHMIDT, J.; NICOLET, D.; POWELL, B.L.; CARTER, T.H.; WETZLER, M.; MOORE, J.O.; KOLITZ, J.E.; BAER, M.R.; CARROLL, A.J.; LARSON, R.A.; CALIGIURI, M.A.; MARCUCCI, G. AND BLOOMFIELD, C.D. (2012). *RUNX1* mutations are associated with poor outcome in

younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and microRNA expression signatures. *Journal of Clinical Oncology*, 30(25): 3109–3118.

NOETZLI, L.; LO, R.W.; LEE-SHERICK, A.B.; CALLAGHAN, M.; NORIS, P.; SAVOIA, A.; RAJPURKAR, M.; JONES, K.; GOWAN, K.; BALDUINI, C.L.; PECCI, A.; GNAN, C.; DE ROCCO, D.; HELLER, P.; GUTIERREZ-HARTMANN, A.; XIAYUAN, L.; PLUTHERO, F.G.; ROWLEY, J.W.; WEYRICH, A.S.; KAHR, W.H.A.; PORTER, C.C. AND DI PAOLA, J. (2015). Germline mutations in *ETV6* are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nature Genetics*, 47(5): 535–538.

NORIS, P.; PERROTTA, S.; SERI, M.; PECCI, A.; GNAN, C.; LOFFREDO, G.; PUJOL-MOIX, N.; ZECCA, M.; SCOGNAMIGLIO, F.; DE ROCCO, D.; PUNZO, F.; MELAZZINI, F.; SCIANGUETTA, S.; CASALE, M.; MARCONI, C.; PIPUCCI, T.; AMENDOLA, G.; NOTARANGELO, L.D.; KLERSY, C.; CIVASCHI, E.; BALDUINI, C.L. AND SAVOIA, A. (2011). Mutations in *ANKRD26* are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood*, 117(24): 6673–6680.

PABST, T.; EYHOLZER, M.; HAEFLIGER, S.; SCHARDT, J. AND MUELLER, B.U. (2008). Somatic *CEBPA* mutations are a frequent second event in families with germline *CEBPA* mutations and familial acute myeloid leukemia. *Journal of Clinical Oncology,* 26(31): 5088–5093.

PABST, T.; MUELLER, B.U.; ZHANG, P.; RADOMSKA, H.S.; NARRAVULA, S.; SCHNITTGER, S.; BEHRE, G.; HIDDEMANN, W. AND TENEN, D.G. (2001). Dominant-negative mutations of *CEBPA*, encoding CCAAT/enhancer binding protein-alpha (C/EBPa), in acute myeloid leukemia. *Nature Genetics*, 27(3): 263–270.

PIPPUCCI, T.; SAVOIA, A.; PERROTTA, S.; PUJOL-MOIX, N.; NORIS, P.; CASTEGNARO, G.; PECCI, A.; GNAN, C.; PUNZO, F.; MARCONI, C.; GHERARDI, S.; LOFFREDO, G.; DE ROCCO, D.; SCIANGUETTA, S.; BAROZZI, S.; MAGINI, P.; BOZZI, V.; DEZZANI, L.; DI STAZIO, M.; FERRARO, M.; PERINI, G.; SERI, M. AND BALDUINI, C.L. (2011). Mutations in the 5′ UTR of *ANKRD26*, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *American Journal of Human Genetics*, 88: 115–120.

POLPRASERT, C.; SCHULZE, I.; SEKERES, M. A.; MAKISHIMA, H.; PRZYCHODZEN, B.; HOSONO, N.; SINGH, J.; PADGETT, R.A.; GU, X.; PHILLIPS, J.G.; CLEMENTE, M.; PARKER, Y.; LINDNER, D.; DIENES, B.; JANKOWSKY, E.; SAUNTHARARAJAH, Y.; DU, Y.; OAKLEY, K.; NGUYEN, N.; MUKHERJEE, S.; PABST, C.; GODLEY, L.A.; CHURPEK, J.E.; POLLYEA, D.A.; KRUG, U.; BERDEL, W.E.; KLEIN, H.U.; DUGAS, M.; SHIRAISHI, Y.; CHIBA, K.; TANAKA, H.; MIYANO, S.; YOSHIDA, K.; OGAWA, S.; MÜLLER-TIDOW, C. AND MACIEJEWSKI, J.P. (2015). Inherited and Somatic Defects in *DDX41* in Myeloid Neoplasms. *Cancer Cell*, 27(5): 658–670.

PORTER, C.C. (2016). Germ line mutations associated with Leukemias. *American Society of Hematology,* 2016(1): 302–308.

PREUDHOMME, C.; RENNEVILLE, A.; BOURDON, V.; PHILIPPE, N.; ROCHE-LESTIENNE, C.; BOISSEL, N.; DHEDIN, N.; ANDRÉ, J.M.; CORNILLET-LEFEBVRE, P.; BARUCHEL, A.; MOZZICONACCI, M.J. AND SOBOL H. (2009). High frequency of *RUNX1* biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. *Blood*, 113(22): 5583–5587.

RADOMSKA, H.S.; HUETTNER, C.S.; ZHANG, P.; CHENG, T.; SCADDEN, D.T. AND TENEN, D.G. (1998). CCAAT / Enhancer Binding Protein α Is a Regulatory Switch Sufficient for Induction of Granulocytic Development from Bipotential Myeloid Progenitors. *Molecular and Cellular Biology*, 18(7): 4301–4314.

RICHARDS, S.; AZIZ, N.; BALE, S.; BICK, D.; DAS, S. AND GASTIER-FOSTER, J. (2015) Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5): 405-424.

ROBINSON, J.T.; THORVALDSDÓTTIR, H.; WINCKLER, W.; GUTTMAN, M.; LANDER,

E.S.; GETZ, G. AND MESIROV, J.P. (2011) Integrative Genomic Viewer. (2013). *Nature Biotechnology*, 29(1): 24–26.

SALIBA, J.; SAINT-MARTIN, C.; DI STEFANO, A.; LENGLET, G.; MARTY, C.; KEREN, B.; PASQUIER, F.; DELLA VALLE, V.; SECARDIN, L.; LEROY, G.; MAHFOUDHI, E.; GROSJEAN, S.; DROIN, N.; DIOP, M.; DESSEN, P.; CHARRIER, S.; PALAZZO, A.; MERLEVEDE, J.; MENIANE, J.C.; DELAUNAY-DARIVON, C.; FUSEAU, P.; ISNARD, F.; CASADEVALL, N.; SOLARY, E.; DEBILI, N.; BERNARD, O.A.; RASLOVA, H.; NAJMAN, A.; VAINCHENKER, W.; BELLANNÉ-CHANTELOT, C. AND PLO, I. (2015). Germline duplication of *ATG2B* and *GSKIP* predisposes to familial myeloid malignancies. *Nature Genetics*, 47(10): 1131–1140.

SCHLEGELBERGER, B. AND HELLER, P. G. (2017). *RUNX1* deficiency (familial platelet disorder with predisposition to myeloid leukemia, FPDMM). *Seminars in Hematology*, 54: 75–80.

SCHMIEDER, R. AND EDWARDS, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6): 863–864.

SCHULZ, E.; VALENTIN, A.; ULZ, P.; BEHAM-SCHMID, C.; LIND, K.; RUPP, V.; LACKNER, H.; WÖLFLER, A.; ZEBISCH, A.; OLIPITZ, W.; GEIGL, J.; BERGHOLD, A.; SPEICHER, M.R. AND SILL, H. (2012). Germline mutations in the DNA damage response genes *BRCA1, BRCA2, BARD1* and *TP53* in patients with therapy related myeloid neoplasms. *Journal of Medical Genetics*, 49(7): 422–428.

SCHWARZ, J.M.; COOPER, D.N.; SCHUELKE, M. AND SEELOW, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, 11(4): 361–362.

SMITH, M.L.; CAVENAGH, J.D.; LISTER, T.A. AND FITZGIBBON, J. (2004). Mutation of *CEBPA* in Familial Acute Myeloid Leukemia. *The New England Journal Of Medicine*, 351(23): 2403–2407.

SONG, W.J.; SULLIVAN, M.G.; LEGARE, R.D.; HUTCHINGS, S.; TAN, X.; KUFRIN, D.; RATAJCZAK, J.; RESENDE, I.C.; HAWORTH, C.; HOCK, R.; LOH, M.; FELIZ, C.; ROY, D.C.; BUSQUE, L.; KURNIT, D.; WILLMAN, C.; GEWIRTZ, A.M.; SPECK, N.A.; BUSHWELLER, J.H.; LI, F.P.; GARDINER, K.; PONCZ, M.; MARIS, J.M. AND GILLILAND, D.G. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nature Genetics*, 23(2): 166–175.

TAWANA, K.; WANG, J.; RENNEVILLE, A.; BÖDÖR, C.; HILLS, R.; LOVEDAY, C.; SAVIC, A.; VAN DELFT, F.W.; TRELEAVEN, J.; GEORGIADES, P.; UGLOW, E.; ASOU, N.; UIKE, N.; DEBELJAK, M.; JAZBEC, J.; ANCLIFF, P.; GALE, R.; THOMAS, X.; MIALOU, V.; DÖHNER, K.; BULLINGER, L.; MUELLER, B.; PABST, T.; STELLJES, M.; SCHLEGELBERGER, B.; WOZNIAK, E.; IQBAL, S.; OKOSUN, J.; ARAF, S.; FRANK, A.K.; LAURIDSEN, F.B.; PORSE, B.; NERLOV, C.; OWEN, C.; DOKAL, I.; GRIBBEN, J.; SMITH, M.; PREDUHOMME, C.; CHELALA, C.; CAVENAGH, J. AND FITZGIBBON, J. (2015). Disease evolution and outcomes in familial AML with germline CEBPA mutations. *Blood*, 126(10): 1214–1223.

THE UNIVERSITY OF CHICAGO HEMATOPOIETIC MALIGNANCIES CANCER RISK TEAM. (2016). How I diagnose and manage individuals at risk for inherited myeloid malignancies. *Blood*, 128(14): 1800–1813.

TSONGALIS, G.J.; PETERSON, J.D.; DE ABREU, F.B.; TUNKEY, C.D.; GALLAGHER, T.L.; STRAUSBAUGH, L.D.; WELLS, W.A. AND AMOS, C. I. (2014). Routine use of the Ion Torrent AmpliSeqTM Cancer Hotspot Panel for identification of clinically actionable somatic mutations. *Clinical Chemistry and Laboratory Medicine*, 52(5): 707–714.

WLODARSKI, M. W.; COLLIN, M. AND HORWITZ, M. S. (2017). *GATA2* deficiency and related myeloid neoplasms. *Seminars in Hematology*, 54(2): 81–86.

XIONG, H.Y.; ALIPANAHI, B.; LEE, L.J.; BRETSCHNEIDER, H.; MERICO, D.; YUEN, R.K.C.; HUA, Y.; GUEROUSSOV, S.; NAJAFABADI, H.S.; HUGHES, T.R.; MORRIS, Q.; BARASH, Y.; KRAINER, A.R.; JOJIC, N.; SCHERER, S.W.; BLENCOWE, B.J. AND FREY, B.J. (2015). The human splicing code reveals new insights into the genetic

determinants of disease. *Science*, 347(6218): 1-20.

ZHANG, M. Y.; CHURPEK, J. E.; KEEL, S.B.; WALSH, T.; LEE, M.K.; LOEB, K.R.; GULSUNER, S.; PRITCHARD, C.C.; SANCHEZ-BONILLA, M.; DELROW, J.J.; BASOM, R.S.; FOROUHAR, M.; GYURKOCZA, B.; SCHWARTZ, B.S.; NEISTADT, B.; MARQUEZ, R.; MARIANI, C.J.; COATS, S.A.; HOFMANN, I.; LINDSLEY, R.C.; WILLIAMS, D.A.; ABKOWITZ, J.L.; HORWITZ, M.S.; KING, M.C.; GODLEY, L.A. AND SHIMAMURA, A. (2015). Germline *ETV6* mutations in familial thrombocytopenia and hematologic malignancy. *Nature Genetics*, 47(2): 180–185.

ZHAO, D.; TAHANEY, W.M.; MAZUMDAR, A.; SAVAGE, M.I. AND BROWN, P. H. (2017). Molecularly targeted therapies for p53-mutant cancers. *Cellular and Molecular Life Sciences*, 74(22): 4171–4187.