SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Medicine and Surgery
PhD program in Translational and
Molecular Medicine XXXIV cycle

Predisposition to hematological malignancies in children and adults: from genetic profiling to clonal evolution

Claudia Saitta
registration number: 798147

Tutor: Prof. Matteo Della Porta
Co-tutor: Prof. Giovanni Cazzaniga
Dr. Grazia Fazio
Coordinator: Prof. Andrea Biondi

ACADEMIC YEAR
2020/2021

A VOI, così folli da essere comunque, senza alcuna costrizione,

la parte migliore della mia 'loquace', disagiata quotidianità…

A TE, che forse, in qualche modo, non mi abbandonerai comunque mai…

Table of contents

Chapter 1

General Introduction

1. Role of genetic predisposition in hematological malignancies

Hematological Malignancies (HM) are a heterogenous group of clonal malignant disorders originating from abnormal transformation of hematological lineages. Their pathogenesis is a multifactorial process, with a conspicuous contribution of several genetic factors that trigger the malignant transformation. In the last decade, genomic analyses have allowed great advances in our understanding of genetic and biological bases of HM, underlying an increasing involvement of hereditary predisposition.[1,2] In the last decades, the number of genetic conditions associated to predisposition to childhood leukemia has been hypothesized to occur in about 8% of pediatric tumors, including leukemias.[3,4] Although several syndromes have already been associated to cancer predisposition, new lines of evidence show that prevalence and spectrum of cancer predisposing mutations among children and adolescents have yet to be recognized.[1,3,5]

A better knowledge of the germline mutational landscape may be crucial to improve the understanding of tumorigenesis, with positive effects on patient care, treatment and follow up, as well as on cancer prevention and genetic counseling for patients and families.[6]

1.1 Predisposition in childhood Acute Lymphoblastic Leukemia

1.1.1 Pediatric B-cell precursor Acute Lymphoblastic Leukemia (BCP-ALL)

Acute Lymphoblastic Leukemia (ALL) is the most frequent cancer in childhood. It is a clonal malignant disorder, in which hematopoietic stem cells of bone marrow lose their ability to differentiate into mature B or T lymphocytes.[7] The aberrant hematopoietic precursors (blasts) maintain an uncontrolled capability to self-renew, suppressing other lineages of the hematopoietic system. In Caucasian population, about 80% of ALL are derived from abnormal proliferation of B-cell precursor (BCP-ALL), while the remaining 20% involves T-cell precursor (TCP-ALL).[7]

The peak incidence is between 3 to 5 years of age, with a slight male predominance (55% males compared to 45% females).[8,9] In Italy, there are about 400 newly diagnosed pediatric ALL cases per year and approximately 15-25% of adult leukemias.[10,11]

Although several progresses have been achieved in term of disease's risk stratification, international multi-agent chemotherapy protocols and personalized 'risk-adapted' therapies, ALL remains the leading cause of death in pediatric age (10% of cases).[12]

Moreover, even in patient with long-term remission, ALL treatment can lead to acute or late severe toxicities and may induce treatment-related second neoplasms.

Pathogenesis

The etiology of ALL is still to be clarified. It is supposed to have a multifactorial origin, including several genetic factors, associated with an increased risk, in addition to exogenous risk factors (infections, ionizing radiation or chemicals).[13] According to 'Greaves delayed infection hypothesis',[9] Leukemia is attributable to a two-step mechanism. A prenatal acquisition of a pre-leukemic clone predisposes to an aberrant response of the immune system to a delayed exposure to common infections. In these susceptible individuals, therefore, contact with any pathogen can determine the development of a second genetic aberration, which can result in an aberrant growth of lymphoid populations.[13]

Etiology and molecular characterization

The genetic landscape plays a key role. Different subtypes of disease can be classified according to genetic abnormalities, that can be divided in two major groups: structural aberrations (translocations and gene rearrangements – *ABL1*, *KMT2A, CRLF2, IGH*) and numerical abnormalities.[9,14] Structural chromosome aberrations [such as translocations t(12;21), t(9;22), t(4;11), t(1;19) etc.] can be detected

through cytogenetic analysis (karyotyping and Fluorescence In Situ Hybridization-FISH) or through molecular screening, such as Next Generation Sequencing approaches[15] or multiplex RT-PCR detection of recurrent break-point regions.

Numerical chromosomal abnormalities are very common in childhood ALL and are investigated through the evaluation of blasts karyotype or DNA content (DNA index). Patients can be classified as normal karyotype, hypodiploid (less than 46 chromosomes) or hyperdiploid (>46 chromosomes).

All these genetic abnormalities are relevant for the prognostic stratification of ALL patients.

1.1.2 Genetic landscape and hereditable susceptibility

Recently, it has been demonstrated that rare germline mutations are responsible of individual genetic susceptibility, mainly linked to familiar cases.[13]

As reported by Inaba and Mullighan,[16] several lines of evidence support the involvement of genetic predisposition in ALL.

These indications include: (i) rare constitutional syndromes; (ii) familial cancer syndromes; (iii) non-coding DNA polymorphisms; and (iv) genes harboring germline non-silent variants.

Among the well-known genetic conditions that predispose to the development of ALL, the Down Syndrome is the most frequent, with

an incidence of 2-3%.[17] Other syndromes associated with an increased ALL risk are Li Fraumeni Syndrome, Bloom syndrome, Ataxia-telangiectasia and Fanconi Anemia.[13,18–20]

As regarding non-syndromic ALL, genomic screenings have showed that a non-negligible number of patients are carrier of germline genetic variants in "key genes", usually disrupted by somatic mutations in disease phase.[21]

Among involved genes, *PAX5* is noteworthy.[22] Germline variants of this gene have been described as predisposing to BCP-ALL in single families, together with somatic cooperating mutations that are needed for the development of the disease.[23–25] Besides, also germline mutations in *ETV6* (~1% of children with apparently sporadic ALL)[26–29], *IKZF1*[30,31], *ANKRD26*, and *GATA2* are linked with ALL risk, alongside with the DNA damage response genes *TP53* and *BRCA1/2*.[21] (Fig.1)

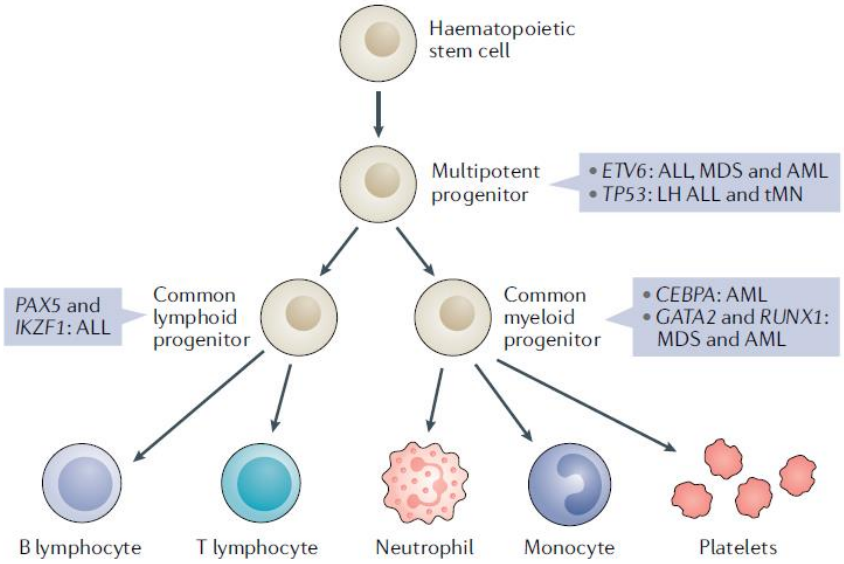


Figure 1. Association between germline mutations and hematopoietic development' hierarchy. Germline alterations in *ETV6* and *TP53* affected both lymphoid and myeloid precursors, whereas *PAX5* is limited to B-lineage. (Adapted from Klco et Mullighan, Nat Rev Cancer. 2021)[4]

The mechanisms of leukemogenesis remain uncertain: the hypothesis is that pathogenic germline variants affect key lymphoid transcription factors and perturb early stages of lymphopoiesis.[31] In this condition of instability, the activation of genes encoding for proteins with a fundamental role in cell cycle checkpoints and DNA repair mechanisms become crucial. The genotoxic damage promotes the emergence of abnormal precursor that are susceptible to the acquisition of additional somatic mutations that lead to malignant transformation.[32,33] (Fig.2)

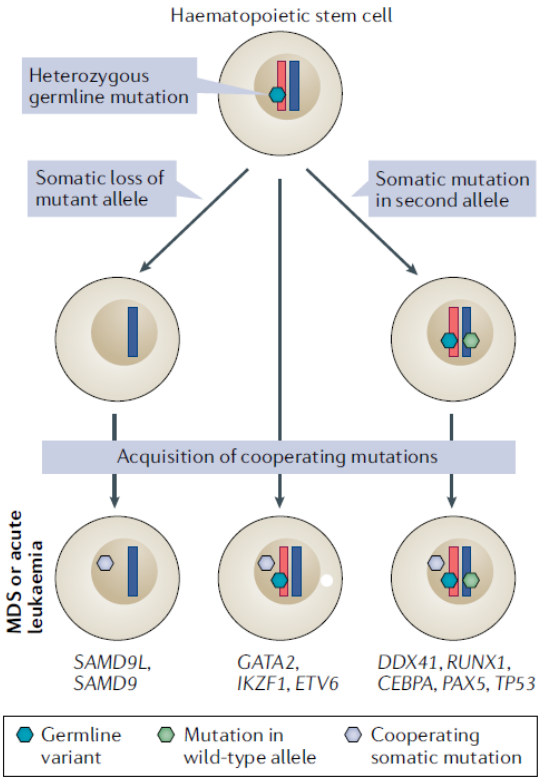


Figure 2. Model of disease development. Progression to disease in patients with myelodysplastic syndrome (MDS) or Acute Leukemia predisposition syndromes takes place through a stepwise process involving loss of the remaining wild-type allele and acquisition of additional cooperating mutations, whereas others appear to maintain the wild-type allele.
(Adapted from Klco et Mullighan, Nat Rev Cancer. 2021)[31]

1.1.3 *TP53* role in childhood ALL – Li Fraumeni Syndrome

Li Fraumeni Syndrome (LFS) is one of the major predisposing conditions, indeed it is a genetic disorder known to be associated with an increased risk for several childhood- and adult-onset malignancies,

including hematological malignancies.[34] The lifetime risk of developing cancer in individuals affected with LFS is ≥70% for men and ≥90% for women. Therefore, clinical surveillance of LFS patients can remarkably improve overall survival (OS), in terms of early tumor detection and reduction of cancer and treatment-related morbidity and mortality.[35,36] Recently, *TP53* germline aberrations, responsible for Li Fraumeni syndrome, have been described in approximately 30-40% of children with low-hypodiploid ALL.[37]

We recently analyzed retrospective cohort of 40 hypodiploid ALL patients. In 20/40 (50%) of the patients we detected a *TP53* variant, that was found to be germline in 13/20 (65%) patients. Among the cohort of patients with a mutated *TP53*, we found a higher incidence of second tumors and a higher frequency of first- or second- degree relatives with a history of cancers in young age (Bettini et al., unpublished data).

Moreover, somatic mutations in *TP53* are the most frequent alterations in cancer. They occur in almost 60% of tumors[38] (10% in AML, 19% in MDS, 15% in ALL, 8% in CLL, 20% in B-cell lymphoma and 13% in myeloma[38–40]) and correlate with aggressive tumors' phenotypes, resistance to therapies and worse overall outcomes. [41]

The role of *TP53* in oncogenesis depends on its biological function: the gene, commonly called "the guardian of the genome", is induced by cellular stress, and, in response, it orchestrates an anti-proliferative

answer through different mechanisms (such as senescence, apoptosis, cell cycle arrest, DNA repair).[42]

During cellular stress, the nuclear transcription factor p53 binds as tetramer to DNA recognition sequences, affecting the transcription factors of its target genes. Thereby, it can guide cell fate outcomes towards survival (DNA repair and cell cycle arrest) or cell death (apoptosis).[43] p53 loss of function alters the normal cell cycle, abrogates checkpoints and apoptosis, generating advantages for carcinogenesis.

Almost 70% of TP53 mutations occur in the DNA Binding Domain and modify its conformation, altering p53 capability to bind and regulate expression of its target genes. Only 10% of *TP53* mutations are loss of function (nonsense, frameshift or deletions), whereas the majority are missense variants.[44]

1.2 Predisposition in Myelodysplastic syndrome and Acute Myeloid Leukemia

1.2.1 Myelodysplastic syndrome (MDS) and Acute Myeloid Leukemia (AML)

Myelodysplasia or Myelodysplastic Syndromes (MDS) are a heterogeneous group of clonal hematopoietic neoplasms characterized by ineffective and dysplastic hematopoiesis, leading to

peripheral blood cytopenia.[45] They are characterized by an increased risk to evolve to Acute Myeloid Leukemia (AML).[46–48]

MDS is the most common cause of acquired bone marrow failure in adults, with an incidence in Italy of 50 cases per 100,000 individuals (70 years of age and older). In pediatric patients, myelodysplastic syndrome accounts for less than 5% of childhood hematological malignancies, with an annual incidence of 1-4 cases per million.[49]

In the last decades, in addition to cytomorphological characteristics, genetic studies have enabled enormous advances in the knowledge of pathogenesis, in term of cytogenetic and molecular landscape and, recently, as regarding genetic background. Next Generation Sequencing (NGS) approaches have allowed the characterization of the complex mutational pattern that drives disease evolution from asymptomatic clonal hematopoiesis to MDS, and, ultimately, to secondary AML.

Although the number of genes involved in MDS predisposing mutations is high, it can be classified into a limited number of subtypes, which correspond to the cellular mode of action involved: RNA splicing (*SF3B1*, *SRSF2*, *ZRSR2*, *U2AF1* genes),[50] DNA methylation (*TET2*, *DNMT3A*, *IDH1/2*), transcription regulation (*RUNX1*), signal transduction (*CBL*, *RAS*), DNA repair (*TP53*), chromatin modification (*ASXL1*, *EZH2*), and Cohesin complex (*STAG2*), as reported in Figure 3.[46]

Up to 90% of patients have been found to have a somatic mutation in

at least one gene: at diagnosis, the majority of MDS patients have 2–4 driver mutations and hundreds of background mutations.[45,46]

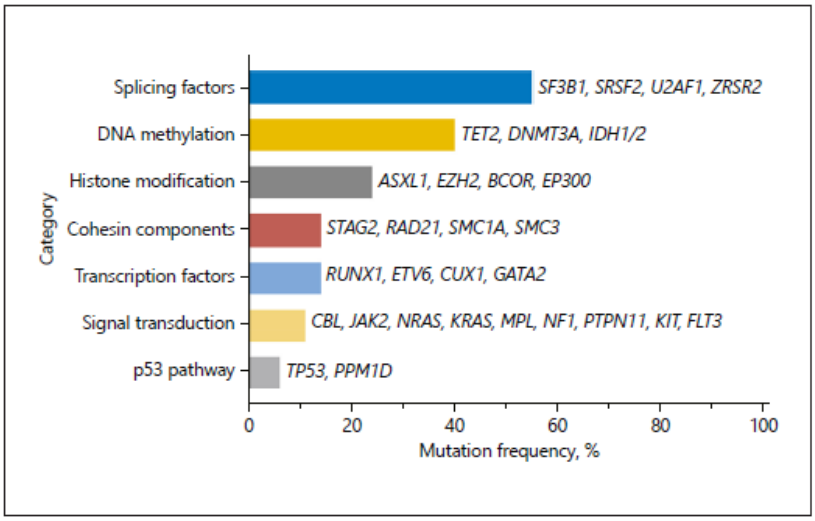


Figure 3. Classification of the subgroups of genes involved in the predisposition to myelodysplastic syndrome, according to cell function. (Adapted from Kennedy et Ebert, J Clin Oncol. 2017)[47]

In more than 20% of affected patients, MDS evolves into secondary AML, that is the most common acute leukemia in adults.[51]

It is a highly heterogeneous disease, characterized by an abnormal expansion and accumulation of clonal myeloid stem cell population.

It is a result of large chromosomal translocations and mutations in genes involved in hematopoietic proliferation and differentiation.

Although patients can be stratified into favorable, intermediate, and adverse risk groups based on their cytogenetic profile, the prognosis within these categories varies widely. The identification of recurrent

genetic mutations, such as *FLT3-ITD*, *NMP1* and *CEBPA*, helped to refine the individual prognosis and guided its management.[52]

Pathogenesis

AML can arise in patients affected by blood disorder or as a consequence of previous therapies (secondary AML). However, in most cases it arises as a *de novo* neoplasm in previously healthy individuals. As regards etiology, chromosomal translocations alter the normal maturation process of myeloid precursor cells. In addition to large chromosomal rearrangements, genetic mutations have been identified in more than 97% of cases, often in the absence of any large chromosomal abnormalities.[53,54]

1.2.2 Genetic landscape and hereditable susceptibility

The crucial role of clonal hematopoiesis and germline syndromes (such as Shwachman-Diamond Syndrome and Diamond-Blackfan Anemia)[55] is now well established in increased risk of MDS/AML. Recent advances have underlined a growing number of non-syndromic familiar MDS predisposition syndrome, caused by mutations in several genes that appear to increase the germline susceptibility even in absence of other clinical manifestations.[56–58] (Tab.1) Although it was a common notion that germline genetic predisposition regards only children and young adults, nowadays it has been demonstrated that it is involved also in leukemogenesis of older adulthood.[59]

To date, several single-gene loci have been classified, when mutated, to predispose to an increased risk of primary MDS and/or AML. Except for *CEBPA* germline mutations, that appear to confer an increased risk only for AML, the others overlap in their associated risks of MDS, AML and thrombocytopenia.[60]

All these conditions require specific clinical practice in term of management and surveillance, that can guarantee better results in in diagnosis and prognosis.[60,61]

Syndrome	Gene	Inheritance	Heme Malignancy
Familial platelet disorder with propensity to myeloid malignancies	*RUNX1*	AD	MDS/AML/T-cell ALL
Thrombocytopenia 2	*ANKRD26*	AD	MDS/AML
Familial AML with mutated DDX41	*DDX41*	AD	MDS/AML, CMML
Thrombocytopenia 5	*ETV6*	AD	MDS/AML, CMML, B-cell ALL, multiple myeloma
Familial MDS/AML with mutated GATA2	*GATA2*	AD	MDS/AML/CMML
Familial aplastic anemia with SRP72 mutation	*SRP72*	AD	MDS
Familial AML with mutated CEBPA	*CEBPA*	AD	AML
Fanconi anemia	Complementation Groups	AR, X-linked	MDS, AML
Telomeropathies (dyskeratosis congenita)	*TERC*, *TERT*, others	AD, AR	MDS/AML

AD, Autosomal dominant; MDS, myelodysplastic syndrome; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; CMML, chronic myelomonocytic leukemia; AR, autosomal recessive.

Tab.1 Genes frequently mutated in familial myelodysplastic syndromes (MDS)/acute leukemia (AL) predisposition syndromes (Adapted from Bannon et DiNardo, Int. J. Mol. Sci. 2021)[60]

RUNX1 is another well characterized gene that causes a rare form of familiar thrombocytopenia (Familiar Platelet Disorder - FPD). Patients affected with FPD show an increased risk of hematological malignancies.

Recent data suggest that clonal hematopoiesis can be detected in >80% of asymptomatic FPD/AML individuals by age 50, with an MDS/AML transformation rate estimated between 20-60%.

Germline mutations of this gene result mainly in a premature protein truncation or are missense variants that affect DNA-binding domain.

The second one is *ANKRD26*, responsible of the *ANKRD26*-related thrombocytopenia. Its mutations are often located at 5'-3' UTR regions; in presence of pathogenic variants, the risk of patients to develop MDS/AML is increased at 4.9% for leukemia and 2.2% for MDS.[62]

DDX41 is one of the most recently gene involved in predisposition. Germline mutations of the DEAD-Box elicase, that leads to altered RNA splicing, cause the so called *DDX41*-associated familial MDS/AML syndrome, where myeloid neoplasms are usually diagnosed in older age (> 60 years of age). For this reason, it may be clinically difficult to distinguish patients with a germline predisposition from those with de

novo MDS/AML. An acquired somatic mutation in the wild-type *DDX41* allele is very frequent in onset of MDS.[60]

Although germline *ETV6* mutations are most frequently associated with pediatric-BCP ALL, they have been observed in additional hematologic malignancies, including MDS and AML, as well as colon-rectal cancer. For this reason, it has been hypothesized that contribute to a more general cancer predisposition syndrome. These mutations, mostly distributed in the DNA binding domain, cause a loss of normal transcriptional repression by *ETV6* and they are responsible of the 'ETV6-associated familial thrombocytopenia and hematologic malignancy'.[28]

Regarding *GATA2*, germline mutations lead to *GATA2* deficiency, a complex multi-systemic disorder in which the risk of MDS/AML is significantly increased approximately at 70% with age of onset about 29 years. The abnormal clonal hematopoiesis is a common event in symptomatic germline mutated *GATA2* patients with MDS, possibly indicating a pre-MDS stage.[56]

Germline mutations in the ribonucleoprotein complex gene *SRP72* have been identified as a rare cause of SRP72-associated MDS. Considering the rarity of these germline variants, only few notions are known about incidence and risk.

Finally, germline *CEBPA* mutations are commonly frameshift or nonsense variants that cause the development of AML with a near complete penetrance, frequently associated with an acquired

mutation in the remaining wild-type allele. Germline mutations are present in about 10% of AMLs with biallelic mutations in *CEBPA* and correlate with a good outcome. Despite mutated patients are prone to second leukemias, they remain sensitive to chemotherapy.[63,64]

Recently, Whole Exome Sequencing (WES) analysis on familial myeloid malignancies allowed the identification of several variants at novel loci, such as *DHX34*.[65]

1.2.3 Clonal hematopoiesis

Clonal Hematopoiesis (CH) refers to an expansion of blood or marrow cells resulting from acquired somatic mutations in leukemia-associated genes detected in healthy individuals. These alterations commonly develop during the hematopoietic system aging, and in most cases, they do not alter stem cell function. In these cases, the phenomenon is defined as benign age-related clonal hematopoiesis (ARCH).[66] Despite this, in some cases these alterations are able to confer a competitive advantage over normal HSCs and cause preferential contribution to mature hematopoietic stem cells (HSC).[67,68] For this reason, CH is a strong predictor of increased risk to evolve into MDS/AML, preceding the disease by many years and conferring an increasing associated risk with the clonal complexity. It represents a premalignant state that can be influenced by germline genetic context (typical of predisposition in younger adults), but it can

also be triggered by exogenous events, even in absence of germline variants.[69,70] (Fig.4)

Several studies have demonstrated that CH is associated with an approximately 10-fold risk increase of developing a hematologic malignancy, with equal proportions of lymphoid and myeloid neoplasms. However, absolute risk of progression of CH to hematologic malignancy is about 0.5% to 1% per year over a follow-up period of 5 to 10 years, and most individuals with CH will not never develop MDS/AML.[68] Several aspects may have a role in the increased risk of malignant transformation, latency and type of blood cancer, such as specific gene mutations, co-mutational patterns and clone sizes. For example, the presence of two or more variants is associated with higher risk. Individuals with a higher mutational complexity is more prone to develop hematological malignancies than age-matched controls.[71–73]

Moreover, the transition from CH to hematologic malignancies involves a complex cross-link between (epi)-genetic alterations of HSC, dysfunctional bone marrow microenvironment and the stepwise acquisition of additional driver mutations.[46]

Depending on the specific mutated genes, the presence of CH is associated with increased risk of both *de novo* and therapy-related AML. Older patients are frequently affected by CH somatic mutations in genes encoding epigenetic modifiers (*TET2*, *DNMT3A*), DNA repair

(*TP53)*, cohesin genes or RNA splicing genes (in particular *SRSF2* and *SF3B1*) and they are responsible of newly diagnosed AML.[53,67,74–76]

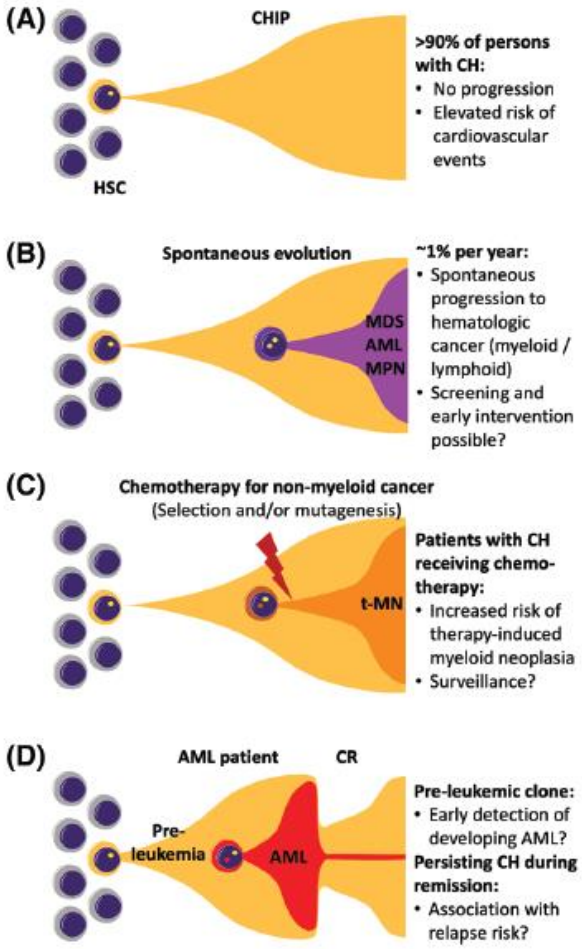


Fig.4 Significance of clonal evolution in various clinical scenario. (Adapted from Hartmann et Metzeler, GenesChr Cancer 2019)[68]

1.2.4 Spliceosome mutations in Clonal Hematopoiesis

Mutations in genes encoding core spliceosomal protein are the most common recurrent lesions in MDS (up to 60% of cases).[77]

They also occur very frequently in AML (*de novo* AML: 10%, secondary AML 10–55%) and comprise 5% of CH mutations.[78]

These heterozygous mutations affect component of the 3' spliceosome, including *SRSF2*, *U2AF1*, *SF3B1* and to a lesser extent *ZRSR2*, leading to an altered function of the splicing machinery. This abnormal splicing is responsible of transcription instability and improper translation, leading to a dysregulation of several transcription factors, crucial in hematopoiesis.[50,66,69]

Among spliceosome genes, *SF3B1* is the most frequently mutated, representing 50% of all mutant cases. It is affected by several aberrations located preferentially in HEAT domains, that correlate to a good prognosis.

SRSF2 is the second most commonly mutated gene (14% of MDS cases and 25% of AML), followed by *U2AF1* (10-15% of patients with MDS and in 4% of patients with AML). Mutations in both genes are associated with an increased risk of AML transformation, independently of clone size,[46] and they correlate with a worse survival.

ZRSR2 mutations affect about 5% of patients with MDS, predominantly males, causing abnormal splicing via intron retention. They have adverse prognostic effect.[79,80]

The incidence of mutations in spliceosome genes, together with the well-known role of Clonal Hematopoiesis, support the 'genetic predestination' hypothesis.[81]

Aberrations affecting spliceosome genes (particularly *SF3B1* and *SRSF2*) are typically early driver mutations in MDS, dictate clonal expansion and future trajectories of disease evolution and correlate with distinct clinical phenotypes. Overall, they have a prognostic significance, and worsening of prognosis correlates with increasing mutational events. Clonal and subclonal mutations equally affect prognosis.[81]

2. Cohesin complex

Cohesin ring is an evolutionary conserved multi-protein complex. In all Eukaryotic organisms, it consists of four core subunits: two subunits SMC (*Structural Maintenance of Chromosomes*), *SMC1A*, *SMC3*, and two subunits SCC (*Sister Chromatid Cohesion*), either *STAG1* or *STAG2* and *RAD21*. The last one is also known as 'double-strand-break repair protein'.[82] (Fig.5)

SMC1 and SMC3 interact each other creating a heterodimer, instead RAD21 bridges both SMC subunits by binding SMC3 through its N-terminal part and by binding SMC1 via its C-terminus. This ring-shape structure interacts with several additional components that regulate its functions and have the capability to encircle chromatin without a direct DNA-binding contact.[83,84]

The canonical role of the complex is related to chromatin cohesion and chromosomes segregation. It maintains the cohesion of sister chromatids from the S-phase until the onset of anaphase, to ensure an equal segregation into the two daughter cells.[85]

It has also a crucial role in the DNA stability and damage response.[86]

Moreover, scientific evidence has recently underlined that Cohesins directly regulate transcription of genes involved in cell proliferation, pluripotency, and differentiation. The involvement of the ring in regulation of gene expression is a more recent discovery and is defined as non-canonical role.[87]

Germline mutations in the genes encoding the core cohesin subunits are implicated in several human developmental disorders, known as 'Cohesinopaties'.[88,89] Among those, Cornelia De Lange Syndrome (CdLS) is the most common.[90,91]

As regards somatic events, cancer genomics analyses have discovered a high frequency of mutations in Cohesins genes, as well as in genes encoding for regulatory factors, in a subset of human tumors including Acute Myeloid Leukemia (AML).[85]

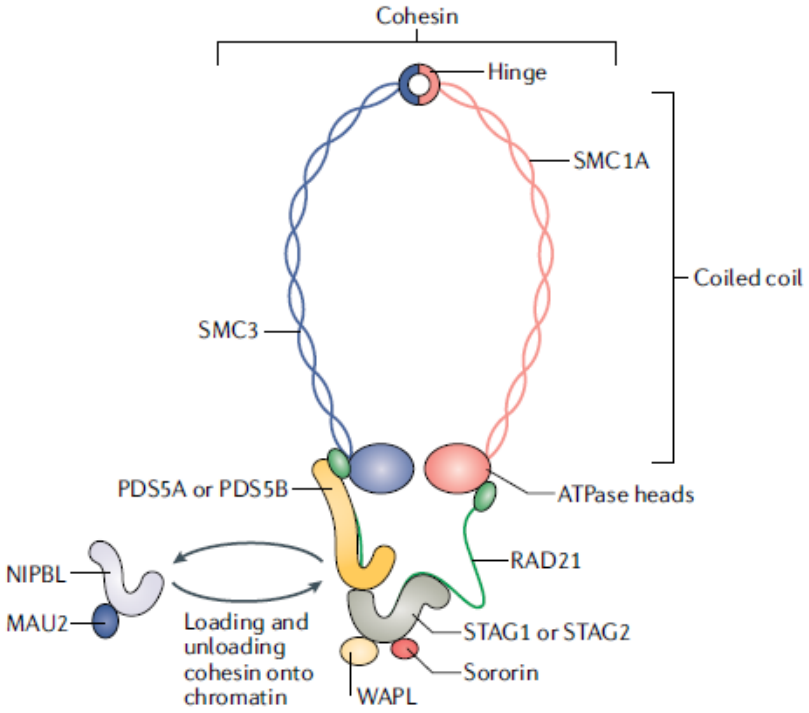


Fig.5 Schematic representation of principal proteins involved in the Cohesin complex. (Adapted from Waldman, Nat. Rev. Cancer 2020)[82]

2.1 Canonical role of the Cohesin complex

2.1.1 Chromatid segregation

Several cohesin regulatory factors are responsible of loading, stability, and cleavage of the complex on chromatin during cell cycle.

In details, cohesin ring loads onto chromatin in G1 phase of the cell cycle, immediately after cytokinesis, and remains bound specifically to

centromeres in prophase. The loading is dependent on the NIPBL-MAU heterodimer (ATPase-dependent manner), and it is promoted by the WAPL, PDS5A and PDS5B proteins, that bind to chromatin-ring complex.[82,92] The acetylation of SMC3 by the acetyl-transferases ESCO1 and ESCO2, as well as the binding of CDCA5 (Soronin), stabilizes the strong interaction and allows the establishment of sister chromatid cohesion during DNA replication in S phase.[93]

Also the activity of STAG proteins promotes the maintenance of the structure: STAG2 subunit is essential for chromatid cohesion at centromeres and along chromosome arms, while STAG1 subunit is essential for chromatid cohesion specifically at telomeres.[94]

At the onset of mitosis in early prophase, phosphorylation of STAG2 by PLK1 drives dissociation of the majority of cohesins along chromosome arms[95], while centromeric cohesion is guaranteed by binding of SGOL1 (Shugoshin). These mechanisms confer their classic X-shape to metaphase chromosomes.

Activation of the anaphase leads to degradation of PTTG1 (Securin) and to activation of ESPL1 (Separase), that cleaves the RAD21 subunit of the remaining chromatin-bound cohesion. Cohesin ring and the centromeric cohesion are so cleaved, allowing sister chromatids to snap apart at the metaphase-to anaphase transition. Thereby chromatids can be separated into daughter cells (Fig.6).[85,96,97]

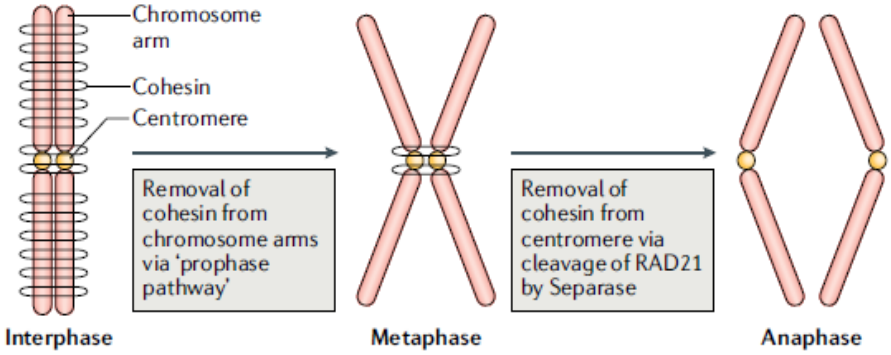


Fig.6 Localization of Cohesins during separation of sister chromatids- Canonical model of cohesin action. (Adapted from Waldman, Nat. Rev. Cancer 2020)[82]

2.1.2 DNA Damage Repair

Among the Cohesin functions, that are crucial to preserve genome integrity, their ring guarantees the correct progression of cell cycle, and it is required for G1, intra-S and G2–M DNA Damage Checkpoints (DDC).[98]

Double strand breaks (DSB) induction leads to Cohesins accumulation near the break site, where the ring is responsible of recruitment and activation of checkpoint/DNA repair proteins. The NIPBL-MAU complex allows Cohesins enrichment around damage site, where the complex activates a network of DNA repair mechanisms to translate checkpoint signals into DNA repair alarms, delaying the progression of cell cycle until the integrity of the double strand is re-established.

First, it promotes an efficient repair by homologous recombination (HR), providing a stable template to the sister chromatids. During S-phase, SMC subunits are phosphorylated by the damage marker ATM,[99] thus the ring limits the synthesis of damaged DNA in order for the replication forks to stall. In G2/M phase, it also prevents the premature entry in mitosis under DNA damage conditions and promotes the formation of sister chromatids junctions. Chromatids are held in close proximity of stalled replication forks, in order to allow timely and efficient resumption and completion of DNA replication.[98]
Finally, it suppresses damage-induced recombination and prevents joining of distal DSBs, responsible of oncogenic chromosomal aberrations.[100,101] (FIG.7)
In support of these data, Watrin et al. demonstrated that cohesin-depleted cells are characterized by a higher number of spontaneous DNA damage events, expression of an increase in activated forms of ATM, CHK1, and H2AX, common DSB markers.[102]
The complex contributes also to the structure of irradiation-induced loci.[103]

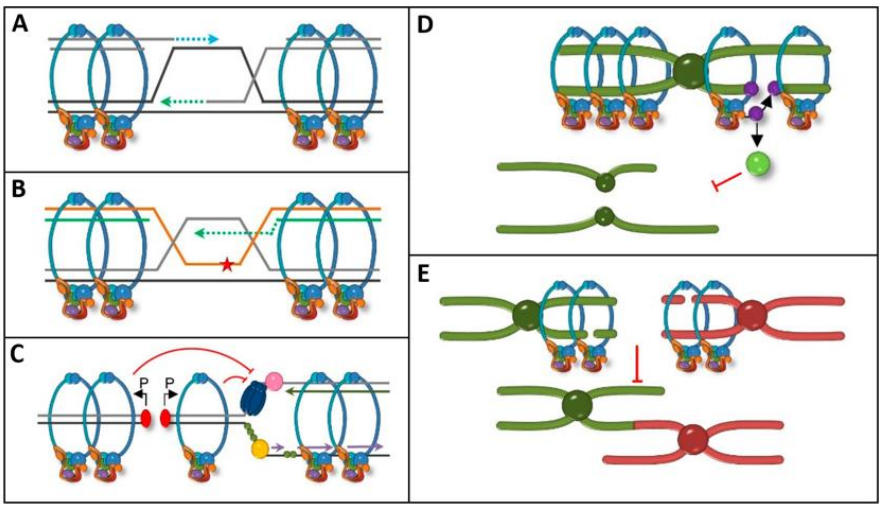


Fig.7 Cohesin functions in DNA damage response. (A) DSB repair by HR; (B) Template switch-mediated gap filing; (C) Inhibition of DNA synthesis in response to DBSs in S-phase; (D) Blocking chromatid separation through G2/M checkpoint activation; (E) Inhibition of chromosome fusion.
(Adapted from Litwin, Pilarczyk et Wysocki, Genes. 2018)[98]

In summary, Cohesins' alterations may compromise chromosomes integrity, increasing the risk of genome instability, that is recognized as an oncogenesis promoting factor.[100,101]

2.2 Non-canonical role of the Cohesin complex: regulation of gene expression

Among the several functions that Cohesin ring carries out, regulation of gene expression is certainly the most complex. Although the mechanisms are not yet fully understood, it is now known that this function is independent of the role in cohesion of sister chromatids. Thanks to their capability to contribute to chromatin architecture,

Cohesins directly regulates transcription of genes involved in cell proliferation, pluripotency and differentiation.[104]

The ring interacts with the CTCF binding factor and other proteins involved in regulation of genes expression, forming, and stabilizing specific topological loops. Thus, it defines spatial conformation of specific loci,[105] allowing long-range interactions between cohesin binding sites and defining communications between enhancers and promoters that reflect specific expression pathways.

Additional evidence shows that CTCF and cohesin contribute differentially to chromatin organization;[106] CTCF brings chromosome loci closer and then Cohesins bind them together and entrap the loop in rings, stabilizing long-range interactions and facilitating transcription, as depicted in Figure 8. Moreover, the ring is also recruited in an CTCF-independent manner, to bind target genes and promote gene transcription.[104,107]

It has been demonstrated that Cohesins depletion is involved in loss of long-range contacts, extensive decompaction of large-scale domains, loss of intra-domain contacts and deregulation of gene expression;[108] in support of this evidence, NIPBL-mutated cells are characterized by a large number of dysregulated genes.[109,110]

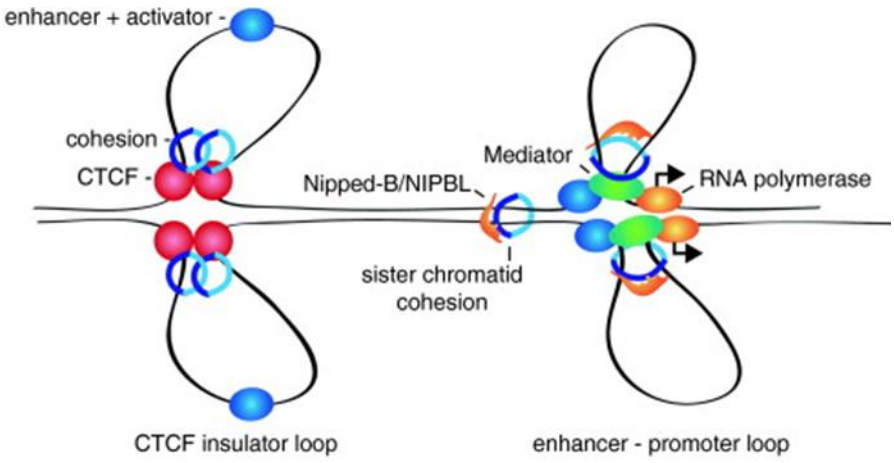


Fig.8 Cohesins capability to facilitate DNA looping.
(Adapted from Dorsett, CurrOpin Genet Dev. 2011)[87]
The diagram illustrates two sister chromatids. On the left, Cohesins support intrachromosomal looping between two CTCF binding sites; the loop functions as an insulator and sequesters a transcriptional enhancer. On the right, the ring stabilizes a loop between an enhancer and promoter, facilitating transcriptional activation. Mediator, a transcriptional coactivator, forms a complex with Cohesin and the cohesin loading factor NIPBL loads the complex at promoters. Mediator and Cohesin occupy different promoters in different cells, thus generating cell-type specific DNA loops linked to the gene expression program of each cell.[111]

2.3 Cohesin genes in genetic syndromes: germline mutations

Mutations in the cohesin complex (both structural and ancillary cohesin genes) cause a multispectrum developmental abnormalities, named “cohesinopathies”. This group of conditions historically included Cornelia de Lange syndrome (OMIM #122,470, # 300,590, # 610,759, # 614,701 and # 300,882), Roberts syndrome (OMIM #

268,300) and Warsaw-Breakage syndrome OMIM #613,398). However, recently, new phenotypes and clinical entities have been described: CHOPS syndrome (OMIM # 616,368) caused by mutation in AFF4, STAG2-related X-linked Intellectual Deficiency (OMIM # 301,022) and CAID (Chronic Atrial and Intestinal Dysrhythmia) syndrome (OMIM # 616,201).[112,113]

Among them, Cornelia de Lange syndrome (CdLS) is the most characterized. It is an autosomal dominant disease, caused by mutation in both ancillary genes, such as *NIPBL* and *HDAC8*, and core cohesion genes, as *SMC1A*, *SMC3*, *RAD21*. Recently also variants in *BRD4* and *ANKRD11* were described in such patients. CdLS patients share typical facial dysmorphism, microcephaly, growth delay and major malformations, such as limb reduction and heart defects. A certain degree of intellectual disability, ranging from severe to mild, is always described.

Biological studies, performed on CdLS patients-derived cell lines and *in vivo* models, don't demonstrated mitotic defects or premature chromatid separation, but showed a dysregulation of some of the cohesin-dependent genes.[114,115]

2.4 Cohesin genes in hematological malignancies

2.4.1 Somatic mutations in Myelodysplastic syndromes (MDS) and Acute Myeloid Leukemia (AML)

Cohesin genes are frequently affected by somatic events in cancer.[82,116] Alterations in the genes encoding the core cohesin subunits or its regulatory factors have been reported in several tumors, included myelodysplastic syndrome (MDS) and Acute Myeloid Leukemia (AML). They occur with high frequency in patients with myeloid neoplasms (12% of cases)[117,118] (Fig.9), where they are often mutually exclusive and lead to decreased function of the Cohesin complex.[119]

Cohesin defects are most prevalent in high-risk MDS and secondary AMLs and are associated with poor overall survival, especially in *STAG2* mutant MDS patients.

Those somatic variants are nonsense and frameshift aberrations that occur early in disease development, in expanding subclones, and co-occur with other mutations known to be drivers of clonal evolution. Analysis for clonal hierarchy performed by Thota et al.[118] demonstrated that mutations in *STAG2*, *SMC3*, and *RAD21* are often ancestral, and they expand to clonal dominance concordant with disease progression. They are often responsible of dysfunction of the checkpoint proteins (as MAD2 and/or BUBR1), with consequent chromosome segregation and DNA repair transaction defects,

exacerbating the genomic instability commonly associated with different type of cancers.[101]

Cohesin mutations are early, but not initiating, genetic lesions during myeloid disease development. They give a clonal advantage and facilitate a positive mutational selection, predetermining the types of additional secondary mutations that result in evident leukemic transformation.[120]

Somatic mutations of the Cohesins alone are often insufficient to impart complete malignant transformation. They act in epigenetic regulation with co-occurring mutations in chromatin modifiers (as *ASXL1)*; however, they have effect on gene expression dysregulation if simultaneously present with aberrations in key transcriptional regulators (*RUNX1*, Ras family genes, and *BCOR*).[118,121,122]

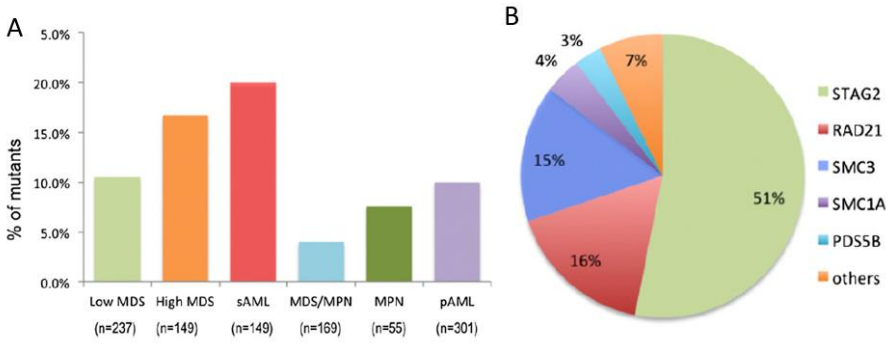


Fig.9 Characterization of Cohesin mutations in patients with a myeloid disease. (A) Frequency of Cohesin mutations in each myeloid malignancy in Thotas' representative cohort: 10.5% of lower-risk MDS patients (25/237),

16.8% of High risk MDS patients (25/149), 20.1% of secondary AML (30/149), 4.1 % of MDS/MPN (7/169), 7.3% of MPN (4/55), 10.6% of primary AML (32/301). (B) Distribution of Cohesin mutations identified across the patient cohort. (Thota et al, Blood.2014)[118]

Therefore, cohesin defects resulted in alteration in chromatin architecture and deregulated expression of genes involved in myeloid development and differentiation, with enhanced effects on self-renewal of hematopoietic stem and progenitor cells. It has been recently demonstrated that depletion of Cohesins severely impairs the expression of *ETV6*, a key transcription factor in self-renewal programs: the failure activation of its repressor Ets abrogates induction of erythroid transcriptional programs of differentiation.[83,123]
So far, somatic Cohesin mutations have not been associated with aneuploidy or complex cytogenetics.[118]
Due to this line of evidence, targeting of cohesin complex is a promising area of drug development, still underexplored.[124]

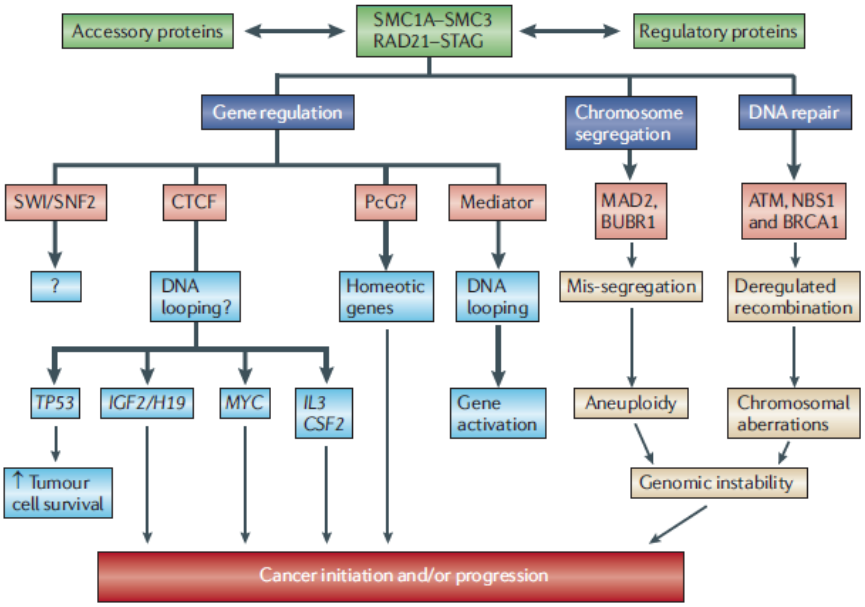


Fig.10 Examples of cancer pathways in which Cohesin and its regulatory and accessory proteins are implicated. Defects in different cohesin functions might contribute to cancer initiation and/or progression. The crosstalk with transcriptional mediators or transcriptional repressors (such as CTCF) influencing lobal gene expression in human cancers. Genes potentially regulated by CTCF–cohesin include proto-oncogenes (as *MYC)*, tumor suppressor gene (as *TP53)*. Deregulation can lead to homeotic gene dysfunction and oncogenesis.
(Adapted from Xu, Tomaszewski et McKay, Nature Reviews. 2011)[101]

2.4.2 Potential role of germline mutations in predisposition to hematological disease

To date, a direct cause-effect correlation between genetic syndromes and oncogenesis is not established yet; in addition, no evidence is

sufficiently significant to define Cohesinopaties cancer-prone syndromes.

Despite this fact, in the last decades, multiple case reports supported a role of Cohesins' germline mutations in predisposition to neoplasms development.

A 23-month-old child affected by Roberts syndrome that developed a sarcoma botryoides,[125] in addition with a case of melanoma in a girl with Roberts-SC Phocomelia Syndrome,[126] suggested a possible correlation between Cohesinopaties and increased risk of malignancy. In hematological field, Vial et al.[127] hypothesized for the first time that germline mutations in Cohesins could constitute a predisposing factor to leukemia. They described a single case of Down syndrome-like Acute Megakaryoblastic Leukemia (AMKL) in a patient with Cornelia de Lange syndrome (CdLS), in which a pre-leukemic clone combines a constitutional *NIPBL* mutation with somatically acquired trisomy 21 and *GATA1* mutation.

Moreover, our group recently described the first CdLS patient with ALL,[128] carrying a *NIBPL* mutation, as described in State of the art 1 paragraph. The analysis of the family indicated a *de novo* origin of this novel deleterious variant.

In support of the connection between Cohesinopaties and Leukemia, in a large cohort of children with CdLS cancer accounted for 2% of deaths, highlighting a slightly increased of cancer risk compared to the healthy population.[129]

Genetic mechanisms through which germline Cohesin mutations could perturb hematopoietic development are not clear yet. Considering the role of the complex in gene expression and DNA repair, loss of Cohesins functions might lead to genetic instability in progenitor cells, that become more susceptible to DNA damage and leukemic transformation. The underlying processes still need to be established and further explored.

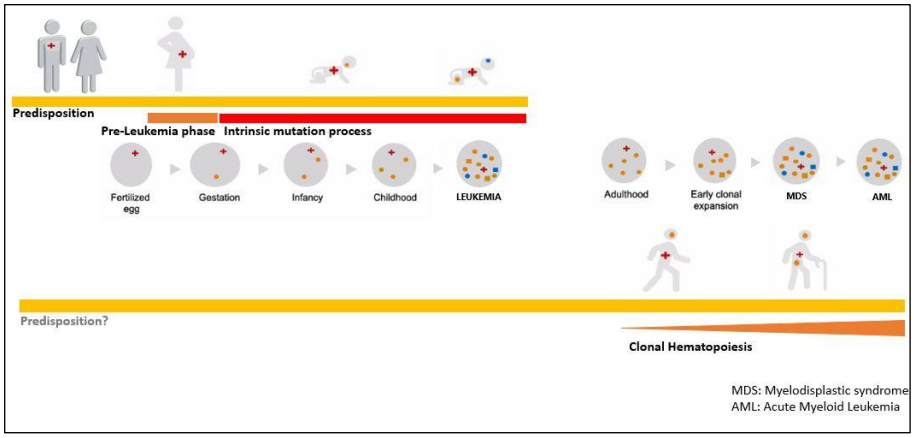


Figure 11. Graphical abstract representing the influence of Predisposition and Clonal Evolution across time of life.
(Adapted from Illumina)

State of the art

State of the art 1

J Clin Pathol. 2019 Aug;72(8):558-561.

First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia

Grazia Fazio,[1] Valentina Massa,[2] Andrea Grioni,[1,3] Vojtech Bystry,[3] Silvia Rigamonti,[1,2] **Claudia Saitta,[1]** Marta Galbiati,[1] Carmelo Rizzari,[4] Caterina Consarino,[5] Andrea Biondi,[1,4] Angelo Selicorni,[6] Giovanni Cazzaniga[1,7]

Abstract

Cornelia de Lange syndrome (CdLS) is a rare autosomal dominant genetic disorder characterized by prenatal and postnatal growth and mental retardation, facial dysmorphism and upper limb abnormalities. Germline mutations of cohesin complex genes *SMC1A*, *SMC3*, *RAD21* or their regulators *NIPBL* and *HDAC8* have been identified in CdLS as well as somatic mutations in myeloid disorders. We describe the first case of a paediatric patient with CdLS with B-cell precursor Acute Lymphoblastic Leukaemia (ALL). The patient did not show any unusual cytogenetic abnormality, and he was enrolled into the high risk arm of AIEOP-BFM ALL2009 protocol because of slow early response, but 3 years after discontinuation, he experienced an ALL relapse. We identified a heterozygous mutation in exon 46 of *NIPBL*, causing

frameshift and a premature stop codon (RNA-Targeted Next generation Sequencing Analysis). The analysis of the family indicated a de novo origin of this previously not reported deleterious variant. As for somatic cohesin mutations in acute myeloid leukaemia, also this ALL case was not affected by aneuploidy, thus suggesting a major impact of the non-canonical role of *NIPBL* in gene regulation. A potential biological role of *NIPBL* in leukaemia has still to be dissected.

State of the art 2

J Clin Oncol. 2021 Apr 10;39(11):1223-1233.

Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes

Matteo Bersanelli, PhD[1,2]; Erica Travaglino, BSc[3]; Manja Meggendorfer, PhD[4]; Tommaso Matteuzzi, PhD[1,2]; Claudia Sala, PhD[1,2]; Ettore Mosca, PhD[5]; Chiara Chiereghin, PhD[3]; Noemi Di Nanni, PhD[5]; Matteo Gnocchi, MSc[5]; Matteo Zampini, PhD[3]; Marianna Rossi, MD[3]; Giulia Maggioni, MD[3,6]; Alberto Termanini, PhD[3]; Emanuele Angelucci, MD[7]; Massimo Bernardi, MD[8]; Lorenza Borin, MD[9]; Benedetto Bruno, MD[10,11]; Francesca Bonifazi, MD[12]; Valeria Santini, MD[13]; Andrea Bacigalupo, MD[14]; Maria Teresa Voso, MD[15]; Esther Oliva, MD[16]; Marta Riva, MD[17]; Marta Ubezio, MD[3]; Lucio Morabito, MD[3]; Alessia Campagna, MD[3]; **Claudia Saitta, MSc[18]**; Victor Savevski, MEng[3]; Enrico Giampieri, PhD[2,19;] Daniel Remondini, PhD[1,2]; Francesco Passamonti, MD[20]; Fabio Ciceri, MD[8]; Niccol `o Bolli, MD[21,22;] Alessandro Rambaldi, MD[23]; Wolfgang Kern, MD[4]; Shahram Kordasti, MD[24,25]; Francesc Sole, PhD[26]; Laura Palomo, PhD[26]; Guillermo Sanz, MD[27,28]; Armando Santoro, MD[3,6]; Uwe Platzbecker, MD[29]; Pierre Fenaux, MD[30]; Luciano Milanesi, PhD[5]; Torsten Haferlach, MD[4]; Gastone Castellani, PhD[2,19]; and Matteo G. Della Porta, MD[3,6]

Abstract

Purpose. Recurrently mutated genes and chromosomal abnormalities have been identified in myelodysplastic syndromes (MDS). We aim to integrate these genomic features into disease classification and prognostication.

Methods. We retrospectively enrolled 2,043 patients. Using Bayesian networks and Dirichlet processes, we combined mutations in 47 genes with cytogenetic abnormalities to identify genetic associations and subgroups. Random-effects Cox proportional hazards multistate modeling was used for developing prognostic models. An independent validation on 318 cases was performed.

Results. We identify eight MDS groups (clusters) according to specific genomic features. In five groups, dominant genomic features include splicing gene mutations (*SF3B1*, *SRSF2*, and *U2AF1*) that occur early in disease history, determine specific phenotypes, and drive disease evolution. These groups display different prognosis (groups with *SF3B1* mutations being associated with better survival). Specific co-mutation patterns account for clinical heterogeneity within *SF3B1*- and *SRSF2*-related MDS. MDS with complex karyotype and/or *TP53* gene abnormalities and MDS with acute leukemia–like mutations show poorest prognosis. MDS with 5q deletion are clustered into two distinct groups according to the number of mutated genes and/or presence of *TP53* mutations. By integrating 63 clinical and genomic variables, we define a novel prognostic model that generates

personally tailored predictions of survival. The predicted and observed outcomes correlate well in internal cross-validation and in an independent external cohort. This model substantially improves predictive accuracy of currently available prognostic tools. We have created a Web portal that allows outcome predictions to be generated for user-defined constellations of genomic and clinical features.

Conclusion. Genomic landscape in MDS reveals distinct subgroups associated with specific clinical features and discrete patterns of evolution, providing a proof of concept for next-generation disease classification and prognosis.

Scope of the thesis

The aim of this study is to dissect the role of predisposition in pediatric hematological malignancies, that occurs in 5-10 % of pediatric cancer, but that is still largely uncharacterized. Among several genes, the Cohesins and the role of their germline variants in promoting cancer-prone conditions were the main focuses of the present study.
In order to investigate the genetic predisposition in cancer, considering the overall time of life, the secondary objective of the work is to investigate how similarly clonal evolution acts in the adult setting, since age-dependent accumulation of somatic mutations increases prevalence of Myelodysplastic syndrome (MDS) among older individuals.

With these purposes, this PhD thesis has been developed in 4 different tasks:

1. Molecular genetic profiling of B-cell precursor Acute Lymphoblastic Leukemia (BCP-ALL) pediatric cohort - the promising role of germline mutations in Cohesin genes.

We identified germline and somatic variants that characterized pediatric children affected by BCP-ALL, assessing the frequency, and classifying them according to pathogenicity prediction. Among 39 genes known to be involved in leukemogenesis, we focused our

attention on germline Cohesin variants, to dissect their promising role in leukemic predisposition.

2. Role of recurrent germline variants in Cohesin *RAD21* in predisposing children to Lymphoblastic Leukemia and Lymphoma.

We characterized a recurrent RAD21 germline variant (Ser298Ala) identified in three children with lymphoblastic leukemia/lymphoma. We collected cancer familiar history and evaluated co-occurrence mutations in these patients, to define the penetrance of the variant and the increased cancer risk across generations. We investigated the effects of *RAD21* variant on gene expression and DNA damage response, in order to evaluate if it could alter mechanisms involved in oncogenesis even in a pre-disease phase.

3. Potential Role of *STAG1* Mutations in Genetic Predisposition to Childhood hematological malignancies.

We investigated the effects of two rare STAG1 germline variants (Arg1167Gln and Arg1187Gln) in two pediatric patients affected by Acute Lymphoblastic Leukemia (ALL) and Myelodysplastic syndrome (MDS), respectively. We characterized their position along gene sequence and protein domains, considering conservation and mutational landscape. Moreover, we dissected the functional consequences of MDS *STAG1* mutation in a lymphoblastoid cell line

(LCL), an *in vitro* preclinical cell model, which allowed the evaluation of chromosomal stability and DNA repair mechanisms. They commonly represent defective processes in oncogenesis and, thus, are essential to evaluate how the variant might contribute to tumor transformation.

4. Clonal hematopoiesis leads to different clinical diseases in the oldest-old population – the role of Splicing genes mutations.

We characterized specific mutational patterns in a healthy cohort of over eighty, describing different risk of developing inflammatory-associated diseases or myeloid neoplasms, even in individuals with unexplained cytopenia. We focused our attention on Splicing genes mutations, the highest predictive value for myeloid neoplasms, as well as driver events that dictate clinical phenotype in Myelodysplastic patients. This approach allowed to define a risk score for developing myeloid neoplasms according to the mutational status, hypothesizing future trajectories of disease evolution.

The joint aim among all the tasks was to improve the understanding of oncogenesis, opening new scenarios regarding the contribution of genetic predisposition and clonal evolution to hematological malignancies.
Overall, a better knowledge and characterization of predisposing genetic alterations could enable targeted surveillance strategies, with

significative effects on both familial genetic counseling and direct patients' care, including regimen conditions and hematopoietic stem cell transplantations (HSTC).

References

1. Ripperger, T. *et al.* Childhood cancer predisposition syndromes—A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *Am. J. Med. Genet. Part A***173**, 1017–1037 (2017).

2. Mangaonkar, A. A. & Patnaik, M. M. Hereditary Predisposition to Hematopoietic Neoplasms: When Bloodline Matters for Blood Cancers. *Mayo Clin. Proc.***95**, 1482–1498 (2020).

3. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.***373**, 2336–2346 (2015).

4. Klco, J. M. & Mullighan, C. G. Advances in germline predisposition to acute leukaemias and myeloid neoplasms. *Nat. Rev. Cancer***21**, 122–137 (2021).

5. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature***555**, 321–327 (2018).

6. Daria V. Babushok Monica Bessler and Timothy S. Olson. Genetic predisposition to myelodysplastic syndrome and acute myeloid leukemia in children and young adults. *Leuk Lymphoma***57**, 520–536 (2016).

7. Cobaleda, C. & Sánchez-García, I. B-cell acute lymphoblastic leukaemia: Towards understanding its cellular origin. *BioEssays***31**, 600–609 (2009).

8. Wood, W. A. & Lee, S. J. Malignant hematologic diseases in adolescents and young adults. *Blood***117**, 5803–5815 (2011).

9. Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *Lancet (London, England)***381**, 1943–1955 (2013).

10. Coccaro, N., Anelli, L., Zagaria, A., Specchia, G. & Albano, F. Next-generation sequencing in acute lymphoblastic Leukemia. *Int. J. Mol. Sci.***20**, (2019).

11. Faderl, S. *et al.* Adult acute lymphoblastic leukemia: concepts and strategies. *Cancer***116**, 1165–1176 (2010).

12. Tran, T. H. & Hunger, S. P. The genomic landscape of pediatric acute lymphoblastic leukemia and precision medicine opportunities. *Semin. Cancer Biol.* (2020) doi:10.1016/j.semcancer.2020.10.013.

13. Pui, C.-H., Robison, L. L. & Look, A. T. Acute lymphoblastic leukaemia. *Lancet (London, England)***371**, 1030–1043 (2008).

14. Harrison, C. J. & Foroni, L. Cytogenetics and molecular genetics of acute lymphoblastic leukemia. *Rev. Clin. Exp. Hematol.***6**, 91–92 (2002).

15. Grioni, A. *et al.* A simple RNA target capture NGS strategy for fusion genes assessment in the diagnostics of pediatric B-cell acute lymphoblastic leukemia. *HemaSphere***3**, 1–9 (2019).

16. Inaba, H. & Mullighan, C. G. Pediatric acute lymphoblastic leukemia. *Haematologica***105**, 2524–2539 (2020).

17. Lee, P., Bhansali, R., Izraeli, S., Hijiya, N. & Crispino, J. D. The biology, pathogenesis and clinical aspects of acute lymphoblastic leukemia in children with Down syndrome. *Leukemia***30**, 1816–1823 (2016).

18. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat. Rev. Cancer***18**, 471–484 (2018).

19. Bhojwani, D., Yang, J. J. & Pui, C. H. Biology of childhood acute lymphoblastic leukemia. *Pediatr. Clin. North Am.***62**, 47–60 (2015).

20. Kratz, C. P., Stanulla, M. & Cavé, H. Genetic predisposition to acute lymphoblastic leukemia: Overview on behalf of the I-BFM ALL Host Genetic Variation Working Group. *Eur. J. Med. Genet.***59**, 111–115 (2016).

21. Porter, C. C. Germ line mutations associated with Leukemias.

*Hematology***2016**, 302–308 (2016).

22. Shah, S. *et al.* A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat. Genet.***45**, 1226–1231 (2013).

23. Hyde, R. K. & Liu, P. P. Germline PAX5 mutations and B cell leukemia. *Nat. Genet.***45**, 1104–1105 (2013).

24. Duployez, N. *et al.* Germline PAX5 mutation predisposes to familial B-cell precursor acute lymphoblastic leukemia. *Blood***137**, 1424–1428 (2021).

25. Auer, F. *et al.* Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. *Leukemia* vol. 28 1136–1138 (2014).

26. Moriyama, T. *et al.* Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet. Oncol.***16**, 1659–1666 (2015).

27. Rampersaud, E. *et al.* Germline deletion of ETV6 in familial acute lymphoblastic leukemia. *Blood Adv.***3**, 1039–1046 (2019).

28. Kirkpatrick, G., Noetzli, L., Di Paola, J. & Porter, C. C. ETV6 mutations define a new cancer predisposition syndrome. *Oncotarget* vol. 6 16830–16831 (2015).

29. Topka, S. *et al.* Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet.***11**, e1005262 (2015).

30. Churchman, M. L. *et al.* Germline Genetic IKZF1 Variation and Predisposition to Childhood Acute Lymphoblastic Leukemia. *Cancer Cell***33**, 937-948.e8 (2018).

31. Brodie, S. A. et al. Pathogenic germline IKZF1 variant alters hematopoietic gene expression profiles. Cold Spring Harb. Mol. case Stud. **7**, (2021).

32. Pui, C. H., Nichols, K. E. & Yang, J. J. Somatic and germline

genomics in paediatric acute lymphoblastic leukaemia. *Nat. Rev. Clin. Oncol.***16**, 227–240 (2019).

33. Sylvester, D. E., Chen, Y., Jamieson, R. V., Dalla-Pozza, L. & Byrne, J. A. Investigation of clinically relevant germline variants detected by next-generation sequencing in patients with childhood cancer: A review of the literature. *J. Med. Genet.***55**, 785–793 (2018).

34. Swaminathan, M. *et al.* Hematologic malignancies and Li-Fraumeni syndrome. *Cold Spring Harb. Mol. case Stud.***5**, (2019).

35. Kratz, C. P. *et al.* Cancer Screening Recommendations for Individuals with Li-Fraumeni Syndrome. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.***23**, e38–e45 (2017).

36. Chompret, A. *et al.* P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *Br. J. Cancer***82**, 1932–1937 (2000).

37. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.***45**, 242–252 (2013).

38. Leroy, B. *et al.* Recommended Guidelines for Validation, Quality Control, and Reporting of TP53 Variants in Clinical Practice. *Cancer Res.***77**, 1250–1260 (2017).

39. Stengel, A. *et al.* The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia***31**, 705–711 (2017).

40. Robles, A. I., Jen, J. & Harris, C. C. Clinical Outcomes of TP53 Mutations in Cancers. *Cold Spring Harb. Perspect. Med.***6**, (2016).

41. Hunter, A. M. & Sallman, D. A. Current status and new treatment approaches in TP53 mutated AML. *Best Pract. Res. Clin. Haematol.***32**, 134–144 (2019).

42. Mantovani, F., Collavin, L. & Del Sal, G. Mutant p53 as a guardian

of the cancer cell. *Cell Death Differ.***26**, 199–212 (2019).

43. Hafner, A., Kublo, L., Tsabar, M., Lahav, G. & Stewart-Ornstein, J. Identification of universal and cell-type specific p53 DNA binding. *BMC Mol. cell Biol.***21**, 5 (2020).

44. Baugh, E. H., Ke, H., Levine, A. J., Bonneau, R. A. & Chan, C. S. Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ.***25**, 154–160 (2018).

45. Chiereghin, C. *et al.* The genetics of myelodysplastic syndromes: Clinical relevance. *Genes (Basel).***12**, (2021).

46. Adam S. Sperling, Christopher J. Gibson, B. L. E. The genetics of myelodysplastic syndrome: from clonal hematopoiesis to secondary leukemia. *Nat Rev Cancer***17**, 5–19 (2017).

47. Kennedy, J. A. & Ebert, B. L. Clinical implications of Genetic mutations in Myelodysplastic syndrome. *J. Clin. Oncol.***35**, 968–974 (2017).

48. Hasserjian, R. P. Myelodysplastic Syndrome Updated. *Pathobiology***86**, 53–61 (2019).

49. Locatelli, F. & Strahm, B. How I treat myelodysplastic syndromes of childhood. *Blood***131**, 1406–1414 (2018).

50. Pellagatti, A. *et al.* Impact of spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated genes/pathways and clinical associations. *Blood***132**, 1225–1240 (2018).

51. Makishima, H. *et al.* Dynamics of clonal evolution in myelodysplastic syndromes. *Nat. Genet.***49**, 204–212 (2017).

52. De Kouchkovsky, I. & Abdul-Hay, M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J.***6**, e441 (2016).

53. Patel, J. P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.***366**, 1079–1089 (2012).

54. Tyner, J. W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature***562**, 526–531 (2018).

55. Crisà, E. *et al.* Genetic predisposition to myelodysplastic syndromes: A challenge for adult hematologists. *Int. J. Mol. Sci.***22**, 1–19 (2021).

56. Park, M. Myelodysplastic syndrome with genetic predisposition. *Blood Res.***56**, 34–38 (2021).

57. Stieglitz, E. & Loh, M. L. Genetic predispositions to childhood leukemia. *Ther. Adv. Hematol.***4**, 270–290 (2013).

58. Feurstein, S., Drazer, M. W. & Godley, L. A. Genetic predisposition to leukemia and other hematologic malignancies. *Semin. Oncol.***43**, 598–608 (2016).

59. Kennedy, A. L. & Shimamura, A. *Genetic predisposition to MDS: Clinical features and clonal evolution*. *Blood* vol. 133 (2019).

60. Bannon, S. A. & Dinardo, C. D. Hereditary predispositions to myelodysplastic syndrome. *Int. J. Mol. Sci.***17**, (2016).

61. Bochtler, T. *et al.* Hematological Malignancies in Adults With a Family Predisposition. *Dtsch. Arztebl. Int.***115**, 848–854 (2018).

62. Noris, P. *et al.* ANKRD26-related thrombocytopenia and myeloid malignancies. *Blood* vol. 122 1987–1989 (2013).

63. Pabst, T., Eyholzer, M., Haefliger, S., Schardt, J. & Mueller, B. U. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.***26**, 5088–5093 (2008).

64. Tawana, K., Drazer, M. W. & Churpek, J. E. Universal genetic testing for inherited susceptibility in children and adults with myelodysplastic syndrome and acute myeloid leukemia: Are we there yet? *Leukemia***32**, 1482–1492 (2018).

65. Rio-Machin, A. *et al.* The complex genetic landscape of familial

MDS and AML reveals pathogenic germline variants. *Nat. Commun.***11**, 1–12 (2020).

66. Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature***559**, 400–404 (2018).

67. Schaefer, E. J. & Lindsley, R. C. Significance of Clonal Mutations in Bone Marrow Failure and Inherited Myelodysplastic Syndrome/Acute Myeloid Leukemia Predisposition Syndromes. *Hematol. Oncol. Clin. North Am.***32**, 643–655 (2018).

68. Hartmann, L. & Metzeler, K. H. Clonal hematopoiesis and preleukemia—Genetics, biology, and clinical implications. *Genes Chromosom. Cancer***58**, 828–838 (2019).

69. Desai, P. & Roboz, G. J. Clonal Hematopoiesis and therapy related MDS/AML. *Best Pract. Res. Clin. Haematol.***32**, 13–23 (2019).

70. Tsai, F., States, U., Lindsley, C. & States, U. Clonal hematopoiesis in the inherited bone marrow failure syndromes Clonal hematopoiesis in the inherited bone marrow failure syndromes Frederick D . Tsai , R . Coleman Lindsley Department of Medical Oncology , Division of Hematologic Neoplasia , Dana-Far. *Blood***136**, 1615–1622 (2021).

71. Young, A. L., Spencer Tong, R., Birmann, B. M. & Druley, T. E. Clonal hematopoiesis and risk of acute myeloid leukemia. *Haematologica***104**, 2410–2417 (2019).

72. Desai, P., Hassane, D. & Roboz, G. J. Clonal Hematopoiesis and risk of Acute Myeloid Leukemia. *Best Pract. Res. Clin. Haematol.***32**, 177–185 (2019).

73. Xie, M. *et al.* Age-related cancer mutations associated with clonal hematopoietic expansion. *Nat Med***20**, 1472–1478 (2014).

74. Koichi Takahashi, M. D. *et al.* Pre-leukemic clonal hematopoiesis and the risk of therapy-related myeloid neoplasm: a case-

control study. *Lancet Oncol.***18**, 100–111 (2017).

75. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.***371**, 2477–2487 (2014).

76. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.***4**, 149ra118 (2012).

77. Inoue, D., Bradley, R. K. & Abdel-Wahab, O. Spliceosomal gene mutations in myelodysplasia: Molecular links to clonal abnormalities of hematopoiesis. *Genes Dev.***30**, 989–1001 (2016).

78. Qiu, J. *et al.* Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. *RNA***22**, 1535–1549 (2016).

79. Visconte, V., Nakashima, M. O. & Rogers, H. J. Mutations in splicing factor genes in myeloid malignancies: Significance and impact on clinical features. *Cancers (Basel).***11**, 1–13 (2019).

80. Wang, X., Song, X. & Yan, X. Effect of RNA splicing machinery gene mutations on prognosis of patients with MDS A meta-analysis. *Med. (United States)***98**, (2019).

81. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood***122**, 3616–3627 (2013).

82. Waldman, T. Emerging themes in cohesin cancer biology. *Nat. Rev. Cancer***20**, 504–515 (2020).

83. Galeev, R. & Larsson, J. Cohesin in haematopoiesis and leukaemia. *Curr. Opin. Hematol.***25**, 259–265 (2018).

84. Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA

molecules. *Nature***454**, 297–301 (2008).

85. Solomon, D. A., Kim, J. S. & Waldman, T. Cohesin gene mutations in tumorigenesis: From discovery to clinical significance. *BMB Rep.***47**, 299–310 (2014).

86. Dorsett, D. & Strom, L. Tha ancient and evolving roles of cohesin in DNA repair and gene expression. *Curr Biol.***10**, 54–56 (2013).

87. Dorsett, D. Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.***21**, 199–206 (2011).

88. Deardorff, M. A. *et al.* RAD21 mutations cause a human cohesinopathy. *Am. J. Hum. Genet.***90**, 1014–1027 (2012).

89. Deardorff, M. A. *et al.* Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.***80**, 485–494 (2007).

90. Sarogni, P., Pallotta, M. M. & Musio, A. Cornelia de Lange syndrome: From molecular diagnosis to therapeutic approach. *J. Med. Genet.***57**, 289–295 (2020).

91. Borck, G. *et al.* NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. *J. Med. Genet.***41**, 1–6 (2004).

92. Nishiyama, T. Cohesion and cohesin-dependent chromatin organization. *Curr. Opin. Cell Biol.***58**, 8–14 (2019).

93. Rankin, S., Ayad, N. G. & Kirschner, M. W. Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. *Mol. Cell***18**, 185–200 (2005).

94. Canudas, S. & Smith, S. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.***187**, 165–173 (2009).

95. Hauf, S. *et al.* Dissociation of cohesin from chromosome arms

and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.***3**, e69 (2005).

96. Horsfield, J. A., Print, C. G. & Mönnich, M. Diverse developmental disorders fromthe one ring: Distinct molecular pathways underlie the cohesinopathies. *Front. Genet.***3**, 1–15 (2012).

97. Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.***22**, 3089–3114 (2008).

98. Litwin, I., Pilarczyk, E. & Wysocki, R. The emerging role of cohesin in the DNA damage response. *Genes (Basel).***9**, (2018).

99. Kim, J. S. *et al.* Specific recruitment of human cohesin to laser-induced DNA damage. *J. Biol. Chem.***277**, 45149–45153 (2002).

100. Covo, S., Westmoreland, J. W., Gordenin, D. A. & Resnick, M. A. Cohesin is limiting for the suppression of DNA damage-induced recombination between homologous chromosomes. *PLoS Genet.***6**, 1–16 (2010).

101. Xu, H., Tomaszewski, J. M. & McKay, M. J. Erratum: Can corruption of chromosome cohesion create a conduit to cancer? (Nature Reviews Cancer (2011) 11 (199-210)). *Nat. Rev. Cancer***11**, 301 (2011).

102. Watrin, E. & Peters, J. M. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *EMBO J.***28**, 2625–2635 (2009).

103. Jessberger, R. Cohesin's dual role in the DNA damage response: Repair and checkpoint activation. *EMBO J.***28**, 2491–2493 (2009).

104. Zhu, Z. & Wang, X. Roles of cohesin in chromosome architecture and gene expression. *Semin. Cell Dev. Biol.***90**, 187–193 (2019).

105. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at

the developmentally regulated IFNG locus. *Nature***460**, 410–413 (2009).

106. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.***111**, 996–1001 (2014).

107. Göndör, A. & Ohlsson, R. Chromatin insulators and cohesins. *EMBO Rep.***9**, 327–329 (2008).

108. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.***32**, 3119–3129 (2013).

109. Liu, J. *et al.* Transcriptional Dysregulation in NIPBLand Cohesin Mutant Human Cells. *PLoS Biol.***7**, (2009).

110. Liu, Y. *et al.* Somatic mutation of the cohesin complex subunit confers therapeutic vulnerabilities in cancer. *J. Clin. Invest.***128**, 2951–2965 (2018).

111. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature***467**, 430–435 (2010).

112. Banerji, R., Skibbens, R. V & Iovine, M. K. How many roads lead to cohesinopathies? *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.***246**, 881–888 (2017).

113. Piché, J., Van Vliet, P. P., Pucéat, M. & Andelfinger, G. The expanding phenotypes of cohesinopathies: one ring to rule them all! *Cell Cycle***18**, 2828–2848 (2019).

114. Liu, J. & Krantz, I. D. Cornelia de Lange syndrome, cohesin, and beyond. *Clin. Genet.***76**, 303–314 (2009).

115. Avagliano, L. *et al.* Chromatinopathies: A focus on Cornelia de Lange syndrome. *Clin. Genet.***97**, 3–11 (2020).

116. Hill, V. K., Kim, J. S. & Waldman, T. Cohesin mutations in human cancer. *Biochim. Biophys. Acta - Rev. Cancer***1866**, 1–11 (2016).

117. Kon, A. *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.***45**, 1232–1237 (2013).

118. Thota, S. *et al.* Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood***124**, 1790–1798 (2014).

119. Thol, F. *et al.* Mutations in the cohesin complex in acute myeloid leukemia: Clinical and prognostic implications. *Blood***123**, 914–920 (2014).

120. Jann, J.-C. & Tothova, Z. Cohesin mutations in myeloid malignancies. *Blood***138**, 649–661 (2021).

121. Leeke, B., Marsman, J., O'Sullivan, J. M. & Horsfield, J. A. Cohesin mutations in myeloid malignancies: underlying mechanisms. *Exp. Hematol. Oncol.***3**, 13 (2014).

122. Fisher, J. B., McNulty, M., Burke, M. J., Crispino, J. D. & Rao, S. Cohesin Mutations in Myeloid Malignancies. *Trends in cancer***3**, 282–293 (2017).

123. Sasca, D. *et al.* Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood***134**, 2195–2208 (2019).

124. Mintzas, K. & Heuser, M. Emerging strategies to target the dysfunctional cohesin complex in cancer. *Expert Opin. Ther. Targets***23**, 525–537 (2019).

125. Wenger, S. L. *et al.* Rhabdomyosarcoma in Roberts syndrome. *Cancer Genet. Cytogenet.***31**, 285–289 (1988).

126. Parry, D. M., Mulvihill, J. J., Tsai, S., Kaiser-Kupfer, M. I. & Cowan, J. M. SC phocomelia syndrome, premature centromere separation, and congenital cranial nerve paralysis in two sisters, one with malignant melanoma. *Am. J. Med. Genet.***24**, 653–672 (1986).

127. Vial, Y. *et al.* Down syndrome-like acute megakaryoblastic

leukemia in a patient with Cornelia de Lange syndrome. *Haematologica* vol. 103 e274–e276 (2018).

128. Fazio, G. *et al.* First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. *J. Clin. Pathol.***72**, 558–561 (2019).

129. Schrier, S. A. *et al.* Causes of death and autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *Am. J. Med. Genet. A***155A**, 3007–3024 (2011).

Chapter 2

(Ready for submission)

Genetic profiling of pediatric Acute Lymphoblastic Leukemia patients

Claudia Saitta[1], Laura Rachele Bettini[1], Stefano Rebellato[1], Daniela Silvestri[2], Simona Songia[1], Maria Grazia Valsecchi[2], Andrea Biondi[1,3], Grazia Fazio[1], Giovanni Cazzaniga[1,4]

[1]Centro Ricerca M. Tettamanti, Pediatrics, University of Milano Bicocca, Monza, Italy.

[2]Center of Biostatistics for Clinical Epidemiology, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy.

[3]Pediatrics, University of Milano Bicocca, Fondazione MBBM/San Gerardo Hospital, Monza, Italy.

[4]Medical Genetics, University of Milano Bicocca, School of Medicine and Surgery, Monza, Italy.

Abstract

Background. Acute Lymphoblastic Leukemia (ALL) is the most common form of leukemia[1] and the first malignancy in childhood, affecting mainly B-lineage vs. T-lineage in Caucasian population. The characterization of genetic mutational pattern is a fundamental step

to implement clinical management of ALL, in terms of diagnosis, accurate risk-stratification and targeted therapy.[3]

Novel evidence supported a role of genetic predisposition in about 5-10% of pediatric tumors, including leukemias.[4] Among genetic syndromes associated with an increased risk of developing leukemia, a possible link has been proposed for Cohesinopathies, like Cornelia de Lange Syndrome (CdLS)[5,6] which is caused by mutations in Cohesin family genes. Somatic mutations in these genes, known for their fundamental role in cell cycle and DNA repair mechanisms,[7,8] are already described in myeloid malignancies[9,10] and solid tumors.[11] We previously described the first case of Acute Lymphoblastic Leukemia (ALL) in a CdLS patient who carries a *NIPBL* variant.[12]

Objectives. The present study aims to characterize genetic profiles of patients with childhood Acute Lymphoblastic Leukemia (ALL), with a particular focus on variants of 39 genes known for their involvement in cancer, including Cohesin genes.

Methods. We set up a target-capture DNA NGS panel, including 39 genes associated to ALL predisposition and classified in 6 pathways. Patients' bone marrow (BM) diagnostic samples have been analyzed, moreover for a subset of patients a germline tissue sample (BM at remission) was also sequenced. Bioinformatic analysis has been performed by Sophia DDM software and variants were classified according to coding consequence and pathogenicity.

Results. Overall, 120 consecutive pediatric ALL cases, including both B- and T-lineage phenotype, have been analyzed through the NGS panel. A total of 229 mutations were identified, distributed as it follows: 63 variants in 'Ras pathway signaling' (27%/total), 62 in 'Chromatin binding and remodeling' (27%/total), 42 in 'Transcription' (18%/total), 37 variants in 'Tumor suppressor and DNA repair' genes (16%/total), 12 in 'Other signaling genes' (15%/total). Surprisingly, 13/229 unique variants (6%) were found in Cohesins family genes exclusively in B-ALL patients; 12/13 were germline, while a unique somatic variant was found in *SMC1A*, a potential second hit in co-presence with another germline variant in a different region on the same gene.

Conclusion. NGS screening allowed the characterization of genetic profiles of a consecutive cohort of pediatric ALL patients, defining the contribution of mutational pattern on leukemogenesis. The results are consistent with literature data, showing the high prevalence of Ras pathway mutations in several pediatric hematological malignancies.[13] Moreover, the screening highlighted a noteworthy frequency of mutations in genes involved in Chromatin remodeling, whose role in hematological malignancies is still emerging.

Finally, the identification of several germline variants among Cohesin genes, with either a potentially pathogenic or unknown significance, supports an involvement in Leukemia predisposition. Their potential role has to be further investigated.

Introduction/Background

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy,[1] the first for incidence. In the 85% of cases, the malignant transformation involved B-lineage lymphocytes, and it is triggered by a variety of conventional genetic aberrations that included chromosomal translocations and alterations in chromosome number.[2] Several genetic changes do impact on prognosis, risk stratification and therapeutic decisions, such as mutations in RAS pathway genes as predictive biomarkers of unfavorable risk.[3,14]

T-ALL accounts for 10-15% of the cases of childhood ALL. The aberrant molecular pathways and the genetic lesions are heterogenous and need to be better characterized. This has an impact on prognosis, that is still less favorable.[15]

In the last decades, several advances in understanding genetic basis of leukemogenesis have been done, with an accurate characterization of somatic structural DNA rearrangements and sequence mutations that commonly perturb lymphoid compartment.[16]

New evidence support an increased role of genetic predisposition in about 5-10% of pediatric tumors, including leukemias, even in non-syndromic patients.[3] However, the prevalence and spectrum of predisposing mutations among children and adolescents remains largely unknown.[4]

Germline variants in some categories of genes, usually affected by somatic variants, are responsible of definition of new leukemia

predisposition syndromes. *ANKRD26*, *GATA2*, *PAX5*, *ETV6*, and *DDX41*, together with DNA damage response genes such as *TP53* and *BRCA1/2*, which are the most representative cases.[17,18]

Regarding the involvement of new genetic syndromes in leukemogenesis, we recently proposed, among others, a connection between ALL and Cornelia de Lange Syndrome (CdLS), describing the first case of a patient affected by both diseases.[12] The CdLS is a developmental disorder caused by germline mutations in Cohesin family genes.[5,19] Somatic mutations in these genes, known for their fundamental role in cell cycle and DNA repair mechanisms,[7,8] are already described in myeloid malignancies (10-20% of AML, 50% of DS-AMKL, 5-15% of MDS and 10% of MPN)[9,10] and solid tumors.[11,20]

So far, the association of ALL with genetic syndromes deserves to be further investigated.

A better knowledge of genomic landscape of childhood ALL may improve the understanding of tumorigenesis, defining important clinical implications in terms of diagnosis, patients' care and genetic counseling for patients and families.[4,17] Moreover, it may represent the starting point to allow an early detection of the disease and to prevent the risk to develop therapy-associated toxicities.[21,22] In this scenario, the present study aims to characterize genetic profiles of patients with childhood Acute Lymphoblastic Leukemia (ALL), with a particular focus on variants of 39 genes known for their involvement in cancer.

Methods

Cohort of patients

The study included 120 consecutive diagnoses of Acute Lymphoblastic Leukemia (ALL) enrolled in the AIEOP-BFM-ALL-2009 protocol in Italian AIEOP centers. All patients were diagnosed between May and September 2016. Of these, 11 patients were T-ALL, 107 B-ALL and 2 cases have mixed phenotype Acute Leukemia. The median age of patients at diagnosis is 4 years old (range:1-17). Patients' characteristics are shown in table 1.

Target-Capture DNA Next Generation Sequencing analysis was performed on DNA extracted from bone marrow at onset of disease. The median percentage of blasts is 78% (range 9.1-95).

For a subgroup of mutated patients, germline variants were screened by sequencing bone marrow DNA defining by a Minimal residual disease (MRD) value below 10^{-4}.[23]

Characteristic	n (%)
Immunophenotype	
B-ALL	107 (89.1)
T-ALL	11 (9.2)
MPAL	2 (1.7)
Risk group	
SR	45 (37.5)
MR	55 (45.8)
HR	17 (14.2)
na	3 (2.5)
Cytogenetics	
t(1;19) positive	6 (5.0)
t(12;21) positive	21 (17.5)
negative	90 (75.0)
na	3 (2.5)
DNA index	
<1	21 (17.5)
=1	66 (55.0)
>1.16	30 (25.0)
na	3 (2.5)
Relapse	
no	99 (82.5)
yes	21 (17.5)
very early	3 (14.3)
early	11 (52.4)
late	7 (33.3)
2nd-3rd relapse	3 (14.3)

Table1. Clinical characteristics of patients of ALL Cohort.
Summary of main clinical characteristics of the analyzed cohort of patients.

Ethic statement

Samples were obtained from patients, after a written informed consent from parents or legal representatives. The study was approved by each institutional review board and conducted in

accordance with the ethical standards of the Declaration of Helsinki and to national and international guidelines.

Custom predisposition genes panel

We designed a custom panel including 39 genes involved in Leukemia predisposition and pathogenesis. They were divided in 6 classes, according to their biological functions. (Fig. 1)

The custom panel was designed with the Integrate DNA Technology (IDT) platform (xGen Predesigned Gene Capture Pools – https://idtdna.com/site/order/ngs), generating high fidelity single strand DNA probes. It consists of 1520 probes and a cumulative targeted region of 141 kb.

Tumor suppressor and DNA repair	Transcription	RAS pathway signaling	Other signaling	Cohesins	Chromatin binding and remodeling
ATM	*AUTS2*	*BRAF*	*ANKRD26*	*BRD2*	*ARID5B*
BLM	*CEBPA*	*KRAS*	*JAK2*	*BRD4*	*CREBBP*
CDKN2A	*CEBPE*	*NF1*	*PIP4K2A*	*HDAC8*	*EP300*
CDKN2B	*ETV6*	*NRAS*	*SH2B3*	*NIPBL*	*EZH2*
NBN	*GATA1*	*PTPN11*		*RAD21*	*NSD1*
TP53	*GATA2*			*SMC1A*	*PHF6*
	GATA3			*SMC3*	
	PAX5			*STAG1*	
	RUNX1			*STAG2*	

Figure 1. Categories of the 39 Cancer Genes analyzed for mutations.
Classification of genes in different classes according to their biological functions.

Target-Capture DNA Next Generation Sequencing Target sequencing was performed on genomic DNA using Nextera Flex for Enrichment by Illumina protocol (#1000000048041 v01). The pool libraries were

paired end (2x150) sequenced on flow cell with v2.5 chemistry on Nextseq550 (Illumina) instrument. FASTQ files were generated by Local Run Manager software.

Data analysis

Bioinformatic analysis was carried out by the Sophia DDM software on FASTQ files (deposited in ArrayExpress database). Alignment was performed against the Human Reference sequence GRCh37/Hg19. Variants were filtered by variant fraction (VF) >5% and coverage at least 500X; Variant Allelic Fraction (VAF) in the population was set at 1%. We included certainly pathogenic, potentially pathogenic and variants of unknown significance (VUS). Novel exonic non-synonymous variants were also retained. Benign/likely benign variants in all databases of prediction are excluded from the results.

The most common databases of prediction were consulted for the interpretation of the pathogenicity, including ClinVar, Clinical Genome, Varsome, InterVar, COSMIC.

Results

TC-DNA NGS Screening Results and Classification of the variants based on Coding consequences.

Among 103 of the 120 patients of the pediatric cohort we identified a total of 229 rare variants, in the 39 genes of the panel. So, 17 samples

resulted negative for variants among all the panel genes (Suppl. Table 1).

Considering the coding consequence, most of the variants are missense (154-67.2%), 8 nonsense (3.5%), 1 no start (0.4%). Moreover, 16 aberrations are in-frame variants (7.0%) and 14 frameshift (6.1%). Further, 13 variants involve splicing: 8 splice-donor (3.5%) and 3 splice-acceptor (1.3%) and 25 mutations (10.9%) are in 5'-3'- UTR regions.

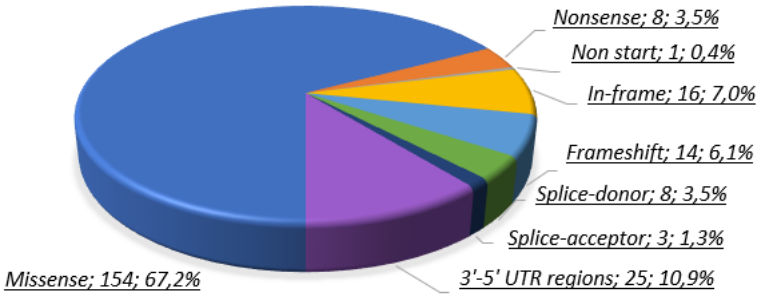


Figure 2. Coding consequences of variants.

Distribution and pathogenicity prediction of mutations.

We described the 229 identified variants affected genes of each biological classes, defining distribution and frequency in the cohort, as well as pathogenicity prediction according to most common databases. Mutations have been classified in 3 different categories of pathogenicity: pathogenic/likely pathogenic, Variant of Uncertain Significance (VUS) - in case of absence of evidence to support-, or

Novel variant, in case of mutation without any annotation in public genetic databases, such as ClinVar, Clinical Genome, Varsome, InterVar, COSMIC. Data are reported in Suppl. Table 1 and Suppl. Figure 1.

Ras pathway signaling.

The results showed a considerable involvement of 'RAS pathway signaling' also considering the role in leukemogenesis. This category is the most frequently affected in the ALL cohort, followed by 'Chromatin binding and remodeling' genes, as shown in Figure 3.

Indeed, we identified a total of 63 variants in genes involved in 'RAS pathway signaling' (27%/total): 25 in *NRAS* (40%), 18 in *KRAS* (29%), 10 in *PTPN11* (16%), 7 in *NF1* (11%) and 3 in *BRAF* (5%), respectively.

Considering *NRAS* gene, 23/25 variants (92%) have a pathogenic/likely pathogenic significance, 1 mutation is a VUS (rs751774011) and very rare in the population ($<10^{-5}$) and 1 is a novel variant in the databases of prediction, with a VAF (Variant Allele Frequency) of 13.2%. The distribution is almost the same for *KRAS*, with 11/18 pathogenic/likely pathogenic variants (61%), 4 VUS and 3 novels. All the variants are characterized by VAF ranging between 5 and 30%, compatible with somatic origin. Moreover, more than half of *PTPN11* mutations (6/10) have pathogenic/likely pathogenic significance, and 4 are uncertain for predictors.

The 7 *NF1* mutations are distributed as follows: 1 pathogenic/likely pathogenic variant, 3 VUS and 3 novels. Somatic mutations in *NF1* are known to be critical drivers in a wide variety of tumors, instead germline mutations are responsible of Neurofibromatosis type 1, an autosomal dominantly inherited tumor predisposition syndrome. In both cases, the aberration result in dysregulation of the RAS/MAPK pathway.[24]

As regarding *BRAF* gene, one variant is novel in databases of prediction, 2/3 mutations have uncertain significance, of which one is annotated as rs377093637. (Fig. 4)

Variants in genes involved in 'Chromatin binding and remodeling' pathways.

The percentage of mutations in Chromatin remodeling genes correspond to 27% of total cases (62 variants). The 47% of these aberrations regards *CREBBP* (29 mutations in 27 patients). The variants in *EP300* are 16 (26%, 14 mutated patients), 9 in *ARID5B* (15%, 8 mutated patients), 5 in *EZH2* (8%) 2 in *PHF6* (3%) and 1 in *NSD1* (2%).

Focusing on pathogenicity, we founded 18 novel variants in *CREBBP*, 8 VUS and 3 pathogenic/likely pathogenic variants (rs200782888 as a single variant and rs587783502 in two different patients). The paralog *EP300* is affected by 1 pathogenic/likely pathogenic variant annotated as rs121434596, 7 novel and 8 VUS, of which the missense rs137935821 with a germline origin. Hereditable variants in *CREBBP*

and *EP300* are responsible of Rubinstein-Taybi syndrome (RSTS), known as a tumor-prone syndrome.[25]

Among the 9 mutations of *ARID5B,* 6/9 are novel, instead the remaining ones have uncertain significance.

We also identified 5 variants in *EZH2*, 3 novel and 2 VUS as annotated in databases (rs151023145 and the conflicting rs6954744), and 2 variants in *PHF6,* 1 novel and 1 nonsense pathogenic variant (VAF 87.5%).

The only mutation identified in *NSD1*, annotated as rs201483724, has a germline origin (VAF 100%) and it is associated with conflicting significance (VUS in InterVar, Benign/likely benign in ClinVar and Varsome) (Fig. 4).

'Trascription' genes are affected by variants.

The genes that characterized 'Transcription' class is mutated in 18% of total aberration (42/229). In 45% of cases *RUNX1* gene is involved, with 19 variants in 18 patients. The mutations in *AUTS2* are 7 (17%, 8 patients), 6 in the B-cell differentiation factor *PAX5* and *ETV6* (14%). We detected a low number of aberrations in binding-protein genes *CEBPA*, *GATA1* (1 variants, 2%) ad *GATA2* (2 variants, 5%). *CEBPE* and *GATA3* are not affected in our cohort. (Fig. 3)

As regarding pathogenicity prediction, variants in *AUTS2* are so distributed: 3 novel and 4 with uncertain significance.

We identified 5 novel variants in *PAX5*, in addition to a pathogenic frameshift mutation rs780753361, located in the trans-activation domain, known to be frequently mutated in ALL.[26] Among the 6 variants on *ETV6*, it is noteworthy the somatic Leu205fs*, reported also by Topka et al.[27] in an ALL cohort.

RUNX1, the most frequent mutated gene of the category, is affected only by novel variants (19/19) that alter the sequence of exon 6 of the genes, in a region known to be mutated in different type of cancers, hematological (Myelodysplastic Syndrome) and non- hematological (Adenocarcinoma) (Pecan.stjude.cloud/proteinapaint/*RUNX1*).

The only variant on *CEBPA* is a germline in frame mutation with uncertain significance (rs746430067). In literature, germline variants in this gene have been recognized as leukemia predisposition syndrome named Familial CEBPA-mutated acute myeloid leukemia.[28]

About *GATA1* and *GATA2* genes, the first is affected by a pathogenic/likely pathogenic variant (rs782632688; VAF 24.3%), instead we identified 2 germline mutations in the second one, both predicted as VUS and annotated as rs749214277 and rs141800945, respectively (VAF ~50% for both cases).

Involvement of 'Tumor suppressors and DNA repair' pathways.

The 37 variants identified among the biological class (16% of total cases) are divided as follow: 14 in *ATM* (38%, 13 patients), 9 in *NBN* (24%), 5 in *BLM* (14%), only 1 variant in *TP53 (3%* of category*)*. In

addition to the aberrations of tumor suppressor genes involved in DNA repair, we identified other 5 variants in *CDKN2A* (14%) and 3 in *CDKN2B* (8%). (Fig.3)

Most mutations have uncertain significance in databases of prediction. For example, *ATM* gene is affected only by VUS. Among these, we checked and confirmed the germline origin for 4 different mutations, reported in databases as rs139552233, rs140856217, rs138398778 and rs137882485. Homozygous germline variants of this gene are responsible of Ataxia Telangiectasia Syndrome, that has an increase of 20-30% of lifetime cancer risk.[29]

We identified 5 variants with conflicting interpretations of pathogenicity in *BLM,* and 1 mutation in exon 5 of *TP53*, a splice acceptor predicted as VUS (rs751253294).

Two among the 9 variants that affected *NBN* are pathogenic/likely pathogenic in databases of prediction: a very rare (MAF $<10^{-6}$) frameshift variant annotated as rs587781891, and a novel nonsense variant. Among the remaining 7 VUS, in 3 cases (rs876659551, rs61753720, rs7558054619) we confirmed their germline origin.

Considering the Cyclin Dependent Kinase Inhibitors, we identified 5 variants in *CDKN2A,* whose 1 pathogenic somatic non-sense variant (rs121913388; VAF 13.8%), 1 VUS (rs45456595) and 3 not present in databases of prediction. The 3 mutations on *CDKN2B* are novel, with a VAF compatible with a germline origin (Fig. 4).

'Other signaling genes'

We overall identified 12 variants affecting *JAK2* (N=5), *ANKRD26* (N=3) and SH2B3 genes (N=3). Only 1 variant alters *PIP4K2A* (Fig. 3).

Dissecting their origin and pathogenicity, we identified 5 variants in *JAK2*: 4/5 have a pathogenic/likely pathogenic prediction. Among these, we checked and confirmed the somatic origin for the well-known Arg683Gly (rs1057519721),[30] as well as the germline origin for Ser919Cys (rs773078489). Also, the VUS Leu113Val, annotated as rs143103233, is a germline mutation.

Bothe the 3 aberrations that affect *SH2B3* and ANKRD26 are of uncertain significance in databases of prediction. One of the *ANKRD26* variants has a germline origin (rs762754151; VAF 45.1% in diagnosis, 44.9% in remission phase). As shown in Figure 4, the only variant on *PIP4KN2A* is novel in databases (VAF 47.81%).

The details of the 229 identified variants are available in supplementary table 1.

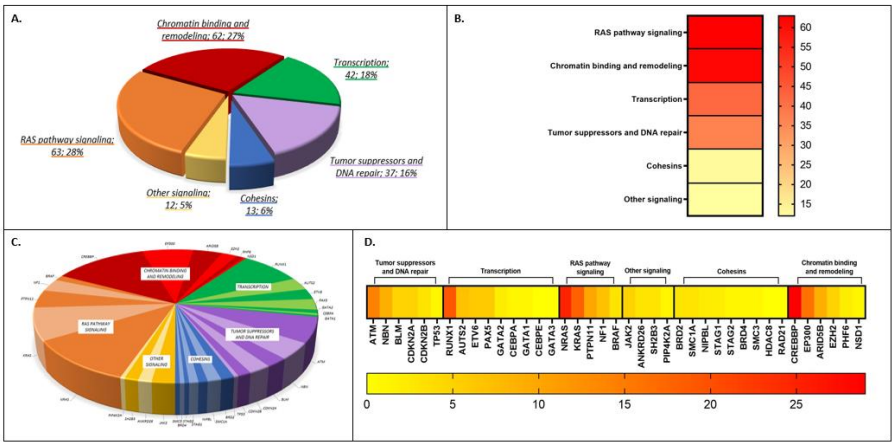


Figure 3. Frequency and distribution of mutations on the 39 genes in the pediatric ALL cohort.

In the panels A-B numbers and percentages of total variants that affected each biological class, represented as pie graph (A) and heatmap (B). In the panels C (pie graph) and D (heatmap) the detailed distribution of mutations in each gene of each category.

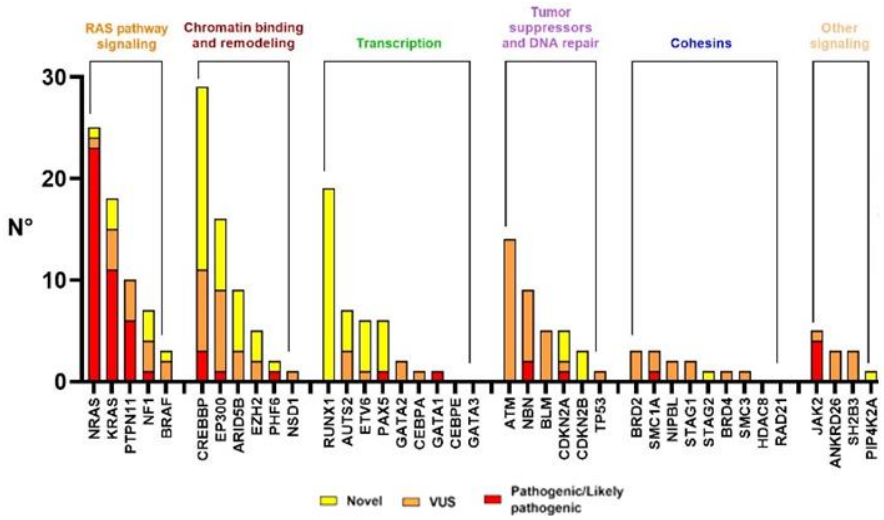


Figure 4. Pathogenicity of mutations on the 39 genes of the pathways in the pediatric ALL cohort.

In yellow all the variants that are novel in databases of prediction; in orange the Variants with Uncertain Significance (VUS). In red the mutations annotated as pathogenic/likely pathogenic in databases of prediction.

Focus on Cohesins variants: distribution, origin, and pathogenicity.

The remaining class includes 'Cohesin genes', that are mutated in 6% of all cases, with 13 variants. The *STAG1* gene is affected by 2 mutations, as *NIPBL* gene (N=2, 15%). 3 mutations were detected in *SMC1A* and *BRD2* (23%), while only 1 variant was identified in *STAG2*, *SMC3* and *BRD4* (8%). No mutations were detected in *HDAC8* and *RAD21* genes.

Almost all the Cohesins variants have uncertain significance (VUS) in databases of prediction (11/13). One pathogenic/likely pathogenic mutation, previously known as rs1569357187, affects *SMC1A*, instead 1 novel variant alters *STAG2* (Fig. 4).

Mutations in these categories have been further investigated in order to define the somatic or germline nature of mutations, as well as their mapping on genes structures. With this purpose, we performed NGS sequencing of remission BM samples of the 10 patients carrying mutations in Cohesins.

The results showed that 12/13 (92.3%) variants are present also in a disease negative sample, confirming their germline origin (Fig. 5, panels A/B).

Moreover, the two *NIPBL* germline mutations are predicted as VUS in databases (rs748290328 - VAF 45.5% in DX, 44.9% in REM; and rs180747605 - VAF 67.5% in DX, 47% in REM).

Recently, a hypothetical link between germline mutations in *NIBPL*-responsible of Cornelia de Lange syndrome (CdLS)- and cancer predisposition has been proposed by our group. We described the first case of a pediatric CdLS patient with B-cell precursor Acute Lymphoblastic Leukemia, characterized by a deleterious *de novo* frameshift variant on the gene. See state of the art 1.[12]

SMC1A is mutated in 2 different patients: the first case carried the rs112727682 germline variant affecting the 3' UTR region (VAF 50.2% in DX, 50.3% in REM). The second patient carried two mutations: one somatic, Asp1141Asn (VAF 13.1%), whose prediction is uncertain, and one germline, annotated as rs1569357187 and correspondent to missense Lys507Arg, likely pathogenic in ClinVar (VAF 99.7%; in homozygous). This is a potential example of double hit mutation.[31]

Only 1 germline VUS has been identified in *SMC3* (VAF 42.6% in DX, 51.1% in REM).

Focusing on *STAG1*, the screening showed 2 germline mutations in two different patients, annotated as rs747617236 and as rs182497205, respectively, with allele frequency compatible with a heterozygous zygosity (VAF 44.6% in DX, 41.3% in REM for the Arg1167Gln variant; 31,9% and 41.4% for the 3'UTR variant). Both have uncertain

significance in databases of prediction and are very rare in general population (MAF $<1^{-5}$ in ExAC).

Finally, a germline splice_acceptor_indel mutation (c.1535-3_1535-2insTA; VAF 68.6% in DX and 76.4% in REM) affected *STAG2*.

BRD2 and *BRD4* variants are not the focus of our study, due to their controversial involvement in the Cohesin ring.

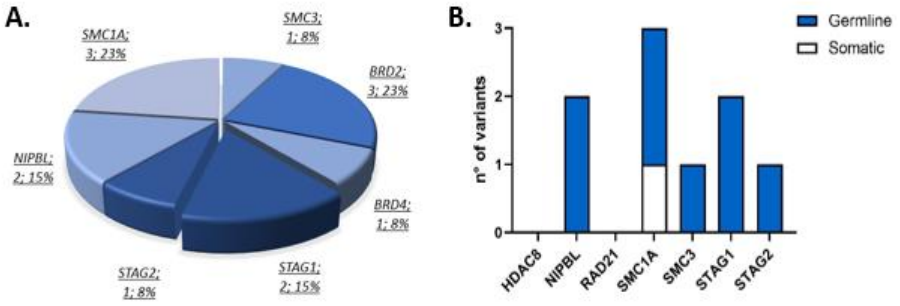


Figure 5. Frequency, distribution, and origin of Cohesins' variants in the pediatric ALL cohort.

In the panel A the distribution of mutations in the different genes of the Cohesins' family. In the panel B the classification of Cohesins' variants according to their origin germline or somatic.

Mapping of variants on Cohesins structures

We mapped the identified variants on each Cohesin, in order to define their specific localization on genes and protein's structure (Fig. 6). Among the 13 aberrations, the majority affected exonic regions, in addition to 3 variants located at 5'-3' UTR regulatory regions. Variants' descriptions are reported in supplementary table 1.

Considering the regions in which they are localized, the Arg1187Glu on *STAG1* and the Asp2755Asn on *NIPBL* variants are both noteworthy, because they affected a frequently mutated domain in several type of carcinomas.

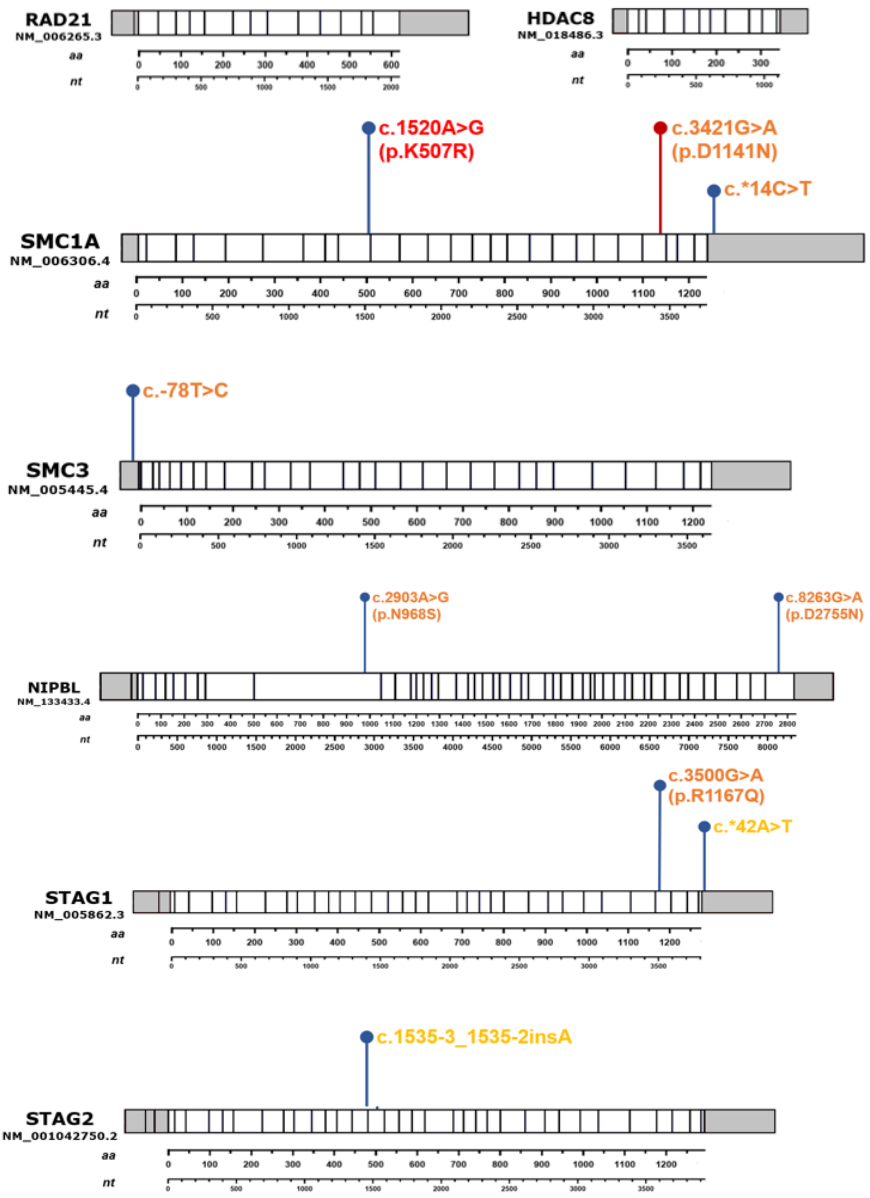


Figure 6. Graphic representation of variants localization on Cohesins.
Blu rods indicate the germline mutations, instead are red in case of somatic origin. In yellow all the variants that are novel in databases of prediction; in orange the VUS. In red the mutations predicted as pathogenic/likely pathogenic.

Cross-validation in a pediatric Danish cohort

To get a cross-validation of identified Cohesins variants and their potential role in pediatric malignancies, we integrated our results with a Danish cohort of 514 pediatric patients with cancer, analyzed by Whole Genome Sequencing (WGS). This screening allowed to identify a total of 25 germline variants in Cohesins, none of which was already found in our cohort (Suppl. Table 2).

Importantly, in two siblings affected by ALL (female) and embryonal carcinoma (male), the *STAG1* mutation Val1041Ile, was identified in the same frequently mutated region of the rare one Arg1167Gln, detected in our ALL cohort.

The genomic position, in addition to the familiar history, could suggest the possible role of *STAG1* germline variants in carcinogenesis. The functional effects of the two *STAG1* variants will be investigated in the next future.

Furthermore, *SMC3* Arg879Leu variant, located in a colon cancer related region, was identified in a child with AML and a family history of leukemia (paternal aunt). Moreover, the *NIPBL* Arg2247Leu variant, located in a HEAT domain of the gene, was detected in an ALL-pediatric patient whose father was affected by testicular cancer.

Discussion

The genetic screening performed in this study allowed us to define the genetic profile of a cohort of 120 consecutive diagnosis of Acute Lymphoblastic Leukemia (ALL).

Among the 39 dissected genes, the results confirmed the crucial role of Ras pathway in leukemogenesis, both considering prevalence and pathogenicity of the mutations (27% of total variants). Abnormal activation of Ras signaling is responsible for alteration of many cellular processes frequently altered in carcinogenesis. In pediatric BCP-ALL, *NRAS*, *NF1* and *PTPN11* variants are associated with high-risk group and correlate with decrease in event free (EFS) and overall survival (OS), as well as chemotherapy resistance and relapses.[13,14,32] Considering the crucial role of RAS pathway in pediatric cancer, in the last years it has been considered one of the most promising therapeutic targets; development of inhibitors of MEK and RAF kinases, duration of response and overcoming of resistance mechanisms, are among the most investigated therapeutic goal.[33]

Moreover, the interesting frequency of mutations in Chromatin remodeling gene (27%), confirms their potential role in pathogenesis of hematological malignancies, that is still emerging. Genetic alterations in histone modifier *CREBBP* and *EP300* genes have been associated with solid tumors and hematological malignancies, such as B-cell Lymphoma.[34,35]

In our ALL cohort, the prevalence of mutations that affect *CREBBP (*47% of the category), support their involvement in leukemogenesis. In particular, *CREBBP* variants have been already described as crucial in relapsed Acute Lymphoblastic Leukemia. These aberrations, present as subclones at diagnosis, become driver in relapse phase, suggesting that they may confer resistance to therapy.[36]
So, *CREBBP* variants can be considered a good marker in leukemia monitoring, in order to improve surveillance strategy in case of higher relapse risk.

Focusing on predisposition to Leukemia, the results of this screening underline the potential contribution of single genes, as well as group of genes, that must be investigated.
First of all, the frequency of germline mutations in *NBN* (3/3 mutated patients with DX/REM analysis) is noteworthy. This gene encodes for a tumor suppressor and DNA repair genes, known to be essential for the maintenance of genome integrity.[37] Homozygous mutations in the gene cause Nijmegen Breakage Syndrome, a well-known chromosome instability cancer-predisposing condition.[38] However, an increase risk of cancer development has been hypothesized also in patients with heterozygous NBN variants.[39]

Furthermore, the involvement of Cohesins gene (6% of total mutations) in ALL is rather novel. Indeed, we hypothesized for the first

time a role of germline mutations in predisposition to hematological malignancies.

Cohesins are found as somatically mutated in various human cancers, such as Myeloid Leukemia, Bladder cancer, Ewing Sarcoma.[20] If affected by germline variants, they are responsible of several disorders called Cohesinopathies. Moreover, the several functions of these genes in cellular mechanisms frequently altered in cancer (sister chromatids cohesion, chromatin structure, gene expression and DNA repair)[11] strongly encourage a potential involvement.

To further support this hypothesis, we recently described the first case of ALL in a Cornelia de Lange patient.[12]

Our results showed that most of the variants that affect Cohesins in ALL are germline. Exclusively one patient carried a somatic second hit (Asp1141Asn) in *SMC1A* in combination to a germline variant (Lys507Arg, rs1569357187). This potential double hit mutations may be the classical example[31] of how a germinal variant may determine the genomic instability that predisposes to additional somatic cooperating mutations, responsible of onset of disease.

Overall, multiple somatic events have been identified in the other biological categories investigated. Only in 1 patient, carrying a germline *SMC3* variant, no other variants were identified.

In summary, a better knowledge of mutational landscape, as well as molecular mechanisms, can be crucial to improve clinical management

of patient, in term of early diagnosis, surveillance strategies and targeted-treatments strategies.

The potential contribution of several genes in leukemogenesis need to be further investigated.

References

1. Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *Lancet (London, England)***381**, 1943–1955 (2013).

2. Coccaro, N., Anelli, L., Zagaria, A., Specchia, G. & Albano, F. Next-generation sequencing in acute lymphoblastic Leukemia. *Int. J. Mol. Sci.***20**, (2019).

3. Inaba, H. & Mullighan, C. G. Pediatric acute lymphoblastic leukemia. *Haematologica***105**, 2524–2539 (2020).

4. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.***373**, 2336–2346 (2015).

5. Piché, J., Van Vliet, P. P., Pucéat, M. & Andelfinger, G. The expanding phenotypes of cohesinopathies: one ring to rule them all! *Cell Cycle***18**, 2828–2848 (2019).

6. Avagliano, L. *et al.* Chromatinopathies: A focus on Cornelia de Lange syndrome. *Clin. Genet.***97**, 3–11 (2020).

7. Brooker, A. S. & Berkowitz, K. M. *The roles of cohesins in mitosis, meiosis, and human health and disease*. *Methods in Molecular Biology* vol. 1170 (2014).

8. Tothova, Z. *et al.* Cohesin mutations alter DNA damage repair and chromatin structure and create therapeutic vulnerabilities in MDS/AML. *JCI Insight***6**, (2021).

9. Thota, S. *et al.* Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood***124**, 1790–1798 (2014).

10. Fisher, J. B., McNulty, M., Burke, M. J., Crispino, J. D. & Rao, S. Cohesin Mutations in Myeloid Malignancies. *Trends in cancer***3**, 282–293 (2017).

11. Solomon, D. A., Kim, J. S. & Waldman, T. Cohesin gene mutations in tumorigenesis: From discovery to clinical significance. *BMB Rep.***47**, 299–310 (2014).

12. Fazio, G. *et al.* First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. *J. Clin. Pathol.***72**, 558–561 (2019).

13. Ney, G. M., McKay, L., Koschmann, C., Mody, R. & Li, Q. The emerging role of ras pathway signaling in pediatric cancer. *Cancer Res.***80**, 5155–5163 (2020).

14. Jerchel, I. S. *et al.* RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. *Leukemia***32**, 931–940 (2018).

15. Bhojwani, D., Yang, J. J. & Pui, C.-H. Biology of childhood acute lymphoblastic leukemia. *Pediatr. Clin. North Am.***62**, 47–60 (2015).

16. Iacobucci, I. & Mullighan, C. G. Genetic basis of acute lymphoblastic leukemia. *J. Clin. Oncol.***35**, 975–983 (2017).

17. Porter, C. C. Germ line mutations associated with Leukemias. *Hematology***2016**, 302–308 (2016).

18. Tran, T. H. & Hunger, S. P. The genomic landscape of pediatric acute lymphoblastic leukemia and precision medicine opportunities. *Semin. Cancer Biol.* (2020) doi:10.1016/j.semcancer.2020.10.013.

19. Sarogni, P., Pallotta, M. M. & Musio, A. Cornelia de Lange syndrome: From molecular diagnosis to therapeutic approach. *J. Med. Genet.***57**, 289–295 (2020).

20. Hill, V. K., Kim, J. S. & Waldman, T. Cohesin mutations in human cancer. *Biochim. Biophys. Acta - Rev. Cancer***1866**, 1–11 (2016).

21. Mangaonkar, A. A. & Patnaik, M. M. Hereditary Predisposition to Hematopoietic Neoplasms: When Bloodline Matters for Blood Cancers. *Mayo Clin. Proc.***95**, 1482–1498 (2020).

22. Pui, C. H., Nichols, K. E. & Yang, J. J. Somatic and germline

genomics in paediatric acute lymphoblastic leukaemia. *Nat. Rev. Clin. Oncol.***16**, 227–240 (2019).

23. Brüggemann, M., Gökbuget, N. & Kneba, M. Acute lymphoblastic leukemia: monitoring minimal residual disease as a therapeutic principle. *Semin. Oncol.***39**, 47–57 (2012).

24. Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N. & Upadhyaya, M. The NF1 somatic mutational landscape in sporadic human cancers. *Hum. Genomics***11**, 13 (2017).

25. Boot, M. V *et al.* Benign and malignant tumors in Rubinstein-Taybi syndrome. *Am. J. Med. Genet. A***176**, 597–608 (2018).

26. Gu, Z. *et al.* PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat. Genet.***51**, 296–307 (2019).

27. Topka, S. *et al.* Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet.***11**, e1005262 (2015).

28. Tawana, K., Rio-Machin, A., Preudhomme, C. & Fitzgibbon, J. Familial CEBPA-mutated acute myeloid leukemia. *Semin. Hematol.***54**, 87–93 (2017).

29. Tiao, G. *et al.* Rare germline variants in ATM are associated with chronic lymphocytic leukemia. *Leukemia***31**, 2244–2247 (2017).

30. Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U. S. A.***106**, 9414–9418 (2009).

31. Lin, M. *et al.* JAK2 p.G571S in B-cell precursor acute lymphoblastic leukemia: a synergizing germline susceptibility. *Leukemia* vol. 33 2331–2335 (2019).

32. Ney, G. M. *et al.* Mutations predictive of hyperactive Ras signaling correlate with inferior survival across high-risk pediatric acute leukemia. *Transl. Pediatr.***9**, 43–50 (2020).

33. Santarpia, L., Lippman, S. M. & El-Naggar, A. K. Targeting the

MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin. Ther. Targets***16**, 103–119 (2012).

34. Attar, N. & Kurdistani, S. K. Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer. *Cold Spring Harb. Perspect. Med.***7**, (2017).

35. Huang, Y.-H. *et al.* CREBBP/EP300 mutations promoted tumor progression in diffuse large B-cell lymphoma through altering tumor-associated macrophage polarization via FBXW7-NOTCH-CCL2/CSF1 axis. *Signal Transduct. Target. Ther.***6**, 10 (2021).

36. Mullighan, C. G. *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature***471**, 235–239 (2011).

37. Ronen, A. & Glickman, B. W. Human DNA repair genes. *Environ. Mol. Mutagen.***37**, 241–283 (2001).

38. Sharma, R., Lewis, S. & Wlodarski, M. W. DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Front. Pediatr.***8**, 570084 (2020).

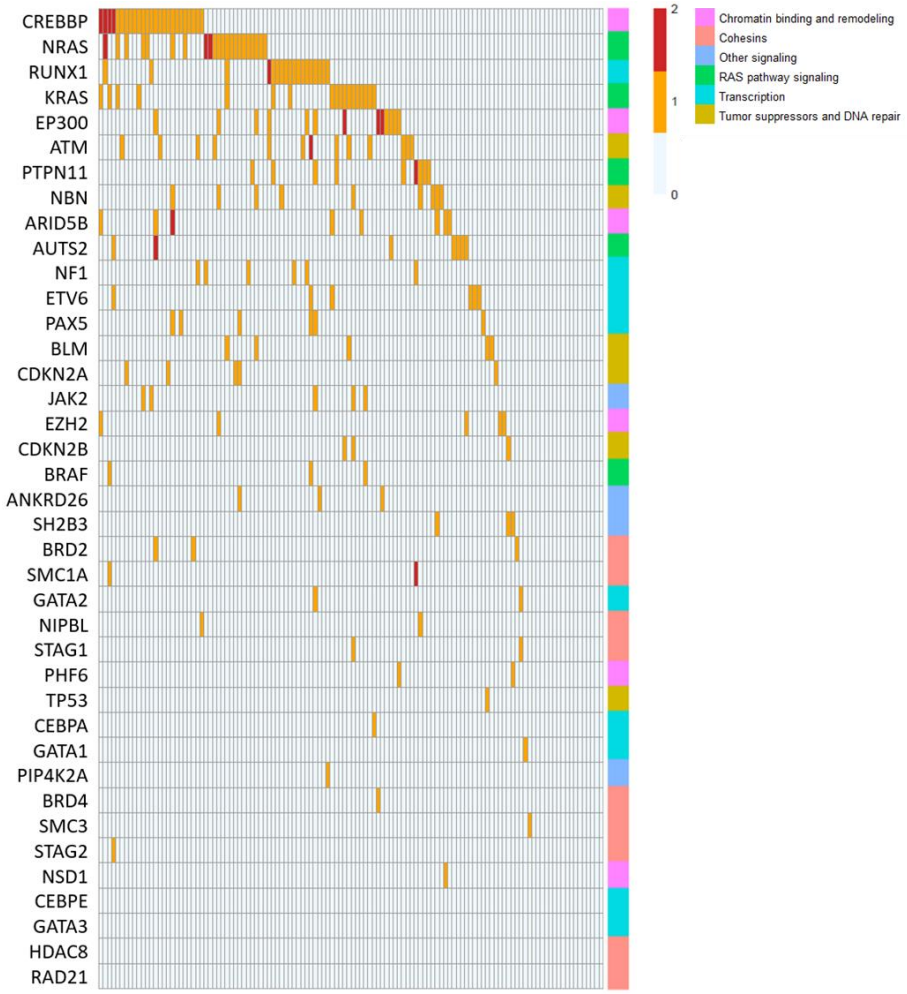
39. Seemanová, E. *et al.* Cancer risk of heterozygotes with the NBN founder mutation. *J. Natl. Cancer Inst.***99**, 1875–1880 (2007).

SUPPLEMENTARY TABLES AND FIGURES

Gene	Chr	Exon	cDNA	Protein	OverlapKnown	CodingConsequence	Pathogenicity	1000gen	Esp5400	GnomAD	ExAc	ALFA	Patients
ANKRD26	10	1	c.3G>A	p.(Met1?)	rs199683454	no-start	VUS		0.0002	0.000207	0.000345		1
n=3		2	c.340A>G	p.(Arg114Gly)	rs762754151	missense	VUS			0.000021	0.000017		1
		34	c.*274T>G		rs191989707	3'UTR	NOVEL	0.0008		0.000264			1
ARID5B	10	2	c.38_39insGTA	p.(Cys13delinsTrpTyr)		inframe_3	NOVEL						1
n=9		2	c.39delinsGGTAATTCGATA	p.(Cys13Trpfs*35)		frameshift	NOVEL						1
		3	c.422A>C	p.(Glu141Ala)	rs150284163	missense	VUS	0.001	0.001	0.001219	0.001293		1
		5	c.400_401del	p.(Ile134Leufs*2)		frameshift	NOVEL						1
		6	c.631_634del	p.(Glu211Lysfs*24)		frameshift	NOVEL						1
		6	c.658_664del	p.(Asp220Glnfs*14)		frameshift	NOVEL						1
		7	c.*70C>A			3'UTR	NOVEL	0.0002					1
		7	c.1441T>A	p.(Asp483Glu)	rs146336312	missense	VUS		0.0002	0.000043	0.000050		1
		7	c.1732C>G	p.(His578Asp)	rs147521761	missense	VUS	0.0008					1
ATM	11	4	c.295A>G	p.(Ser99Gly)	rs137882485	missense	VUS		0.0001	0.000185	0.000188		1
n=14		5	c.482A>C	p.(Gln161Pro)	rs587780625	missense	VUS					0.000	1
		8	c.1009C>T	p.(Arg337Cys)	rs138398778	missense	VUS		0.0001	0.000171			1
		11	c.1700A>G	p.(Asn567Ser)	rs786203230	missense	VUS	none	none	none	none	none	1
		12	c.1837G>T	p.(Val613Leu)	rs200124136	missense	VUS			0.000043	0.000043		1
		20	c.2932T>C	p.(Ser978Pro)	rs139552233	missense	VUS	0.0006	0.0003	0.000335	0.001247		2
		31	c.4709T>C	p.(Val1570Ala)	rs140856217	missense	VUS	0.0002	0.0005	0.000464	0.000487		2
		41	c.6067G>A	p.(Gly2023Arg)	rs11212587	missense	VUS	0.0012	0.0024	0.001241	0.001575		1
		47	c.6860G>A	p.(Gly2287Glu)	rs1800061	missense	VUS					0.00115	1
		50	c.7390T>C	p.(Cys2464Arg)	rs55801750	missense	VUS		0.0005	0.000378	0.000569		1
		58	c.8560C>T	p.(Arg2854Cys)	rs201958469	missense	VUS			0.000128	0.000181		1
		63	c.*236C>T		rs3092834	3'UTR	VUS	0.0028		0.006405			1
AUTS2	7	18	c.3305_3308delinsC	p.(His1102_His1103delinsPro)		inframe_3	NOVEL						1
n=7		18	c.3317A>C	p.(His1106Pro)	rs865932880	missense	VUS					0.000	1
		18	c.3320A>C	p.(His1107Pro)	rs867872684	missense	VUS					0.000	1
		18	c.3422C>T	p.(Thr1141Met)	rs773568699	missense	VUS			0.000014			1
		18	c.3527C>T	p.(Thr1176Ile)		missense	NOVEL						1
		18	c.*56G>A		rs772352925	missense	VUS	0.0004					1
		18	c.*159_*162dup			3'UTR	NOVEL						1
BLM	15	3	c.191A>T	p.(Asp64Val)	rs140382474	missense	VUS		0.0003	0.000257	0.000316		1
n=5		4	c.842A>C	p.(His281Pro)	rs202042636	missense	VUS		0.0003	0.000143	0.000091		1
		7	c.1601A>G	p.(Asn534Ser)	rs35224686	missense	VUS	0.0018	0.0012	0.000919	0.000372		1
		17	c.3358G>C	p.(Gly1120Arg)		missense	VUS						1
		21	c.4076+4T>G		rs183176301	splice_donor_+4	VUS	0.0014	0.0034	0.004356	0.003819		1
BRAF	7	4	c.608+2T>C	p.(?)		splice_donor_+2	NOVEL						1
n=3		10	c.1237G>A	p.(Val413Met)	rs377093637	missense	VUS		0.0001		0.000016		1
		18	c.*203A>G			3'UTR	VUS						1
BRD2	6	3	c.314C>T	p.(Ala105Val)	rs138290874	missense	VUS		0.0003	0.000214	0.000243		1
n=3		10	c.1795G>C	p.(Ala599Pro)	rs55952113	missense	VUS	0.0034	0.0025	0.002531	0.004153		1
		13	c.2365T>C	p.(Ser789Pro)		missense	VUS						1
BRD4	19	12	c.2227C>T	p.(Arg743Cys)	rs368124974	missense	VUS	0.0006	0.0002	0.000093	0.000570		1
CDKN2A	9	1	c.187G>C	p.(Gly63Arg)	rs45456595	missense	VUS	0.0012	0.003	0.002117	0.002152		1
n=5		1	c.193+1_193+2insAG	p.(?)		splice_donor_indel	NOVEL						1
		2	c.151-1G>A	p.(?)		splice_acceptor_-1	NOVEL						1
		2	c.187_188insG	p.(Leu63Argfs*57)		frameshift	VUS						1
		2	c.238C>T	p.(Arg80*)	rs121913388	nonsense	PATH/LIKELY PATH				0.000009		1
CDKN2B	9	1	c.-201C>G			5'UTR	NOVEL						1
n=3		1	c.209G>A	p.(Arg70Gln)		missense	NOVEL						1
		2	c.157G>T	p.(Val53Phe)		missense	NOVEL						1
CEBPA	19	1	c.201_209del	p.(Pro68_Pro70del)	rs746430067	inframe_9	VUS					0.000005	1
CREBBP	16	1	c.-507>G			5'UTR	NOVEL						4
n=29		1	c.-45G>C			5'UTR	NOVEL						1
		4	c.1108C>T	p.(Arg370*)		nonsense	NOVEL						1
		11	c.2083C>T	p.(Gln695*)		nonsense	NOVEL						1
		12	c.2189G>T	p.(Arg730Leu)		missense	NOVEL						1
		18	c.3497A>T	p.(Tyr1166Phe)	rs200346970	missense	VUS		0.0001	0.000135	0.000126		1
		19	c.3596G>A	p.(Cys1199Tyr)		missense	NOVEL						1
		20	c.3718_3722+14delinsTTCTCCC	p.(?)		splice_donor_cdsindel	NOVEL						1
		20	c.3722+1G>A	p.(?)	rs200782888	splice_donor_+1	PATH/LIKELY PATH	none	none	none	none	none	1
		20	c.3722+2_3722+3insGCCCTC			splice_donor_+3	NOVEL						1
		24	c.4118G>T	p.(Gly1373Val)		missense	NOVEL						1
		24	c.4118G>A	p.(Gly1373Glu)	rs201186078	missense	VUS	none	none	none	none	none	1
		25	c.4235A>C	p.(Tyr1412Ser)		missense	NOVEL						1
		26	c.4301G>T	p.(Trp1434Leu)		missense	NOVEL						1
		26	c.4441T>C	p.(Tyr1481His)		missense	NOVEL						1
		27	c.4555_4556ins75	p.(Leu1518_Glu1519ins25)		inframe_75	NOVEL						1
		29	c.4925_4927del	p.(Ser1642del)	rs587783502	inframe_3	PATH/LIKELY PATH	none	none	none	none	none	2
		30	c.5796_5798del	p.(Gln1932del)		inframe_3	VUS						3
		30	c.5963T>A	p.(Leu1988His)	rs748565790	missense	VUS			0.000007	0.00001		1
		30	c.6169C>T	p.(Gln2057*)		nonsense	NOVEL						1
		30	c.6492_6494del	p.(Gln2178del)	rs753500354	inframe_3	VUS					0.00007	3
EP300	22	1	c.-96A>C			5'UTR	NOVEL						2
n=16		1	c.-88C>T			5'UTR	NOVEL						1
		1	c.32C>T	p.(Pro11Leu)	rs121434596	missense	PATH/LIKELY PATH				0.000008		1
		2	c.680C>T	p.(Ser227Phe)	rs773141352	missense	VUS				0.000008		1
		3	c.741G>A	p.(Met247Ile)		missense	NOVEL						1
		15	c.2870C>G	p.(Thr957Ser)	rs774604668	missense	VUS			0.000014	0.000008		1
		28	c.4617+4_4617+13delinsGCCTAGG			splice_donor_+4	NOVEL						1
		29	c.4724A>G	p.(Asn1575Ser)	rs144547088	missense	VUS		0.0003	0.000264	0.000148		1
		31	c.5582A>G	p.(Gln1861Arg)	rs75046918	missense	NOVEL			0.009917			1
		31	c.5636C>T	p.(Pro1879Leu)	rs141530780	missense	VUS		0.0002	0.000014	0.000033		1
		31	c.5833G>A	p.(Val1945Met)	rs774508702	missense	VUS				0.000008		1
		31	c.6627_6638del	p.(Asn2209_Gln2213delinsLys)	rs587778256	inframe_12	VUS				0.001715		1
		31	c.6934C>T	p.(Pro2312Ser)	rs137935821	missense	VUS	0.0002	0.0002	0.000078	0.000124		1
		31	c.7114A>G	p.(Met2372Val)	rs768061933	missense	VUS	none	none	none	none	none	1
		31	c.*111G>A			3'UTR	NOVEL						1
ETV6	12	1	c.-20G>C		rs200485978	5'UTR	NOVEL	0,0002	0,0003	0.001934	0.001409	0.00068	1
n=6		3	c.315dupA	p.(Ser106Ilefs*6)		frameshift	NOVEL						1
		4	c.338T>G	p.(Leu113Arg)		missense	NOVEL						1
		5	c.496G>A	p.(Val166Met)	rs142603082	missense	VUS	0,0046	0,0007	0.000371		0.000162	1
		5	c.613dupC	p.(Leu205Profs*12)		frameshift	NOVEL						1
		8	c.*23C>T		rs72550782	3'UTR	NOVEL	0,0078	0,0144	0.014860	0.014487	0.01690	1

Gene	Chr	Exon	cDNA	Protein	OverlapKnown	CodingConsequence	Pathogenicity	1000gen	Esp5400	GnomAD	ExAc	ALFA	Patients
EZH2	7	9	c.943C>G	p.(Leu315Val)	rs6954744	missense	VUS			0.000007			1
n=5		9	c.950A>G	p.(Asn317Ser)	rs151023145	missense	VUS	0,0002	0,0003	0.000321	0.000305	0.000508	1
		17	c.2010C>G	p.(Asn670Lys)		missense	NOVEL						1
		19	c.2176delinsGGCCTTC	p.(Tyr726delinsGlyLeuHis)		Inframe_6	NOVEL						1
		20	c.*271A>C			3'UTR	NOVEL						1
GATA1	X	4	c.727C>T	p.(Arg243Trp)	rs782632688	missense	PATH/LIKELY PATH				0.00002		1
GATA2	3	3	c.40C>T	p.(Pro14Ser)	rs749214277	missense	VUS			0.000007	0.00003		1
n=2		4	c.829A>G	p.(Ser277Gly)	rs141800945	missense	VUS	0,0002	0,0002	0.000093	0.000086		1
JAK2	9	4	c.337C>G	p.(Leu113Val)	rs143103233	missense	VUS	0,0002	0,0016	0.000692	0.000636		1
n=5		13	c.1711G>A	p.(Gly571Ser)	rs139504737	missense	PATH/LIKELY PATH		0,0006		0.000480		2
		16	c.2047A>G	p.(Arg683Gly)	rs1057519721	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		20	c.2755A>T	p.(Ser919Cys)	rs773078489	missense	PATH/LIKELY PATH					0.00000	1
KRAS	12	2	c.34G>T	p.(Gly12Cys)	rs121913530	missense	PATH/LIKELY PATH			0.000007	0.000020		1
n=18		2	c.35G>A	p.(Gly12Asp)	rs121913529	missense	PATH/LIKELY PATH			0.000014	0.000020		2
		2	c.35G>T	p.(Gly12Val)	rs121913529	missense	PATH/LIKELY PATH			0.000014	0.000020		1
		2	c.35G>C	p.(Gly12Ala)	rs121913529	missense	VUS			0.000014	0.000020		1
		2	c.38G>A	p.(Gly13Asp)	rs112445441	missense	PATH/LIKELY PATH			0.000014		0.00000	4
		2	c.38_40dup	p.(Gly13dup)		Inframe_3	NOVEL						1
		2	c.40_41insGCCCTGGGGGTGGCG	p.(Gly13_Val14insGlyProGlyGlyGly)		Inframe_15	NOVEL						1
		2	c.68T>G	p.(Leu23Arg)	rs730880472	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		2	c.82T>G	p.(Phe28Val)		missense	NOVEL						1
		3	c.182A>C	p.(Gln61Pro)	rs121913240	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		4	c.351A>C	p.(Lys117Asn)	rs770248150	missense	PATH/LIKELY PATH				0.000008		1
		4	c.436G>A	p.(Ala146Thr)	rs121913527	missense	VUS					0.00000	2
		5	c.*60C>A		rs1229307469	3'UTR	VUS			0.000029			1
NBN	8	2	c.119C>T	p.(Ser40Leu)	rs587781530	missense	VUS			0.000007			1
n=9		2	c.123del	p.(Ser42Alafs*7)	rs587781891	frameshift	PATH/LIKELY PATH			0.000014		0.00007	1
		4	c.37G>A	p.(Asp13Asn)	rs61753720	missense	VUS	0,0008	0,0022	0.001883	0.001858	0.003066	1
		6	c.265A>G	p.(Ile89Val)	rs61754966	missense	VUS	0,0004	0,001	0.001455	0.001394		1
		7	c.382G>T	p.(Val128Phe)	rs61754796	missense	VUS	0,0002	0,0011	0.000471	0.000482		1
		7	c.451A>T	p.(Lys151*)		nonsense	PATH/LIKELY PATH						1
		12	c.1171C>A	p.(Gln391Lys)	rs755805461	missense	VUS				0.000009	0.00008	1
		13	c.1667C>A	p.(Ser556Tyr)	rs756036331	missense	VUS				0.00004		1
		16	c.1981_1982del	p.(Leu661Phefs*2)	rs876659551	frameshift	VUS			0.000004			1
NF1	17	4	c.304A>G	p.(Met102Val)	rs756370324	missense	VUS			0.000007	0.000008	0.000043	1
n=7		6	c.620T>C	p.(Val207Ala)		missense	NOVEL						1
		29	c.3883A>G	p.(Thr1295Ala)	rs143836226	missense	VUS		0,0001	0.000036	0.000099	0.000090	1
		30	c.4046C>T	p.(Ala1349Val)	rs1229215269	missense	PATH/LIKELY PATH			0.000012			1
		36	c.5203_5204InsGCCCA	p.(Lys1735Serfs*11)		frameshift	NOVEL						1
		42	c.6373_6388delinsGGA	p.(Lys2125Glyfs*4)		frameshift	NOVEL						1
		57	c.*125C>T		rs973008062	3'UTR	VUS			0.000014			1
NIPBL	5	10	c.2903A>G	p.(Asn968Ser)	rs180747625	missense	VUS	0,0004	0,0002			0.000264	1
n=2		47	c.8263G>A	p.(Asp2755Asn)	rs748290328	missense	VUS			0.000036	0.000033	0.00007	1
NRAS	1	2	c.34G>A	p.(Gly12Ser)	rs121913250	missense	PATH/LIKELY PATH	none	none	none	none	none	4
n=25		2	c.34G>T	p.(Gly12Cys)	rs121913250	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		2	c.35G>A	p.(Gly12Asp)	rs121913237	missense	PATH/LIKELY PATH			0.000014	0.000008	0.00008	3
		2	c.37G>T	p.(Gly13Cys)	rs121434595	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		2	c.38G>A	p.(Gly13Asp)	rs121434596	missense	PATH/LIKELY PATH				0.000008	0.0001	11
		3	c.182A>T	p.(Gln61Leu)	rs11554290	missense	PATH/LIKELY PATH					0.000	1
		3	c.183A>T	p.(Gln61His)	rs121913255	missense	PATH/LIKELY PATH	none	none	none	none	none	2
		4	c.436G>A	p.(Ala146Thr)		missense	NOVEL						1
		4	c.412G>A	p.(Gly138Arg)	rs75174011	missense	VUS				0.000008		1
NSD1	5	11	c.4564G>A	p.(Asp1522Asn)	rs201483724	missense	VUS	0,0004	0,0001			0.00013	1
PAX5	9	3	c.239C>G	p.(Pro80Arg)		missense	NOVEL						1
n=6		5	c.575_576dup	p.(Asp193Profs*86)		frameshift	NOVEL						1
		5	c.600dupC	p.(Glu201Argfs*42)		frameshift	NOVEL						1
		6	c.778C>T	p.(Gln260*)		nonsense	NOVEL						1
		7	c.889C>T	p.(Gln297*)		nonsense	NOVEL						1
		8	c.963dupC	p.(Ala322Argfs*23)	rs780753361	frameshift	PATH/LIKELY PATH				0.000129		1
PHF6	X	7	c.673C>T	p.(Arg225*)		nonsense	PATH/LIKELY PATH						1
n=2		9	c.835-2A>G	p.(?)			NOVEL						1
PIP4K2A	10	10	c.*155dupA			3'UTR	NOVEL						1
PTPN11	12	1	c.-101C>T		rs1324685370	5'UTR	VUS	none	none	none	none	none	1
n=10		3	c.215C>T	p.(Ala72Val)	rs121918454	missense	VUS	none	none	none	none	none	3
		3	c.218C>T	p.(Thr73Ile)	rs121918462	missense	PATH/LIKELY PATH					0.000	1
		3	c.226G>A	p.(Glu76Lys)	rs121918464	missense	PATH/LIKELY PATH					0.00000	2
		3	c.182A>T	p.(Asp61Val)	rs121918461	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		4	c.417G>C	p.(Glu139Asp)	rs397507520	missense	PATH/LIKELY PATH	none	none	none	none	none	1
		13	c.1520G>C	p.(Gly507Ala)	rs397507546	missense	PATH/LIKELY PATH			0.000007		0.000031	1
RUNX1	21	6	c.1182_1184delinsGGC	p.(Glu395Ala)		missense	NOVEL						1
n=19		6	c.1184A>G	p.(Glu395Gly)		missense	NOVEL						2
		6	c.1187G>A	p.(Arg396His)		missense	NOVEL						2
		6	c.1189T>C	p.(Ser397Pro)		missense	NOVEL						7
		6	c.1189_1191delinsGCC	p.(Ser397Ala)		missense	NOVEL						4
		6	c.1189_1191delinsCCC	p.(Ser397Pro)		missense	NOVEL						2
		6	c.1189T>G	p.(Ser397Ala)		missense	NOVEL						1
SH2B3	12	4	c.334G>A	p.(Glu112Lys)	rs548009613	missense	VUS	0,0002		0.000021	0.000025	0.00007	1
n=3		5	c.598G>A	p.(Val200Met)	rs755796482	missense	VUS			0.000008	0.000008	0.00000	1
		7	c.820C>T	p.(Leu274Phe)		missense	VUS	0,0004	0,0005				1
SMC1A	X	11	c.1520A>G	p.(Lys507Arg)	rs156935187	missense	PATH/LIKELY PATH			0.000005			1
n=3		24	c.3421G>A	p.(Asp1141Asn)		missense	VUS	none	none	none	none	none	1
		26	c.*14C>T		rs112727682	3'UTR	VUS	0,0034	0,0065	0.004356	0.00451	0.00463	1
SMC3	10	1	c.-787C>T			5'UTR	VUS						1
STAG1	3	31	c.3500G>A	p.(Arg1167Gln)	rs747617236	missense	VUS				0.000008	0.00000	1
n=2		34	c.*61A>T		rs182497205	3'UTR	NOVEL	0,0002		0.000014	0.000008	0.00000	1
STAG2	X	17	c.1535-3_1535-2insTA	p.(?)		splice_acceptor_indel	NOVEL						1
TP53	17	5	c.376-2dupA	p.(?)	rs751253294	splice_acceptor_indel	VUS			0.000007	0.000017	0.00001	1

Suppl. Table 1 Details of variants identified in the pediatric ALL cohort.



Suppl. Figure 1 Heatmap of the variants of pediatric ALL cohort.

Genes are order according to mutational load.

Gene	Chr	Exon	cDNA	Protein	OverlapKnown	CodingConsequence	Intervar	ClinVar	Varsome	VAF	Alt Allele Freq (AF)	pop-max (AF)	Disease	Family history
NIPBL	5	3	c.190C>A	p.Gln64Lys		missense	VUS			59,6%			ALL	2° relatives with cancer in the 7th decade
n=6	5	10	c.1832G>A	p.Ser611Asn		missense	VUS			48,8%			ALL	No cancer in family
	5	10	c.2294G>A	p.Arg765Lys		missense	VUS	LB/VUS		52,2%	0.0005548919	0.0001034	ALL	Father's sister breast cancer at 33yo
	5	12	c.3335A>T	p.Asp1112Val		missense	VUS			44,4%			LCH	3° relatives with adult onset cancer
	5	17	c.4087+3T>C	?	rs202064847	splice region variant -		LB		44,8%	8,22354	0.00001592	ALL	3° relatives with adult onset cancer
	5	39	c.6740G>T	p.Arg2247Leu		missense	VUS			40,0%			ALL	Father testicular cancer
SMC1A	X	16	c.2508C>A	p.His836Gln		missense	VUS			95,8%	1,12112	0.00003241	ALL	Mother melanoma
n=2	X	5	c.771C>G	p.Asp257Glu		missense	VUS			50,0%			JPA	No information, hypopigmentation
SMC3	10	9	c.702G>T	p.Glu234Asp		missense	VUS			61,5%			AML	3° relatives with adult onset cancer
n=4	10	23	c.2636G>T	p.Arg879Leu		missense	LP			58,3%			AML	Father's sister died of leukemia at 18yo, hypopigmentation at arm
	10	25	c.3106-8_3106-4delGTTTT	?		splice region variant -				45,2%	0,04897	0.00001283	DIPG	3° relative with leukemia <30yo
	10	26	c.3259G>A	p.Val1087Ile		missense	VUS			46,2%	0.0002033376	0.00004113	SARCOMA	2° relative with adult onset cancer
RAD21	8	4	c.1583A>C	p.Glu528Ala		missense	VUS			62,5%	0,4067755	0.00003268	ALL	2° relatives with adult onset cancer
n=4										47,1%			MDS	2° and 3° degree relatives with adult onset cancer, no pattern
	8	4	c.1544T>G	p.Leu515Arg		missense	VUS			36,7%			JPA	2° relative with adult onset cancer
	8	4	c.1532A>G	p.Asn511Ser		missense	VUS			37,9%			DIPG	2° relative with adult onset cancer
	8	2	c.790A>G	p.Met264Val		missense	VUS			43,9%			SARCOMA	No data
STAG2	X	5' UTR	c.-98+7A>G			splice region variant -				37,8%			ALL	Many 2° and 3° relatives with cancer on both sides > 50yo
n=4	X	16	c.1535-12_1535-3delTTTTTTTTTT	?		splice region variant -				72,7%			JPA	2° relative with myelomatosis
	X	31	c.3344T>C	p.Met1115Thr		missense	VUS			47,4%	0,559939	0.00001223	ALL	No cancer in family
	X	32	c.3497G>A	p.Arg1166Lys		missense	VUS			100,0%			LCH	No cancer in family
STAG1	3	8	c.3121G>A	p.Val1041Ile		missense	VUS			47,8%	0,16252	0.0003263	ALL	Brother hembryonal carcinoma
n=3										61,1%			EC	Sister ALL
	3	7	c.2336G>C	p.Cys779Ser		missense	VUS			57,4%			CNS	No cancer in family
	3	6	c.2009A>G	p.Asn670Ser	rs201418951	missense	VUS	VUS/LP		45,0%	8,12605	0.0001202	ALL	No cancer in family

*LP=LIKELY PATHOGENIC
*VUS=VARIANT OF UNKNOWN SIGNIFICANCE
*LB=LIKELY BENIGN

*ALL=Acute Lymphoblastic Leukemia
*LCH=Langerhans Cell Histiocytosis
*JPA=Juvenile Pilocytic Astrocytoma
*AML=Acute Myeloid Leukemia
*DIPG=Diffuse Intrinsic Pontine Glioma
*MDS=MyeloDysplastic Syndrome
*CNS=Central Nervous System Tumours
*EC=Embryonal Carcinoma

Suppl. Table 2 Details of identified germline variants in Cohesin genes in pediatric Danish cohort.

Chapter 3

Submitted to Blood Cancer Discovery, under revision

Recurrent germline variant in the cohesin complex gene RAD21predisposes children to Lymphoblastic Leukemia and Lymphoma

Anne Schedel[1], Ulrike Friedrich[1], Rabea Wagener[2], Juha Methonen[3], Titus Watrin[2], **Claudia Saitta[4]**, Triantafyllia Brozou[2], Pia Michler[1], Carolin Walter[5], Asta Försti[6,7], Arka Baksi[8], Maria Menzel[1], Peter Horak[9], Nagarajan Paramasivam[10], Grazia Fazio[4], Robert J Autry[6,7], Stefan Fröhlin[9], Meinolf Suttorp[1], Christoph Gertzen[11], Holger Gohlke[11,12], Sanil Bhatia[2], Karin Wadt[13], Kjeld Schmiegelow[14], Martin Dugas[5], Daniela Richter[15], Hanno Glimm[16], Merja Heinäniemi[3], Rolf Jessberger[8], Gianni Cazzaniga[4,17], Arndt Borkhardt[2], Julia Hauer[1,18], Franziska Auer[18]

[1]Pediatric Hematology and Oncology, Department of Pediatrics, University Hospital "Carl Gustav Carus", TU Dresden, Dresden, Germany.

[2]Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine University Duesseldorf, Medical Faculty, Duesseldorf, Germany.

[3]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Yliopistonranta 1, FI-70211, Kuopio, Finland.

[4] Tettamanti Research Center, Pediatrics, University of Milan Bicocca, Fondazione MBBM/San Gerardo Hospital, Monza, Italy;

[5]Institute of Medical Informatics, University of Muenster, Muenster, Germany.

[6]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany.

[7]Hopp Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany.

[8]Institute of Physiological Chemistry, Medical Faculty “Carl Gustav Carus”, TU Dresden, Dresden, Germany

[9]Division of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg and German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany.

[10]Computational Oncology, Molecular Diagnostics Program, National Center for Tumor Diseases (NCT), Heidelberg, Germany.

[11]Institute for Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225, Düsseldorf, Germany.

[12]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), and Institute of Biological Information Processing (IBI-7: Structural Biochemistry), ForschungszentrumJülich GmbH, 52425 Jülich, Germany.

[13]Department of Clinical Genetics, University hospital of Copenhagen, Faculty of health and 34 Medical Sciences, University of Copenhagen, Denmark.
[14]Department of Paediatrics and Adolescent Medicine, Copenhagen University Hospital 36 Rigshospitalet, Copenhagen, Denmark.
[15]Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Dresden and German Cancer Research Center (DKFZ), Germany; German Cancer Consortium (DKTK) Dresden, Germany.
[16]Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Dresden and German Cancer Research Center (DKFZ), Germany; Center for Personalized Oncology, National Center for Tumour Diseases (NCT) Dresden and University Hospital Carl Gustav Carus Dresden at TU Dresden, Dresden, Germany; Translational Functional Cancer Genomics, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany; German Cancer Consortium (DKTK) Dresden, Germany.
[17] Medical Genetics, Dept. Of Medicine and Surgery, University of Milan Bicocca, Monza, Italy.
[18]National Center for Tumor Diseases (NCT), Dresden, Germany: German Cancer Research Center (DKFZ), Heidelberg, Germany; Faculty of Medicine and University Hospital Carl Gustav Carus, TechnischeUniversität Dresden, Dresden, Germany; Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Dresden, Germany.

Keywords:

Acute lymphoblastic leukemia, TRIO-Sequencing, germline cancer predisposition, *RAD21*, cohesin complex

Abstract

Somatic loss of function mutations in Cohesin genes are frequently associated with various cancer types, but Cohesin disruption in the germline causes Cohesinopathies like Cornelia-de-Lange syndrome (CdLS) and CdLS patients are generally not known to be tumor-prone. Here, we present the discovery of a recurrent heterozygous *RAD21* germline variation at amino acid position 298 (p.P298S/A) identified in three children with Lymphoblastic leukemia and Lymphoma in a total dataset of 482 pediatric cancer patients. While RAD21 p.P298S/A did not disrupt the formation of the Cohesin complex, it altered gene expression and DNA damage response and primary patient fibroblasts showed increased G2/M arrest after irradiation and Mitomycin-C treatment. Subsequent single-cell RNA-sequencing analysis of healthy human bone marrow confirmed the upregulation of distinct Cohesin gene patterns during hematopoiesis, highlighting the importance of *RAD21* expression within proliferating B- and T-cells. Our clinical and functional data therefore suggest that *RAD21* germline variants can predispose to childhood Lymphoblastic Leukemia and Lymphoma without displaying a CdLS phenotype.

Significance

The identification and assessment of variants within Cohesin complex genes (such as *RAD21*) that predispose to hematopoietic malignancies in childhood is essential to identify children at risk and enable

surveillance strategies. This is especially important in the context of incomplete penetrance and in cases without a classical syndromal phenotype.

Introduction

The Cohesin complex is a cogwheel of ordered chromosome alignment and segregation during cell division, homologous recombination-driven DNA repair and regulation of gene expression.[1–6] By connecting the SMC1 and SMC3 cohesin subunits, and thereby generating the functional ring-like structure of cohesin (Figure 1A), RAD21 is essential for this machinery and thus for life. Paradoxically, while *RAD21*-inactivating heterozygous somatic mutations are a well-established correlate of various human cancers, such as Acute Myeloid Leukemia (AML),[7–9] germline mutations[10] Lead to Cohesinopathies (e.g. Cornelia-de-Lange syndrome (CdLS)) which are not primarily known as cancer predisposition syndromes. However, an index case with a germline *NIPBL* mutation and concomitant Acute Lymphoblastic Leukemia (ALL) was recently described.[11] Nevertheless, the relationship between germline mutations of Cohesin genes and cancer predisposition in non-syndromic cases is not known.[12,13]

Materials and Methods

Patients

Patients ≤19 years of age were unselectively recruited at the Pediatric Oncology Department, Dresden (years 2019 2020), or as previously described.[1–3] Consent of the families was obtained according to the Ethical Vote EK 181042019 (Dresden) and in line with the Declaration of Helsinki. For the IntReALL cohort, patients' parents or their legal guardians gave informed consent to genetic analyses in the context of add on studies linked to the clinical protocol to which patients were enrolled.

Whole exome sequencing (WES)

Germline DNA was extracted from the patient's fibroblasts using AllPrep DNA/RNA Mini Kit (Qiagen) and from PBMCs of the parents and the remaining patient's using the QIAamp DNA Blood Mini Kit (Qiagen). Sequenceable next-generation libraries for WES were generated with the SureSelect Human All Exon V7 kit (Agilent). The libraries were sequenced on a NovaSeq 6000 platform (Illumina) in paired-end mode (2x150bp) and with a final on-target coverage of ≥ 100x. Processing of the WES data was performed as previously described.[2]

In short, after the generation of read files in fastq format using bcl2fastq v2.19.0, 120 trimmomatic v0.33 was used to remove adapter and low-quality sequences.[4] The alignment to the human reference

genome GRCh37 was performed using BWA-MEM v0.7.12[5] and Samtools v1.2.[6] The tool Peddy 0.4.6[7] performed gender and relatedness analyses to validate the correct sample assignment and the expected relationship of the patient's data with the corresponding parents' data. Single nucleotide variants (SNVs) and insertion/deletions (indels) were called using GATK v4.1.4.1 and VarScan2 v2.3.9,[8] applying the trio-mode. Functional annotation of variants was done using Ensembl Variant Effect Predictor v98.3.[9] For in silico prediction of the effect of the variants, SIFT, Polyphen and CADD were applied. The COSMIC database (downloaded 25.03.2019 130 https.//cancer.sanger.ac.uk/cosmic/download) was used to identify variants located in somatic mutational hotspots. In addition, we used the ClinVar database (download 132 02/12/2019), the IARC *TP53* germline database and the LOVD database for *MSH2*, *MSH6*, *APC and NF1* in order to identify previously reported pathogenic variants. Furthermore, we used the dbNSFP 3.5 plugin to annotate the conservation scores based on GERP++ and phastCons100way_vertebrate. For in silico prediction of the effect of splice site variants, we applied the dbscSNV v1.1 plugin for VEP,[10] which annotated the ada-and rf-scores to the splice variants.

The initial variant interpretation was carried out with the CPSR pipeline,[14] which classified the variants as pathogenic, likely pathogenic, variant of unknown significance (VUS), likely benign, or benign. The additional variant interpretation was manually performed

(e.g. by taking CADD scores into account)[12] as well as by utilizing an extended cancer gene list.

Sanger Sequencing Validation

RAD21p.P298S and RAD21p.P298A were validated via PCR and subsequent Sanger 146 sequencing using the following primers (5'->3'):

Name	Sequence (5' → 3')	Patient
hRAD21_Ex8_F	ATTGGGTTCAAGTCTGGCGG	Case-18
hRAD21_Ex8_R	TTCTGGGAACCCCAGGAGAC	Case-18
hRAD21_Ex8_F	ATTGGGTTCAAGTCTGGCGG	TRIO-DD_017
hRAD21_Ex8_R	TTCTGGGAACCCCAGGAGAC	TRIO-DD_017
hATR_Ex45_F	CAGAAATTGTGCCATTTCGCC	TRIO-DD_017
hATR_Ex45_R	AGGGGCCAATAATTATATTCGAGG	TRIO-DD_017

***RAD21* Variation Analysis**

RAD21, transcript ID ENST00000297338 was analyzed. Minor allele frequencies of all coding germline variants present in *RAD21* in a global, non-cancer population, taken from the gnomAD exome r.2.1.1 dataset (n=118.479), were summed up codon-wise. The variants had to be VEP annotated to one of the following consequences for inclusion: start lost, missense variant, inframe insertion, inframe deletion, stop gained, frameshift variant, coding sequence variant,

stop lost, incomplete_terminal_codon_variant, transcript ablation, transcript amplification, protein altering variant. Somatic, coding variants reported for adult cancer patients derive from COSMIC, GRCh37 Release 91 (CosmicCodingMuts.normal.vcf.gz, n=1,443,198 samples) and were similarly combined for each codon along *RAD21*. Both collected datasets were smoothed using the LOWESS algorithm (fraction: 0.06, 160 iterations: 3) prior to plotting.

In-silico Modeling

To assess the structural impact of the P298S and P298A substitutions in RAD21, we aimed at structurally modeling the 50 residues on each side of the substitutions. However, several homology modeling and *ab initio* modeling approaches failed to generate a secondary structure for this region. This reflects that the substitution site is part of a very flexible and likely intrinsically disordered region. MFDp2,[13] a disorder 168 predictor, also predicts this part to be disordered, (predicted disorder content of the entire RAD21: 51.7%). Accordingly, a recent cryo-electron microscopy (cryo-EM) structure contains the largest structurally resolved part of RAD21 in addition to (partially) resolved binding partners STAG1, SMC1, SMC3, and NIPBL and double-stranded DNA (PDB ID: 6WG3).[15] However, the part of RAD21 containing the substitution site was not resolved because of the high flexibility of this region. From the cryo-EM structure, estimating with which residues P298 and the substitutions P298S and P298A may interact results in

many putative interaction partners in STAG1, SMC3, NIPBL and the double-stranded DNA. Hence, we aimed at narrowing down putative interaction partners based on sequence analyses. We used the HMMER suite[3] to produce a multiple sequence alignment, applying jackhammer and using an E-value cut-off of 10-6 and minimum coverage of 75%. The alignment reveals that proline and serine can occur at amino acid 298. Alanine also occurs at this position, but with a lower probability than proline or serine. Next, a possible co-evolution of P298 within RAD21 was determined via GREMLIN,[4] using the full-length sequence and, first, default parameters as well as, second, an E-value cut-off of 10-6, minimum coverage of 25%, and gap removal of 50% using HHblits. Both searches did not leave enough sequences to generate a co-evolution analysis.

Cell culture

Primary fibroblasts were initially cultivated in BIO-AMF™-2 Medium (Biological-189 Industries) up to a passage of 5. For experimental analysis, fibroblasts were cultured in Dulbecco's Modified Eagle Medium (DMEM; GIBCO) with 20% fetal calf serum (FCS; 191 GIBCO), 1% Penicillin/Streptomycin (P/S; 10,000 units/ml; GIBCO) and 1% MEM Non-essential Amino Acids (NEAA; GIBCO) up to a passage of 13.

HEK293T cells were cultured in DMEM with 10% FCS, 1% P/S and 1% NEAA, and selected with 200 µg/ml Hygromycin. All cells were kept at 37°C and 5% CO2.

HEK293T cell transfection

HEK293T cells were seeded at a density of $4x10^5$ cells and stably transfected with 4 µg of Vector[14] (R32-hRAD21 or R32-hRAD21 p.P298S or R32-hRAD21 p.P298A using Lipofectamine2000 (Invitrogen) and selected with Hygromycin (Invitrogen) at a concentration of 200 µg/ml for 7 days.

Immunoblotting

For whole-cell lysates, $7x10^6$ HEK293T cells stably overexpressing RAD21 WT, p.P298S or p.P298A were lysed in RIPA Buffer (50 mM TRIS, 150 mM NaCl, 0.5% Sodium Deoxycholate, 1% TRITON and 0.1% SDS 20%, with 10x PhosSTOP (PS, Roche) and 25x PIC (Protease Inhibitor Cocktail, Roche) freshly added), for 30 min on ice, while vortexing every 5 min. Whole-cell lysates of leukemia cells (8-10 x 10^6 cells) were lysed equally (3 biological replicates). For MycTaq validation HEK293T, HEK293T RAD21 WT, HEK293T RAD21 p.P298S and HEK293T RAD21 p.P298A were arrested with colchicine as described in "Immunoprecipitation" and lysed with 10 mM TRIS-HCl (pH 8.0), 1 mM EDTA, 100 mM NaCl, 2 mM MgCl2, 10% NP40, 25x PIC and 10x PS kept 30 min on ice with vortexing every 5 min. Protein concentration was measured with the Bradford protein assay (Roti-Quant, Roth) by determining the OD_{595nm}. Cytoplasmic and nuclear lysates were

prepared from $10x10^6$ cells using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermofisher, adapted for $10x10^6$ cells). 12 µg (for HEK cells samples) or 15 ug (for leukemia cell samples and MycTaq validation blot) were heated for 10 min at 95 °C while shaking at 350 rpm and loaded accordingly onto BIORAD Mini-Protean TGX Gel 4-20%. The blot was run cold (only 219 HEK samples) for 20 min at 70 V, following ~90 min at 130 V. Transfer was performed using the Trans-Blot Turbo Transfer System (High molecular weight, BIO-RAD, Trans-Blot Turbo 1x Transfer Buffer). The immunoblot was blocked in 5% cow milk (diluted in 222 TBS-T) at room temperature for 1h. After 3 washes with 1x TBS-T the HEK cell blot was incubated overnight at 4 °C with the following antibodies diluted in 5% Bovine Serum Albumin (Sigma): Myc-Taq (Cell Signaling #2278S, 1:250) and GAPDH (Cell Signaling #5174S, 1:1,000). The leukemia lysate blot was incubated with the following antibodies diluted in 5% bovine serum albumin (Sigma): WAPL (Cell Signaling #77428, 1:1000), RAD21 (Bethyl #A300-080A, 1:10,000), HSP90 (Cell Signaling #4877S, 1:1,000) and β-Actin (Santa Cruz Technology #B0719, 1:1,000). The Myc-Tag validation blot was incubated with Myc-Tag (Cell Signaling #71D10, 1:250) and β-Actin (Santa Cruz Technology #B0719, 1:1,000). The following day, the secondary antibody depending on primary species was applied after 3 consecutive washes (Cell Signaling Anti-Rabbit IgG #7075 1:1,000, Cell signaling Anti-Mouse IgG #7076) for 1 h in the dark, at room temperature diluted in 5% milk. After 3 consecutive washes the blot

was imaged after application of HRP linked solution (SuperSignal West Pico PLUS Chemiluminescent Substrate, Thermofisher). After stripping (Millipore Reblot Plus Strong Solution 10x) and re-blocking with 5% milk, the membrane was incubated with RAD21 (Bethyl #A300-080A 1:10,000) and Lamin B1 (Cell Signaling #12586, 1:1,000).

Immunoprecipitation

HEK293T cells stably overexpressing RAD21 WT, p.P298S or p.P298A were arrested in the metaphase. Cells growing exponentially were FCS deprived for 24h (cultured only in DMEM+ 1% Penicillin/Streptomycin) and then treated with 5 µg/ml colchicine (dissolved in DMSO). After 2 h all cells were collected and washed 3x with PBS. Nuclear lysates were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermofisher). Protein concentration was evaluated using Bradford assay. 100 µg per lysate was used for immunoprecipitation. Immunoprecipitation was performed using RAD21 Antibody (Bethyl concentration: 1,000 µg/ml, 4 µg per 100 µg of lysate) or Myc-Tag (Bethyl #A191-101, concentration 250 µg/ml, 2 µg per 100 µg of lysate) in 500 µl PBS with freshly added 25x Protease Inhibitor Cocktail (PIC) and 10x PhosSTOP (PS, Roche). Immunoprecipitation samples were rotated overnight at 4 °C. 16 h later 7 µl of each Dynabeads Protein A (Thermofisher #10001D) and Protein G (Thermofisher #10003D) were added and rotated at 4 °C for another 4 h. IP was performed using DynaMac-2 (Thermofisher #12321D) with 3 consecutive washes of

PBS/PIC/PS. Immunoprecipitation samples were heated like input samples to 95 °C for 10 min and with 350 rpm rotation. Immunoblotting was performed as described above using 15 µg of nuclear lysate as input control. Flow through and washes were immunoblotted equally for technical validation. After transfer and blocking, the membrane was incubated with the primary antibodies overnight at 4 °C: RAD21 (Bethyl #A300-080A 1:10,000), Myc-Tag (Bethyl #A191-101, 1:1000), Lamin B1 (Cell Signaling #12586, 1:1,000 or 1:500), SMC1 (Bethyl #A300-055A, 1:1,000), SMC3 (Bethyl #A300-060A, 1:5,000), SA1 (Bethyl #A300-157A, 1:1,000), SA2 (Bethyl #A300-158A, 1:1,000), PDS5B (Bethyl #A300-537A, 1:250), WAPL (Cell Signaling #D9J1U, 1:1,000).

Secondary antibodies were added as described above. Stripping of blots was performed with Reblot Plus Strong Solution (Millipore #2504) for 22 min.

Microarrays

Stably transfected HEK293T cells overexpressing RAD21 either WT, p.P298S or p.P298A were seeded onto 10 cm plates in a density of $2x10^6$ cells in quadruplicates. After 48 h, control cells were harvested and $6x10^6$ cells were pelleted and stored at -80 °C for later RNA extraction. RNA was extracted using the RNeasy Mini Kit (Qiagen #74106) with 350 µl of RLT Buffer+ BME using QIAshredder (#79656)

and RNAse-Free 273 DNase Set (Qiagen #79254). RNA was stored at -80 °C.

RNA samples were sent to Macrogen Europe B.V. (Amsterdam, Netherlands) for gene expression analysis using the SurePrint G3 Human Gene Expression 8x60K v3 276 microarray (Agilent, Inc., Santa Clara, CA). In short words, Cy3-labeled cRNA was prepared from 1~5 µg total RNA (Quick Amp Labeling Kit, Agilent), subsequently fragmented and (1.65 µg) hybridized to the microarray. Scanning was performed by the SureScan Microarray Scanner System G4900DA (Agilent).

For analysis, raw data were extracted using the software provided by Agilent Feature Extraction Software (v11.0.1.1). The raw data for the same probe was summarized automatically in the Agilent feature extraction protocol to provide expression data for each gene probed on the array. Flag A-tagged probes were filtered out and the remaining gProcessedSignal values were log transformed and quantile normalized.

Furthermore, all technical replicates (n=4) of one sample were combined and samples were compared pairwise by fold-change values: RAD21 p.P298A vs. WT, RAD21 p.P298S vs. WT and RAD21 p.P298A vs. RAD21 p.P298S. The p-value calculated with an independent Student‘s t-test was corrected for multiple testing and used to define the significance of these pairwise comparisons. Genes with an absolute fold-change of 1.5 or more and an adjusted p-value

below 0.05 were considered as significantly up-or down-regulated. These data (n=995 probes) were used to perform a two-dimensional hierarchical clustering using Euclidean distance and complete linkage. Results were represented as heat map (seaborn.clustermap with prior optimal leaf ordering, Python 3.6). The same analysis was performed for a smaller set (n=83 probes), which were differentially expressed in both mutants RAD21 p.P298A and p.P298S vs. WT was similarly analyzed and represented.

GO-Term Analysis

Gene Ontology (GO) term analysis was performed using the web server EnrichR (https://maayanlab.cloud/Enrichr/).[16] The set of genes significantly regulated in both mutants RAD21 p.P298A and p.P298S, comprised 83 probes and 69 annotated genes and was used as query data set for the analysis. GO terms of the categories “Molecular Function”, “Biological Pathway”, “Cellular Component” and “KEGG” were analyzed and results with an adjusted p-value < 0.05 are represented.

Irradiation and Cell Cycle Analysis

Fibroblasts were seeded in T25 cm2 bottles at a density of 1-2x10^5 cells and cultured at 37 °C and 5% CO_2. After reaching 70% confluency cells were irradiated with 6 Gy and 309 10 Gy. Negative control (0 Gy) was kept outside the incubator in the meanwhile. After 48 h apoptosis was

analyzed by propidium iodide staining and flow cytometry adapted from Riccardi and Nicoletti.[17] In short: cells were trypsinized (GibcoTrypLE Express), washed with PBS (Dulbecco's Phosphate Buffered Saline), centrifuged at 200xg for 5 min at RT and fixated with 70% ice-cold Ethanol/PBS for at least 20 min (up to 1 h) at -20 °C. Cells were then centrifuged (all centrifugation steps at 400xg, 4 °C) and washed with cold PBS. After centrifugation cells were treated with 500 µl of DNA extraction buffer (Nicoletti) and 500 µl of cold PBS. After 5 min incubation on ice cells were again centrifuged and resuspended in staining solution (PBS, 100 µg propidium iodide (BioLegend #421301), 1 mg DNA-free RNAse (Thermo Scientific #EN0531) and 2 mM EDTA (Nicoletti). Cells were incubated protected from light for at least 30 min before FACS analysis.

Mitomycin-C Treatment and Cell Cycle Analysis

Fibroblasts were seeded in 10 mm plates or T25 cm^2 bottles at a density of 1-3 x 10^5 cells and cultured at 37 °C and 5% CO_2. After reaching 70% confluency, cells were washed with PBS and treated with Mitomycin-C from *Streptomyces caespitosus* (Sigma #M4287) diluted in PBS to a final concentration of 0.1 mg/ml and incubated at room temperature for 5 min. The negative control was kept in PBS only. After 5 min cells were washed 3x with PBS and cultured with fresh media at 37 °C and 5% CO_2 for 24 h, before apoptosis was analyzed by

propidium iodide staining and flow cytometry as explained for "Irradiation and Cell Cycle Analysis".

Immunofluorescence Staining

Hek293T cells were plated onto Poly-L-Lysine pre-coated coverslips in 24-well-plates at a density of 6-8 x 10^4 cells and cultured for 48 h at 37 °C and 5% CO_2. Cells were fixed for 15 min in 3% formaldehyde/PBS, blocked with 0.25% Triton X-100/PBS and blocked in 1% bovine serum albumin/PBS for 30 min. Samples were incubated with primary antibodies for 1 h at RT. yH2AX and 53BP1 foci were detected using a mouse polyclonal anti-phospho-histone H2AX (Ser139) antibody (Millipore #05-636) at a dilution of 1:100 and a rabbit 53BP1 antibody (Novusbio #NB100-304) at a dilution of 1:1,000. Coverslips were further stained with secondary antibodies for 1 h at room temperature in the dark. The goat anti-mouse Alexa Fluor 488 IgG antibody (Invitrogen #A-11029) and the goat anti-rabbit Alexa Fluor 594 IgG antibody (Invitrogen #A-11037), were used as secondary antibodies, each at a dilution of 1:200. Slides were mounted in ProLong Diamond Antifade medium containing DAPI. Wide field microscopy was performed with a Zeiss Axio Observer microscope (CFCI, TU Dresden) using a Plan Apochromat objective. The DAPI images were used to detect signals inside the nuclei.

Single-cell RNA Sequencing

Healthy human bone marrow scRNA-seq data from eight donors was downloaded from 350 Human Cell Atlas (https://data.humancellatlas.org/explore/projects/cc95ff89-2e68-4a08-a234-480eca21ce79) and aligned to hg19 using Cell Ranger v3.0.0. Scanpy (https://doi.org/10.1186/s13059-017-1382-0) was used to characterize the cell types in the data, correcting for possible batch effects with Mutual Nearest Neighbors (https://doi.org/10.1038/nbt.4091) and filtering for outliers using median absolute deviation. Cell clusters found with Louvain clustering (https://zenodo.org/record/1054103) were mapped to cell types using known marker genes. Cell cycle phases were annotated by scoring cell cycle marker gene sets from https://doi.org/10.1101/gr.192237.115. Two-dimensional visualization was done with UMAP (https://doi.org/10.1038/nbt.4314).

Quantitative Real-Time Analysis

RNA was extracted from primary fibroblasts ($X10^7$, Case-18, and TRIO-DD_017 2.25- $5x10^6$ cells) using RNaeasy Mini Kit (Qiagen #74106) with 350 µl of RLT Buffer+ BME using QIAshredder (#79656) and RNAse-Free DNase Set (Qiagen #79254) or using the gDNA eliminator spin column. 3 biological seedings and RNA extractions were performed. RNA was equally extracted from HEK293T RAD21 WT, p.P298S and p.P298A from $4x10^6$ cells (3 biological replicates for HEK293T RAD21

WT, p.P298S, 3 technical replicates for p.P298A). 1 µg of RNA was transcribed into cDNA using the QuantiTect Reverse Transcription Kit following manufacturer's instructions. qRT-PCR was performed using TaqMan Universal Master Mix II following manufacturer's instructions (Thermofisher #PN4428173) for 20 µl reaction with 1.5 µl of cDNA. The following assays were used: TBP (Hs00427620_m1), HPRT1 (Hs02800695) and RAD21 (Hs01085854_mH).

Results

To add a novel piece to the understanding of Cohesins in cancer predisposition, we analyzed Whole Exome Sequencing data of an unselected German parent-child cohort of children with cancer (n=60, TRIO-DD), as well as a recently published parent-child pediatric cancer cohort (n=158, TRIO-D)[14] for germline variants in Cohesin complex genes (Supplementary Table 1). Overall, in both childhood cancer cohorts, 13 variants (minor allele frequency (MAF) <0.1%; gnomAD non-cancer database) in seven different Cohesin genes were identified (Figure 1B). All were transmitted from one of the parents, heterozygous, mutually exclusive and significantly enriched in Leukemia (6=lymphatic, 2=myeloid) and Lymphoma (n=3) patients as compared to patients with solid tumors within the cohort (Fisher's exact test; p=0.0081) (Figure 1C, Supplementary Figure 1). Thereof, CdLS phenotypes were observed in one AML patient carrying NIPBL

p.(G998E) (TRIO-D, Case-92) and in one precursor B-cell Acute Lymphoblastic Leukemia (BCP-ALL) patient harboring MAU2 p.(N410S) (TRIO-D, Case-74) (Supplementary Table 2). Interestingly, among all cohesin complex variants, one recurrently mutated nucleotide leading to an amino acid (AA) exchange at position of RAD21 was identified in 2 families (one per cohort) (Supplementary Figure 2, Supplementary Table 3 and 4). While the affected pediatric cancer patients carrying the recurrent *RAD21* variation did not show signs of CdLS, both three-generation pedigrees displayed a remarkable family history of early-in-life cancer (Figure 1D). In family I (Case-18), the heterozygous RAD21p.P298S variant was identified in a 13-year-old boy with T-ALL. His father died from breast cancer at the age of 41. Family II (TRIO-DD_017) displayed an alternative AA substitution at the same protein position (RAD21p.P298A), which was detected in a 2-year-old patient with precursor B-cell Lymphoblastic Lymphoma (pB-LBL). Here, the variant was inherited from the healthy father, whose brother had died from pediatric cancer of an unknown subtype at 8 years old. Deploying the Cancer Predisposition Sequencing Reporter (CPSR)[14] algorithm on the complete gene dataset of both families harboring RAD21p.P298S/A, no known cancer-related pathogenic variants were identified (Supplementary Figure 3).

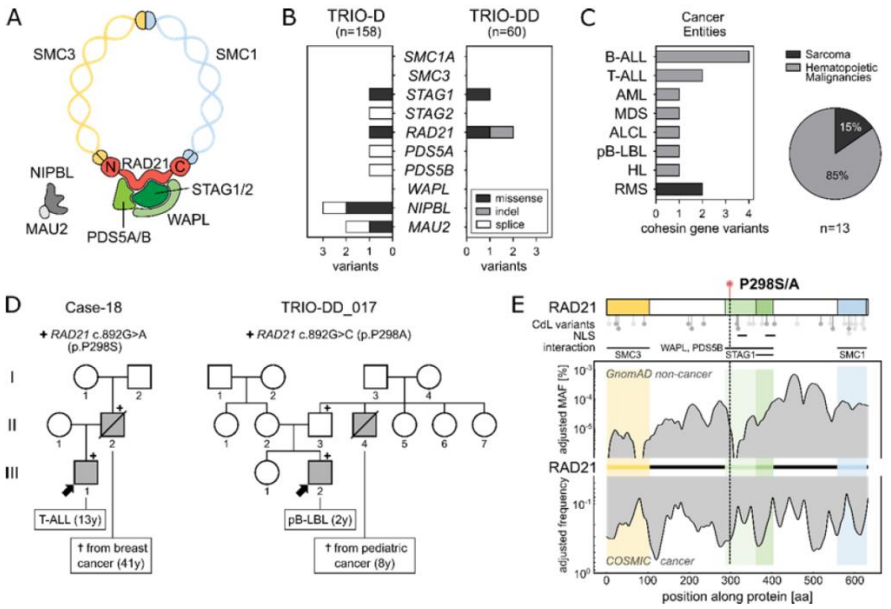


Figure 1. Identification of a recurrent RAD21 germline variation (p.P298A/S)

A: The cohesin complex is formed by the 4 main core units SMC1 and SMC3 connected by RAD21 and STAG1 or STAG2. WAPL and PDS5 as co-factors and NIPBL and MAU2 as loaders are depicted.

B: Two patient cohorts (TRIO-D: n=158 and TRIO-DD n=60) were analyzed for germline variants within Cohesin genes as depicted in Supplementary Table 1. Only non-synonymous variants with a MAF <0.1% 412 (gnomAD non-cancer population) were included.

C: Tumor entities of patients carrying a coding variant in one of the cohesin genes as shown in B (both cohorts combined, n=13). Hematological malignancies account for 84.6% of cancers in the patients with germline cohesin variants. Further cohesin variants were identified in 2 patients with rhabdomyosarcoma. ALL: Acute Lymphoblastic Leukemia, AML: Acute Myeloid Leukemia, MDS: Myelodysplastic syndrome, ALCL: Anaplastic large-cell Lymphoma, pB-LBL: precursor B-cell Lymphoblastic Lymphoma, HL: Hodgkin Lymphoma, RMS: Rhabdomyosarcoma

D: Family pedigrees of patients carrying the heterozygous germline RAD21variant p.P298S/A. Index patients are marked with an arrow. Family members affected by pediatric cancer are highlighted in grey. Variant carriers are marked with “+”.

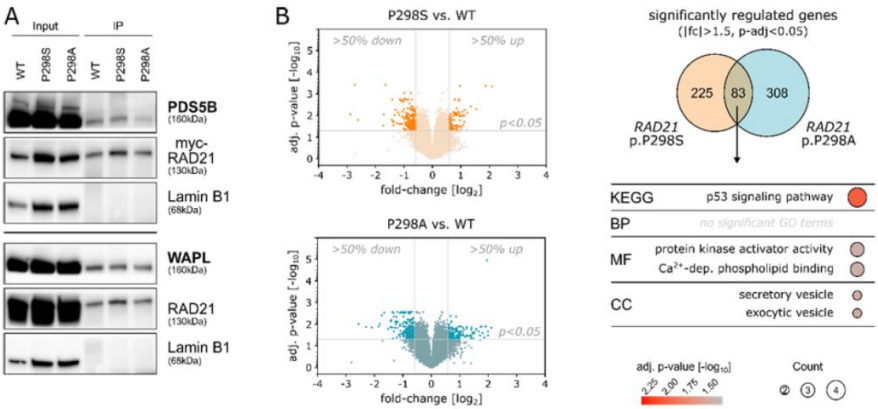
E: Upper: RAD21 protein structure displaying the interaction domains with SMC3 (1-103 amino acids 423 (AA)), WAPL and PDS5B (287-403AA), STAG1/STAG2 (362-403AA) and SMC1 (558-628AA). Lollipops below depict the positions of variants known in Cornelia de Lange (CdL) syndrome patients, adapted from *Krab et. al, 2020,*[16] with light gray representing missense variants and in-frame deletions and darker gray representing protein truncations. Lower: Distribution of variant frequencies along RAD21, based on two databases: The top shows the adjusted MAF (%) of *RAD21* germline variants in the gnomAD non-cancer database, while the bottom shows the adjusted frequency of variants in the COSMIC (somatic cancer mutations) database.

RAD21p.P298S/A is evolutionarily conserved across species (GERP-score 5.61, phastCons=1), located within the WAPL/PDS5B binding domain, and has not yet been reported in individuals with CdLS (16) (Figure 1E, Supplementary 434 Table 5). A link between RAD21p.P298S/A and cancer is supported by the general allele frequency of the AA position across non-tumor and tumor cohorts. While a low MAF at RAD21 p.P298 and its surrounding AA indicates that these positions are rarely mutated in the germline of the non-cancer population (gnomAD database n=118,479; MAF RAD21p.P298S <10-6 and p.P298A <10^{-5}), high somatic frequencies (COSMIC database n=37,221) are observed at the end of the SMC3 interaction domain and the start of the WAPL/PDS5B interacting domain, where the variants are located (Figure 1E). To further assess the structural impact of RAD21p.P298S/A, we aimed to generate a computational model of the 50 adjacent residues on each side. However, several homology modeling and *ab-initio* modeling approaches failed to generate a

secondary structure for this region, reflecting the substitution site as part of a very flexible and intrinsically disordered region (predicted disorder content of RAD21: 51.7%)[17] (Supplementary Figure 4).

Given that RAD21 p.298S/A affects a hyper-flexible domain, we next aimed to investigate its interaction with Cohesin complex partners at the protein level. Therefore, the identified *RAD21* variants were cloned and transfected into HEK293T cells. Compared to RAD21 WT, the variants RAD21 p.P298S/A did not affect protein expression or nuclear localization (Supplementary Figure 5)

Immunoprecipitation assays of the nuclear fraction showed binding of RAD21 with WAPL and PDS5B for the WT, as well as for both mutant proteins RAD21 p.298S/A, respectively (Figure 2A).



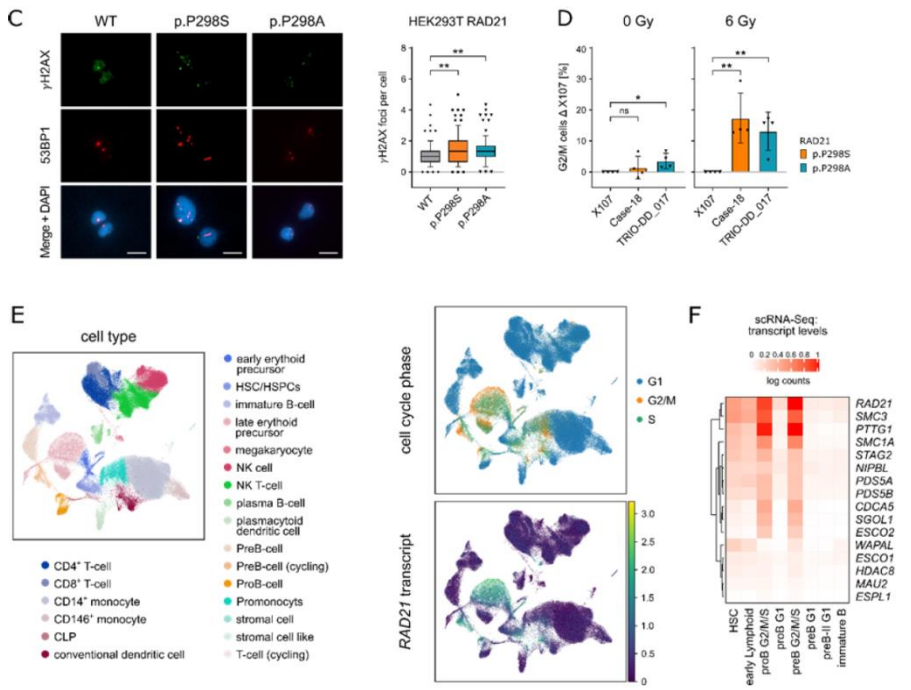


Figure 2. RAD21p.P298A/S alters gene expression and DNA damage response

A: Immunoprecipitation was performed on HEK293T cells overexpressing cMyc-tagged RAD21 WT, RAD21 p.P298S or RAD21 p.P298A. Cells were FCS deprived and after 24 h arrested with colchicine (0.5 µg/ml) for 2 h, while the nuclear fraction was used for immunoprecipitation with the cMyc-tag.

B: Volcano plot of average gene expression based on microarray data. Fold-change and adjusted p-values are calculated by comparing RAD21 p.P298A to WT (blue, left panel) and RAD21p.P298S to WT (orange, panel). Probes with >50% up- or downregulation and an adjusted p-value <0.05 are considered as differentially expressed (DE) and highlighted in dark blue (RAD21 p.P298A, left panel) or dark orange (RAD21 p.P298S, right panel). DE genes are compared between RAD21 p.P298S vs. WT and RAD21 p.P298A vs. WT and show an overlap >20%. GO-term analysis of shared DE genes from the previous analysis identified enriched GO-terms. All GO-terms that exceed the significance (Benjamini-Hochberg FDR < 0.05) are represented.

C: Left: representative images of gammaH2AX (green) and 53BP1 (red) foci. DAPI (blue) was used for DNA labelling. Scale bar: 10 µm. Right: quantification of gammaH2AX foci per cell in HEK293T RAD21WT, p.P298A and p.P298S cells. Experiments were performed as 3 independent replicates. Values are expressed in boxplots with whiskers from percentile 10-90. For the statistical analysis, Student's t-test was performed (**=p≤0.01).
D: X107 (healthy control), Case-18 (RAD21 p.P298S), and TRIO-DD_017 (RAD21 p.P298A) primary fibroblasts were subjected to irradiation with 6 Gy (n=4) and the cycle analyzed using propidium iodide staining. For indicated p-values, Student's t-testing was performed (*=p≤0.05; **=p≤0.01). Case-18 and TRIO-DD_017 were adjusted to $X10^7$ as a baseline response.
E: Left: UMAP-visualization of the healthy human bone marrow scRNA-seq data. Right: Cell cycle stages colored on the UMAP-visualization (upper) and *RAD21* gene expression colored on the UMAP-visualization (lower).
F: Heat map indicating the Cohesin complex genes' expression levels in cells of the different stages of B-cell differentiation.

Furthermore, the interaction of RAD21 WT and RAD21 p.P298S/A to SMC1 and STAG2 was not affected (Supplementary Figure 6), suggesting that RAD21 p.P298S/A does not perturb the formation of the Cohesin complex.

Since one major function of the complex is the control of transcriptional regulation through genome-wide chromatin organization,[18–20] we next tested the effect of RAD21 p.P298S/A on gene expression by microarray analysis in the cell line system described above. Hierarchical clustering of differentially expressed genes (fc >1.5, adj. p-value <0.05) showed a clear clustering of replicates and a separation of each condition (Supplementary Figure 7). In total, 308 and 391 genes were differentially regulated (fc >1.5, adj. p-value <0.05) in cells carrying the RAD21variants p.P298S/A. 83

genes were significantly up-/down-regulated in both RAD21 cell line models (Figure 2B, Supplementary Figure 8, Supplementary Table 6). GO term analysis of these genes identified “p53 signaling pathway” as the most prominent among enriched signaling pathways (Figure 2B). In line with these observations, HEK293T cells carrying RAD21 p.P298S/A showed an increased number of gammaH2AX and 53BP1 co-localized foci indicating the dimension of DNA double-strand breaks compared to the WT (**=p≤0.01; Student’s t-test) (Figure 502 2C) (Figure 2C).

Based on these results, we questioned whether patients carrying RAD21p.P298S/A would also display signaling abnormalities during normal and cellular stress conditions. Therefore, primary patient fibroblasts carrying the respective RAD21 p.P298S/A variants in comparison to RAD21 WT control fibroblasts were challenged through DNA damage and their response assessed via cell-cycle analysis. Both fibroblasts carrying RAD21 p.P298A and RAD21 p.P298S displayed significant differences to a WT control after ionizing irradiation, as calculated by Student’s t-test (298S: p=0.0049 [6Gy]; p=0.0026 511 [10Gy]; 298A: p=0.0054 [6Gy]; p=0.0006 [10Gy]) (Figure 2D; Supplementary Figure 9). Likewise, upon treatment with the DNA cross-linking agent Mitomycin-C (MMC), RAD21 p.P298S fibroblasts arrested more cells at the S/G2/M cell-cycle stage (p=0.0033; Student’s t-test) (Supplementary Figure 10).

Although RAD21p.P298S/A variants were only found in pediatric Lymphoblastic Leukemia/Lymphoma patients, the Cohesin complex is ubiquitously expressed. Furthermore, *RAD21* is expressed in human Leukemias, as observed in gene expression data across various hematological malignancies (Supplementary Figure 11). Thus, to identify vulnerable populations during hematopoietic differentiation, which are dependent on high *RAD21* expression and are potentially susceptible to RAD21 p.P298S/A, single-cell RNA-sequencing (scRNA-Seq) data of healthy human bone marrow from the Human Cell Atlas[21] was analyzed for Cohesin complex gene expression. In line with its essential role in mitosis, *RAD21* expression was primarily up-regulated in actively dividing cells within the G2/M or S-phase compared to cells in G1 ($p<2.2e-16$, Wilcoxon test) (Figure 2E, Supplementary Figure 12). Particularly high *RAD21* transcript levels clustered with *SMC3*, *PTTG1*, and *SMC1A* transcripts and were detected in pre- and pro-B-cells, while *RAD21* expression in common lymphoid progenitors (CLPs) and hematopoietic stem and progenitor cells (HS/PCs) was significantly lower ($p<2.2e-16$, Wilcoxon test) (Figure 2F and Supplementary Figure 13).

To confirm a correlation between germline RAD21 p.P298S/A and pediatric Leukemia/Lymphoma, we analyzed an additional unpublished pediatric cancer cohort of 150 children with relapsed ALL who were enrolled into the Italian IntReALL standard risk study (R-ALL) for RAD21p.P298S/A. Here, we identified a third case with

RAD21p.P298A in a boy who was diagnosed with B-cell precursor ALL (BCP-ALL) at 12 years old and had a combined bone marrow/CNS relapse 5 years later (Table 1).

		TRIO-DD	TRIO-D	R-ALL
Cohort	Number of patients	**n=60** pediatric	**n=158** pediatric	**n=150** pediatric
	% Hematopoietic malignancies	38.3%	51.3%	100%
	Inclusion criteria	Primary diagnosis	Primary diagnosis	IntReALL SR
Patient	Sex	Male	Male	Male
	Age	2	13	12
	Tumor	pB-LBL	T-ALL	BCP-ALL
	Risk group	SR	HR	SR
RAD21 variant p.P298	Protein exchange	ENSP00000297338.2 p.P298A	ENSP00000297338.2 p.P298S	ENSP00000297338.2 p.P298A
	Base exchange	ENST00000297338.2 c.892C>G	ENST00000297338.2 c.892G>A	ENST00000297338.2 c.892C>G
	SNP ID	rs148308569	rs148308569	rs148308569
	MAF GnomAD	10^{-5}	10^{-6}	10^{-5}
	MAF within the cohort	1.7×10^{-2}	6.5×10^{-3}	6.7×10^{-3}
Genetic history	Genetic counselling*	+	+	unknown
	Family history	+	+	unknown
2nd Hit	Somatic Mutations	unknown	*KRAS* p.Q61R	*KRAS* p.G12C

SR = standard risk
HR = high risk
* based on criteria from Jongmans*et al.,* Eur J Med Genet 59 (2016) 116-125 und Rippger*et al.*, Am J Med Genet A. (2017)

Table 1. Recurrent RAD21p.P298A/S variants in pediatric cancer cohorts
Cohort descriptions and identified *RAD21* variants analyzed in the context of clinical phenotypic and pathogenic findings. HR = High risk, SR = Standard risk, pB-LBL = precursor B-cell lymphoblastic lymphoma, T-ALL = T-cell Acute Lymphoblastic Leukemia, BCP-ALL = precursor B-cell Acute Lymphoblastic Leukemia.

In a fourth cohort including 114 children with therapy refractory leukemia and lymphoma (INFORM), no germline indels or missense variants affecting *RAD21* were identified, suggesting no enrichment in the relapsed or therapy refractory patients. To further cross-validate RAD21 p.P298S/A in a non-pediatric cancer setting, a cohort of 2300 young adults (<51 years) with cancer was mined (MASTER program). In this extensive sample collection, only one patient harboring RAD21 p.P298A with a solid tumor was identified (Supplementary Table 7). Therefore, amongst all cohorts, RAD21 p.P298S/A was found to be enriched in pediatric vs. adult cancers (3/479 vs. 1/2299; Fisher's exact test; p=0.018).

Discussion

Overall, within all analyzed datasets, comprising in total 368 primary pediatric cancer patients, 114 therapy refractory pediatric cancers and 2300 adult cancers as controls, we present three children with lymphoblastic leukemia/lymphoma all carrying a recurrent *RAD21* germline variation at position 298. None of the patients displayed a CdLS phenotype, which is in line with *RAD21* mediated CdLS phenotypes being rather mild.[22] The observed familial cancer history in two of the patients demonstrates an increased cancer risk across generations. Nevertheless, the healthy father in family II carrying RAD21 p.P298A also indicates incomplete penetrance of the variant,

highlighting the fact that additional factors (e.g. synergizing germline mutations or external influences) are necessary on top of the *RAD21* predisposition to drive tumor evolution. In agreement, we could show that the described variants caused differential regulation of p53 signaling with increased cell cycle arrest in primary patient cells. Likewise, *RAD21* variants have been previously described in radiosensitive cancer patients[23] and CdLS patients display increased DNA damage sensitivity.[24,25] Interestingly, in two patients carrying RAD21 p.P298A/S we identified known pathogenic somatic *KRAS* hot-spot mutation, which is in line with a recently published association between Cohesin complex mutations and RAS signaling in cancer progression.[26]

Taken together, in addition to loss-of-function *RAD21* germline and somatic variants that lead to Cohesinopathies and predominantly myeloid cancers, respectively, our data propose a third category of *RAD21* variants that mediate germline predisposition to lymphoblastic malignancies in childhood.

References

1. Uhlmann, F. SMC complexes: from DNA to chromosomes. *Nat. Rev. Mol. Cell Biol.***17**, 399–412 (2016).

2. Peters, J.-M. & Nishiyama, T. Sister chromatid cohesion. *Cold Spring Harb. Perspect. Biol.***4**, (2012).

3. Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat. Rev. Cancer***14**, 389–393 (2014).

4. Ström, L., Lindroos, H. B., Shirahige, K. & Sjögren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell***16**, 1003–1015 (2004).

5. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature***544**, 503–507 (2017).

6. Xu, H. *et al.* Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. *PLoS One***5**, e12112 (2010).

7. Fisher, J. B., McNulty, M., Burke, M. J., Crispino, J. D. & Rao, S. Cohesin Mutations in Myeloid Malignancies. *Trends in cancer***3**, 282–293 (2017).

8. Kon, A. *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.***45**, 1232–1237 (2013).

9. Fang, C., Rao, S., Crispino, J. D. & Ntziachristos, P. Determinants and role of chromatin organization in acute leukemia. *Leukemia***34**, 2561–2575 (2020).

10. Piché, J., Van Vliet, P. P., Pucéat, M. & Andelfinger, G. The expanding phenotypes of cohesinopathies: one ring to rule them all! *Cell Cycle***18**, 2828–2848 (2019).

11. Fazio, G. *et al.* First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. *J. Clin. Pathol.***72**, 558–561 (2019).

12. Waldman, T. Emerging themes in cohesin cancer biology. *Nat. Rev. Cancer***20**, 504–515 (2020).

13. Cheng, H., Zhang, N. & Pati, D. Cohesin subunit RAD21: From biology to disease. *Gene***758**, 144966 (2020).

14. Nakken, S. *et al.* Cancer Predisposition Sequencing Reporter (CPSR): A flexible variant report engine for high-throughput germline screening in cancer. *Int. J. cancer***149**, 1955–1960 (2021).

15. Wagener, R. *et al.* Comprehensive germline-genomic and clinical profiling in 160 unselected children and adolescents with cancer. *Eur. J. Hum. Genet.***29**, 1301–1311 (2021).

16. Krab, L. C. *et al.* Delineation of phenotypes and genotypes related to cohesin structural protein RAD21. *Hum. Genet.***139**, 575–592 (2020).

17. Mizianty, M. J., Peng, Z. & Kurgan, L. MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disord. proteins***1**, e24428 (2013).

18. Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell***154**, 801–813 (2013).

19. Merkenschlager, M. & Nora, E. P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.***17**, 17–43 (2016).

20. Nishiyama, T. Cohesion and cohesin-dependent chromatin organization. *Curr. Opin. Cell Biol.***58**, 8–14 (2019).

21. Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L. & Salomonis, N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp. Hematol.***68**, 51–61 (2018).

22. Yuan, B. *et al.* Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genet. Med.***21**, 663–675 (2019).

23. Severin, D. M. *et al.* Novel DNA sequence variants in the hHR21 DNA repair gene in radiosensitive cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.***50**, 1323–1331 (2001).

24. Revenkova, E. *et al.* Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. *Hum. Mol. Genet.***18**, 418–427 (2009).

25. Vrouwe, M. G. *et al.* Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Hum. Mol. Genet.***16**, 1478–1487 (2007).

26. Martín-Izquierdo, M. *et al.* Co-occurrence of cohesin complex and Ras signaling mutations during progression from myelodysplastic syndromes to secondary acute myeloid leukemia. *Haematologica***106**, 2215–2223 (2021).

Chapter 4

(Ready for submission)

Potential Role of *STAG1* Mutations in Genetic Predisposition to Childhood hematological malignancies

Claudia Saitta[1], Stefano Rebellato[1], Laura Rachele Bettini[1], Giovanni Giudici[1], Nicolò Panini[2], Eugenio Erba[2], Valentina Massa[3], Julia Hauer[4,5], Andrea Biondi[1,6], Grazia Fazio[1], Giovanni Cazzaniga.[1,7]

[1]Centro Ricerca M. Tettamanti, Pediatrics, University of Milano Bicocca, Monza, Italy.

[2]Department of Oncology Research, Istituto di ricerche farmacologiche Negri, Milan, Italy.

[3] Department of Health Sciences, University of Milan, Milan, Italy.

[4] Pediatric Hematology and Oncology, Department of Pediatrics, University Hospital "Carl Gustav Carus", TU Dresden, Dresden, Germany.

[5] National Center for Tumor Diseases (NCT), Dresden, Germany: German Cancer Research Center (DKFZ), Heidelberg, Germany; Faculty of Medicine and University Hospital Carl Gustav Carus, TechnischeUniversität Dresden, Dresden, Germany; Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Dresden, Germany.

[6]Pediatrics, University of Milano Bicocca, Fondazione MBBM/San Gerardo Hospital, Monza, Italy.
[7]Medical Genetics, University of Milano Bicocca, School of Medicine and Surgery, Monza, Italy.

Abstract

Background. Cohesins are a multi protein-complex that plays an essential role in sister chromatids segregation, post-replicative DNA repair and transcriptional regulation. Among Cohesins, STAG1/2 are fundamental for the maintenance of the structure.[1]
Germline mutations of Cohesin genes lead to Cohesinopathies, while somatic mutations have been described in myeloid malignancies and solid tumors.
Objectives. This study aims to define the association of germline variants of the Cohesin genes with predisposition to pediatric hematological malignancies.
Methods. A Lymphoblastoid Cell Line (LCL) was obtained from peripheral blood of a MDS patient affected by the Arg1187Gln variant in *STAG1*, then used for functional studies.
Genomic stability and X-ray induced DNA-damage repair were evaluated in the *STAG1*-mutated LCL by Sister Chromatids Exchange (SCE) assay, cell cycle analysis (Propidium Iodide) and pH2AX staining.
Results. We identified the STAG1 Arg1167Gln and the Arg1187Gln variants, located in a highly conserved region and frequently mutated

in solid tumors. We evaluated the effects of STAG1 Arg1187Gln on chromosomal stability, observing a higher number of abnormal chromatids exchanges in mutated LCL compared to control LCLs (*STAG1* patient mean = 4.8 exchanges/metaphase vs. 4 control LCLs mean = 3.05; p <0.0001), such as an increased percentage of double exchanges (*STAG1* patient = 22,17% vs. four control LCLs mean= 12,66%; p=0.0069).

Moreover, *STAG1* mutated LCL showed a significantly lower capability to repair DNA after an ionizing radiation. γH2AX phosphorylation status appears higher also in basal condition (T0: 1.7X, p <0.01; T24:2.2X, p <0.0001; T48: 2.4X, p <0.0001; MFI *STAG1* over MFI control LCLs) and it remains at higher levels 48 hours after irradiation (T48: 3.7X, p<0.001 [3Gy]; 4.1X, p<0.0001 [6Gy]; MFI *STAG1* over MFI control LCLs). These data are confirmed by the ratio of the *STAG1*-mutated LCL median MFI over controls (T48/T0: 3.42 vs. 0.92, p<0.0001 [3Gy]; T48/T0: 2.33 vs. 0.64, p <0.01 [6Gy]).

Conclusion. Our study confirms a considerable effect of a *STAG1* germline variant on genetic instability that could lead to oncogenesis, providing new strong evidence for Cohesins' contribution to genetic predisposition to hematological malignancies.

Introduction

Cohesin ring is a multi-protein complex that plays an essential role in a wide range of cellular processes: besides the canonical role in sister chromatids cohesion and segregation,[2] the complex gives a fundamental contribution in efficient DNA repair and maintenance of genome integrity.[3] It is also involved in transcriptional regulation and gene expression.[4,5] According to the protein they encode, Cohesin genes are classified as core cohesin subunits encoding (*SMC1A*, *SMC3*, *RAD21* and the paralog *STAG1*/*STAG2*, and cohesin regulatory factors (e.g. *NIPBL*, *HDAC8* and others).[6]

Among these, *STAG1* is a key subunit of the complex, essential for chromatids cohesion.[2] It is involved in the separation of sister telomeres during mitosis and have the capacity to bind DNA independently from the cohesin ring.[7]

Germline mutations of Cohesins lead to Cohesinopathies,[8] while recurrent somatic mutations in multiple component of the complex are known to represent genetic drivers in myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML),[9–11] as well as solid tumors.[12] Evidence suggested that mutations of Cohesin genes cause alterations in the balance between hematopoietic stem and progenitor cells (HSPC), distracting transcriptional programs of differentiation.[13]

A correlation between Cohesinopaties and cancer predisposition has not been established yet. However, the report of one single case of Down syndrome-like Acute Megakaryoblastic Leukemia (AMKL)[14] and

the first case with Acute Lymphoblastic Leukemia (ALL) in two Cornelia de Lange patients (CdLS),[15] supports the hypothesis that germline mutations in Cohesins could constitute a predisposing factor to leukemia.

Thus, the present study aims to identify and characterize germline Cohesins variants in pediatric patients affected by hematological diseases, such as B-ALL and MDS. Moreover, these findings could provide new insights in understanding of hematological oncogenesis, with potential applications in clinical practice in term of patients' management and surveillance.

Material and methods

Ethic statement

Samples were obtained from healthy donors and patients, with a written informed consent from patients or legal representatives. The study has been conducted in accordance with the ethical standards of the Declaration of Helsinki and to national and international guidelines. The study is approved by each institutional review board.

Next Generation Sequencing and Bioinformatic data analysis

Next Generation Sequencing (NGS) experiments have been performed by a targeted custom Nextera Flex DNA panel, on bone marrow (BM)

or peripheral blood (PB) samples of hemato-oncological samples referred to our institution. Germline variants on Cohesin genes have been investigated both in disease and remission samples. Sequencing has been performed by Nextseq550 (Illumina) in 2x150 paired end. FASTQ files are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress).

Bioinformatic analysis was carried out by Sophia DDM software. Variants were filtered by variant fraction (VF) >5% and coverage at least 500X; Variant Allelic Fraction (VAF) in the population was set at 1%. We included certainly pathogenic, potentially pathogenic and variants of unknown significance (VUS).

The most common databases of prediction were consulted for the interpretation of the pathogenicity, including: ClinVar, Clinical Genome, Varsome, InterVar, COSMIC. Benign/likely benign variants in all databases of prediction were excluded from the results. (Update September 2021).

Lymphoblastoid Cell Lines

Lymphoblastoid cell lines (LCLs) were derived from *in vitro* transformation and immortalization of B lymphocytes in fresh peripheral blood by Epstein Barr virus (EBV), by BioBank Service at Gaslini Hospital (Genova, Italy).

Cells were grown in T25 flasks in RPMI medium added with 10% FBS, 1% Pen-Strep and 1% L-glutamine, in standard incubation conditions (37 °C, 5% CO_2).

RT-PCR and variants validation

RT-PCR was performed using forward primer and reverse primer of the specific *STAG1* and *STAG2* involved exons, to amplify the mutated regions (Supplemental Table S1).

All the RT-PCR reactions were performed in the following conditions: one step (2' at 94°C), thirty-five cycles of amplification (30 s at 94°C, 30 s at 60°C, 60 s at 72°C), using Platinum SuperFi II DNA Polymerase–High-Fidelity PCR Enzyme (Life Technologies, Thermofisher, Carlsbad, CA, USA).

Phenotype characterization

A flow cytometry antibody panel was developed to characterize LCL B-cell phenotype, including specific antibodies for B-cells, T-cells and myeloid cells markers, such as CD10 (APC, #332777, Becton Dickinson™, Franklin Lakes, New Jersey, US), CD19 (FITC, #11-0199-42, eBioscience™), CD45 (PO, #MHCD4530, Invitrogen™, Waltham, Massachusetts, US), CD3 (Alexa700, #557943, Becton Dickinson™), CD13 (PE, #347406, Becton Dickinson™) and CD33 (PeCy7, #333952, Becton Dickinson™) in addition to the stemness marker CD34 (PerCPCy5.5, #347222, Becton Dickinson™). After 30' incubation (RT,

in the dark), cells were washed and resuspended in 200 µl of PBS and analyzed with BD LSRFortessa™ X-20 Flow Cytometer and BD FACSDiva™ software (BD Biosciences).

Growth Curve

A growth curve experiment has been set up to evaluate and compare the growth rate of different lymphoblastoid cell lines. LCLs were seeded at different concentrations according to their growth characteristics, previously established: $0.1x10^6$/ml for CTR3-8F_LCL, $0.22x10^6$/ml for CTR6-9M_LCL and $0,18x10^6$/ml for STAG1_LCL (MW6 plates). Cells were collected after 72, 96 and 120 hours after seeding, considering independent wells for each timepoint. Live cells were counted by Trypan Blue exclusion both through Countess Automated Cell Counter (Thermo Fisher, Carlsbad, California, United States) and Burker' counting chamber at optical microscope, in parallel. Detailed data in supplementary Fig. S2.

X-ray irradiation

LCLs underwent a cycle of X-ray irradiation performed with the RADGIL instrument (Gilardoni SpA, Mandello del Lario, Italy). Two irradiation conditions were selected: 3 Gy (190 V, 12 A, 5.5') and 6 Gy (190 V, 12 A, 11').

Cell Cycle Assay

Basal and irradiated cells were collected in polypropylene tubes in quantities of $2x10^6$/ml. After a centrifugation (1800 rpm, 5') the pellet was resuspended on ice in 1 mL of GM saline buffer (Glucose 1.1 g/l, NaCl 8 g/l, KCl 0.4 g/l, $Na_2HPO_4.2H_2O$ 0.2 g/l, KH_2PO_4 0.15 g/l, EDTA 0.5M 0.2 g/l). 1.5 mL of 96% Ethanol were then added under stirring for each sample. The fixed samples, stored at + 4 ° C or at -20 ° C, were centrifuged (1200 rpm, 10') and then washed with 1 mL PBS. They were subsequently incubated overnight in the dark with 1 mL solution of Propidium Iodide (2.5 µg/mL) and 12.5 µl of RNase (1 mg/mL). Flow cytometry analysis was executed using the BD LSRFortessa™ X-20 instrument and BD FACSDiva™ software. Cell cycle analysis was performed on at least 20.000 cells for each. Cell cycle phase distribution was calculated as percentages by a Gaussian-modified method.[16] Detailed data in supplementary Fig. S3.

Sister Chromatids Exchange Assay

1 ml of LCLs cell culture suspension ($1x10^6$ cells in 5 ml RPMI 10% FBS at conc. $0.3x10^6$/ml) was added with 7 ml of medium and 250 µl of Phytohemagglutinin that stimulates the growth of T lymphocytes. After an incubation at 37°C for 24h, 80 µl of a 1 µg/µl BrdU stock solution was added, and samples were incubated at 37°C for 48h.

During the following two days, cells grow and replicate, new synthesized DNA won't be marked with BrdU, allowing the

visualization of chromosomal exchanges. Colchicine must be added to block the mitotic spindle during the metaphase. Samples were incubated for 1,5 h and then transferred into 15 mL Falcon tubes and centrifuged at 1800 rpm (10'). Cells were resuspended in 7 mL of hypotonic solution (KCl 0.08 M) and incubated at 37°C for 15'. 1 mL of fixative solution (methyl alcohol and acetic acid in ratio 3: 1) was added to the samples which were centrifuged at 1800 rpm (10') and then resuspended in 7 mL of fixative. In the end, the pellet was resuspended in 2 mL of fixative solution and smeared on a cold glass slide.

The slides are stained with 10 µl of Hoechst (1:5000) each and incubated for 20'. Now it is possible to visualize the frequency of abnormal chromatids exchanges within the single chromosomes through a fluorescence microscope. The entire protocol lasts 5 days.

pH2AX level evaluation

To investigate the capability of LCLs to repair after DNA double-strand breaks (DSBs) induced by an ionizing radiation, we evaluated the phosphorylation level of γH2AX, a DSB marker, by FACS analysis.

Cells were seeded in MW6 at the same concentration and conditions used for the growth curves, in order to perform the experiments in exponentially growing phase.

Approximately $1x10^6$ of basal or irradiated cells per sample was collected in FACS tubes and centrifuged at 1200 rpm (5'). Cells were

resuspended in 1 mL of PBS and 2 mL of fixative solution (4.5% PFA/PBS, 3% final concentration). Samples were incubated for 10 min (RT). After a centrifugation (1200 rpm, 5') the pellet was resuspended in 3 mL of cold Ethanol 70% and vortexed briefly.

To remove the ethanol, pellet was washed 3 times in 3 mL of washing solution (0.5% BSA/PBS) and cells were resuspended with Phospho-Histone H2AX antibody (Alexa Fluor® 488 Conjugate - BD #9719) and incubated for 1h (RT). Samples were centrifuged at 1800 rpm for 5 minutes and resuspended in 200 µl of PBS.

Flow cytometry analysis was performed using the BD LSRFortessa ™ X-20 instrument and BD FACSDiva software.

Statistical analysis was performed by Graphpad Prism software ver.9. 2way ANOVA test with Bonferroni's multiple comparisons is shown as *p<0.05, **p<0.01, ***p<0.001.

Results

Identification and Sanger validation of *STAG1/2* variants

Based on results obtained by our previous NGS screening of 120 consecutive diagnoses of ALL and sporadic cases with familial recurrence of cancer or other hematological disorders, we focused on *STAG1* gene. We identified 2 germline variants in a patient affected by B-ALL and in child with MDS, respectively. Clinical characteristics of patients are summarized in Table 1.

	Gender	Age (y)	Diagnosis	Kariotype	Translocation	Relapse	HSCT
Pt 1	Male	2	ALL-BII	Euploid	Negative	Yes	Yes
Pt 2	Male	14	MDS EB-1	Trisomy Chr8, 16q deletion	Negative	n.a.	Yes

Table 1 Clinical characteristics of *STAG1* mutated patients.
(ALL: Acute Lymphoblastic Leukemia; MDS: Myelodysplasia; EB-I: Excess of Blast; n.a.: not applicable; HSCT: Hematopoietic Stem Cells Transplantation)

Interestingly, the variants are located in a highly conserved region of the gene, frequently affected by mutations already known as implicated in oncogenesis in databases of prediction (https://pecan.stjude.cloud/STAG1) (Fig.1, A).

The Arg1167Gln (c.3500G>A; rs747617236) missense variant was identified in a B-ALL patient; it is classified as a variant of uncertain significance (VUS) in InterVar and Varsome (not reported in ClinVar).

The Arg1187Gln (c.3560G>A; rs777032446) missense variant was found in a MDS pediatric patient, and it is predicted as likely pathogenic variant in Varsome/VUS in InterVar (not reported in ClinVar).

The MDS patient is characterized also by the somatic Arg953* variant (c.2857C>T) in the paralog *STAG2* gene, annotated in InterVar, Varsome and COSMIC as pathogenic (not reported in ClinVar) and responsible of different cancer types (Fig.1, E).

In order to set up an *in vitro* model to investigate the potential role in predisposition of germline Arg1187Gln on *STAG1* gene, a

Lymphoblastoid Cell Line (LCL) was generated through the immortalization of peripheral blood-B lymphocytes.

We amplified and analyzed *STAG1* and *STAG2* mutated sequences by qualitative PCR and Sanger Sequencing, comparing bone marrow sample of *STAG1/2* mutated patient, *STAG1* mutated LCL and a LCL obtained from a healthy donor (CTR1-42M_LCL as representative control). Thereby we confirmed that STAG1_LCL have maintained the genetic profile after immortalization. As show in chromatograms, *STAG1* c.3560G>A germline variant (rs777032446) affects bone marrow of mutated patient (panel B), and it is conserved in STAG1_LCL (panel C) and it is absent is CTR1_LCL (panel D).

The somatic c.2857C>T variant on *STAG2* gene is present only in patient' bone marrow (panel E), whereas the sequence is wild type in STAG1_LCL (panel F) and CTR1-42M_LCL (panel G). The low variant fraction of somatic mutation doesn't allow the immortalization in the mutated LCL. This aspect ensures that potential mutational effects on biological mechanisms are caused only by the *STAG1* germline variant.

Moreover, to evaluate the correlation between the variants and cancer, we considered the general allele frequency of the mutated positions across non-tumor and tumor cohorts. As showed in figure 1I, MAF at STAG1 p.1167 and p.1187 indicates that these mutations are rare in general population (gnomAD database MAF *STAG1* Arg1167Gln $<10^{-5}$ and Arg1187Gln $<10^{-6}$), while they are relatively frequent in

cancer cells. Analogously, a low MAF at STAG2 R953* somatic variants indicates that this position is rarely mutated in the germline of the non-cancer population (gnomAD database MAF STAG2 R953* $<10^{-5}$), instead is high the somatic frequencies on COSMIC database. (Fig.1L)

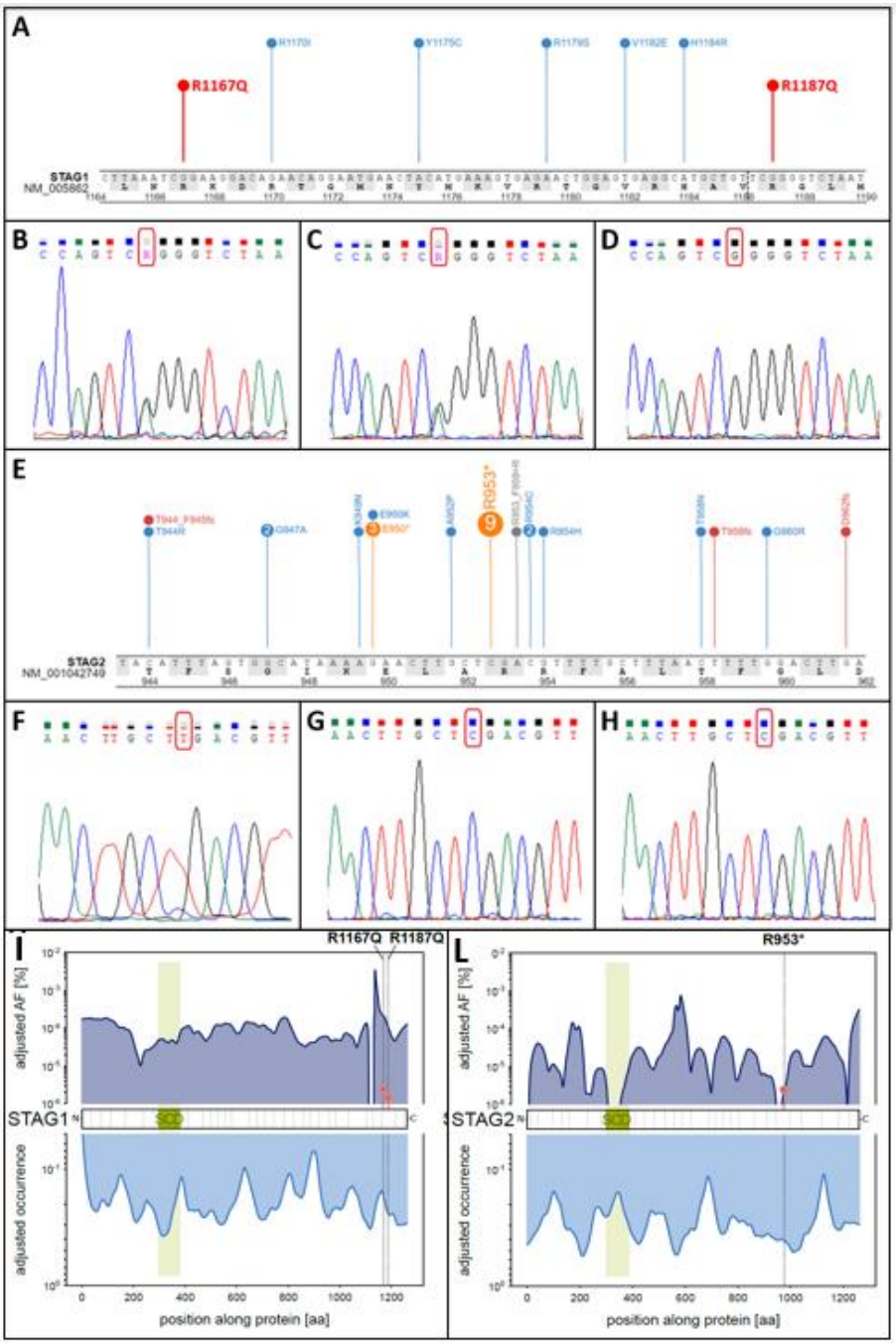

A
R1167Q
R1187Q
STAG1
NM_005862
B
C
D
E
R953*
STAG2
NM_001042749
F
G
H
I
R1167Q R1187Q
adjusted AF [%]
STAG1
adjusted occurrence
position along protein [aa]
L
R953*
STAG2
adjusted occurrence
position along protein [aa]

Figure 1. Graphic representation of variants in *STAG1* and *STAG2* region tables and validation on Sanger sequencing.
In panel A, the ALL-mutation (R1167Q) and the MDS-mutation (R1187Q) (in red) and the other variants previously described (in blue) on STAG1. In panels B-C-D, chromatograms of bone marrow MDS patient, STAG1_LCL and wild type CTR1_42M-LCL samples respectively. In panel E, somatic STAG2 variant (R953*) together with the others present in COMISC database. In panels F-G-H, chromatograms of bone marrow MDS patient, wild type STAG1_LCL and wild type CTR1_42M-LCL samples respectively. In panel I-L distribution of variants frequencies along *STAG1* (on the left) and *STAG2* (on the right), based on two databases: the top shows the adjusted MAF (%) of variants in the gnomAD non-cancer database, while the bottom shows the adjusted frequency of variants in the COSMIC (somatic cancer mutations) database.

Increased number of abnormal chromatid exchanges in *STAG1*_LCL

To investigate the functionality of Cohesins' complex on DNA stability, we firstly evaluated the status of chromatin exchanges during the second mitotic division.

All LCLs have been stained with *BrdU* (T24) and thus, blocked in metaphase (T72). Fluorescence microscopy after *Hoechst* staining showed that STAG1_LCL are characterized by a higher number of abnormal chromatin exchanges. The average number of exchanges in each nucleus is equal to 4,8 for STAG1_LCL, while the four control LCLs mean is 3,05 (range 2,66 for CTR6-9M_LCL-3,50 for CTR3-8F_LCL). Statistical analysis confirmed the significance of the result ($p<0,0001$-Bonferroni's multiple comparison Test) (Fig.2, A).

The results are significative also considering the percentage of cells that had one or more chromosomes with double exchanges. The

percentage of double exchanges in STAG1_LCL is equal to 22,17% and it is higher than the mean of percentage of the other control LCLs (mean 12,66%; range from 4,57% for CTR6-9M_LCL to 19,09% for CTR5-39M_LCL, p=0,0069) (Fig.2, B).

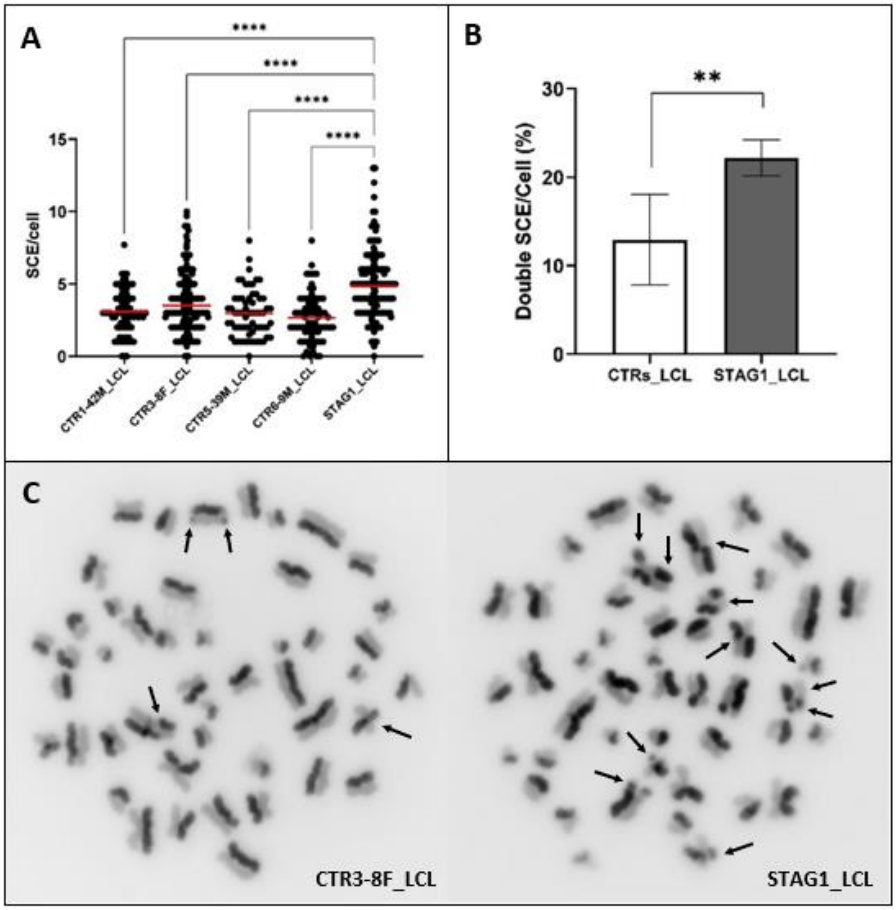


Figure 2. Sister Chromatids Exchange (SCE) incidence in LCL cells.
Panel A shows the higher number of abnormal SCE in STAG1_LCL compared to controls LCLs. In panel B the percentage of cells with double exchanges significantly higher in mutated cells. Panel C shows representative

metaphases with single/double abnormal chromatid exchanges observed at fluorescence microscopy (on the left CTR3_8F vs STAG1_LCL on the right). Average of 76 metaphases for each line. (Statistical analysis performed by One-way Bonferroni's multiple comparison correction. * <0,05; **<0,01 ***<0,001; ****<0,0001)

Increased STAG1_LCL γH2AX phosphorylation status reflect a defective capability of cells to repair DNA after an ionizing radiation.

We evaluated the phosphorylation level of histone γH2AX, a common marker of DNA double strand breaks damage.[3,17] γH2AX phosphorylation status in *STAG1* mutated LCL appears higher in basal condition (T0: 1.7X, p <0.01; T24: 2.2X, p <0.0001; T48: 2.4X, p <0.0001; MFI *STAG1* over MFI control LCLs) (Fig.3, A). This differential status is emphasized after an ionizing radiation and we applied different doses, ranging from 3Gy to 6Gy. In Figure 3B, a representative experiment at 3Gy is shown, demonstrating a significantly lower capability of *STAG1* mutated cells to repair after a DNA damage compared to controls' LCLs (T48: 3.7X, p<0.001 [3Gy]; MFI *STAG1* over MFI control LCLs).

These data are confirmed by the ratio of the STAG1_LCL median MFI over controls, significantly higher at timepoint 48 hours after irradiation (T48/T0: 3.42 vs 0.92, p<0.0001 [3Gy]).

So, while controls' LCLs present a reduction of γH2AX phosphorylation 48 hours after irradiation (successful DNA damage repair), STAG1_LCL has an increased phosphorylation during the days after irradiation, supporting a defective DNA repair capability, as shown in Figure 3, C.

The evaluation of pH2AX positive cells percentage further confirmed that *STAG1* mutated cells exibited higher phosphorylation levels after irradiation compared to control LCLs (T24: 1.6X, p=0.01 [3Gy]; $pH2AX^{+}$ cells *STAG1* over $pH2AX^{+}$ cells CTR6-9M_LCL) (T48: 1.7X, p <0.01 [3Gy]; $pH2AX^{+}$ cells *STAG1* over $pH2AX^{+}$ cells control LCLs) (Fig.3, D).

In addition, a highly positivepH2AX subpopulation (namely $pH2AX^{++}$) can be discriminated only in STAG1_LCL even in absence of irradiation (10.5X, p <0.01; $pH2AX^{++}$ cells *STAG1* over $pH2AX^{++}$ cells control LCLs). The results obtained after irradiation showed that $pH2AX^{++}$ population displays an increasing level of phosphorylation, expression of an inefficient capability to repair after DNA damage. (T0: 6.2X, p <0.01; T24: 10.6X, p <0.0001; T48: 14.1X, p <0.0001; $pH2AX^{++}$ cells *STAG1* over $pH2AX^{++}$ cells control LCLs), as shown Figure 3, E.

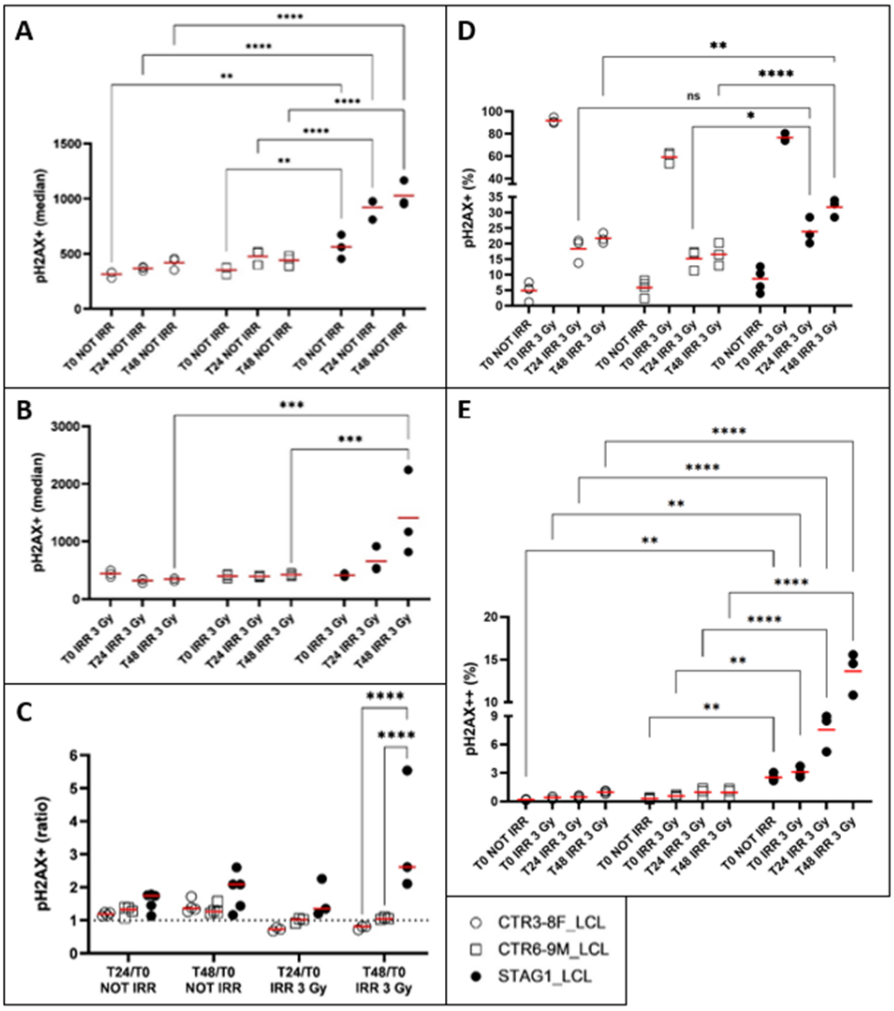


Figure 3. γH2AX phosphorylation status before and after an X-ray irradiation.

Panel A shows that γH2AX phosphorylation status of STAG1_LCL in basal conditions is higher and increases during timepoints compared to control LCLs. It remains at higher levels also after an irradiation [3Gy] (panel B). This data is confirmed by ratio values (panel C). * Comparable results are obtained expressing the γH2AX phosphorylation levels as percentage of $pH2AX^{+}$ cells (panel D). The $pH2AX^{++}$ subpopulation, present only in STAG1_LCL either in absence or after irradiation, shows the same trend (panel E).

(Statistical analysis performed by One-way Bonferroni's multiple comparison correction. * <0,05; **<0,01 ***<0,001; ****<0,0001)
*(Ratio levels around 1 indicates similar between the two timepoints considered (T0/T24 or T0/T48), instead ratio levels > 1 indicates that γH2AX phosphorylation levels are higher at T24 or T48 over T0).

These same findings have been obtained also after 6Gy irradiation (data shown in supplementary data file, figure S4). We assessed the consequences of a higher X-ray dose irradiation (10Gy) and we detected a too toxic effect with a higher pH2AX phosphorylation in STAG1_LCL population; however, the irradiation intensity caused an intolerable DNA damage that drives cells to apoptosis instead of DNA repair mechanisms, both in control and STAG1 LCLs (data not shown).

Discussion

We identified two germline variants on *STAG1* gene that are located in a highly conserved region characterized by numerous variants involved in solid tumors (https://pecan.stjude.cloud/STAG1). The *STAG1* variant (Arg1187Gln) that affect a pediatric MDS patient was further characterized in LCL with the aim to investigate the potential effects on DNA instability that are common in oncogenesis.
In support of our hypothesis, it has been recently demonstrated that depletion of *STAG1*- in *STAG2*-mutated cancer cells results in increased susceptibility to DNA damage and defects in DNA repair mechanisms,

confirming the strict interaction between the two paralogs and their role in DNA stability.[18]

We are the first to explore specifically the role of germline *STAG1* variants in oncogenesis, evaluating how they can corrupt a pre-leukemic clone, making it genetically instable and more prone to further somatic mutations.

Our results demonstrated that *STAG1*-mutated LCL has a higher number of abnormal chromatids exchanges, both single and double exchanges, compared to control LCLs. These characteristics are a common indicator of poor chromosomal strength and spontaneous chromosome instability, associated with failure of DNA repair mechanisms and accumulation of DNA damage. Similarly, SCE have been already found increased in other kinds of familiar cancers, such as *BRCA1/2* breast cancer.[19]

Moreover, mutated cells display increased DNA damage sensitivity, with a significantly lower DNA repair capability after X-ray irradiation. The phosphorylation status of γH2AX, a double strand breaks (DSBs) marker, is higher even in absence of a damage stimulus, confirming that *STAG1* mutation is responsible of the increased vulnerability and lowered response rate to exogenous and endogenous agents. Our results are consistent with what described by Bauerschmidt et al. Specifically the authors demonstrated that repair of radiation-induced DNA DSBs was reduced in *SMC1*- or *RAD21*-depleted cells.[20]

Even if the variant seems to not affect Cohesin capability to regulate cell cycle, the effects of DNA stability and DNA damage repair mechanisms, commonly altered in oncogenesis, have been proven. Considering that these are predisposing and non-founding variants in leukemogenesis, it seems justifiable that they have an effect more on genetic instability rather than biological mechanisms, such as growth capacity and cell cycle, which are associated with aberrant proliferative states typical of full-blown disease.

Taken together, our study provides strong evidence in support of the involvement of *STAG1* germline variants in predisposition to onco-hematological disease in childhood.

References

1. Canudas, S. & Smith, S. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.***187**, 165–173 (2009).

2. Solomon, D. A., Kim, J. S. & Waldman, T. Cohesin gene mutations in tumorigenesis: From discovery to clinical significance. *BMB Rep.***47**, 299–310 (2014).

3. Litwin, I., Pilarczyk, E. & Wysocki, R. The emerging role of cohesin in the DNA damage response. *Genes (Basel).***9**, (2018).

4. Brooker, A. S. & Berkowitz, K. M. *The roles of cohesins in mitosis, meiosis, and human health and disease*. *Methods in Molecular Biology* vol. 1170 (2014).

5. Tothova, Z. *et al.* Cohesin mutations alter DNA damage repair and chromatin structure and create therapeutic vulnerabilities in MDS/AML. *JCI Insight***6**, (2021).

6. Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.***22**, 3089–3114 (2008).

7. Romero-Pérez, L., Surdez, D., Brunet, E., Delattre, O. & Grünewald, T. G. P. STAG Mutations in Cancer. *Trends in Cancer***5**, 506–520 (2019).

8. Piché, J., Van Vliet, P. P., Pucéat, M. & Andelfinger, G. The expanding phenotypes of cohesinopathies: one ring to rule them all! *Cell Cycle***18**, 2828–2848 (2019).

9. Jann, J.-C. & Tothova, Z. Cohesin mutations in myeloid malignancies. *Blood***138**, 649–661 (2021).

10. Thota, S. *et al.* Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood***124**, 1790–1798 (2014).

11. Kon, A. *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.***45**, 1232–1237 (2013).

12. Solomon, D. A. *et al.* Frequent truncating mutations of STAG2 in bladder cancer. *Nat. Genet.***45**, 1428–1430 (2013).

13. Sasca, D. *et al.* Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood***134**, 2195–2208 (2019).

14. Vial, Y. *et al.* Down syndrome-like acute megakaryoblastic leukemia in a patient with Cornelia de Lange syndrome. *Haematologica* vol. 103 e274–e276 (2018).

15. Fazio, G. *et al.* First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. *J. Clin. Pathol.***72**, 558–561 (2019).

16. Ubezio, P. Microcomputer experience in analysis of flow cytometric DNA distributions. *Comput. Programs Biomed.***19**, 159–166 (1985).

17. Sharma, R., Lewis, S. & Wlodarski, M. W. DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Front. Pediatr.***8**, 570084 (2020).

18. Liu, Y. *et al.* Somatic mutation of the cohesin complex subunit confers therapeutic vulnerabilities in cancer. *J. Clin. Invest.***128**, 2951–2965 (2018).

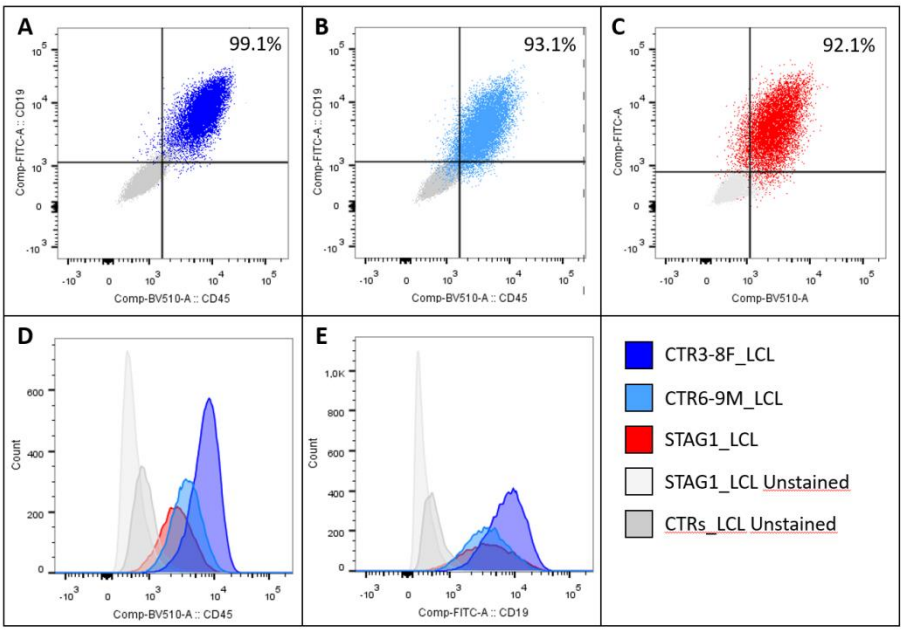
19. De Pascalis, I. *et al.* Sister chromatid exchange: A possible approach to characterize familial breast cancer patients. *Oncol. Rep.***33**, 930–934 (2015).

20. Bauerschmidt, C. *et al.* Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Res.***38**, 477–487 (2010).

SUPPLEMENTARY RESULTS

LCL phenotype characterization

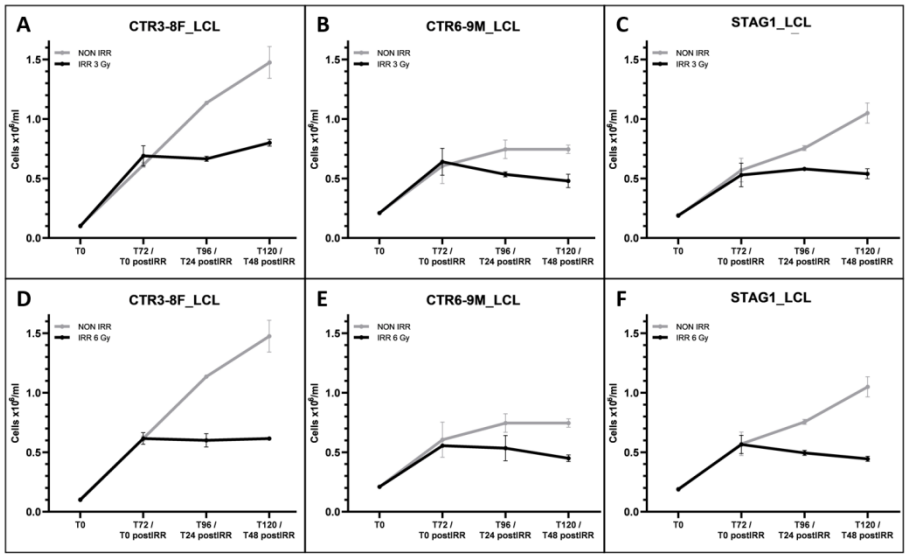
In order to assess that LCLs have maintained the B-lineage profile after EBV immortalization, a flow cytometry antibody panel was developed to characterize their phenotype. We evaluated specific markers of hematopoietic subpopulations, including common lymphocyte markers (the pan-leukocyte hCD45, hCD19 and hCD10 for B-cells and hCD3 for T-cells), myeloid markers (hCD13 and hCD33) and a stemness marker (hCD34). The results show a marked positivity against hCD45 and hCD19 antibodies, confirming the immortalization of the B-cell subpopulation. The results are comparable in all LCLs tested, derived both from healthy donors and mutated patient (Fig. S1).



Suppl. Fig.1 Phenotype characterization on LCLs. In panels A-C dotplots of $hCD19^+/hCD45^+$ cells in CTR3-8F_LCL, CTR6-9M and STAG1_LCL respectively. In panels D-E are represented overlay histograms of $hCD45^+$ and $hCD19^+$ cells in each LCL, compared to unstained LCLs.

STAG1_LCL and control LCLs growth is affected by X-ray irradiation.

To evaluate the different grow rates of LCLs, firstly, we tested different seeding cells for each line, in order to identify the best individual conditions that guarantee for each the exponential phase in the same timepoint. On the bases of the results, we set up the experimental conditions to compared STAG1_LCL with mean of control LCLs and we demonstrated that in basal condition the growth ratio for each timepoint over the previous one is comparable between the cell lines. (Ratio T24/T0 1.41 for STAG1_LCL over 1.50 for controls' mean, $p> 0.05$ n.s.; T48/T0: 1.29 for STAG1_LCL over 1.14 for controls' mean, $p>0.05$ n.s. One-sample T-Test). After X-ray irradiation, both STAG1_LCL and control LCLs are characterized by a remarkable reduction in term of growth capability in response to the damage stimulus, but the trend remains similar as shown in Fig. S2.



Suppl. Fig.2 Growth curves of LCLs, before and after X-ray irradiation. The reduction of cells growth rate is comparable between STAG1_LCL and the mean of the two control LCLs, either after a 3Gy irradiation (A-B-C) or after a 6Gy irradiation (D-E-F).(Ratio T24/T0 0.99 for STAG1_LCL over 0.96 for controls' mean, p> 0.05 n.s.; T48/T0: 0.92 for STAG1_LCL over 0.95 for controls' mean, p>0.05 n.s. [3Gy]; Ratio T24/T0 0.90 for STAG1_LCL over 1.01 for controls' mean, p> 0.05 n.s.; T48/T0: 0.87 for STAG1_LCL over 0.82 for controls' mean, p>0.05 n.s. [6Gy].(Statistical analysis performed by One-sample T-Test. * <0,05; **<0,01 ***<0,001; ****<0,0001)

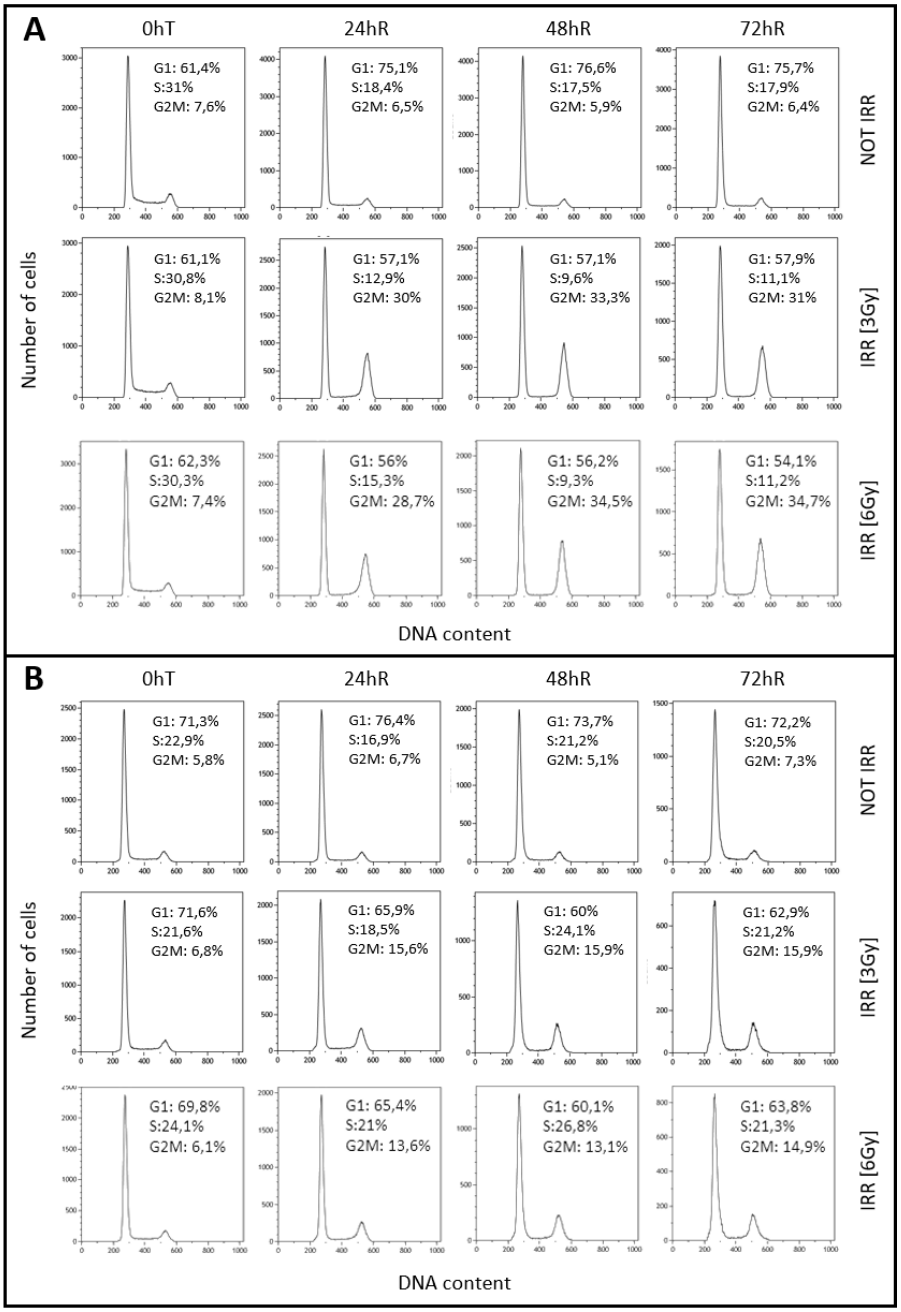
STAG1_LCL and control LCLs cell cycle is affected by X-ray irradiation -G2M block

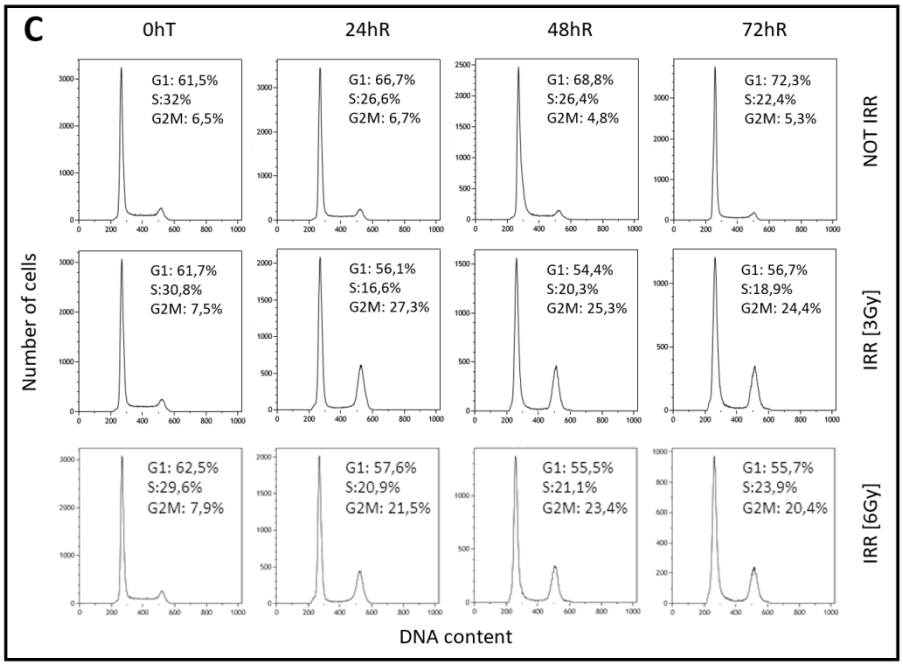
To assess the different distribution in cell cycle phases, we evaluated the percentage of cells in each phase (G0/G1; S; G2/M) for each LCLs. We didn't appreciate any significative different between CTRs and

STAG1 mutated cells in basal condition. This trend is comparable with the growth rate observed in cell growth curves.

Only when referring to timepoints 24h and 48h, we found a slight difference in G0/G1 and S phases in STAG1_LCL compared to CTRs, where mutated cells seem to have higher percentage of cell in S phase. This trend does not persist in the ulterior timepoints. (G0/G1: CTRs mean 75.8% vs 66.7% STAG1_LCL, T24; CTRs mean 75.2% vs 68.8% STAG1_LCL, T48) (S: CTRs mean 17.7% vs 26.6% STAG1_LCL, T24; CTRs mean 19.4% vs 26.4% STAG1_LCL, T48).

Even after X-ray irradiation, cell cycle perturbations are comparable across LCLs lines. As shown in figure S3, a G2M block induced by 3 Gy and 6 Gy was detected both in control LCLs and STAG1_LCL.


A
0hT
24hR
48hR
72hR
G1: 61,4%
S:31%
G2M: 7,6%
G1: 75,1%
S:18,4%
G2M: 6,5%
G1: 76,6%
S:17,5%
G2M: 5,9%
G1: 75,7%
S:17,9%
G2M: 6,4%
NOT IRR
G1: 61,1%
S:30,8%
G2M: 8,1%
G1: 57,1%
S:12,9%
G2M: 30%
G1: 57,1%
S:9,6%
G2M: 33,3%
G1: 57,9%
S:11,1%
G2M: 31%
IRR [3Gy]
G1: 62,3%
S:30,3%
G2M: 7,4%
G1: 56%
S:15,3%
G2M: 28,7%
G1: 56,2%
S:9,3%
G2M: 34,5%
G1: 54,1%
S:11,2%
G2M: 34,7%
IRR [6Gy]
Number of cells
DNA content
B
0hT
24hR
48hR
72hR
G1: 71,3%
S:22,9%
G2M: 5,8%
G1: 76,4%
S:16,9%
G2M: 6,7%
G1: 73,7%
S:21,2%
G2M: 5,1%
G1: 72,2%
S:20,5%
G2M: 7,3%
NOT IRR
G1: 71,6%
S:21,6%
G2M: 6,8%
G1: 65,9%
S:18,5%
G2M: 15,6%
G1: 60%
S:24,1%
G2M: 15,9%
G1: 62,9%
S:21,2%
G2M: 15,9%
IRR [3Gy]
G1: 69,8%
S:24,1%
G2M: 6,1%
G1: 65,4%
S:21%
G2M: 13,6%
G1: 60,1%
S:26,8%
G2M: 13,1%
G1: 63,8%
S:21,3%
G2M: 14,9%
IRR [6Gy]
Number of cells
DNA content



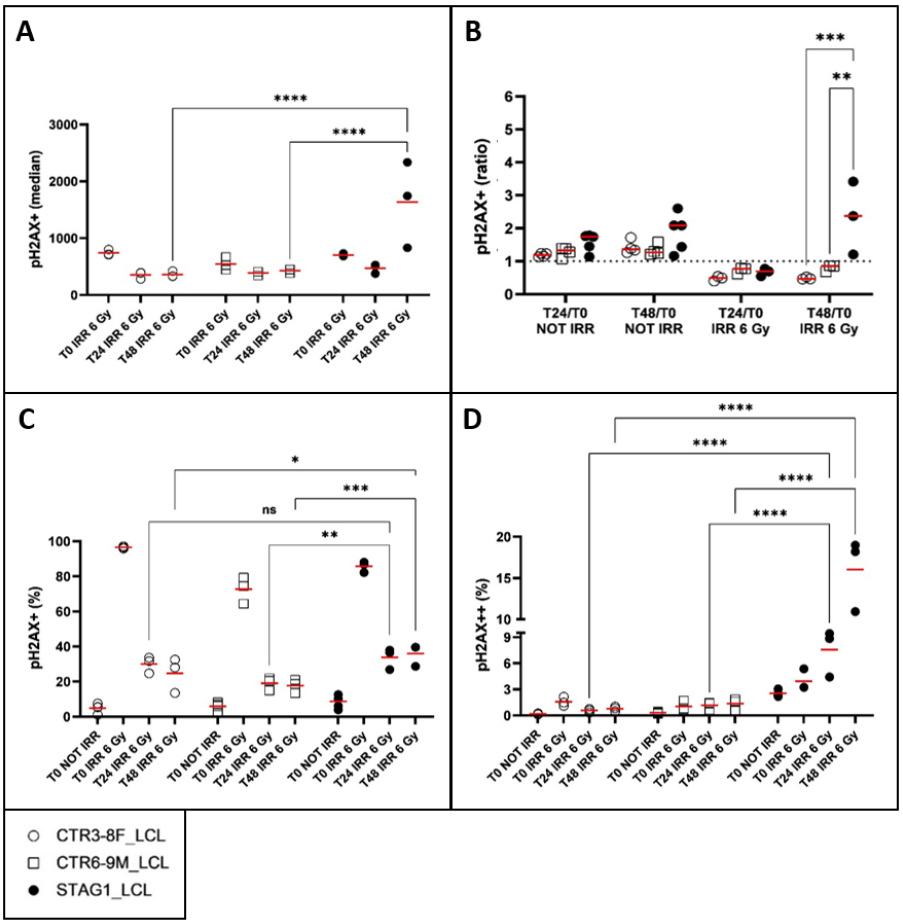
Supplementary Fig.3 Effect of X-ray irradiation on the cell cycle in LCLs. Cell cycle phase perturbations induced by [3Gy] [6Gy] irradiation on CTR3-8F_LCL (A), CTR6-9M_LCL (B) and STAG1_LCL (C) after 24, 48 and 72 h after damage stimulus.

Defective capability of STAG1_LCL to repair DNA after an ionizing radiation at 6Gy

The γH2AX phosphorylation status of STAG1_LCL remains at higher levels than control LCLs also after a higher ionizing radiation [6Gy] (T48: 4.2X, $p<0.0001$ [6Gy]; MFI *STAG1* over MFI control LCLs) (Fig. S4, A). These data confirmed a significantly lower capability of STAG1 mutated cells to repair after a DNA damage, compared to controls' LCLs. Moreover, they demonstrated that more intense X-ray dosage

causes a higher DNA damage, thus mutated cells are more impaired in repairing. The ratio of the *STAG1*_LCL median MFI over controls shown the same trend (T48/T0: 2.33 vs 0.64, p <0.01 [6Gy]) (Fig. S4, B).

The percentage of phophorilation confirm that the value after irradiation is higher in STAG1 mutated both considering the $pH2AX^{+}$ cells (T24: 1.8X, p<0.01 [6Gy]; $pH2AX^{+}$ cells *STAG1* over $pH2AX^{+}$ cells CTR6-9M_LCL) (T48: 1.7X, p <0.05 [6Gy]; $pH2AX^{+}$ cells *STAG1* over $pH2AX^{+}$ cells control LCLs) (Fig. S4, C) and the $pH2AX^{++}$ subpopulation (T24: 8.8X, p <0.0001; T48: 15.2X, p <0.0001; $pH2AX^{++}$ cells *STAG1* over $pH2AX^{++}$ cells control LCLs) (Fig. S4, D).

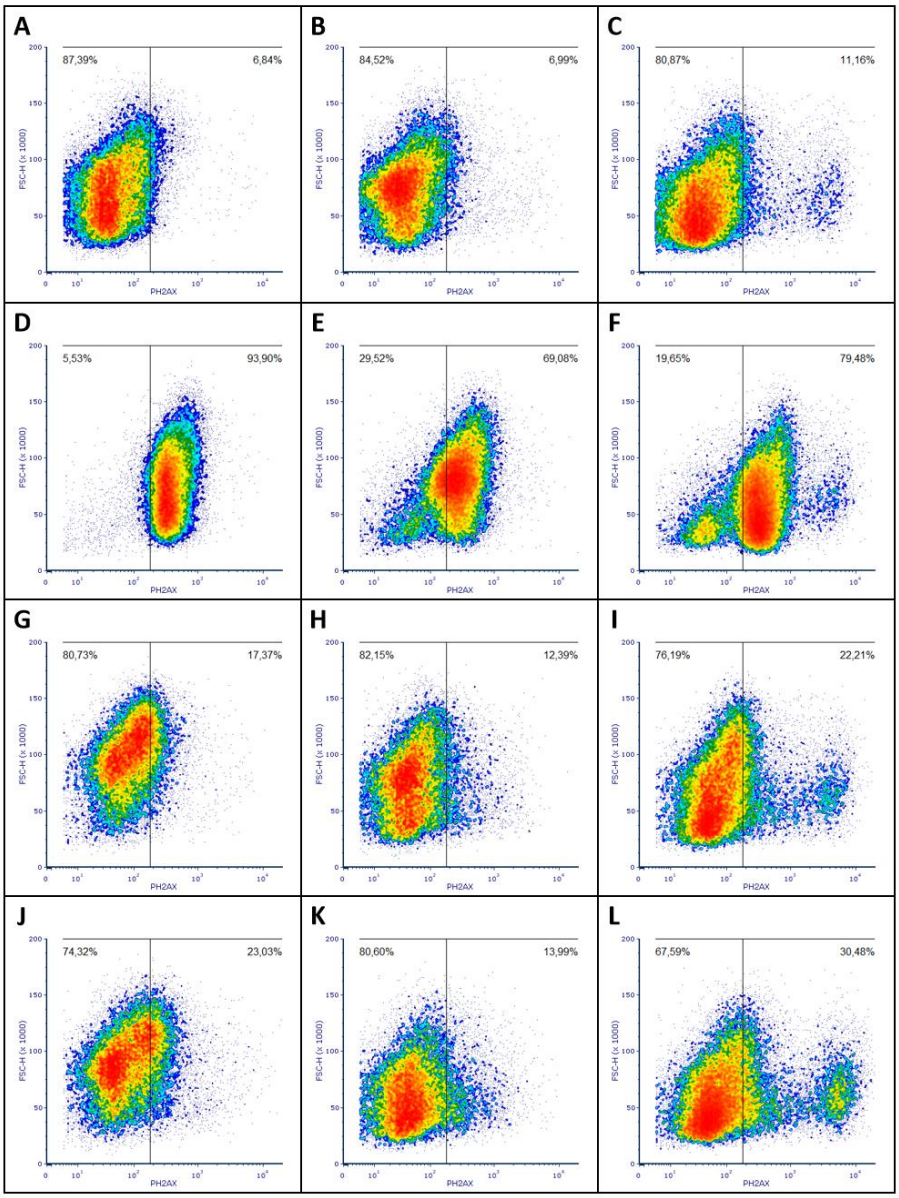


Supplementary Fig.4 γH2AX phosphorylation status before and after an X-ray irradiation [6Gy].

After a higher X-ray irradiation [6 Gy] γH2AX phosphorylation status remains at higher levels in STAG1_LCL compared to control LCL (panel A). The ratio value confirms the previous data both in absence and after the irradiation [6Gy] (panel B). The percentage of $pH2AX^{+}$ cells (panel C) and subpopulation $pH2AX^{++}$ (panel D) show comparable results.

(Statistical analysis performed by One-way Bonferroni's multiple comparison correction. * <0,05; **<0,01 ***<0,001; ****<0,0001)

*(Ratio levels around 1 indicates similar between the two timepoints considered (T0/T24 or T0/T48), instead ratio levels > 1 indicates that γH2AX phosphorylation levels are higher at T24 or T48 over T0).



Suppl. Fig.5 Representative dot plots of γH2AX phosphorylation status before and after X-ray irradiation [3Gy] – Timepoints 0h, 24h and 48h. CTR3-8F_LCL and CTR6-9M_LCL in the first two column and STAG1_LCL in the third one. In the panels A-B-C not irradiated cells; in the panels D-E-F phosphorylation status at T0 after irradiation; in the panels G-H-I and J-K-L reduction of $pH2AX^{+}$ cells at T24 and T48 can be appreciated, expression of different capability to repair after a DNA damage between control LCLs and STAG1_LCL.

Gene	Left Primer	Right Primer
STAG1_exon32	AGTCTGTAGGTCATGATTAGAAGGT	GATCAATAACCATGGTGTCCTCAAA
STAG2_exon28	ACATGCTTTCTTTCTTTCCAAACAG	GCATGGCAATGGCTTCTCTTG

Suppl. Table 1. *STAG1* and *STAG2* mutations validation RT-PCR primers.

Chapter 5

Blood. 2021 Jun 14; blood.2021011320.

Clinical relevance of clonal hematopoiesis in the oldest-old population

Marianna Rossi MD[1], Manja Meggendorfer PhD[2], Matteo Zampini PhD[1], Mauro Tettamanti PhD[3], Emma Riva PhD[3], Erica Travaglino BsC[1], Matteo Bersanelli PhD[4], Sara Mandelli PhD[3], Alessia Antonella Galbussera PhD[3], Ettore Mosca PhD[5], Elena Saba PhD[1], Chiara Chiereghin PhD[1], Nicla Manes PhD[1], Chiara Milanesi PhD[1], Marta Ubezio MD[1], Lucio Morabito MD[1], Clelia Peano PhD[1,6], Giulia Soldà PhD[1,4], Rosanna Asselta PhD[1,4], Stefano Duga PhD[1,4], Carlo Selmi MD[1,4], Maria De Santis MD[1], Karolina Malik MD[4], Giulia Maggioni MD[1,4], Marilena Bicchieri PhD[1], Alessia Campagna MD[1], Cristina A Tentori MD[1,4], Antonio Russo MD[1,4], Efrem Civilini MD[1,4], Paola Allavena PhD[1], Rocco Piazza MD[7], Giovanni Corrao PhD[8], Claudia Sala, PhD[9,10], Alberto Termanini PhD[1], Laura Giordano PhD[1], Paolo Detoma MD[11], Aurelio Malabaila MD[11], Luca Sala MD[12], Stefano Rosso MD[13], Roberto Zanetti MD[13], **Claudia Saitta BsC[1,7]**, Elena Riva BsC[1,7], Gianluigi Condorelli MD[1,4], Francesco Passamonti MD[14], Armando Santoro MD[1,4], Francesc Sole PhD[15], Uwe Platzbecker MD[16], Pierre Fenaux MD[17], Niccolo' Bolli

MD[18,19], Gastone Castellani PhD[9,10], Wolfgang Kern MD[2], George S Vassiliou PhD[20], TorstenHaferlach MD[2], Ugo Lucca PhD[3] and Matteo G Della Porta MD[1,4].

[1] IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano, Milan, Italy

[2] MLL Munich Leukemia Laboratory, Max-Lebsche-Platz 31, 81377 Munich, Germany

[3] Laboratory of Geriatric Neuropsychiatry, Department of Neuroscience, Istituto di RicercheFarmacologiche Mario Negri IRCCS, Via Via Mario Negri, 2, 20156 Milan, Italy

[4]Humanitas University, Department of Biomedical Sciences, Via Rita Levi Montalcini 4, 20090 Pieve Emanuele – Milan, Italy

[5] Institute of Biomedical Technologies, National Research Council (CNR), Via Fratelli Cervi 93, 20090 Segrate- Milan, Italy

[6] Institute of Genetic and Biomedical Research, National Research Council, via Manzoni 56, 20089 Rozzano – Milan, Italy

[7] Department of Medicine and Surgery, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

[8] Unit of Biostatistics, Epidemiology, and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

[9] Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2 40138 Bologna, Italy

[10] Experimental, Diagnostic and Specialty Medicine – DIMES, Via Zamboni 33, 40126 Bologna, Italy

[11] Laboratory of Analysis, Ospedaledegli Infermi, Via dei Ponderanesi 2, 13875 Ponderano – Biella, Italy

[12]Dipartimento di Prevenzione ASL Biella, Via dei Ponderanesi 2, 13875 Ponderano – Biella, Italy

[13] Piedmont Cancer Registry, Centre for Epidemiology and Prevention in Oncology, Via Santena 7, 10126 Turin, Italy

[14] Hematology, ASST Sette Laghi, Ospedale di Circolo of Varese & University of Insubria, VialeBorri 57, Varese, Italy

[15]Institut de Recerca Contra la LeucèmiaJosep Carreras, Camí de les Escoles, s/n, 08916 Badalona – Barcelona, Spain

[16] Medical Clinic and Policlinic 1, Hematology and Cellular Therapy, University Hospital Leipzig, Liebigstrasse 22, 04103 Leipzig, Germany

[17] Service d'Hématologie Séniors, Hôpital Saint-Louis, Assistance Publique-Hôpitaux de Paris and Université Paris, 1 avenue Claude Vellefaux, 75010 Paris, France

[18] Department of Oncology and Hemato-Oncology, University of Milan, Via Festa del Perdono 7, 20122 Milan, Italy

[19] Hematology, Fondazione IRCCS Ca' GrandaOspedale Maggiore Policlinico, Via F. Sforza 35, 20122 Milan, Italy

[20] Department of Haematology, Cambridge University Hospitals National Health Service (NHS) Trust, Trinity Lane, Cambridge, United Kingdom.

Running head: Clonal hematopoiesis in oldest-old people

Key Points

- In oldest-old population, specific mutational patterns define different risk of developing myeloid neoplasms *vs.* inflammatory-associated diseases.
- In oldest-old individuals with unexplained cytopenia, mutational status identifies subjects with presumptive evidence of myeloid neoplasms.

Abstract

Clonal hematopoiesis of indeterminate potential (CHIP) is associated with increased risk of cancers and inflammation-related diseases. This phenomenon becomes very common in oldest-old individuals, in whom the implications of CHIP are not well defined. We performed a mutational screening in 1794 oldest-old individuals enrolled in two population-based studies and investigate the relationships between CHIP and associated pathologies. Clonal mutations were observed in one third of oldest-old individuals and were associated with reduced survival. Mutations in *JAK2* and splicing genes, multiple mutations (*DNMT3A, TET2, ASXL1* with additional genetic lesions) and variant allele frequency ≥0.096 had positive predictive value for myeloid neoplasms. Combining mutation profiles with abnormalities in red blood cell indices improved the ability of myeloid neoplasm prediction. On this basis, we defined a predictive model that identifies 3 risk

groups with different probabilities of developing myeloid neoplasms. Mutations in *DNMT3A, TET2, ASXL1* or *JAK2* (most occurring as single lesion) were associated with coronary heart disease and rheumatoid arthritis. Cytopenia was a common finding in oldest-old population, the underlying cause remaining unexplained in 30% of cases. Among individuals with unexplained cytopenia, the presence of highly-specific mutation patterns was associated with myelodysplastic-like phenotype and a probability of survival comparable to that of myeloid neoplasms. Accordingly, 7.5% of oldest-old subjects with cytopenia had presumptive evidence of myeloid neoplasm. In conclusion, specific mutational patterns define different risk of developing myeloid neoplasms *vs.* inflammatory-associated diseases in oldest-old population. In individuals with unexplained cytopenia, mutational status may identify those subjects with presumptive evidence of myeloid neoplasms.

Introduction

Exome sequencing studies have identified the frequent age-dependent clonal expansion of somatic mutations in the hematopoietic system.[1-5] Clonal hematopoiesis of indeterminate potential (CHIP) describes individuals with hematologic malignancy-associated mutations in blood or marrow, but without other diagnostic criteria for a hematologic malignancy[5] and is associated with increased

risk of cancers (in particular myeloid neoplasms) and chronic inflammatory diseases (coronary heart disease).[1-6]

The phenomenon of CHIP becomes very common in oldest-old population (people aged 80+ years),[7-10] that represents the fastest growing age segment in developed countries.[9,11] In these individuals, clinical implications of CHIP are expected to be relevant but are not well defined.[7-8]

The incidence of solid cancers and myeloid neoplasms increases with age and mortality is higher after the age of 75.[12] The risk of several chronic inflammatory diseases is age-related as well, leading to large prevalence of frailty and disability among oldest-old people.[13] We hypothesized that the study of oldest-old population can contribute to define the relationship between specific mutational patterns in the hematopoietic system and the individual risk of developing cancers *vs.* other adverse events (chronic inflammatory diseases).

Anemia is a common finding in the elderly and is associated with worse cognitive and functional outcomes and increased mortality.[14-18] Underlying cause of anemia remained unexplained in 30% of cases, and a proportion of unexplained cytopenia may account for myeloid neoplasms.[15,18] We hypothesized that the study of CHIP may contribute to determine specific causes of anemia in elderly people and to define personalized treatment strategies to mitigate anemia-related negative sequela.

The primary objective of the present study was to evaluate the prevalence of CHIP and the relationships between CHIP and associated pathologies in oldest-old population. The secondary objective was to analyze clinical outcome of patients affected with unexplained cytopenia and to define the clinical effect of clonal abnormalities among these individuals. The definitions of Idiopathic Cytopenia of Unknown Significance (ICUS) and Clonal Cytopenia of Unknown Significance (CCUS) were applied to identify individuals with non-clonal *vs.* clonal unexplained cytopenia.[5,19]

Patients and methods

Study procedures are in accordance with the Declaration of Helsinki. Ethics Committees of Humanitas Research Hospital and Mario Negri Pharmacological Institute, Milan Italy approved the study. Written informed consent was obtained prior to blood sampling. (ClinicalTrials.gov number: NCT03907553)

Study population

Study procedures are described in Supplementary_File_1.

We analyze subjects included in the two population-based studies enriched in oldest-old individuals (80+ years), i.e., "Health_&_Anemia"19,21 and "Monzino_80+".[24]

"Health_&_Anemia" is a prospective study (2003-2018) aimed to investigate clinical consequences of anemia in the elderly.[15,17] We

studied 1059 oldest-old subjects (median age 83 years, range 80-105) in whom peripheral blood samples collected at study enrollment were available for mutational screening. At study enrollment (May 2003) clinical history was collected and complete blood count was performed. When a hemoglobin concentration was below WHO reference criteria for anemia (<12g/dL in women and <13g/dL in men), further investigations were made to define specific causes of anemia. Follow-up was updated to December 2018. A total of 344,565 laboratory tests were available during follow-up. Data on hospitalization and mortality were available for all subjects. Diagnosis of chronic inflammatory diseases and cancers were defined according to International Classification of Diseases for Oncology ninth edition CM (ICD-9-CM) codes and local cancer registry data. Diagnosis of myeloid neoplasms was based in addition on information provided by revision of bone marrow biopsy reports provided by Hematology Unit of Biella Hospital. Three investigators (MR, ER and MGDP) have reviewed independently all this information and a final diagnosis of myeloid neoplasms was provided by a consensus meeting.

As a second cohort, we analyzed 735 individuals (median age 90 years, range 80-104) enrolled in the "Monzino_80+" prospective study (2002-2018) aimed at investigating relationships between age, cognitive decline and dementia.[20]

Due to the fact that data collection on myeloid neoplasms diagnosis was less accurate in "Monzino_80+" population with respect to

"Health_&_Anemia" cohort (see Supplementary_File_1), we analyzed in addition 727 subjects aged ≥75<80 years from "Health_&_Anemia" study, to specifically validate the predictive value of clinical and mutational features on the risk of developing myeloid neoplasms.

Finally, to compare clinical features and outcome of oldest-old subjects with CHIP to those of patients with myeloid neoplasms, we analyzed a sex- and age-matched population of patients affected with myelodysplastic syndrome from the retrospective EuroMDS database. (2000-2018, ClinicalTrials.gov number:NCT04174547) Each subject with CHIP was matched with 5 patients with the same year of birth and sex; overall, 255 patients with myeloid neoplasms were included in this analysis.

Mutation screening

Using peripheral blood DNA we looked for mutations in 47 genes related to myeloid neoplasms. Gene list is available in *Supplementary_Table_1*, sequencing procedures and variant calling are available in *Supplementary_File_2)*[5]

CHIP was defined as the presence of a clonal blood cell population associated with a hematologic malignancy-related mutation at a variant allele frequency (VAF)≥0.01. Median coverage was 3455×. Mutations with VAF<0.10 were re-sequenced on an independent platform. The effectiveness of re-sequencing in confirming genomic variants with VAF≥0.01 and <0.10 was of 96.5%, while DNA sequencing

was significantly less performant and reproducible below the threshold of 0.01(P=0.02). The technique we used missed clonal skewing in the absence of mutations in putative myeloid neoplasms driver genes.[22]

Statistical analysis

Numerical variables were summarized by median and range; categorical variables were described with count and relative frequency (%) of subjects in each category.

Survival analyses were performed with Kaplan-Meier method and differences between groups were evaluated by log-rank test. Cox models were built to estimate hazard ratio (HR, with 95% CI) for probability of overall survival and risk of developing coronary heart disease and chronic inflammatory diseases (only incident cases were considered in the analyses).

The accuracy of mutational factors in predicting the risk of developing myeloid neoplasms was analyzed (only incident cases were considered in the analyses). The accuracy of categorical variables (presence *vs.* absence of mutations in a specific gene) was estimated by calculating time-dependent positive and negative predictive value (PPV, NPV).[23] For variables measured on a continuous scale (variant allele frequency, VAF), time-dependent ROC curve, the corresponding area under the ROC curve (AUC) and the optimal cutoff point were calculated by using online available *cenROC*R package.[24,25]

To define a risk score for developing myeloid neoplasms, HR from a multivariable Cox analysis on "Health_&_Anemia" population including age, sex, mutational status and non-mutational parameters as covariates were used. A diagnosis of myeloid neoplasm was considered as event; subjects were censored at the end of follow-up or at time of death. The goodness of concordance for the predictive score was measured by concordance index (C-index), internal 5-folds cross-validation and independent external validation.

Cumulative incidence of myeloid neoplasms was calculated by Kaplan-Meier method (death for any cause was considered as competing-event in the estimation of cumulative incidence function).[26] Left truncation was applied when calculating the cumulative incidence of myeloid neoplasms with age as the time scale.[27]

Data Sharing

Data are found under accession number PRJNA736552.

Results

Prevalence and clinical effect of CHIP in oldest-old population

We studied prevalence of CHIP and relationship between CHIP and probability of survival in oldest-old population. Analyses were performed on both "Health_&_Anemia" and "Monzino_80+" cohorts.

Mutations were observed in 32.6%[95%CI 29.4-35.5] and 26.0%[22.9-29.8] of subjects enrolled in “Health_&_Anemia” and “Monzino_80+” cohorts, respectively. The majority of variants in both cohorts occurred in three genes: *DNMT3A, TET2 a*nd *ASXL1.* (prevalence of mutated genes is reported in *Figure_1*, Supplementary_Figure_1 and Supplementary_Table_2) CHIP was more common in males *vs.* females (P=0.02 and P=0.005, respectively) and its prevalence increased with age (P=0.001 and P=0.03, respectively).

Considering genes grouped according to functional patterns in both cohorts, we observed a significant increase of mutations with age in epigenetics and cohesin complex-related genes (P=0.01 and P=0.02, respectively). Considering single genes, we observed a significant increase in the prevalence of *TET2* and *ASXL1* mutations after the age of 90 (P=0.021 and P=0.032, respectively). After testing for gender bias, we observed two genes significantly more mutated in males than in females: *ZRSR2* and *U2AF1*.

We focused on centenarians in both cohorts (n=44) stratified according to the presence of comorbidity (including heart disease, diabetes, stroke, cancer, osteoporosis, thyroid condition, Parkinson's disease and chronic obstructive pulmonary disease).[28] Individuals with chronic age-related illness before the age of 100 were 33 *vs.* 11 who remained disease-free at age 100. Prevalence of CHIP was higher in patients with *vs.* without comorbidity (62% *vs.* 20%, P=0.015).

Subjects with CHIP were older with respect to individuals without CHIP (median age of individuals without CHIP *vs.* those with 1 mutation *vs.* those with ≥2 mutations was 83y *vs.* 84y *vs.* 85y, respectively, in the "Health_&_Anemia" cohort, P=0.019; and 90y *vs.* 91y *vs.* 94y, respectively, in the "Monzino_80+" cohort, P<0.001). The presence of CHIP was associated with a lower probability of survival (P<0.001 in both cohorts) and prognosis was even poorer in subjects carrying ≥2 mutations. (P<0.001 in both cohorts, Figure_2). The independent association between CHIP and increased mortality was maintained in a multivariable analysis including age, sex and cytopenia as covariates (HR 1.28[1.1-1.9], P=0.009 and HR 1.37[1.2-1.71],P=0.006, in the "Health_&_Anemia" and "Monzino_80+" cohort, respectively) and when focusing on cancer-related death (HR 1.91[1.44-2.21],P=0.001 and HR 2.13[1.94-2.39],P=0.002) and non cancer-related mortality (HR 1.43[1.29-1.77],P=0.02 and 1.56[1.31-1.8],P=0.01) as separate outcomes.

Focusing on subjects carrying ≥2 mutations in both cohorts, the independent effect of carrying ≥2 mutations on mortality was maintained in a multivariate analysis including age, sex and cytopenia as covariates (HR 1.4[1.19-2.31, P=0.008]. Considering specific causes of death, a higher prevalence of cancer-related deaths was observed in individuals carrying ≥2 mutations *vs.* both subjects carrying 1 mutation and those without CHIP (P=0.028 and P=0.009, respectively)

CHIP and risk of developing myeloid neoplasms in oldest-old population

We investigated the relationship between CHIP and risk of developing myeloid neoplasms. Prevalent and incident cases were 16 and 25, respectively. We calculated the time-dependent PPV and NPV for developing myeloid neoplasms at the age of 95y for most relevant genomic features. (Supplementary_Figure_2)

Absence of mutations had high NPV (0.89[0.86-0.92]), while the presence of CHIP *per se* had low PPV (0.11[0.09-0.16]). Among investigated genes, splicing genes (*SF3B1, SRSF2, U2AF1, ZRSR2*) and *JAK2* have the highest PPV (0.58[0.41; 0.63] and 0.70[0.41-0.98], respectively). To evaluate the impact of multiple mutations in the same individual, we focused on the three most commonly mutated genes (*DMNT3A, TET2, ASXL1*) comparing single mutations with co-mutation patterns. PPV of mutations in *TET2, DNMT3A* or *ASXL1* with co-mutation patterns was higher than PPV of single lesions (0.28[0.14; 0.41] vs. 0.08[0.02-0.23],P=0.001). (Supplementary_Figure_2) Overall, mutations in splicing genes, mutations in *JAK2* gene and co-mutation patterns involving *TET2, DNMT3A* and *ASXL1* accounted for 75% of myeloid neoplasms diagnosed in subjects with prior CHIP.

We then explored the best cutoff value of VAF for developing myeloid neoplasms. VAF showed a significant accuracy as evaluated by time-dependent ROC curve (AUC was 0.88[0.81-0.94],P<0.001). At the age

of 95y, a VAF of 0.096 was found as optimal cutoff value (sensitivity 0.84, specificity 0.83).

In a multivariable analysis including the number of mutations per subject and the most frequently mutated genes, having co-mutation patterns involving *TET2, DNMT3A* and *ASXL1* (HR 4.64[2.11-12.23] ,P<0.001), carrying splicing mutation (HR 10.63[5.23-17.68],P<0.001) or having VAF >0.096 (HR 2.35[1.22- 5.48],P=0.021) were independent predictors for developing myeloid neoplasms.

Finally, we aimed to study whether non-mutational factors may improve the capability to capture individual risk of developing myeloid neoplasms. We focused on changes in red blood cell (RBC)-indices (mean corpuscular volume [MCV] and red blood cell distribution width [RDW]) that occur as early phenotypic abnormalities in subjects who later develop myeloid neoplasms.[4]

In "Health_&_Anemia" cohort, we found that high MCV (>98fl) and RDW (>14) values at study enrollment were associated with reduced survival. (P=0.01 and P=0.008, respectively, Supplementary_Figure_3)

In a multivariable analysis including splicing mutations, co-mutation patterns involving *TET2, DNMT3A* and *ASXL1* and VAF as covariates, abnormal RBC-indices are associated with higher risk of developing myeloid neoplasms, independently from mutational features (HR 2.02[1.18-4.7], P<0.001).

Definition of a risk score for developing myeloid neoplasms according to mutational status and RBC-indices

We aimed to define a risk score for developing myeloid neoplasms according to mutational status and RBC-indices. Subjects from "Health_&_Anemia" cohort entered this analysis.

HR from multivariable Cox analysis including age, sex, mutational status and RBC-indices as covariates were used to define a risk score (details are reported in *Figure_3*). A score of 1 was assigned for abnormalities in RBC-indices and VAF>0.096, a score of 2 was assigned for co-mutation patterns involving *TET2, DNMT3A* and *ASXL1,* and a score of 5 was assigned for splicing mutations. Risk groups were defined as low (score 0-1), intermediate (score 2-4) and high (score ≥5). A simplified risk classification was also provided. Cumulative incidence of myeloid neoplasms was significantly different among these three risk groups (P<0.001). In particular, in high-risk individuals (3% of the whole oldest-old general population), cumulative incidence of myeloid neoplasm was 14%, 34% and 42% at the age of 85, 90 and 95 years, respectively. The accuracy of the predictive score was good (C-index 0.851) and was confirmed by internal 5- folds cross validation (mean C-index in test sets was 0.849).

We performed in addition an external validation on an independent cohort of 727 subjects aged ≥75<80y from "Health_&_Anemia" study. (Figure_3 and Supplementary_Figure_4) The analyses performed on

the validation cohort confirmed a high concordance of the model (C-index was 0.889), thus suggesting a high generalizability of the results.

Clonal evolution in old-oldest population with multiple samples available

We studied clonal evolution in 96 subjects from "Health_&_Anemia" cohort, in which multiple samples were available over a period of 4 years.

CHIP was found at baseline in 22 cases (23%): during follow up, 2 individuals acquired additional mutations, 10 displayed ≥0.05 VAF increase, while in 3 cases CHIP was lost. In 13/74 subjects without mutations at baseline, CHIP was acquired during follow-up (17.5%). We identified 2 subjects in whom clonal evolution preceded a diagnosis of a myeloid neoplasm (myelodysplastic syndrome in both cases). (Figure_4 and Supplementary_Table_3)

Relationship between CHIP, coronary heart disease and chronic inflammatory diseases in oldest-old population

We aimed to define the relationship between CHIP and risk of developing chronic inflammatory diseases. We analyzed "Health_&_Anemia" cohort, in which the diagnosis of chronic inflammatory diseases was systematically recorded.

Coronary heart disease was defined as a history of myocardial infarction or coronary revascularization after the time of DNA

collection. Prevalent and incident cases were 57 and 27, respectively. We defined *ASXL1, TET2, DNMT3A* and *JAK2* mutations as high-risk for vascular events.[6] Participants with CHIP had a significantly higher risk of coronary heart disease with respect to those without mutations (HR 1.61[1.28-3.21], P=0.02). When considering patients with high-risk mutations, HR increased to 2.21[1.45-4.01], P=0.006 and the effect was maintained when adjusting for sex, age smoking, hypertension and hyperlipidemia (not shown). Mutations in splicing genes were not associated with an increased risk of coronary heart disease (HR 0.91[0.79-1.65], P=0.84).

As a further step, we investigated the possible role of CHIP in other chronic inflammatory diseases (stroke, diabetes, arthritis and autoimmune diseases). We observed preliminary evidence of increased risk of developing rheumatoid arthritis (12 prevalent and 6 incident cases) in participants with *vs.* without high-risk mutations.6 (HR 4.79[1.9-17.62], P=0.039) However, due to the low number of cases observed, this finding should be interpreted with caution.

Clinical relevance of CHIP in oldest-old individuals with unexplained cytopenia.

We aimed to specifically analyze clinical outcome of oldest-old patients affected with unexplained cytopenia and to define the clinical effect of clonal abnormalities among these individuals. The definitions of Idiopathic Cytopenia of Unknown Significance (ICUS) and Clonal

Cytopenia of Unknown Significance (CCUS) were applied to identify individuals with non-clonal *vs.* clonal unexplained cytopenia.

Individuals from both "Health_&_Anemia" and "Monzino_80+" cohorts entered this analysis. The most common cytopenia reported at study enrollment was anemia (14.9% and 29.5% respectively).[15-18] The underlying cause of persistent (>6 months) cytopenia was unexplained in 29% and 34% of cases, respectively. (*Table_1*) Prevalence of CHIP was not significantly different in oldest old subjects with *vs.* without cytopenia (30.5% *vs.* 29.6%, respectively, P=0.24). Focusing on subjects affected with anemia stratified according to the underlying cause, no significant difference on the prevalence of CHIP was observed among different groups of patients. We noticed an enrichment of splicing gene mutations in patients with unexplained anemia with respect to subjects with anemia associated with specific underlying cause (P=0.031). We considered subjects with unexplained cytopenia from both studies (n=133), stratified according to the presence of mutations as ICUS (n=82) *vs.* CCUS (n=51,38%).[19] Subjects with CCUS showed a significantly lower probability of survival compared with ICUS (P=0.002), while no significantly different probability of survival was noticed between ICUS and individual without cytopenia. (Figure_5) As we observed a difference in survival among CCUS *vs*. ICUS, we tested the hypothesis that highly-specific mutation patterns for myeloid neoplasms may provide presumptive

evidence of hematological malignancy in patients with unexplained cytopenia even in absence of definitive morphological criteria.

According to our findings, highly-specific mutation patterns for myeloid neoplasms were defined as the presence of mutations of splicing factors, co-mutation patterns involving *TET2, ASXL1* or *DNMT3A* and/or mutations with VAF>0.096.

CCUS with highly-specific mutation patterns frequently showed macrocytic anemia and/or multilineage cytopenia (30 of 39 cases) consisting with a myelodysplastic syndrome phenotype. Two subjects died because of acute myeloid leukemia, while the most frequent cause of death was cardiac disease (25 subjects).

We then compared clinical features and outcomes of subjects with CCUS with those of an age- and sex-matched population affected with myeloid neoplasms (myelodysplastic syndromes) reported to retrospective EuroMDS database. (n=255, Supplementary_Table_4) No significant differences were observed in probability of survival between CCUS with highly-specific mutation patterns with respect to patients with myeloid neoplasms, while CCUS without highly-specific mutation patterns showed higher probability of survival with respect to CCUS with highly-specific mutation patterns. (HR 2.05[1.61- 4.64], P=0.06 *Figure_5*).

Considering CCUS with highly-specific mutation patterns as “potential myeloid neoplasms”, 7.5% of oldest-old subjects with cytopenia had presumptive evidence of myeloid neoplasm.

Discussion

We observed that CHIP is a common finding in oldest-old population (30% of individuals) and that specific mutational profiles are associated with distinct clinical outcomes.

Mutations in splicing genes (mostly occurring as single genetic lesion)[29] showed the highest predictive value for myeloid neoplasms,[1-4,30] while they are not significantly associated with increased risk of developing chronic inflammatory diseases.

By contrast, the positive predictive value for myeloid neoplasms of isolated mutations in *TET2, DNMT3A* and *ASXL1* genes was lower, and additional genetic events are required to give rise to a myeloid neoplasm.[1-4,30] These mutations are mostly linked to increased risk of myocardial infarction and arthritis.[1-4,6,31]

Focusing on prediction of myeloid neoplasms, both number of mutations per subject and size of the mutant clone (VAF) had significant positive predictive value.[30] We observed in addition that abnormalities in RBC-indices significantly improve the capability of molecular features to capture individual risk of developing myeloid neoplasms.[4] By combining specific mutational patterns, size of mutant clone and abnormalities in RBC-indices, we defined three groups of individuals with different risk of developing myeloid neoplasms. In details, we identified a small population (3% of individuals aged 80 years or older) in which cumulative incidence of myeloid neoplasm was 14%, 34% and 42% at the age of 85, 90 and 95 years, respectively.

Importantly, the predictive power of the score was confirmed by an external, independent validation.

Clonal evolution was a frequently observed event in oldest-old subjects with sequential samples available and in some cases preceded the occurrence of an overt myeloid neoplasm, suggesting that serial analysis may improve clinical monitoring.[30]

We had the opportunity to study CHIP in centenarian according to different morbidity profiles. CHIP is common in centenarians with age-associated diseases, while rarely observed those who attain their 100th birthday without comorbidity. The possible relationship between absence of CHIP and exceptional longevity should be addressed by specific investigations.[32]

Finally, we hypothesized that myeloid neoplasms could be underdiagnosed in oldest-old population, especially in cases with unexplained cytopenia.[14,32] We observed that among individual with unexplained cytopenia, the presence of a highly-specific mutation pattern for myeloid neoplasms[30,33] is associated with reduced survival. According to mutational features, 7.5% of oldest old subjects with cytopenia may have presumptive evidence of myeloid neoplasm. The study of CHIP is therefore expected to contribute to determine specific causes of cytopenia in elderly people and to define personalized treatment strategies.

Since CHIP was described, caution is suggested against adopting mutational testing in clinical practice. In fact, the presence of

mutations "*per se*" in a given individual has only limited predictive power as conversion to overt diseases is rare regardless of mutation status.[1-3,34] Our findings improve the capability to capture clinical information at individual patient level with respect to the presence of specific mutation patterns. These data support the rationale for prospective studies including CHIP as a part of conventional laboratory investigations to evaluate the health general status of elderly people and drive strategies to prevent adverse events.

Tables

	"Health_&_Anemia" study	"Monzino_80+" study	P
Number of subjects	1059	735	
Women (%)	697 (65.8)	546 (74.3)	<0.001
Men (%)	362 (34.2)	189 (25.7)	
Age y median (range)	82.8 (80-105)	90.5 (80.5-103.9)	<0.001
≥80<85 y (%)	725 (68.5)	128 (17.4)	<0.001
≥85<90 y (%)	235 (22.2)	215 (29.3)	
≥90<95 y (%)	71 (6.7)	247 (33.6)	
≥95<100 y (%)	24 (2.3)	105 (14.3)	
≥100 y (%)	4 (0.3)	40 (5.4)	
Death (%)	683/1059 (64.5)	695/735 (94.5)	<0.001
Blood count at study enrollment			
Hb g/dl median (range)	13.6 (7.3-18)	13.1 (6.7-16.8)	<0.001
HCT % median (range)	42.4 (22.9-55.5)	39.2 (22-50.6)	<0.001
MCV fL median (range)	95.9 (61.9-123.3)	92.2 (58.6-121.8)	<0.001
RDW fL median (range)	14.4 (12.3-21.6)	13.4 (11.3-22.2)	<0.001
WBC x10^9/L median (range)	6.4 (1.6-40.3)	6.5 (2.3-78.9)	0.664
ANC x10^9/L median (range)	3.8 (0.6-21.2)	3.8 (1.2-15.2)	0.059
PLT x10^9/L median (range)	222 (52-1117)	231 (25-915)	0.044
Anemia at study enrollment			
All (%)	158/1059 (14.9)	217/735 (29.5)	<0.001
Women (%)	94/697 (13.5)	142/546 (26)	<0.001
Men (%)	64/362 (17.7)	75/189 (39.7)	<0.001
Anemia by age (years)			

≥80<85 (%)	81/725 (11.2)	21/128 (16.4)	0.142
≥85<90 (%)	41/235 (17.4)	49/215 (22.8)	0.249
≥90<95 (%)	25/71 (35.2)	87/247 (35.2)	0.930
≥95<100 (%)	9/24 (37.5)	40/105 (38.1)	0.971
≥100 (%)	2/4 (50)	20/40 (50)	1.000
Severity of anemia			
Mild (normal value to 10 g/dl) (%)	141/158 (89.3)	197/217 (90.7)	0.910
Moderate (9.9-8 g/dl) (%)	16/158 (10.1)	16/217 (7.3)	0.389
Severe (<8 g/dl) (%)	1/158 (0.6)	4/217 (2)	0.320
Causes of anemia			
Iron deficiency (%)	9/158 (5.7)	14/217 (6.4)	0.777
Beta thalassemia (%)	12/158 (7.6)	5/217 (2.3)	0.021
B12/folate deficiency (%)	18/158 (11.4)	41/217 (18.8)	0.091
Renal (%)	29/158 (18.4)	49/217 (22.6)	0.419
Chronic disease (%)	16/158 (10.1)	16/217 (7.3)	0.389
Multifactorial anemia (%)	23/158 (14.6)	18/217 (8.2)	0.087
Unexplained anemia (%)	41/158 (25.9)	74/217 (34.1)	0.216
Neutropenia at study enrollment			
All (%)	10/1059 (1)	26/735 (3.5)	<0.001
Unexplained (%)	4/10 (40)	11/26 (42.3)	0.936
Thrombocytopenia at study enrollment			
All (%)	51/1059 (4.8)	56/735 (7.6)	0.021
Unexplained (%)	9/51 (17.6)	16/56 (28.5)	0.294
Unexplained cytopenia	48/1059 (4.5)	85/735 (11.5)	<0.001

Table 1. Prevalence and leading causes of anemia and other cytopenias at study enrollment in subjects aged 80y or older from "Health_&_Anemia" and "Monzino_80+" cohorts.

The classification of anemia and other cytopenias (based on the hematologic findings) was supported by the clinical conditions and pharmacological therapies of the elderly (criteria are defined on *Supplementary_File_1*). Anemias and other cytopenias that could not be classified into any of the previously defined categories were considered to be of "unexplained origin". A panel of the three physicians reviewed, discussed and reached a final consensus for each case of cytopenia with discrepant classification.

Figure Legends

Figure 1. Clonal hematopoiesis of indeterminate potential (CHIP) in oldest-old subjects from "Health_&_Anemia" and "Monzino_80+" cohorts.

Panels *A* and *C* show the prevalence of most frequently mutated genes in the two cohorts (considering mutated and unmutated patients). Panels *B* and *D* show the number of persons with 1, 2, or more than 2 variants.

Figure 2. Panels *A* and *C* show cumulative probability of overall survival according to the presence of CHIP in oldest-old subjects from "Health_&_Anemia" and "Monzino_80+" cohorts, respectively.

Panels *B* and *C* show cumulative probability of overall survival according to the number of variants (0 *vs.* 1 *vs.* 2 or more) in old oldest subjects from "Health_&_Anemia" and "Monzino_80+" cohorts, respectively.

Figure 3. (A) Definition of a score based on specific mutational patterns (CHIP) and red blood cell (RBC)-indices to predict the risk of developing myeloid neoplasms.

To define a risk score for developing myeloid neoplasms, we used HR from a multivariable Cox analysis on "Health_&_Anemia" cohort (learning cohort) adjusted for age and sex, including mutational status (splicing mutations, co-mutation patterns involving *TET2, DNMT3A* and *ASXL1*, and VAF>0.096) and non-mutational parameters (RBC-indices) as covariates. A diagnosis of myeloid neoplasm (including myelodysplastic syndrome and acute myeloid leukemia) was considered as event; subjects were censored at the end of follow-up or at time of death. (***B***) Cumulative incidence (CI) of myeloid neoplasms for oldest-old individuals stratified into 3 risk categories (learning cohort). Cumulative incidence was calculated by Kaplan-Meier method (death for any cause was considered as competing-event in the estimation of CI function). Left truncation was applied to calculate CI of myeloid neoplasms with age as the time scale. (***C***) Validation of the score on an independent cohort of 727 subjects aged ≥75<80y from "Health_&_Anemia" study.

Figure 4. (A) Clonal evolution in subjects from "Health_&_Anemia" cohort with multiple sample available (n=96); (B) Clonal evolution - Subject #269.

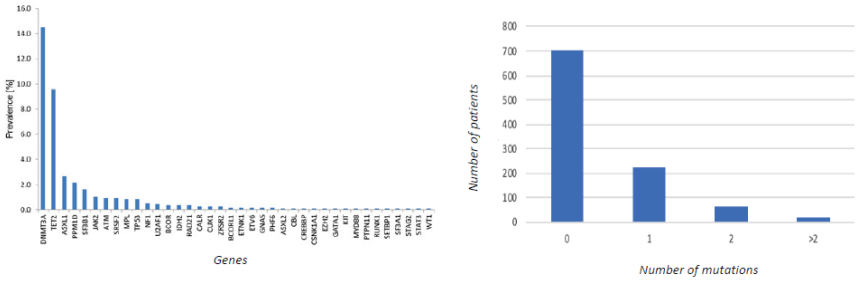
This female subject was born on 1921. In 1999, she displayed normal blood count. In 2003 a first mutational screening was performed with evidence of a mutation in SF3B1 gene with 0.02 VAF. At this time, the subject showed normal hemoglobin level (12.1 g/dl), RDW (12) and MCV (87). Since 2003 this subject experienced increasing in RDW and MCV, and in 2007 a mild anemia was observed (10 g/dl). At this time, mutational screening showed increase in SF3B1 VAF (0.24). In 2008, a diagnosis of myelodysplastic syndrome (MDS) with ring sideroblasts was performed. (**C**) Clonal evolution -Subject #1145. This male subject was born on 1922. In 1999 blood count was normal. In 2003 a first mutational screening revealed a mutation in *TET2* gene with 0.02 VAF. At this time, the subject showed a mild anemia (12.7 g/dl), with increased RDW (14.2) and normal MCV (89). In 2007 hemoglobin level decreased to 11.8 g/dl, and mutation analysis showed an increase in *TET2* mutation VAF to 0.16. From 2007 to 2012 a further decrease in hemoglobin level together with increasing in RDW and MCV value was noticed. In 2013, a diagnosis of myelodysplastic syndrome (MDS) with unilineage dysplasia was performed.

Figure 5. (*A*) Overall survival of subjects with unexplained cytopenia form both "Health_&_Anemia" and "Monzino_80+" cohorts stratified according to the presence of mutations as idiopathic (non-clonal) cytopenia of undetermined significance (ICUS) *vs.* clonal cytopenia of undetermined significance (CCUS).

Probability of survival of individuals without cytopenia was also reported; (***B***) Overall survival of subjects with CCUS with highly specific mutational patterns for myeloid neoplasms *vs.* CCUS without specific mutational patterns and *vs.* age- and sex-matched patients affected with myeloid neoplasms (myelodysplastic syndromes, from EuroMDS database).

Fig.1

A) Prevalence of most frequently mutated genes in the "Health_&_Anemia" cohort B) Number of persons with 1, 2, or more than 2 variants in "Health_&_Anemia" cohort



C) Prevalence of most frequently mutated genes in the “Monzino_80+” cohort D) Number of persons with 1, 2 or more than 2 variants in the “Monzino_80+” cohort

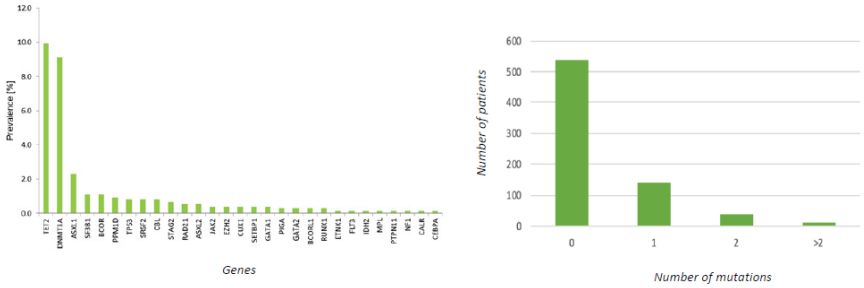
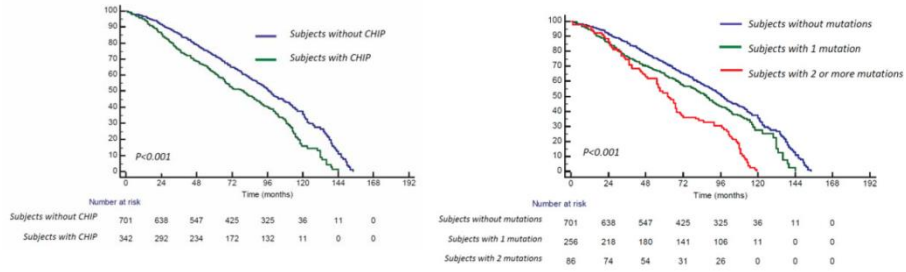
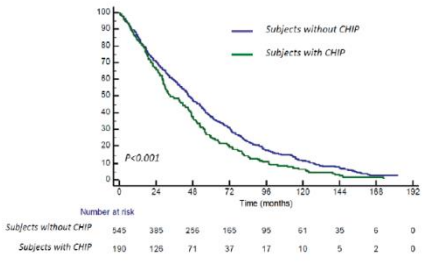


Fig. 2

A) Probability of survival according to the presence of CHIP in “Health_&_Anemia” cohort B) Probability of survival according to the number of variants in “Health_&_Anemia” cohort



C) Probability of survival according to the presence of CHIP in “Monzino_80+” cohort D) Probability of survival according to the number of variants in “Monzino_80+” cohort



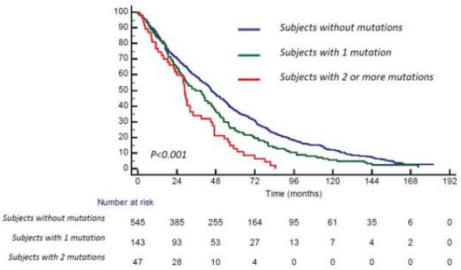


Fig.3

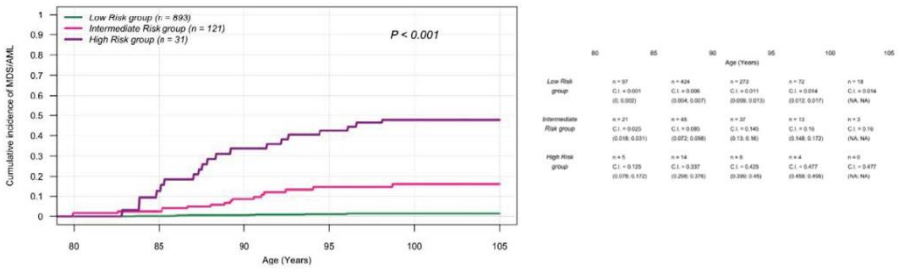
A)Score to predict individual risk of developing myeloid neoplasms

Parameter	HR (multivariable Cox regression analysis) [95%CI]	Score Value*
Presence of abnormal RBC-indices as defined as increased MCV (>98) and/or RDW (>14) value	2.61 [1.18-6.21]	1
Somatic mutation related to CHIP with VAF >0.096	3.24 [1.32-7.91]	1
Presence of TET2, DNMT3A, or ASXL1 mutations combined with other genetic lesions	5.02 [1.82-11.57]	2
Presence of spicing gene mutations	12.3 [6.78-23.85]	5

** Risk groups are defined as:*

RISK GROUP	SCORE VALUE	RISK FACTORS
Low Risk	0-1	Absence of risk factors; Presence of one of the following risk factors: abnormal RBC-indices or CHIP with VAF >0.096
Intermediate Risk	2-4	≥2 of the following risk factors: abnormal RBC-indices; CHIP with VAF >0.096; TET2, DNMT3A, or ASXL1 mutations combined with other genetic lesions
High Risk	≥5	Spicing gene mutations with or without additional risk factors

B) Cumulative incidence [95% CI] of myeloid neoplasms for oldest-old individuals stratified into three risk categories (learning cohort)



C) Validation of the score in an independent cohort of 727 subjects aged ≥75<80 years from “Health_&_Anemia” study.

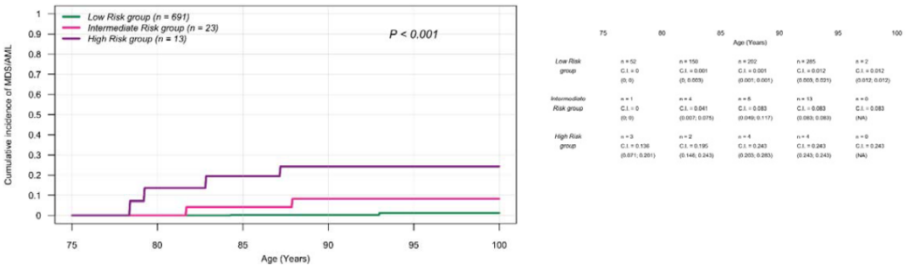
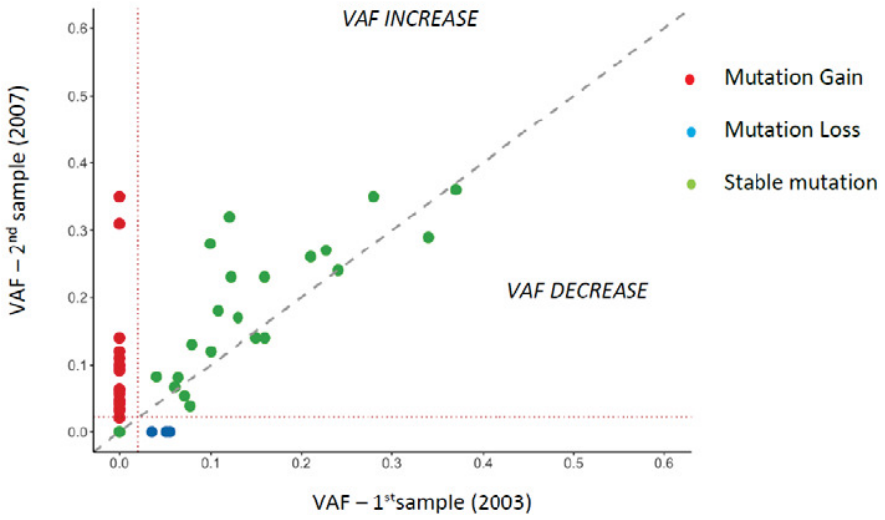
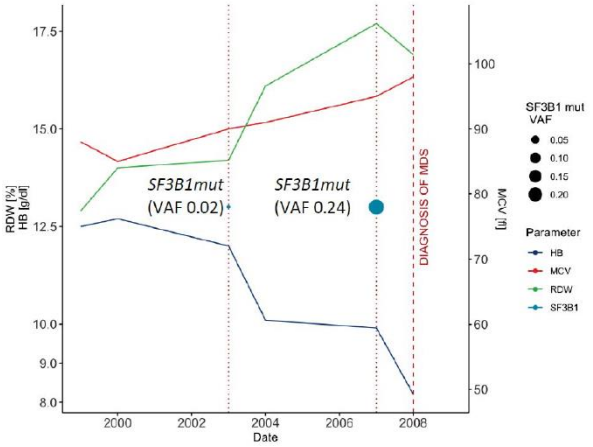


Fig.4

A) Clonal evolution in subjects from “Health_&_Anemia” cohort with multiple sample available



B) Clonal evolution - Subject #269



C) Clonal evolution -Subject #1145

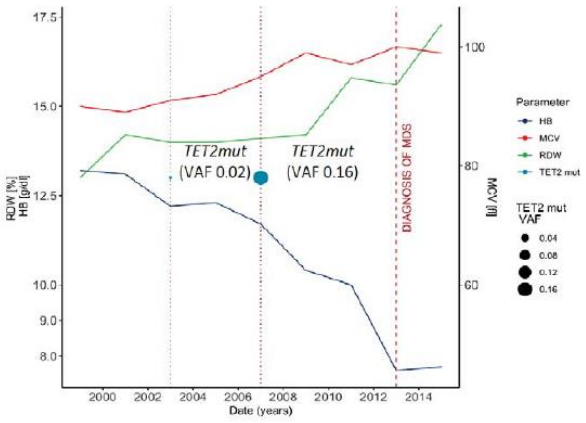
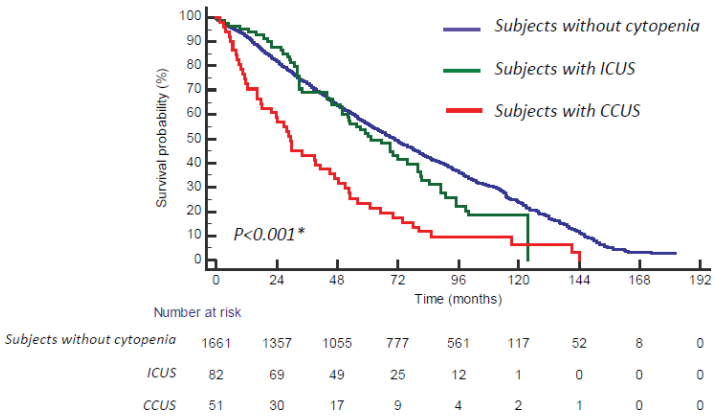


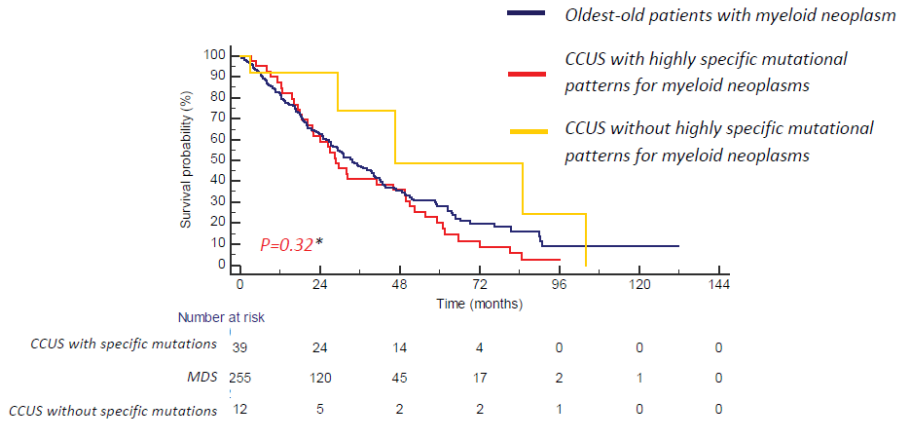
Fig. 5

A) Overall survival of subjects with unexplained cytopenia stratified according to the presence of mutations



* ICUS *vs.* subjects without cytopenia HR 0.81[0.75-0.97], P=0.35; CCUS *vs.* subjects without cytopenia HR 2.34 [1.98-3.04], P<0.001; CCUS *vs.* ICUS HR 2.06 [1.77-2.78], P=0.002

B) Overall survival of CCUS with and without highly specific mutational pattern for myeloid neoplasms vs. oldest-old patients with myeloid neoplasms (myelodysplastic syndromes)



* CCUS with specific mutations *vs.* myeloid neoplasms HR 0.84 [0.75-0.97], P=0.29; CCUS with specific mutations *vs.* CCUS without specific mutations HR 2.05 [1.61-4.64], P=0.06)

References

1. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477–2487

2. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488–2498

3. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014;20(12):1472-1478

4. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature.* 2018;559(7714):400-404

5. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood.* 2015;126(1):9-16

6. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med.* 2017;377(2):111-121

7. Shlush LI. Age-related clonal hematopoiesis. *Blood.* 2018;131(5):496-504

8. Warren JT, Link DC. Clonal hematopoiesis and risk for hematologic malignancy. *Blood.* 2020;136(14):1599-1605

9. Survey of Health, Ageing and Retirement in Europe http://www.share-project.org/

10. Chung SS, Park CY. Aging, hematopoiesis, and the myelodysplastic syndromes. *Blood Adv.* 2017;1(26):2572-2578

11. Christensen K, Doblhammer G, Rau R, et al: Ageing populations: the challenges ahead. *Lancet.* 2009;374(9696):1196-1208

12. Rollison DE, Howlader N, Smith MT, et al. Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood.* 2008;112(1):45-52

13. Wilkins JT, Ning H, Berry J, et al. Lifetime risk and years lived free of total cardiovascular disease. *JAMA.* 2012;308(17):1795-801

14. Stauder R, Valent P, Theurl I. Anemia at older age: etiologies, clinical implications, and management. *Blood.* 2018;131(5):505-514

15. Tettamanti M, Lucca U, Gandini F, et al. Prevalence, incidence and types of mild anemia in the elderly: the "Health and Anemia" population-based study. *Haematologica*. 2010;95(11):1849-56

16. Penninx BW, Pahor M, Woodman RC, et al. Anemia in old age is associated with increased mortality and hospitalization. *J Gerontol A Biol Sci Med Sci.* 2006;61(5):474-479

17. Riva E, Tettamanti M, Mosconi P, et al. Association of mild anemia with hospitalization and mortality in the elderly: the Health and Anemia population based study. *Haematologica.* 2009;94(1):22-28

18. Guralnik JM, Eisenstaedt RS, Ferrucci L, et al. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood.* 2004;104(8):2263-2268

19. Valent P. ICUS, IDUS, CHIP and CCUS: Diagnostic Criteria, Separation from MDS and Clinical Implications. *Pathobiology.* 2019;86(1):30-33

20. Lucca U, Tettamanti M, Logroscino G, et al. Prevalence of dementia in the oldest old: the Monzino 80-plus population-based study. *Alzheimers Dement.* 2015;11(3):258-270

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

22. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood.* 2017;130(6):742-752

23. Moskowitz CS, Pepe MS. Quantifying and comparing the accuracy of binary biomarkers when predicting a failure time outcome. *Stat Med.* 2004;23(10):1555-1570

24. Liu, X., Jin, Z. Optimal survival time-related cut-point with censored data. *Stat Med.* 2015;34(3):515-524

25. cenROC R package: functions to estimate a smoothed and a non-smoothed (empirical) time-dependent ROC curve (receiver operating characteristic curve), the corresponding area under the ROC curve (AUC) and the optimal cutoff point for the right and interval censored survival data. https://cran.rproject. org/web/packages/cenROC/cenROC.pdf

26. Chen BE, Kramer JL, Greene MH, et al. Competing Risks Analysis of Correlated Failure Time Data. *Biometrics.* 2008;64(1):172-179

27. Klein J, Moeschberger M. Survival Analysis Techniques for Censored and Truncated Data. New York: Springer, 1997.

28. Evert J1, Lawler E, Bogan H, et al. Morbidity profiles of centenarians: survivors, delayers, and escapers. *J Gerontol A Biol Sci Med Sci.* 2003;58(3):232-237

29. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. *Blood.* 2013;122(25):4021-4034

30. Malcovati L, Gallì A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood.* 2017;129(25):3371-3378

31. Fuster JJ, MacLauchlan S, Zuriaga MA et al, Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science.* 2017;355(6327):842-847

32. Garagnani P, Marquis J, Delledonne M, et al. Whole-genome sequencing analysis of semi-supercentenarians. *Elife*. 2021;10: e57849.

33. van Zeventer IA, de Graaf AO, Wouters HJCM, et al. Mutational spectrum and dynamics of clonal hematopoiesis in anemia of older individuals. *Blood.* 2020;135(14):1161-1170

34. Bersanelli M, Travaglino E, Meggendorfer M, et al. Classification and personalized prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. *J Clin Oncol*. 2021;39(11):1223-1233.

Chapter 5.1

Additional data

Clinical relevance of Splicing genes mutations in myeloid malignancies

As described in the previous publications, mutations in Splicing genes showed the highest predictive value for myeloid neoplasms.

Considering that Splicing genes are responsible of presumptive evidence of myeloid neoplasm in healthy over-eighty subjects, these results are consistent with literature data in which splicing mutations are driver early events in myelodysplastic disease evolution.[1]

We also previously demonstrated that five dominant genomic features, including Splicing genes mutations, determined drive disease evolution in Myelodisplatic syndromes (See State of the art 2).[2]

SF3B1 mutations, in particular, have been identified in individuals with clonal hematopoiesis of indeterminate potential (CHIP), and they are known to promote tumorigenesis of various neoplastic diseases.[3]

In order to confirm this hypothesis, we screened a cohort of 2215 onco-hematological patients, so distributed: 875 MDS subjects, 200 with MDS/MPN overlapping features, 1090 secondary AML and 50 MF patients, respectively (Tab.1).

MDS	MDS/MPN	sAML	MF	TOTAL PTs
875	200	1090	50	2215

*MDS: Myelodysplastic syndrome; MDS/MPN: Myelodysplastic syndrome/Myeloproliferative neoplasms overlapping features; sAML: secondary Acute Myeloid Leukemia; MF: Myelofibrosis

Table 1. Distribution of patients of the cohort across hematological malignancies.

We focused our attention on Splicing genes mutations, evaluating the distribution according to different hematological malignancies, the timing in oncogenesis, the clonality of each identified mutations and how they affect overall survival (OS).

Incidence of Splicing mutations across different hematological malignancies

We evaluated how frequently Splicing genes are affected by mutations across different myeloid malignancies. As show in table 2, these aberrations are very frequent in MDS and MDS/MPN patients, that are mutated in 47,53% (875 mutated vs 966 wild type) and 64,31% (200 mutated vs 111 wild type) of cases, respectively. The percentages of Splicing genes mutations in sAML and MF cohorts are about 20%. (AML: 264 mut vs 826 wt, 24,22%; MF: 50 mut vs 177 wt, 22,03%).

In each malignancy, Splicing mutations occur mainly as single events,[4] without co-occurrence aberration in the same genes category, as

follows: 44,16% of MDS, 62,7% of MDS/MPN, 22,66% of sAML, 20,7% of MF patients, respectively.

nMut (RNA splicing)	MDS	MDS/MPN	AML	MF
0	966 (52.47%)	111 (35.69%)	826 (75.78%)	177 (77.97%)
1	813 (44.16%)	195 (62.7%)	247 (22.66%)	47 (20.7%)
2	59 (3.2%)	5 (1.61%)	15 (1.38%)	3 (1.32%)
3	3 (0.16%)	0 (0%)	2 (0.18%)	0 (0%)
TOT	1841	311	1090	227
Number of samples with at least one splicing mutation	875	200	264	50
Fraction of samples with at least one splicing mutation	47,53%	64,31%	24,22%	22,03%

Table 2. Frequency of Splicing genes mutations across myeloid malignancies.

Splicing genes are mutated early in MDS disease evolution

We calculated a global ranking of analyzed genes revealing how early in diseases evolution they are mutated (Fig. 1). The results showed that mutations in genes involved in RNA splicing, such as Chromatin and histone modifiers, occur early and are driver events in MDS, MDS/MPN and sAML diseases progression (Fig.1A,B,C). These aberrations define the trajectories of clonal evolution of the cancer, acting with co-occurent genetic lesions to drive disease evolution. On the contrary, in MF patients splicing gene such as *SRSF2* and *U2AF1* are typically mutated late (Fig.1D).

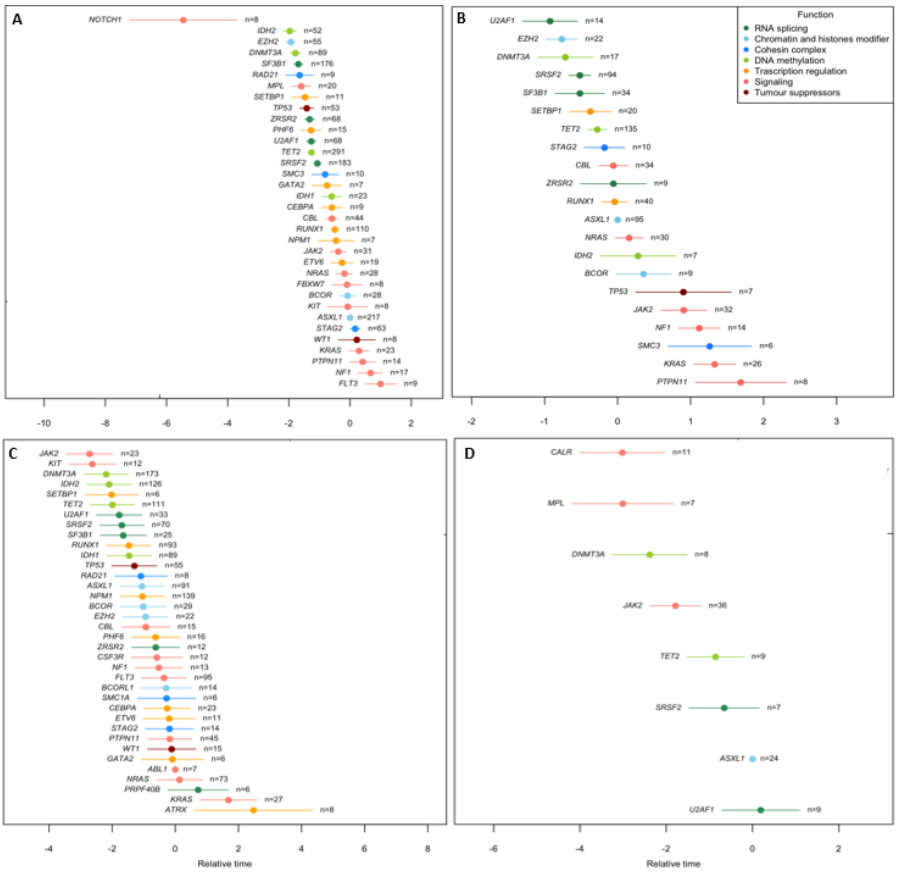


Figure 1. Results of a Bradley-Terry model showing the relative temporal order of genes involved. Genes are classified and colored according to their biological function.
Mutations in RNA splicing gene are early events in pathogenesis of MDS (panel A), MDS/MPN (panel B), and sAML (panel C), respectively. In panel D the timing of MF mutations show that splicing gene aberrations occur late, after Signaling and DNA methylation categories.

Only *SF3B1* clonal mutations correlate with a good prognosis in MDS patients

Finally, we evaluated the effects of mutations in Splicing genes on prognosis ad overall survival of patients. Overall, they correlate with a worse prognosis in any hematological malignancies, especially if they affect a majority clone (Fig. 3).

Only *SF3B1* aberrations in MDS and MDS/MPN cases have a positive effect on prognosis, with a significative value in the second group.

Kaplan-Meyer estimator allowed to consolidate the previous data, underling how *SF3B1* as a single mutational events affect only MDS patient, with a percentage of 19.6% of cases (Fig.4 A).

Moreover, the confidence interval for survival distribution underlines the worsening effects of cumulative splicing genes mutations mostly in MF cohort (Fig.4 D).

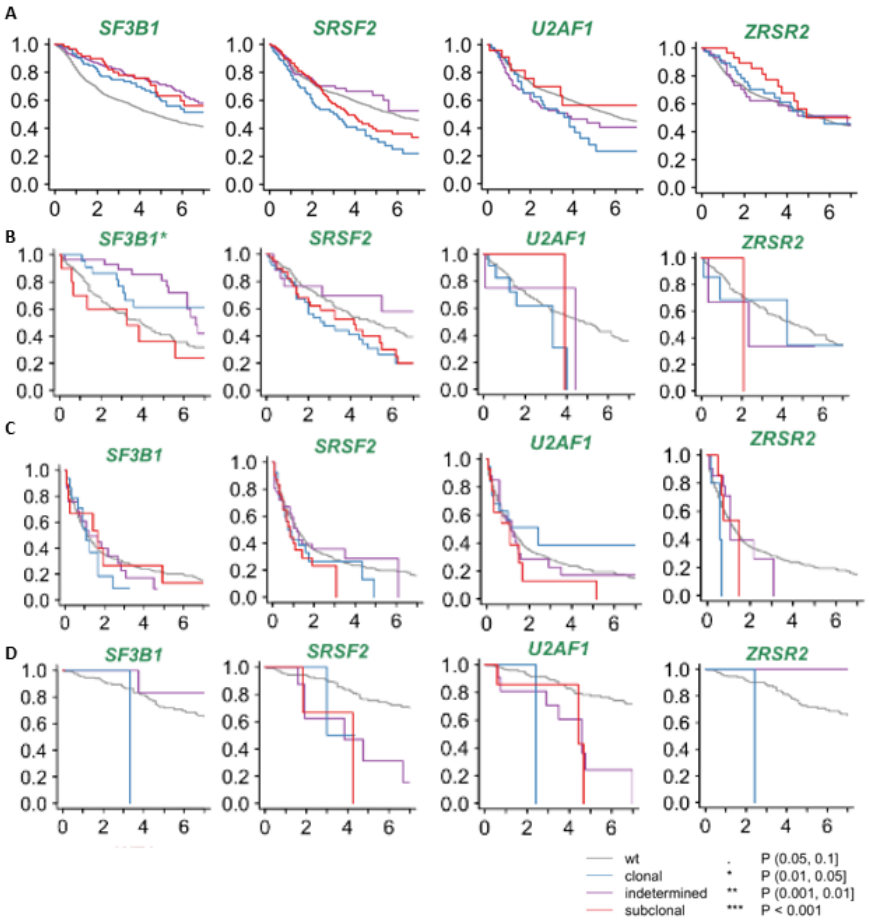


Figure 3. Survival analysis for hematological malignancies patients according to Splicing genes mutational status.
Kaplan–Meier estimates overall survival in patients harboring clonal or subclonal*SF3B1*, *SRSF2*, *U2AF1* and *ZRSF2* mutations. Except for *SF3B1* in MDS and MDS/MPN patients, mutations in these genes correlate with a worse prognosis in each malignancy (panel A: MDS; panel B: MDS/MPN; panel C: sAML; panel D: MF).

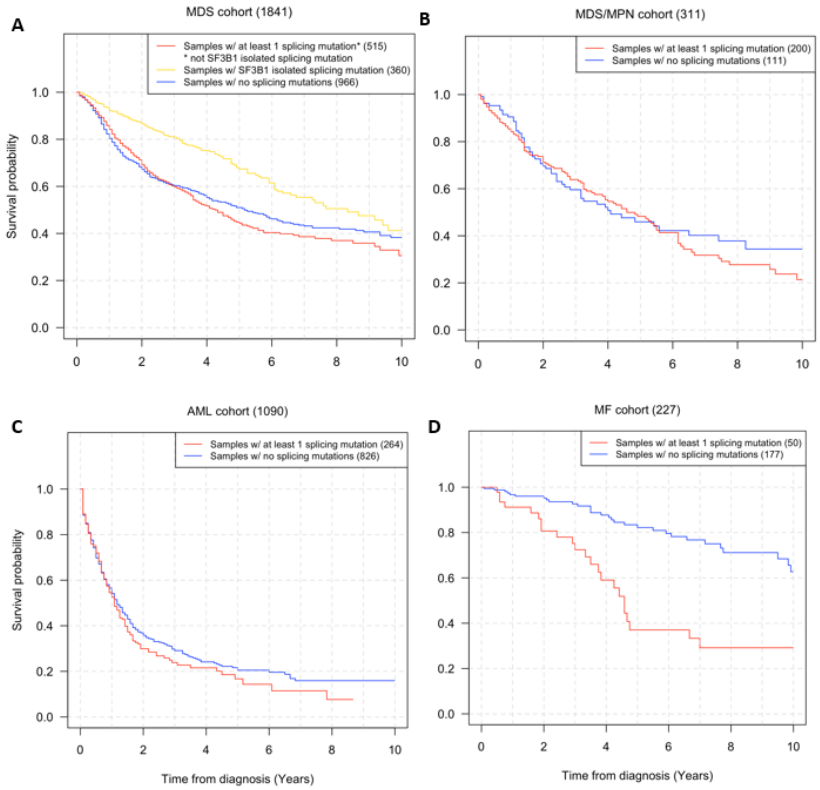


Figure 4. Cumulative probability of overall survival according to the presence of Splicing gene mutations, calculated by Kaplan-Meier method. Effects on MDS (panel A), MDS/MPN (panel B), sAML (panel C), and MF (panel D) cohorts, respectively. Only *SF3B1* aberration as a single event have positive effects on MDS patients' prognosis (panel A, yellow line).

Discussion

The results of this study represent a systematical characterization of the clonal and subclonal mutational landscape in Splicing genes across different myeloid malignancies. We confirmed that these aberrations occur typically early in MDS, MDS/MPN and sAML patients, instead they are late events in MF clonal hematopoiesis. As early driver events in clonal hierarchy, they dictate disease evolution with distinct clinical phenotypes.

Both as early or late events, mutations in Splicing genes have prognostic significance and correlate with unfavorable prognosis in each hematological malignancy. Only *SF3B1* aberrations significantly predict for a better outcome in MDS and MDS/MPN patients.[5]

Overall, these preliminary data strongly support the predisposing role of Splicing genes mutation in early phases of clonal hematopoiesis.

Additional methods

Bradley-Terry model – Mutation acquisition order

The Bradley–Terry (BT) model was used to deduce the relative temporal order of mutation acquisitions within patients, which reflects how early in disease progression the genes are mutated. It allows to assess the prognostic value of clonal *vs.* subclonal mutations.

In order to determine the relative order of mutation acquisition, comparisons were made for each pair of mutations in each patient. The model guarantees to infer the relative order in which two events occurred even without a time course experiment. It was used to determine the relative probabilities of a gene mutation occurring first or second. For each patient the proportions of cells carrying each mutation, the variant allele fractions corrected for any copy number change at the site of the variant were considered.

References

1. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.***122**, 3616–3627 (2013).

2. Bersanelli M, Travaglino E, Meggendorfer M, et al. Classification and personalized prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. *J Clin Oncol*. 2021;**39**(11):1223-1233.

3. Pangallo, J. et al. Rare and private spliceosomal gene mutations drive partial, complete, and dual phenocopies of hotspot alterations. Blood **135**, 1032–1043 (2020).

4. Inoue, D., Bradley, R. K. & Abdel-Wahab, O. Spliceosomal gene mutations in myelodysplasia: Molecular links to clonal abnormalities of hematopoiesis. *Genes Dev.***30**, 989–1001 (2016).

5. Wang, X., Song, X. & Yan, X. Effect of RNA splicing machinery gene mutations on prognosis of patients with MDS A meta-analysis. *Med. (United States)* **98**, (2019).

Chapter 6

Summary, Conclusions and Future Perspectives

In the present project, we focused our attention in dissecting the role of genetic predisposition in both childhood and adult hematological malignancies. With this purpose, we planned and developed our study through several tasks, characterized by the joint purpose of improving knowledge about biological mechanisms of early stage of oncogenesis. We made a focus on cancer-prone genetic alterations that act in pre-leukemic phase.

Actually, despite predisposition occurs in 5-10 % of pediatric cancer,[1,2] it is still a nebulous field, that has to be better characterized. In the adult setting, clonal evolution acts similarly, and the age-dependent accumulation of somatic mutations increases with prevalence of myeloid neoplasms among older individuals. Also in this context, the lines of evidence are recent and the mechanisms by which specific co-occurrence of somatic events predispose to hematological malignancies, must be further clarified.

At first, we screened a cohort of 120 consecutive diagnosis of pediatric patients affected by Acute Lymphoblastic Leukemia. Genetic profiling confirmed literature data regarding the crucial role in Leukemogenesis of target genes, like those belonging to Ras pathway, both in term of incidence and pathogenicity. Moreover, it shed light on novel pathways, such as Cohesin genes. Their germline mutations, usually associated to genetic syndromes (e.g., Cornelia de Lange S.), are not random or sporadic events, but occur with a frequency (6%) that is not

negligible and worthy of further study to understand their impact on tumor transformation.

Therefore, we investigated the contribution of Cohesin germline variants in the specific hematological setting, such as the pre-leukemic phase, dissecting biological mechanisms by which an aberrant genetic landscape can promote cancer prone conditions in pediatric patients, evolving to leukemia. We found variants in Cohesin genes, in a not negligible percentage of cases. Most of them were missense, in contrast to what observed in the Cohesinopathies, in which the most of the mutations are frameshift.[3] Missense variants are consistent with a milder phenotype, as the one expected in cancer predisposing conditions. Among the identified variants in the present study, we focused our attention to *STAG1* and *RAD21* genes, due to their role in biological mechanisms fundamental for cells' integrity and survival.[4,5] Overall, we can conclude that germline variants in these genes lead to a poor chromosomal strength and promote spontaneous instability, resulting in a lowered response to exogenous and endogenous agents, commonly altered in oncogenesis.[6,7] Moreover, the failure of DNA damage repair mechanisms worsens the accumulation of damage and aggravates the risk of somatic events, responsible of the disease' onset.

Familiar cancer histories of two *RAD21* mutated patients highlighted how the germline variants increase cancer risk but need additional factors to drive tumor evolution. Classical examples are synergizing

germline mutations and external influences, that affect the genetic profile, or onset of somatic events responsible of disease.
Overall, these results confirmed that germline Cohesins variants alter cellular mechanisms involved in oncogenesis even in a pre-disease phase, setting up the ideal cancer prone conditions that lead to canonical second hits.

In order to evaluate overall the contribution of genetic predisposition in cancer during the time of life, we investigated the role of clonal evolution in the adult setting. Despite it is considered normal in aging, it is also significantly associated with cardiovascular disease, as well as solid tumors, and hematological malignancies (MDS and both de novo /therapy-related AML).[8,9]
Our results demonstrated that the phenomenon characterizes one third of healthy oldest-old individuals (over-eighty), defining specific mutational patterns that outline a different risk of developing inflammatory-associated diseases or myeloid neoplasms.

Our line of evidence allows the definition of a predictive model that classified 3 risk groups, according to the mutational landscape of the patient. In particular, mutations either in splicing genes or *JAK2* or the presence of multiple mutations (*DNMT3A, TET2, ASXL1* with additional genetic lesions), as well as variants with allele frequency ≥0.096, have a positive predictive value for myeloid neoplasms. Single events in *DNMT3A, TET2, ASXL1* or *JAK2* occur in patients with coronary heart

disease and rheumatoid arthritis. Moreover, the presence of highly specific mutation patterns in cases with unexpected cytopenia tracks a myelodysplastic-like phenotype that influence the overall survival.
Finally, we confirmed that splicing genes variants, that are the highest predictive value for myeloid neoplasms, are driver events that dictate phenotype, prognosis and overall survival of myeloid patients.

We are aware that this study contributes in a preliminary manner to a broader scenario, that need to be explored in the near future.

In conclusion, in the context of genetic predisposition in childhood, we demonstrated not only that the phenomenon exists and alter genetic conditions creating cancer-prone conditions, but also that involves genes that are not classically related to full-blown stage of hematological disease.
On the same way, in the adult context, the mutational screening of 1794 oldest-old individuals established how the differential age-dependent accumulation of somatic mutations increases prevalence of myeloid malignancies among older individuals. We underlined the role of Splicing genes mutations not only as early events in pathogenesis, but also in a previous phase, as key players in determine future direction towards the onset of myelodysplastic disease.

Overall, the knowledge of these alterations has potentially different impacts: first, it will improve the understanding of tumorigenesis,

opening new scenarios regarding the contribution of genetic predisposition and clonal evolution to hematological malignancies.

Secondly, a better knowledge and characterization of predisposing genetic alterations could have significative effects on both patients' care and familial genetic counseling, enabling targeted surveillance strategies and tailored therapeutical adjustments.

References

1. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).

2. Inaba, H. & Mullighan, C. G. Pediatric acute lymphoblastic leukemia. *Haematologica* **105**, 2524–2539 (2020).

3. Mannini, L., Cucco, F., Quarantotti, V., Krantz, I. D. & Musio, A. Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. *Hum. Mutat.* **34**, 1589–1596 (2013).

4. Canudas, S. & Smith, S. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.* **187**, 165–173 (2009).

5. Cheng, H., Zhang, N. & Pati, D. Cohesin subunit RAD21: From biology to disease. *Gene* **758**, 144966 (2020).

6. De Pascalis, I. *et al.* Sister chromatid exchange: A possible approach to characterize familial breast cancer patients. *Oncol. Rep.* **33**, 930–934 (2015).

7. Bauerschmidt, C. *et al.* Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Res.* **38**, 477–487 (2010).

8. Desai, P., Hassane, D. & Roboz, G. J. Clonal Hematopoiesis and risk of Acute Myeloid Leukemia. *Best Pract. Res. Clin. Haematol.* **32**, 177–185 (2019).

9. Desai, P. & Roboz, G. J. Clonal Hematopoiesis and therapy related MDS/AML. *Best Pract. Res. Clin. Haematol.* **32**, 13–23 (2019).

Additional publications

1. Chiereghin C, Travaglino E, Zampini M, Saba E, **Saitta C**, Riva E, Bersanelli M, Della Porta MG. The Genetics of Myelodysplastic Syndromes: Clinical Relevance. Genes (Basel). 2021

2. Bersanelli M, Travaglino E, … **Saitta C**, … Della Porta MG. Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. J Clin Oncol. 2021

3. Mazzola M, Pezzotta A, Fazio G, Rigamonti A, Bresciani E, Gaudenzi G, Maria Pelleri MC., **Saitta C**, Ferrari L, Parma M, Fumagalli M, Biondi A, Cazzaniga G, Marozzi A, Pistocchi A. Dysregulation of NIPBL leads to impaired RUNX1 expression and haematopoietic defects. J Cell Mol Med. 2020

4. Fazio G, Massa V, Grioni A, Bystry V, Rigamonti S, **Saitta C**, Galbiati M, Rizzari C, Consarino C, Biondi A, Selicorni A, Cazzaniga G. First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia. J Clin Pathol. 2019

5. Mazzola M, Deflorian G, Pezzotta A, Ferrari L, Fazio G, Bresciani E, **Saitta C**, Ferrari L, Fumagalli M, Parma M, Marasca F, Bodega B, Riva P, Cotelli F, Biondi A, Marozzi A, Cazzaniga G., Pistocchi A. NIPBL: a new player in myeloid cell differentiation. Haematologica. 2019

6. Grioni A, Fazio G, Rigamonti S, Bystry V, Daniele G, Dostalova Z, Quadri M, **Saitta C**, Silvestri D, Songia S, Storlazzi CT, Biondi A, Darzentas N, Cazzaniga G. A Simple RNA Target Capture NGS Strategy for Fusion Genes Assessment in the Diagnostics of Pediatric B-cell Acute Lymphoblastic Leukemia. Hemasphere. 2019

Recently submitted manuscripts

1. Fazio G, Bresolin S, Silvestri D, **Saitta C**, Quadri M,… Cazzaniga G. *PAX5* fusions genes are frequent in poor risk childhood Ph-like acute lymphoblastic leukemia and can be efficiently targeted by the kinase inhibitor Nintedanib. Submitted to Blood as BLD-2021-014118.
2. Palmi C, Bresolin S, Junk S, Fazio G,…**Saitta C**,…Cazzaniga G. Definition and prognostic impact of Ph-like and *IKZF1*plus features in children with Down Syndrome Acute Lymphoblastic Leukemia.