

# UNIVERSITA' DEGLI STUDI DI ROMA TOR VERGATA



MACROAREA DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA MAGISTRALE IN BIOLOGIA CELLULARE E MOLECOLARE E  
SCIENZE BIOMEDICHE

TESI DI LAUREA

## **Deregolazione del *pathway* del TGF- $\beta$ in pazienti affetti da sindrome mielodisplastica con mutazione del gene *SF3B1***

**Relatore:**

*Prof.ssa Bianca Maria Ciminelli*

**Candidato:**

*Giorgia Silvestrini*

**Relatore esterno:**

*Prof.ssa Maria Teresa Voso*

**Correlatore:**

*Dott. Emiliano Fabiani, PhD*

*Dott.ssa Giulia Falconi, PhD*

Anno Accademico 2019/2020

*“Non ti arrendere mai.  
Di solito è l'ultima chiave del mazzo  
quella che apre la porta.”  
Paulo Coelho*

## INDICE

<b>1. INTRODUZIONE</b> .....	<b>9</b>
1.1. Sindromi mielodisplastiche .....	9
1.2. Patogenesi .....	10
1.3. Aspetti citogenetici .....	12
1.4. Aspetti genetici .....	14
1.5. Classificazione WHO .....	25
1.6. Classificazione IPSS e IPSS-R .....	26
1.7. Trattamento .....	28
1.8. Microambiente midollare e Cellule staminali ematopoietiche .....	30
1.9. TGF- $\beta$ <i>signaling pathway</i> .....	32
1.9.1. Ruolo del TGF- $\beta$ nell'emopoiesi e nelle mielodisplasie .....	35
<b>2. SCOPO DELLA TESI</b> .....	<b>39</b>
<b>3. MATERIALI E METODI</b> .....	<b>40</b>
3.1. Popolazione di studio .....	40
3.2. Separazione degli elementi mononucleati del sangue mediante gradiente di densità .....	41
3.3. Analisi del genoma .....	42
3.3.1. Estrazione del DNA genomico .....	42
3.3.2. Valutazione qualitativa e quantitativa del DNA .....	43
3.3.3. <i>Next Generation Sequencing</i> (NGS) .....	45
3.3.3.1. Il flusso di lavoro in NGS .....	46
3.3.4. <i>Whole Genome Sequencing</i> (WGS) .....	49
3.4. Analisi del trascrittoma .....	50
3.4.1. Estrazione dell'RNA totale .....	50
3.4.2. Quantificazione spettrofotometrica dell'RNA .....	51
3.4.3. RNA-Seq .....	51
3.4.4. Retrotrascrizione dell'RNA in cDNA .....	52
3.4.5. Real-Time PCR (qRT-PCR) .....	52

3.5. Analisi bioinformatica .....	54
3.5.1. Analisi differenziale dell'espressione genica e di arricchimento dei geni .....	55
3.6. Analisi statistica .....	56
<b>4. RISULTATI .....</b>	<b>57</b>
4.1. Analisi del profilo mutazionale della popolazione di studio tramite <i>targeted</i> -NGS .....	57
4.2. Analisi di WGS .....	59
4.3. Identificazione e analisi dei geni differenzialmente espressi .....	62
4.4. Validazione dei risultati di RNA-Seq .....	69
<b>5. DISCUSSIONE .....</b>	<b>72</b>
<b>6. BIBLIOGRAFIA .....</b>	<b>77</b>
<b>7. RINGRAZIAMENTI .....</b>	<b>86</b>

## INDICE DELLE ABBREVIAZIONI

<i>ABL = ABL proto-oncogene</i>	<i>CSNK1A1 = Casein kinase 1 alpha 1</i>
<i>ABCB7 = ATP binding cassette subfamily B member 7</i>	<i>CUL7 = Cullin 7</i>
<i>ACLY = ATP-citrate lyase</i>	<i>DCN = Decorin</i>
<i>ACTN3 = Actinin alpha 3</i>	<i>DEG = Differentially expressed genes</i>
<i>ARSD = Arylsulfatase D</i>	<i>DNAH = Dynein axonemal heavy chain</i>
<i>ASXL1 = ASXL transcriptional regulator 1</i>	<i>DNMT3A = DNA methyltransferase 3 alpha</i>
<i>ATXN3 = Ataxin 3</i>	<i>ETV6 = ETS variant transcription factor 6</i>
<i>BFU = Burst Forming Units</i>	<i>EZH2 = Enhancer of zeste 2 polycomb repressive complex 2 subunit</i>
<i>BM-MNCs = Bone Marrow-Mononuclear Cells</i>	<i>FAM20C = FAM20C golgi associated secretory pathway kinase</i>
<i>BMP = Bone Morphogenetic Protein</i>	<i>FLT3 = Fms related receptor tyrosine kinase 3</i>
<i>BPS = Branching Point Site</i>	<i>FN1 = Fibronectin</i>
<i>BRAF = B-Raf proto-oncogene, serine/threonine kinase</i>	<i>FST = Follistatin</i>
<i>CACNA1H = Calcium voltage-gated channel subunit alpha1 H</i>	<i>GDF = Growth differentiation factor 11</i>
<i>CALR = Calreticulin</i>	<i>HEAT = Huntingtin, Elongation factor 3, protein phosphatase 2A, Targets of rapamycin 1</i>
<i>CBL = Cbl proto-oncogene</i>	<i>IDH1/2 = Isocitrate dehydrogenase</i>
<i>CCUS = Clonal Cytopenia of Undetermined Significance</i>	<i>IGF1 = Insulin-like growth factor I</i>
<i>CEBPA = CCAAT enhancer binding protein alpha</i>	<i>IPSS = International Prognostic Scoring System</i>
<i>CFU = Colony Forming Units</i>	<i>IPSS-R = International Prognostic Scoring System-Revised</i>
<i>CHIP = Clonal Hematopoiesis of Indeterminate Potential</i>	<i>ITGA10 = Integrin subunit alpha 10</i>
<i>CSE = Cellula staminale ematopoietica</i>	<i>JAK2 = Janus kinase 2</i>
<i>CSF3R = Colony stimulating factor 3 receptor</i>	<i>KDM4B = Lysine demethylase 4B</i>

*KIT = KIT proto-oncogene, receptor tyrosine kinase*

*LAM = Leucemia acuta mieloide*

*LAP = Latency-associated Peptide*

*LLC = Large Latent Complex*

*LNP1 = Leukemia NUP98 fusion partner 1*

*LTBP = Latent TGF- $\beta$  Binding Protein*

*MMP2 = Matrix metalloproteinase 2*

*MP = Mature Protein*

*MPL = MPL proto-oncogene, thrombopoietin receptor*

*NF-K $\beta$  = Nuclear factor kappa B*

*NGS = Next Generation Sequencing*

*NLGN2 = Neuroligin 2*

*NPM1 = Nucleophosmin 1*

*OS = Overall Survival*

*PFS = Progression Free Survival*

*PIEZO1 = Piezo type mechanosensitive ion channel component 1*

*PI3K/Akt = Phosphatidylinositol 3-kinase/serine/threonine-specific protein kinase*

*PLAU = Plasminogen activator, urokinase*

*PTPN11 = Protein tyrosine phosphatase non-receptor type 11*

*RAS = RAS proto-oncogene, GTPase*

*RPS14 = Proteina ribosomiale 14*

*RUNX1 = RUNX family transcription factor 1*

*SETBP1 = SET binding protein 1*

*SF = Splicing Factor*

*SF3B1 = Splicing factor 3B subunit 1*

*SH3TC1 = SH3 domain and tetratricopeptide repeats 1*

*SLC = Small Latent Complex*

*SLC7A8 = Solute carrier family 7 member 8*

*SMAD = Small Mother Against Decapentaplegic*

*SMD = Sindrome mielodisplastica*

*SNP = Single Nucleotide Polymorphism*

*SNV = Single Nucleotide Variant*

*SP = Signal Peptide*

*SR = Serine/Arginine-rich proteins*

*SRSF2 = Serine and arginine rich splicing factor 2*

*TET2 = Tet methylcytosine dioxygenase 2*

*TGF- $\beta$  = Transforming growth factor  $\beta$*

*TIF1 $\gamma$  = Transcriptional intermediary factor 1 gamma*

*TP53 = Tumor protein p53*

*TTN = Titin*

*U2AF1 = U2 small nuclear RNA auxiliary factor 1*

*VAF = Variant allele frequency*

*WGS = Whole Genome Sequencing*

*ZRSR2 = Zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2*

*WT1 = WT1 transcription factor*

## ABSTRACT

Myelodysplastic syndromes (MDS) represent a heterogeneous group of hematologic malignancies affecting the hematopoietic stem cell, characterized by ineffective hematopoiesis, cytopenias in peripheral blood, cytogenetic and/or molecular abnormalities, and a variable risk of transformation into acute myeloid leukemia (AML). These disorders generally affect elderly patients and may occur *de novo*, without prior exposure to carcinogens, or be secondary to exposure to cytotoxic agents or treatments (MDS therapy-related).

MDS subgroups are defined by the number of dysplastic lineages, presence or absence of ring sideroblasts, percentage of bone marrow and peripheral blood blasts and type of cytogenetic alterations. In myelodysplastic syndromes, the myeloid stem cell undergoes changes at the DNA level, resulting in the production of dysplastic cells, which fail to complete their differentiation program until maturation in blood cells. This results in ineffective erythropoiesis, defined as the inability to produce an adequate number of mature erythroid cells. The result is hyperplasia of immature erythroid cells in the bone marrow and anemia in the peripheral blood.

Gene alterations are frequent in patients with MDS in association or not with cytogenetic abnormalities. Genetic alterations affect epigenetic regulation, which involves DNA methylation, the splicing process, transcription factors and the cytokine signaling pathway. *SF3B1*, *SRSF2* and *U2AF1* represent the most commonly mutated splicing factors in MDS. Mutations identified in MDS patients alter the function of these factors, leading to defects in the process of RNA maturation and the production of variants with a different or impaired biological activity, resulting in an altered gene expression profile.

The aim of this dissertation was to analyze the gene expression profile in patients with myelodysplastic syndrome, with particular reference to genes involved in RNA splicing.

MDS patients enrolled for the study (n=50, 30 male, 20 female, median age at diagnosis 75 years old, range 40-91 years) were selected from an initial cohort of 300 patients, of which we had paired DNA and RNA samples extracted from bone marrow mononuclear

cells. In detail, 32 out of 50 patients were selected by targeted-NGS (Next Generation Sequencing), based on the presence of mutations in splicing genes, while in 18 cases the mutational profile was unknown. Bone marrow samples from 64 hospital patients without hematologic disease, matched for age and sex with MDS patients, were used as control.

In order to assess the functional role of additional mutations in the three subgroups of patients with mutations in splicing genes, we performed a Whole Genome Sequencing (WGS) analysis in the entire study cohort (n=50).

WGS showed that *FAM20C* and *KDM4B* were mutated at high frequency (60.87% and 52.17%, respectively) in all three categories of patients considered, whereas *ACLY* was mutated at high frequency in *SF3B1* mutated (90%) and *SRSF2* mutated (50%) patients, but not in patients harboring *U2AF1* mutations (0%).

Data obtained from transcriptome analysis, performed by RNA-seq, showed an altered expression profile in MDS patients (n=46) compared to controls (n=64). In particular, 300 genes were up-regulated, while 3921 genes were down-regulated. Stratification of the analysis, according to the presence of mutations in *SF3B1*, *SRSF2* and *U2AF1*, led to the identification of differentially expressed genes (DEGs) in each patient subgroup. In particular, 622 DEGs were identified in *SF3B1* mut patients, 13 DEGs were identified in *SRSF2* mut patients, and only 4 genes were differentially expressed in *U2AF1* mut patients.

DEGs, selected based on the chosen cut-offs ( $p_{adj} < 0.05$  and  $\log_2FC \geq 2$  and  $\leq -2$ ), were used to perform an enrichment analysis using the DAVID bioinformatics tool, which identified a list of 37 deregulated pathways in MDS patients harboring *SF3B1* mutation, whereas no pathways were highlighted for *SRSF2* and *U2AF1* mutated patients.

We then chose to investigate the TGF- $\beta$  pathway due to the recent approval of a new drug, which is a recombinant protein (*Luspatercept*), able to modulate signal transduction, and restore normal erythropoiesis in patients with ineffective hematopoiesis.

We found 6 DEGs belonging to the TGF- $\beta$  pathway (*CHRD*, *DCN*, *SMAD9*, *FST*, *ID4*, *PITX2*) and 10 DEGs which are known to modulate TGF- $\beta$  activity (*BDNF*, *BMPER*, *COL1A1*, *COL1A2*, *FN1*, *GLI2*, *IGF1*, *MMP2*, *NFIB* and *PLAU*), and validated by Q-RT-PCR the expression of 7 genes, confirming in all cases the gene deregulation identified by RNA-Seq.

Our study showed a significant increase in the mRNA expression levels of *DCN*, *FST*, *SMAD9*, *FN1*, *IGF1*, *MMP2* and a significant decrease in the mRNA expression level of *PLAU* in patients harboring *SF3B1* mutation compared with wild-type patients for splicing factors, and bone marrow mononuclear cells isolated from healthy donors.

To date, we are collecting a prospective cohort of MDS patients (*SF3B1* mutated and *SF* wild-type) in order to validate our results in an independent study cohort.

In conclusion, my dissertation shows that MDS patients display an altered and extremely heterogeneous gene expression profile. This heterogeneity could be linked to the presence of specific mutations in genes involved in splicing process. In particular, patients harboring *SF3B1* mutation show a deregulation of genes belonging to TGF- $\beta$  pathway, suggesting for some of them a possible role as molecular targets.

## 1. INTRODUZIONE

### 1.1. Sindromi mielodisplastiche

Il termine “displasia” (dal greco antico *δυσ-* “cattivo” e *πλάσις* “formazione”) è un termine usato in patologia per indicare la formazione anomala di organi o tessuti. Nel 1982, il termine “mielodisplasia” è stato utilizzato dal gruppo FAB (*French-American-British*) per descrivere le alterazioni morfologiche delle linee cellulari mieloidi in condizioni pre-leucemiche, alterazioni che hanno portato alla definizione delle sindromi mielodisplastiche (SMD) (Bennett et al., 1982).

Le sindromi mielodisplastiche sono un gruppo eterogeneo di neoplasie ematologiche tipiche del paziente anziano, caratterizzate da emopoiesi inefficace, citopenie nel sangue periferico, alterazioni citogenetiche e/o molecolari e da una tendenza all’evoluzione in leucemia acuta mieloide (LAM) (Arber et al., 2016). Si tratta di patologie acquisite della cellula staminale ematopoietica (CSE), che si presenta alterata nelle sue due funzioni fondamentali di proliferazione e di differenziamento. Da qui, la ridotta produzione di elementi maturi, responsabile della citopenia periferica, e la comparsa di cellule indifferenziate atipiche denominate blasti, che possono aumentare progressivamente di numero nel midollo osseo ematopoietico (Avvisati, 2019).

Le SMD sono patologie relativamente rare, ma in costante aumento negli ultimi anni, e colpiscono soggetti di età avanzata: l’età media si attesta intorno ai 65-70 anni, mentre la percentuale di soggetti di età inferiore ai 50 anni alla diagnosi è di circa il 10%.

Le sindromi mielodisplastiche sono state tradizionalmente suddivise in SMD *de novo*, qualora insorgano nel paziente senza che vi sia una precedente esposizione dello stesso a un trattamento antitumorale, e SMD *therapy-related*, se si sviluppano in seguito a trattamento chemio/radioterapico adottato per eradicare un tumore primario, una malattia autoimmune oppure ad esposizione ad agenti tossici (Vardiman et al., 2009). In merito a questa seconda categoria, stanno assumendo sempre maggior rilievo fattori ambientali quali il benzene e i suoi derivati, i pesticidi, i solventi organici e i derivati ammoniacali, e il tempo di esposizione al fattore ambientale sembra essere il parametro più importante nel predisporre all’insorgenza della SMD.

La diagnosi si basa su criteri essenzialmente morfologici a cui fanno da supporto dei criteri molecolari. Infatti, conditio sine qua non per la diagnosi è che i precursori emopoietici midollari appartenenti a una o più linee siano caratterizzati da anomalie maturative che si manifestano in alterazioni morfologiche, definite displastiche, che interessano tutte le serie midollari (diseritropoiesi, disgranulopoiesi, dismegacariocitopoiesi) e che rappresentano l'espressione visibile dell'alterata proliferazione e differenziamento del clone mielodisplastico. Per rilevare le alterazioni morfologiche sono necessari l'esame citomorfologico dello striscio di sangue venoso periferico e dello striscio di sangue prelevato con l'aspirato midollare e successiva osservazione al microscopio con colorazione *May Grünwald Giemsa* in almeno il 10% delle cellule di ciascuna linea emopoietica interessata (Giuliani and Olivieri, 2020). L'esame dello striscio di sangue prelevato con l'aspirato midollare mostra una cellularità aumentata, ad indicare una proliferazione abbondante del clone mielodisplastico con difetti di maturazione tali da portare ad apoptosi di gran parte delle cellule (emopoiesi inefficace) (Avvisati, 2019).

## **1.2. Patogenesi**

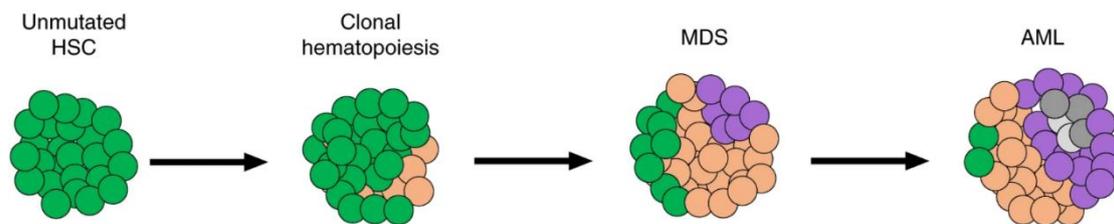
Mutazioni preesistenti che portano all'espansione clonale, e quindi a una condizione di pre-malignità, si trovano molto spesso a precedere lo sviluppo della maggior parte delle neoplasie. La crescita e l'espansione di un clone mutato a livello somatico rappresentano un modello di sviluppo delle SMD, infatti, l'evoluzione biologica della malattia avviene secondo un processo cosiddetto *multistep* (Avvisati, 2019) (**Figura 1**). Il vantaggio selettivo di un clone deriva dall'acquisizione di alterazioni genetiche definite mutazioni *driver*, ovvero quelle mutazioni ricorrenti di uno stesso gene che vengono selezionate positivamente e conferiscono un vantaggio di crescita che è biologicamente importante per determinare lo sviluppo del tumore (Weinberg, 2016). La mutazione iniziale interessa la cellula staminale ematopoietica che è in grado di auto-rinnovarsi (*self-renewal*), mentre ulteriori mutazioni associate con la progressione clonale potrebbero verificarsi nelle cellule progenitrici ematopoietiche (CPE), conferendo loro la medesima capacità (Clarke, 2019).

Il primo *step* nel modello di propagazione delle sindromi mielodisplastiche è rappresentato dal verificarsi di una mutazione *driver* in una cellula staminale ematopoietica, che genera un clone di cellule staminali mutate e cellule progenitrici e precursori anormali.

Il secondo *step*, noto come emopoiesi clonale di potenziale indeterminato (CHIP, *Clonal Hematopoiesis of Indeterminate Potential*), è caratterizzato dalla migrazione delle cellule staminali mutate e dal propagarsi dell'emopoiesi clonale. Infatti, le cellule staminali mutate migrano gradualmente attraverso il sangue periferico e si stabiliscono in altri distretti midollari, formando cloni locali.

Quando le cellule ematopoietiche portatrici della mutazione arrivano a costituire circa il 4% di tutte le cellule presenti nel midollo osseo, allora si parla di CHIP (VAF, *Variant Allele Frequency*,  $\geq 2\%$ ). La maggior parte dei pazienti con CHIP presenta mutazioni somatiche nei geni regolatori epigenetici (*DNMT3A*, *TET2*, *ASXL1*) e solo una piccola parte possiede mutazioni nei geni dello *splicing* (*SF3B1*, *SRSF2*, *U2AF1*) (Buscarlet et al., 2017; Jaiswal et al., 2014). Ciò è coerente con l'ipotesi per cui soggetti che hanno mutazioni a livello dello spliceosoma hanno una maggiore tendenza a sviluppare rapidamente una sindrome mielodisplastica o una citopenia clonale di significato incerto (CCUS, *Clonal Cytopenia of Undetermined Significance*), mentre soggetti con mutazioni in geni coinvolti in meccanismi epigenetici possono rimanere nella fase di CHIP per anni (Malcovati et al., 2017).

Il terzo *step* prevede che l'emopoiesi clonale si espanda progressivamente e diventi dominante nel midollo osseo ed è comunemente associata all'insorgenza di mutazioni somatiche addizionali. Infine, l'ultimo *step* è rappresentato dall'evoluzione in una fase leucemica, un passaggio non obbligatorio ma molto frequente, alla cui base verosimilmente vi sono l'acquisizione di ulteriori mutazioni *driver* o la comparsa di mutazioni preesistenti che portano alla selezione di subcloni di cellule ematopoietiche con una compromessa capacità differenziativa. Quando la proporzione di blasti cellulari raggiunge il 20%, allora si può porre diagnosi di LAM (Cazzola, 2020).



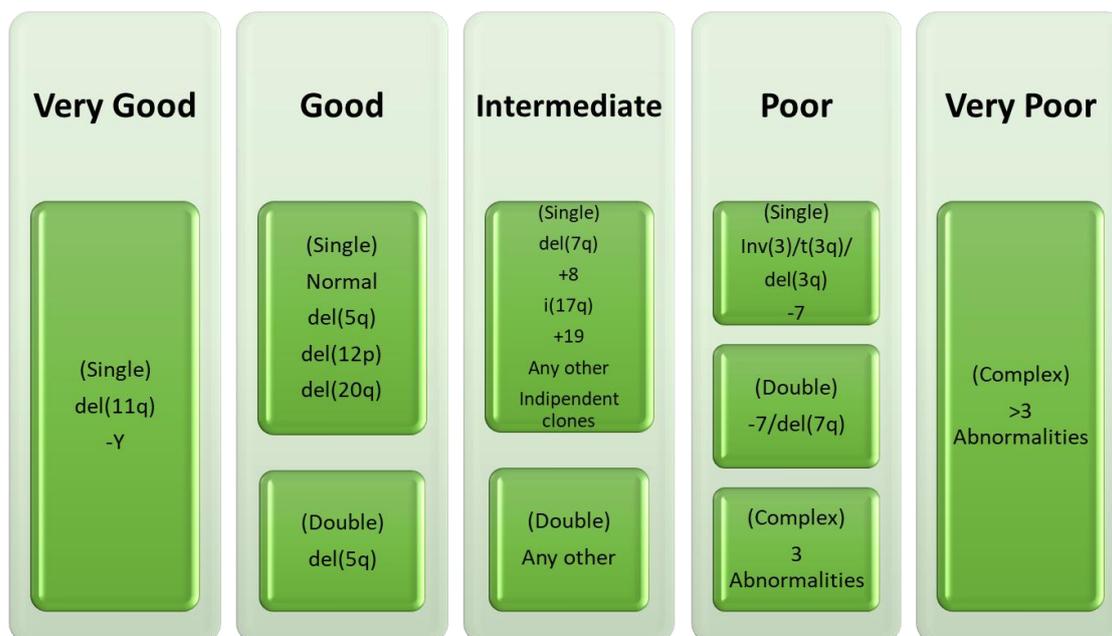
**Figura 1.** Modello di crescita e propagazione delle sindromi mielodisplastiche (Chen et al., 2019).

In un modello di evoluzione clonale, dunque, che culmina con una neoplasia ematologia franca, la transizione a SMD coinvolge presumibilmente un complesso gioco tra alterazioni epigenetiche nelle cellule staminali ematopoietiche, un microambiente midollare alterato e una graduale acquisizione di mutazioni *driver* addizionali (Arber et al., 2016). L'emopoiesi inefficace e le citopenie che definiscono le sindromi mielodisplastiche potrebbero sembrare un paradosso dal momento che le modificazioni genetiche conferiscono alle CSE un vantaggio clonale e l'espansione delle cellule progenitrici porta a un midollo osseo ipercellulare. Tuttavia, la proliferazione, e quindi il vantaggio di crescita, si verifica a livello delle cellule staminali ematopoietiche, mentre lo "svantaggio" si ha a livello dei precursori ematopoietici, determinando un blocco nel differenziamento terminale che porta a una aumentata apoptosi e a citopenie periferiche (Cazzola et al., 2013).

### 1.3. Aspetti citogenetici

Lo studio del cariotipo completo è obbligatorio nel percorso diagnostico delle sindromi mielodisplastiche e, oltre a confermare la diagnosi, fornisce importanti dati prognostici. Le anomalie cromosomiche sono presenti nel 50% dei casi di SMD *de novo* e nell'80% dei casi di SMD *therapy-related*. Le anomalie includono delezioni di un braccio o di un intero cromosoma, inversioni e guadagni di porzioni di cromosoma. Anomalie citogenetiche bilanciate, tra cui traslocazioni reciproche, inversioni e inserzioni, sono prevalenti nelle leucemie mieloidi, come la LAM o la leucemia mieloide cronica (LMC), ma sono poco comuni nelle SMD, in cui le anomalie cromosomiche sbilanciate che riflettono guadagni o perdite di materiale cromosomico sono maggiormente prevalenti (Haase et al., 2007). La presenza di più di tre anomalie cromosomiche definisce il

cariotipo complesso. Il diverso peso prognostico delle anomalie cromosomiche ricorrenti nelle SMD ha permesso di individuare 5 classi di rischio: *very good*, *good*, *intermediate*, *poor* e *very poor* (Schanz et al., 2012) (Figura 2).



**Figura 2.** Classificazione delle alterazioni citogenetiche ricorrenti nelle SMD e correlate alla prognosi (adattato da Schanz et al., 2012).

L'anomalia cromosomica più comune rilevata nelle SMD è la delezione del braccio lungo del cromosoma 5 (del(5q)), che si osserva nel 9%-14% dei casi. La delezione isolata di questo cromosoma è associata a una prognosi favorevole. Al contrario, la presenza di questa aberrazione in combinazione con altre è legata a una malattia più aggressiva e a un esito peggiore. La regione più comunemente deleta del cromosoma 5 contiene almeno 40 geni, tutti comunemente espressi nelle cellule staminali ematopoietiche. Vari studi mostrano la sotto-espressione di geni come la proteina ribosomiale 14 (RPS14) e la caseina chinasi 1 $\alpha$  (CSNK1A1) (Boulwood et al., 2007), e la perdita di diversi microRNA. In particolare, *CSNK1A1* codifica per una serina/treonina chinasi e svolge un ruolo di rilievo nell'emopoiesi. Inoltre, la delezione in omozigosi di questo gene determina l'arresto del ciclo cellulare tramite l'attivazione di *TP53* (Schneider et al., 2014). La perdita di vari microRNA, invece, è associata all'attivazione sia del pathway PI3K/AKT (*Phosphatidylinositol 3-kinase/ serine/threonine-specific protein kinase*) sia di

NF- $\kappa$ B (*Nuclear factor kappa B*), nonché all'attivazione inappropriata di segnali immunitari innati che risultano in trombocitosi, displasia megacariocitaria e neutropenia lieve e, probabilmente, nella patogenesi della sindrome della delezione del cromosoma 5 (5q-).

La seconda anomalia cromosomica più frequente è rappresentata dalle aberrazioni che coinvolgono il cromosoma 7, che includono la perdita parziale del cromosoma (del(7q)) o la monosomia (-7). Le due delezioni si riscontrano rispettivamente nell'8% e nel 4% dei casi di SMD *de novo* e sono indipendentemente associate a una prognosi sfavorevole (Haase et al., 2007; Schanz et al., 2012). Tra i geni che mappano sul cromosoma 7 troviamo *EZH2* (*Enhancer of zeste 2 polycomb repressive complex 2 subunit*), che codifica per un istone metiltransferasi implicata nella trascrizione genica. Di conseguenza, la perdita di questo gene è correlata a una prognosi sfavorevole e a una ridotta sopravvivenza, anche se la delezione del cromosoma 7 (del(7q)) non è predittiva della perdita dell'espressione del gene *EZH2* (Shallis et al., 2018).

La trisomia del cromosoma 8 è stata identificata nel 5% circa dei pazienti affetti da SMD ed è associata a un rischio intermedio di sviluppare la malattia. Tale alterazione cariotipica è responsabile di meccanismi di evasione dell'apoptosi, grazie alla overespressione di geni antiapoptotici che determinano un vantaggio selettivo rispetto alle normali cellule staminali ematopoietiche. Altre alterazioni cromosomiche nelle SMD includono la trisomia del cromosoma 13 o del 21, la delezione del braccio lungo del cromosoma 20 (del(20q)), la monosomia del cromosoma 21 e la perdita del cromosoma X (Shallis et al., 2018).

#### **1.4. Aspetti genetici**

Con l'avanzamento delle tecnologie *high-throughput* per il sequenziamento del DNA, con particolare riferimento al sequenziamento di nuova generazione (NGS, *Next Generation Sequencing*), *Whole-genome sequencing* (WGS) e *Whole-exome sequencing* (WES), è stato possibile comprendere la patogenesi alla base delle SMD. Nel dettaglio, sono state identificate mutazioni somatiche che consentono di classificare tutte quelle alterazioni che sfuggono all'inquadramento del paziente in un cariotipo normale e che

sono correlate a risvolti clinici importanti (Bejar et al., 2011). Esse interessano i geni implicati nello *splicing* dell'RNA (Damm et al., 2012; Inoue et al., 2016), i geni regolatori epigenetici, i quali modificano le caratteristiche chimiche della cromatina (Voso et al., 2013a), fattori di trascrizione e proteine che regolano la trasduzione del segnale. Mutazioni somatiche nella maggior parte di questi geni (**Tabella 1**) sono state identificate in circa il 78%-89% dei pazienti affetti da SMD (con una media di tre mutazioni somatiche per paziente) (Shallis et al., 2018).

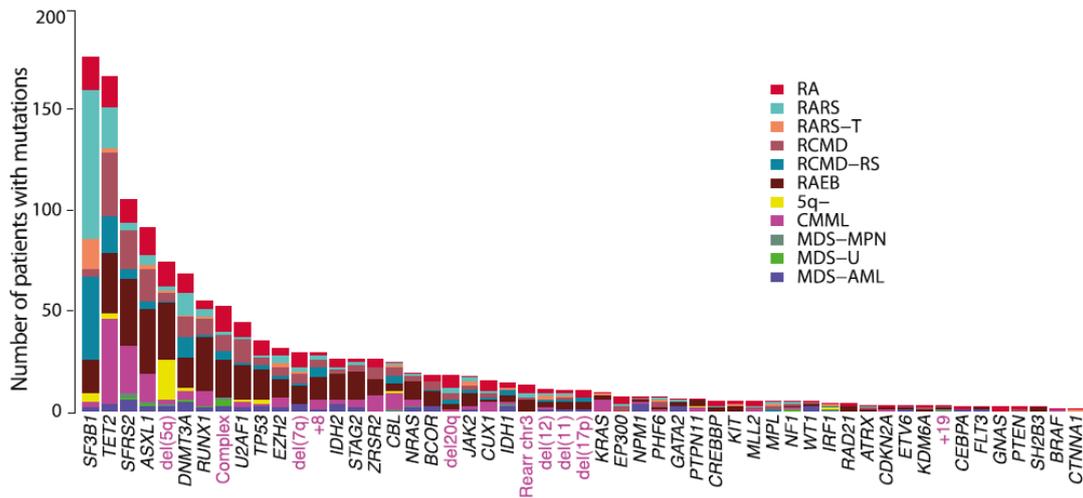
Function/mutated gene	Incidence	Prognostic impact	Phenotype and other associations
<b>RNA splicing</b>			
SF3B1	20%-28%	↑OS, ↑LFS, ↑EFS	Predicts MDS-RS and MDS-MLD, strong association with presence of ring sideroblasts, associated with DNMT3A mutations
SRSF2	12%	↓OS, ↑LT	Predicts CMML phenotype; associated with RUNX1, TET2, and IDH1 mutations; associated with RUNX1 overexpression
U2AF1	7%-8%	↓OS	Associated with ASXL1 and IDH2 mutations
ZRSR2	3%	Unknown	Associated with TET2 mutations
<b>DNA methylation</b>			
TET2	22%	Unknown	Found in 60% of CMML, predicts response to DNMTi
DNMT3A	8%	↓OS, ↑LT	Associated with IDH2 mutations
IDH1/IDH2	5%	Unknown	Associated with DNMT3A, ASXL1, and SRSF2 mutations
<b>Histone modification</b>			
ASXL1	11%-14%	↓OS	Associated with U2AF1 mutation
EZH2	6%	↓OS	
<b>DNA transcription/signal transduction</b>			
TP53	8%-13%	↓OS	Associated with higher marrow myeloblast percentage, higher-risk disease and complex karyotype; rarely associated with spliceosome mutations
RUNX1	5%-9%	↓OS	Associated with higher marrow myeloblast percentage, strongly associated with SRSF2 mutations
NRAS	4%-6%	↓OS, ↑LT	Associated with higher marrow myeloblast percentage
SETBP1	3%-9%	↓OS, ↑LT	Associated with del(7q) and ASXL1 mutations
JAK2	3%-5%	Unknown	Detected in 50% of cases of MDS/MPN-RS-T
ETV6	3%-4%	↓OS	
<b>Chromatid cohesion</b>			
STAG2	7%	↓OS	Rarely identified as a founding clone

**Tabella 1.** Mutazioni somatiche più ricorrenti identificate nelle sindromi mielodisplastiche.

EFS: Event-free survival; LFS: Leukemia-free survival; LT: Leukemic transformation; OS: Overall survival (Shallis et al., 2018).

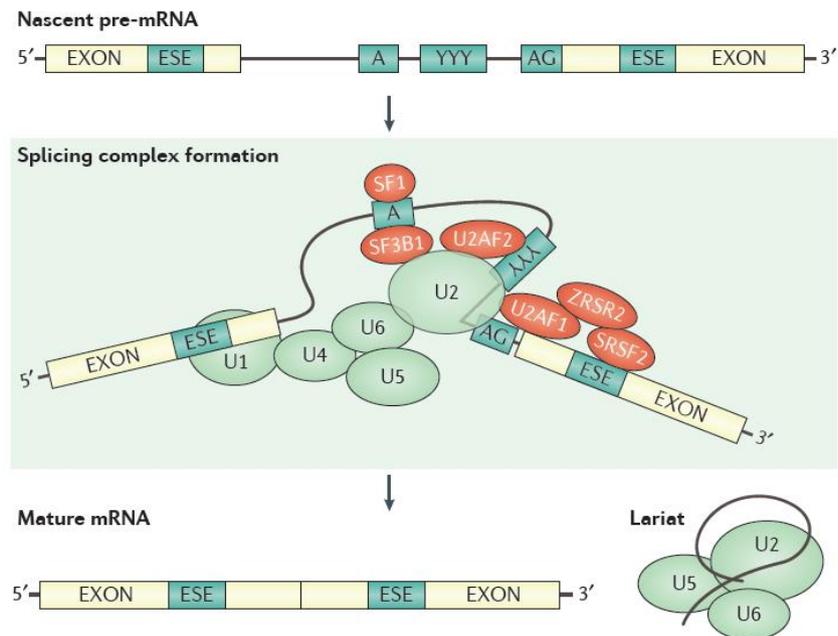
Papaemmanuil *et al.*, sfruttando la tecnologia del sequenziamento massivo di una coorte di 738 pazienti con SMD e un pannello di 111 geni implicati nella patogenesi delle

neoplasie mieloidi, hanno riscontrato la presenza di mutazioni oncogeniche in 43 geni (Papaemmanuil et al., 2013) (**Figura 3**). In particolare, solo 4 geni sono risultati mutati in più del 10% dei pazienti, con ulteriori 3 geni portatori di mutazioni *driver* nel 5-10% dei pazienti, mentre 36 geni sono stati trovati mutati in meno del 5% dei pazienti, contribuendo a un totale del 33,5% di tutte le mutazioni genetiche identificate.



**Figura 3.** Frequenza delle mutazioni *driver* e delle alterazioni citogenetiche in una coorte di 738 pazienti affetti da SMD. Sono rappresentati in colori diversi i vari sottogruppi di SMD (Papaemmanuil et al., 2013).

Haferlach *et al.* hanno messo in luce una correlazione significativa tra le alterazioni molecolari e la prognosi, identificando almeno 25 mutazioni che incidono in maniera significativa sulla prognosi di pazienti affetti da SMD, suggerendo in tal senso una revisione dell'IPSS in accordo con i *pattern* molecolari alterati (Haferlach et al., 2014). Le mutazioni più comunemente osservate nelle SMD riguardano geni che codificano per le proteine che fanno parte dello spliceosoma che, se modificate, portano ad un anomalo processo di *splicing* dell'RNA e quindi ad una anormale sintesi proteica. Tali mutazioni sono distribuite lungo il complesso maggiore dello spliceosoma (U2 snRNP, *U2 small nuclear ribonucleoprotein*) dove giocano un ruolo importante nel riconoscimento del sito di *splicing* al 3' (*3'SS*, *3' Splicing Site*) (**Figura 4**) e sono situate nei geni *SF3B1* (*Splicing factor 3B subunit 1*), *SRSF2* (*Serine and arginine rich splicing factor 2*), *U2AF1* (*U2 small nuclear RNA auxiliary factor 1*) e *ZRSR2* (*Zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2*) (**Figura 6**).



**Figura 4.** Rappresentazione del legame dei fattori di splicing al complesso di splicing (Sperling et al., 2016).

Lo *splicing* dell'RNA è un processo altamente regolato messo in atto dallo spliceosoma maggiore (U2-dipendente) e minore (U12-dipendente) al fine di rimuovere le regioni non codificanti (introni) dal precursore dell'RNA messaggero (pre-mRNA) e di promuovere la giunzione degli esoni prima della traduzione proteica.

Il processo ha inizio con il riconoscimento delle giunzioni esone-introne al 5' e al 3', a cui segue il reclutamento di tutte le componenti del macchinario di *splicing* che è costituito da 5 piccoli RNA nucleari (snRNA), U1, U2, U4, U5 e U6, e da circa 150 proteine che catalizzano il processo stesso. Le informazioni che definiscono le regioni di *splicing* sono incluse in piccole sequenze conservate localizzate al sito di *splicing* al 5' (5'SS), al 3' (3'SS) e al sito di ramificazione (BPS, *Branching Point Site*) (Wahl et al., 2009).

Lo *splicing* vede il coinvolgimento di fattori trans-agenti (SF, *Splicing Factor*) che si legano a elementi regolatori localizzati negli esoni e negli introni per promuovere o reprimere l'attività di *splicing* o contribuire allo *splicing* alternativo (Liu et al., 2007), ossia il meccanismo attraverso il quale uno stesso pre-mRNA può subire eventi di *splicing*

differenti che portano alla creazione di diversi mRNA alternativi, che a loro volta possono codificare differenti proteine.

Le mutazioni a carico dei geni che codificano per le proteine spliceosomali sono mutualmente esclusive, anche se recentemente sono emerse delle evidenze per cui non sempre questo si verifica. Infatti, Taylor *et al.* hanno dimostrato come nello 0,85% dei pazienti con neoplasie mieloidi vi sia la presenza concomitante di mutazioni nei fattori di *splicing*, sebbene queste interessino codoni meno comuni (Taylor et al., 2020).

Di particolare importanza è l'identificazione di mutazioni nel gene *SF3B1*, che codifica per una subunità chiave dello spliceosoma, presenti con una incidenza del 20%-28% nei pazienti affetti da SMD (Papaemmanuil et al., 2011) e strettamente associate alla presenza di sideroblasti ad anello, dove sono state riscontrate ad una frequenza di circa l'80-85% (Malcovati et al., 2015; Papaemmanuil et al., 2011). Questa scoperta ha portato all'inclusione di *SF3B1* come unica mutazione diagnostica nell'ultima classificazione WHO.

Le mutazioni di *SF3B1* in pazienti affetti da SMD con sideroblasti ad anello originano nel compartimento midollare delle cellule staminali ematopoietiche e solamente queste cellule possiedono la capacità, rispetto ad altre popolazioni di cellule progenitrici, di propagare il clone *SF3B1* mutato, con la risultante presenza a valle della mutazione nelle cellule progenitrici mieloidi ed eritroidi (Mortera-Blanco et al., 2017).

L'alterazione del macchinario di *splicing* che ne consegue porta ad una anormale sintesi proteica che, nel contesto del compartimento delle CSE, determina una emopoiesi inefficace. Tuttavia, resta non del tutto chiaro il meccanismo attraverso cui le cellule staminali ematopoietiche mutate per lo spliceosoma vengano selezionate per l'espansione clonale (Malcovati et al., 2011).

La proteina SF3B1 è un componente essenziale della ribonucleoproteina U2 (U2snRNP) ed è responsabile del riconoscimento del sito di ramificazione (BPS), ovvero un motivo di sequenza degenerato contenente una adenosina (A) e localizzato 21-34 paia di basi a monte di un tratto polipirimidinico che precede il sito accettore AG al 3'. SF3B1 media e stabilizza il legame della ribonucleoproteina U2 sul pre-mRNA.

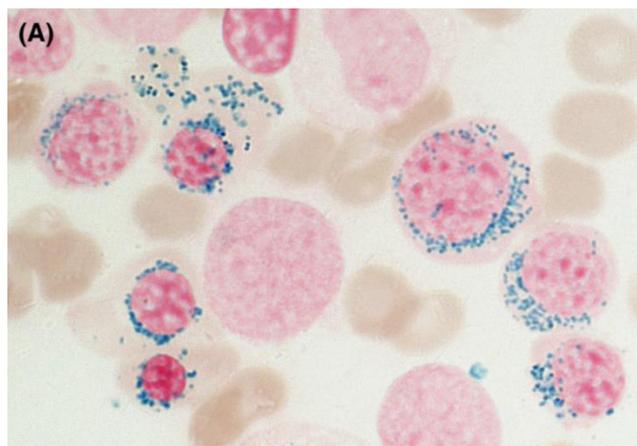
A livello strutturale, il gene *SF3B1* codifica per una proteina che possiede una regione idrofilica all'estremità N-terminale coinvolta nelle interazioni RNA-proteina e proteina-proteina, e una regione C-terminale che consta di 22 ripetizioni HEAT (*Huntingtin*, *Elongation factor 3*, *protein phosphatase 2A*, *Targets of rapamycin 1*) non identiche.

Le mutazioni oncogeniche in *SF3B1* sono mutazioni missenso con tre *hotspot* principali che interessano rispettivamente la quinta, la sesta e la settima ripetizione HEAT in posizione R625, K666 e K700, corrispondenti a una regione genica che va dal dodicesimo al quindicesimo esone (Papaemmanuil et al., 2011).

È interessante notare come le mutazioni K700 siano di gran lunga le più frequenti in tumori maligni ematopoietici e interessino residui che sono spazialmente vicini l'uno all'altro e quindi potrebbero avere un impatto funzionale simile (Papaemmanuil et al., 2013).

Inoltre, un'analisi di *RNA-sequencing* (RNA-Seq) ha mostrato lo stesso tipo di alterazione dello *splicing* in tutti quei tumori che presentano mutazioni nel gene *SF3B1*, alterazione che consiste nell'utilizzo di un sito di *splicing* criptico al 3' (in seguito chiamato AG'), situato dalle 10 alle 30 basi a monte del normale sito di *splicing* al 3' (3'SS), tuttavia il meccanismo non è stato ancora compreso.

Il gene *SF3B1* mutato, inoltre, è responsabile della *down*-regolazione di *pathway* che regolano il trasporto del ferro a livello mitocondriale. In particolare, si assiste a una riduzione dell'espressione del gene *ABCB7* (*ATP binding cassette subfamily B member 7*), un trasportatore dell'eme, mutazione che determina la distruzione dell'ATP binding cassette e la conseguente fuoriuscita degli ioni ferro dal mitocondrio e che è coinvolta nell'anemia sideroblastica. Questo determina una anomala deposizione di ioni ferro (a forma di anello) nei mitocondri perinucleari degli eritroblasti (Dolatshad et al., 2016). Le sindromi mielodisplastiche che presentano tale anomalia vengono definite sindromi mielodisplastiche con sideroblasti ad anello e possono essere identificate mediante colorazione di *Perls* (Malcovati and Cazzola, 2016) (**Figura 5**).



**Figura 5.** Sideroblasti ad anello della sede midollare di un paziente con sindrome mielodisplastica (Malcovati and Cazzola, 2016).

Il secondo gene appartenente al macchinario di *splicing* più frequentemente mutato nelle SMD è *SRSF2* che normalmente riconosce gli *enhancer* esonici (ESE, *Exonic Splicing Enhancer*) all'interno del pre-mRNA con un motivo consenso SSNG (dove S rappresenta Citosina o Guanina, N qualsiasi nucleotide e G rappresenta la Guanina) per promuovere il riconoscimento degli esoni (Fu and Ares, 2014). Nello stato *wild-type*, *SRSF2* si lega e promuove lo *splicing* degli ESE ricchi di C e G allo stesso modo (GGTG o CCAG), mentre *SRSF2* mutato ha una preferenza alterata per gli ESE ricchi in C (CCAG) (Kim et al., 2015; Kon et al., 2018). Più specificamente, le cellule che esprimono *SRSF2* mutato utilizzano preferenzialmente gli *enhancer* esonici con motivi CCNG.

Le mutazioni di *SRSF2* si verificano nel 20-30% delle SMD (Papaemmanuil et al., 2013). Si tratta per la maggior parte di mutazioni missenso in eterozigosi, che interessano soprattutto la prolina 95 (P95) e alterano il riconoscimento degli esoni contenenti specifiche sequenze ESE.

Queste mutazioni sono associate a una prognosi sfavorevole e a tassi più elevati di progressione leucemica. Tali mutazioni sono state frequentemente identificate in associazione con mutazioni nei geni *TET2* (*Tet methylcytosine dioxygenase 2*) e *ASXL1* (*ASXL transcriptional regulator 1*).

*U2AF1* è un membro della famiglia delle proteine ricche in serina e arginina (SR, *Serine/Arginine-rich proteins*) e rappresenta una subunità della snRNP U2 che è

responsabile del riconoscimento del dinucleotide AG nel sito di *splicing* al 3' (3'SS) (Inoue et al., 2016; Wang et al., 2008). Mutazioni in questo gene sono responsabili *in vitro* di un implementato processo di *splicing* e del "salto dell'esone" (ES, *Exon Skipping*) e interessano esclusivamente due codoni, S34 e Q157, che sono localizzati in due distinti domini definiti "a dita di zinco" (*Zinc finger domain*) (Graubert et al., 2012).

Le mutazioni di *U2AF1* alterano la preferenza per il sito di *splicing* al 3' in modo sequenza-specifico basato sulla tipologia di nucleotide che delimita il dinucleotide AG che costituisce il *core* della sequenza consenso 3'SS.

Nel dettaglio, nel caso della mutazione S34F/Y si avrà il riconoscimento della sequenza 3'SS che possiede una C o una A immediatamente a monte di AG e la repressione di quella che possiede una T in questa posizione (Shirai et al., 2015).

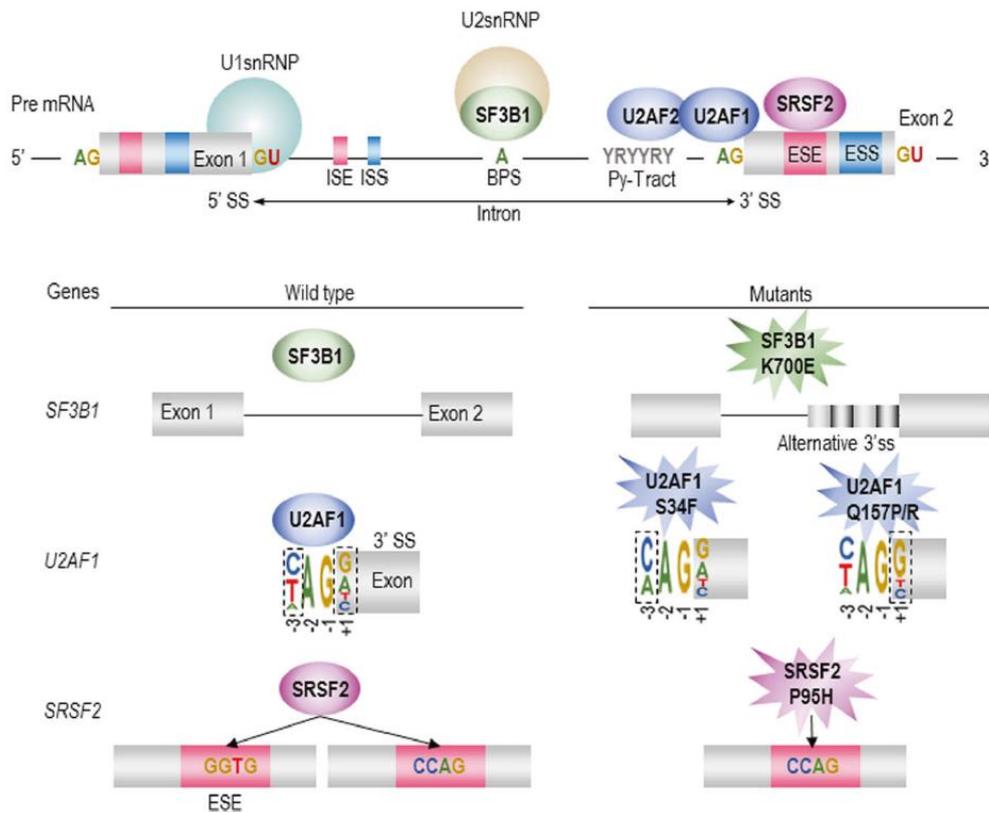
Al contrario, le mutazioni Q157P/R promuovono il riconoscimento della sequenza 3'SS che possiede una G immediatamente a valle del dinucleotide AG e reprimono quelle sequenze che portano una A in questa posizione (Ilagan et al., 2015; Inoue et al., 2016). Questi dati suggeriscono che le mutazioni S34F e Q157P/R hanno diversi effetti, anche se causano un'alterazione nei medesimi segnali di *splicing*. Le mutazioni nel gene *U2AF1* sono correlate a una ridotta sopravvivenza nei pazienti affetti da SMD e ad un aumentato rischio di progressione in leucemia acuta mieloide (Graubert et al., 2012).

*ZRSR2* è localizzato sul cromosoma X e il suo prodotto interagisce con altri componenti di assemblaggio del pre-spliceosoma, tra cui l'eterodimero *U2AF2/U2AF1* e *SRSF2*. È inoltre un componente della ribonucleoproteina U12 e prende contatto con la sequenza 3'SS degli introni di tipo U12, ossia una rara classe di introni che vengono rimossi dallo spliceosoma minore e comprendono solo lo 0,5% circa di tutti gli introni umani.

È interessante notare come un sottoinsieme di introni di questo tipo va incontro a *splicing* in modo inefficiente rispetto alla controparte di tipo U2. Questo *splicing* inefficiente può risultare in una ritenzione nucleare (IR, *Intron Retention*) di mRNA contenenti introni di tipo U12 non sottoposti a *splicing*, eventualmente seguita da decadimento nucleare, con conseguente meccanismo di regolazione analogo a quello osservato nell'eritropoiesi. Studi recenti dimostrano che gli eritroblasti elaborano un

programma di ritenzione degli introni ampio e diversificato che comprende numerosi geni eritroidi essenziali, compresi quelli che codificano per fattori di *splicing* e proteine coinvolte nell'omeostasi del ferro. Queste evidenze rafforzano il concetto per cui una regolazione fine del processamento dell'RNA gioca un ruolo chiave nel differenziamento eritroide terminale (Madan et al., 2015).

È importante notare come le mutazioni in questo gene non siano presenti solo in specifiche localizzazioni, come riportato per i geni precedenti, ma siano diffuse sull'intera sequenza esonica di *ZRSR2*, rendendo difficile l'interpretazione funzionale di ognuna di esse.



**Figura 6.** Splicing dell'RNA e ruolo delle proteine frequentemente mutate nelle SMD.

*ISE: Intronic splicing enhancer; ISS: Intronic splicing silencer; BPS: Branch point site; Py-Tract:*

*Polypyrimidine Tract; ESE: Exonic splicing enhancer; ESS: Exonic splicing silencer*

*(Nagata and Maciejewski, 2019).*

Oltre alle mutazioni che interessano lo *splicing*, sono state identificate mutazioni in geni che intervengono nella regolazione epigenetica, ossia nel controllo dell'espressione

genica tramite processi che non comportano cambiamenti di sequenza del DNA. La regolazione epigenetica assolve a un ruolo cruciale nella normale emopoiesi, ovvero nel *commitment* delle cellule progenitrici verso una specifica linea ematopoietica e i principali meccanismi coinvolti includono la metilazione del DNA e la modificazione degli istoni.

La metilazione del DNA avviene ad opera dei membri della famiglia enzimatica delle metiltransferasi del DNA, che catalizzano l'aggiunta di un gruppo metile in posizione 5' dell'anello citosinico nell'ambito dei dinucleotidi CpG (Holliday and Grigg, 1993) (**Figura 7**). Tali dinucleotidi sono fortemente sottorappresentati nel genoma umano rispetto alla frequenza attesa a causa della deaminazione spontanea cui va incontro la base instabile 5-metilcitosina che si trasforma in timina e viene quindi rimossa in modo efficiente dal DNA (ad opera dei sistemi di riparazione) dando luogo ad eventi mutazionali.

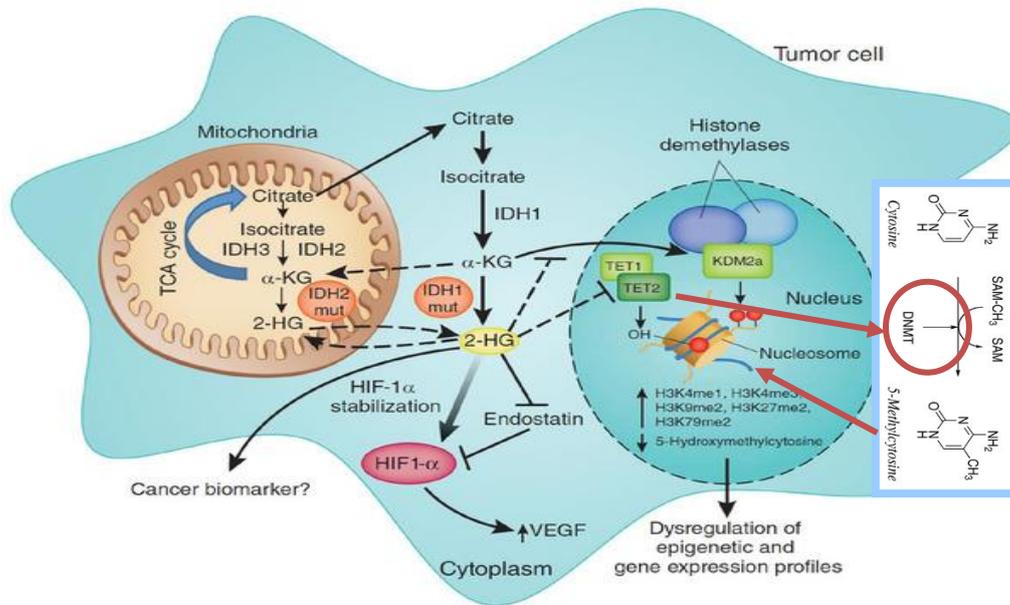
La densità dei dinucleotidi CpG è tuttavia particolarmente alta in specifiche regioni del genoma denominate isole CpG, che sono localizzate in corrispondenza dei promotori e delle regioni codificanti al 5' di circa la metà dei geni noti ed a livello di sequenze ripetitive del DNA.

Il risultato finale della metilazione del DNA è l'assemblaggio della cromatina in una struttura compatta, ostacolando così il legame dei fattori di trascrizione fino al silenziamento genico. La mutazione missenso R882 nel gene *DNMT3A* comporta una diminuzione dell'attività catalitica e l'ipometilazione del DNA. È interessante notare come la mutazione *hotspot* R882 sia risultata essere la più comune, ma anche capace di influenzare la risposta alla terapia (Lin et al., 2018).

Il gene *TET2* codifica per proteine coinvolte nel controllo epigenetico attraverso il processo di demetilazione. Di conseguenza, le mutazioni in questo gene provocano un aumento della metilazione e il silenziamento di geni che sarebbero normalmente espressi. Le mutazioni di *TET2* sono rilevate nel 15% dei casi di SMD, sono associate a una prognosi favorevole e correlano positivamente con la risposta agli agenti ipometilanti (Cedena et al., 2017).

Le mutazioni che colpiscono le deidrogenasi *IDH1* e *IDH2* provocano la perdita simultanea della loro normale attività catalitica, la produzione di  $\alpha$ -chetoglutarato ( $\alpha$ -

KG) e l'acquisizione di una nuova funzione, cioè la produzione di 2-idrossiglutarato (2-HG). Il 2-HG agisce come un antagonista dell' $\alpha$ -KG per inibire competitivamente più diossigenasi  $\alpha$ -KG-dipendenti, compresa la famiglia TET delle idrossilasi del DNA. (Yang et al., 2012).



**Figura 7.** Enzimi coinvolti nella metilazione del DNA (adattato da Prensner and Chinnaiyan, 2011).

*TP53 (Tumor protein p53)* è il gene *tumor suppressor* più frequentemente mutato nei tumori umani e va incontro a mutazioni ricorrenti nelle sindromi mielodisplastiche, mutazioni che interessano fino al 13% dei pazienti, percentuale che aumenta se si prendono in considerazione solo i pazienti affetti da SMD ad alto rischio e quelli con un cariotipo complesso. Infatti, diversi studi hanno dimostrato l'importanza prognostica indipendente di questa mutazione, che è fortemente associata a una minore sopravvivenza generale (OS, *Overall Survival*) e sopravvivenza libera da progressione (PFS, *Progression Free Survival*), nonché a un più alto tasso di ricaduta. P53 media la risposta allo stress cellulare aumentando l'espressione di geni coinvolti nell'arresto del ciclo cellulare e nell'apoptosi (Kaneko et al., 1995; Lowe et al., 2004). Le mutazioni in questo *pathway* portano a un ingresso inappropriato nel ciclo cellulare prima che il meccanismo di riparazione del danno al DNA sia portato a termine e ad alterazioni nel processo stesso di riparazione del DNA, con conseguente accumulo di alterazioni genetiche e instabilità cromosomica (Sperling et al., 2017).

### 1.5. Classificazione WHO

Nel 1976, il Dott. John M. Bennett e i colleghi del gruppo Franco-Americano-Britannico (FAB) hanno pubblicato un documento che non solo ha fornito la prima classificazione della leucemia acuta mieloide, ma ha anche sensibilizzato la comunità ematologica circa le caratteristiche cliniche e morfologiche delle sindromi mielodisplastiche (Vardiman, 2012). Il gruppo FAB ha introdotto il termine "sindrome mielodisplastica" in una pubblicazione del 1982, che comprendeva linee guida dettagliate per la diagnosi e una classificazione basata sulla morfologia e sulla percentuale di blasti nel sangue e nel midollo osseo.

La classificazione FAB identificava cinque diversi sottogruppi: leucemia mielomonocitica cronica (LMMC), anemia refrattaria (AR), anemia refrattaria con eccesso di blasti (AREB), anemia refrattaria con sideroblasti ad anello (ARSA) e anemia refrattaria con eccesso di blasti in trasformazione (AREB-t). Secondo i criteri FAB, era necessaria la presenza di almeno il 5% di blasti in una sola linea displastica, tuttavia i medici dovettero affrontare la difficoltà di inquadramento di pazienti con displasia multilineare che non soddisfacevano questo criterio.

Attualmente, viene utilizzata la classificazione dell'Organizzazione Mondiale della Sanità (WHO, *World Health Organization*), che si basa sulle caratteristiche cliniche, sulla percentuale di blasti nel sangue periferico (PB, *Peripheral Blood*) e nel midollo osseo (BM, *Bone Marrow*) e sull'analisi citogenetica.

Essa prevede l'abbassamento della soglia dei blasti per la diagnosi di LAM dal 30% al 20%, eliminando così la categoria AREB-t.

Un'altra novità è stata la definizione di due sottocategorie AREB (I e II) in base alla percentuale di blasti nel midollo osseo (5-9% e > 10-19%) e alla presenza o assenza di Corpi di Auer.

Le nuove informazioni raccolte dal 2008 in poi e i progressi ottenuti nel campo della biologia molecolare e cellulare hanno portato a una nuova classificazione (**Tabella 2**), pubblicata nel 2016, che introduce dati di citogenetica (SMD con del5q) e molecolari (SMD con mutazione *SF3B1*). Inoltre, anche se la soglia per definire la displasia rimane il 10% di cellule displastiche in qualsiasi linea ematopoietica, si stabilisce che la displasia

oltre il 10% possa essere trovata in individui normali e in casi di citopenia di natura non neoplastica.

Per quanto riguarda le SMD con sideroblasti ad anello, queste sono frequentemente associate a mutazioni nel gene dello spliceosoma *SF3B1*. Studi recenti hanno dimostrato che in questo specifico sottotipo, la percentuale effettiva di sideroblasti ad anello identificati non è così rilevante come si pensava in passato. Questo ha portato ad una rivalutazione dei criteri diagnostici con un sostanziale abbassamento della percentuale di sideroblasti ad anello identificati (5%) e ad attribuire maggiore importanza alla presenza della mutazione del gene. Tuttavia, in mancanza della suddetta mutazione, è ancora richiesto almeno il 15% di sideroblasti ad anello.

Name	Dysplastic lineages	Cytopenias*	Ring sideroblasts as % of marrow erythroid elements	BM and PB blasts	Cytogenetics by conventional karyotype analysis
MDS with single lineage dysplasia (MDS-SLD)	1	1 or 2	<15% <sup>‡</sup> / <sup>‡</sup> <5%	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with multilineage dysplasia (MDS-MLD)	2 or 3	1-3	<15% <sup>‡</sup> / <sup>‡</sup> <5%	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
<b>MDS with ring sideroblasts (MDS-RS)</b>					
MDS-RS with single lineage dysplasia (MDS-RS-SLD)	1	1 or 2	≥15% <sup>‡</sup> / <sup>‡</sup> ≥5%	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS-RS with multilineage dysplasia (MDS-RS-MLD)	2 or 3	1-3	≥15% <sup>‡</sup> / <sup>‡</sup> ≥5%	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with isolated del(5q)	1-3	1-2	None or any	BM <5%, PB <1%, no Auer rods	del(5q) alone or with 1 additional abnormality except -7 or del(7q)
<b>MDS with excess blasts (MDS-EB)</b>					
MDS-EB-1	0-3	1-3	None or any	BM 5%-9% or PB 2%-4%, no Auer rods	Any
MDS-EB-2	0-3	1-3	None or any	BM 10%-19% or PB 5%-19% or Auer rods	Any
<b>MDS, unclassifiable (MDS-U)</b>					
with 1% blood blasts	1-3	1-3	None or any	BM <5%, PB = 1%, <sup>‡</sup> no Auer rods	Any
with single lineage dysplasia and pancytopenia	1	3	None or any	BM <5%, PB <1%, no Auer rods	Any
based on defining cytogenetic abnormality	0	1-3	<15% <sup>§</sup>	BM <5%, PB <1%, no Auer rods	MDS-defining abnormality
Refractory cytopenia of childhood	1-3	1-3	None	BM <5%, PB <2%	Any

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

<sup>‡</sup>If *SF3B1* mutation is present.

<sup>‡</sup>One percent PB blasts must be recorded on at least 2 separate occasions.

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

**Tabella 2. Classificazione WHO 2016 delle Sindromi mielodisplastiche (Arber et al., 2016).**

## 1.6. Classificazione IPSS e IPSS-R

L'identificazione di fattori prognostici e la creazione di un accurato schema di punteggio hanno influenzato le tipologie di trattamento, consentendo una stratificazione dei pazienti sulla base del rischio (Greenberg et al., 2012; Voso et al., 2013b). *L'International*

*Prognostic Scoring System (IPSS, range 0-3,5)* nasce nel 1997 come frutto di un *workshop* internazionale, che identifica nella percentuale dei blasti presenti nel midollo osseo, numero e grado di citopenie e alterazioni del cariotipo, i più potenti fattori prognostici nelle SMD. Per calcolare il rischio è necessario sommare il punteggio di queste tre variabili, dalla cui combinazione sono emersi 4 gruppi di rischio applicabili per stimare sia la sopravvivenza in generale sia la progressione a leucemia mieloide acuta: basso, intermedio-1, intermedio-2 e alto rischio (Schanz et al., 2012) (**Tabella 3**).

Risk Group	IPSS score	Median Overall Survival (years)	Median time to AML evolution
Low	0	5.7	9.4
Intermediate-1	0.5-1	3.5	3.3
Intermediate-2	1.5-2	1.2	1.1
High	2.5	0.4	0.2

**Tabella 3.** Gruppi di rischio dell'International Prognostic Scoring System (IPSS) nelle sindromi mielodisplastiche (adattato da Schanz et al., 2012).

Nel 2012, l'IPSS ha subito delle modifiche che hanno interessato le tre variabili precedentemente elencate, determinando un ampliamento a 5 categorie di rischio e portando alla nascita del *Revised International Prognostic Scoring System (IPSS-R)*. Tra le modifiche apportate troviamo una suddivisione del gruppo di blasti midollari < 5% in due sottogruppi: 0 - 2% e > 2 - < 5%; una maggiore diversificazione delle alterazioni citogenetiche (16 specifiche alterazioni cariotipiche ripartite in 5 sottogruppi) e stratificazione delle citopenie sulla base dei valori di emoglobina, piastrine e neutrofili (Greenberg et al., 2012) (**Tabella 4 e 5**).

Table 3. IPSS-R prognostic score values

Prognostic variable	0	0.5	1	1.5	2	3	4
Cytogenetics	Very good	—	Good	—	Intermediate	Poor	Very poor
BM blast, %	≤ 2	—	> 2% - < 5%	—	5% - 10%	> 10%	—
Hemoglobin	≥ 10	—	8 - < 10	< 8	—	—	—
Platelets	≥ 100	50 - < 100	< 50	—	—	—	—
ANC	≥ 0.8	< 0.8	—	—	—	—	—

— indicates not applicable.

**Tabella 4.** Schema di punteggio IPSS-R nelle sindromi mielodisplastiche. BM: Bone marrow; ANC: Absolute neutrophil count. (Greenberg et al., 2012).

Risk Group	IPSS-R Score	Median Overall Survival (years)	Median time to 25 percent AML evolution (years)
Very low	≤ 1.5	8.8	>14.5
Low	> 1.5 to 3.0	5.3	10.8
Intermediate	> 3 to 4.5	3.0	3.2
High	> 4.5 to 6	1.6	1.4
Very high	> 6	0.8	0.7

*Tabella 5. Gruppi di rischio del Revised International Prognostic Scoring System (IPSS-R) nelle sindromi mielodisplastiche (adattato da Greenberg et al., 2012).*

### 1.7. Trattamento

Le sindromi mielodisplastiche sono delle patologie croniche clonali, non eradicabili farmacologicamente. Solamente il trapianto allogenico di cellule staminali ematopoietiche rappresenta una cura valida per i pazienti affetti da SMD con insufficienza midollare. Tuttavia, le SMD rappresentano una condizione tipica dell'adulto anziano e questo fa sì che la maggior parte dei pazienti non possa essere arruolata per questa tipologia di trattamento particolarmente invasiva (Tura et al., 2018).

Anche se la maggior parte dei pazienti avrà un decorso della malattia poco aggressivo, con solo bisogno di cure di supporto, in circa un terzo di essi la malattia progredirà in sindrome mielodisplastica ad alto rischio e in leucemia acuta mieloide secondaria. I pazienti con SMD ad alto rischio hanno una aspettativa media di vita minore di due anni. Per questi pazienti, il trattamento è finalizzato non solo a migliorare le citopenie, ma anche a prevenire l'evoluzione a LAM e quindi a prolungare la sopravvivenza (Malcovati et al., 2013). In questi casi, al paziente vengono somministrati dei farmaci ipometilanti, quali la azacitidina e la decitabina, che hanno la capacità di indurre l'ipometilazione del DNA inibendo l'enzima DNA metiltransferasi e riducendo i gruppi metilici che si legano ai siti promotori di molti geni, impedendone l'espressione. Mediante tale azione, essi sono in grado di ripristinare la normale emopoiesi (Fenaux et al., 2009).

Per diverso tempo, le cure per i pazienti a basso rischio e a rischio intermedio hanno riguardato trasfusioni di globuli rossi e piastrine e agenti stimolanti l'eritropoietina.

Questa tipologia di trattamenti è riservata ai pazienti sintomatici, mentre non è prevista alcun tipo di terapia nei soggetti asintomatici, senza una anemia importante, con meno del 5% di blasti midollari e con un cariotipo a basso rischio.

In questi casi, sono raccomandate delle visite di controllo periodiche al fine di monitorare il decorso e la progressione della malattia (approccio *watch and wait*) e, qualora il trattamento si renda necessario, esso è volto a migliorare la citopenia, l'anemia e più in generale la qualità di vita.

Al contrario, nei pazienti con anemia sintomatica è necessario iniziare un trattamento con eritropoietina (EPO) che, nel caso in cui si rivelasse inefficace, lascerebbe il posto alla trasfusione (Malcovati et al., 2013).

Trasfusioni frequenti sono la causa di quello che viene definito "sovraccarico marziale" o sovraccarico di ferro secondario, che si verifica quando gli eritrociti trasfusi vengono distrutti dai macrofagi e il ferro libero viene rapidamente trasportato nel fegato (Tura et al., 2018).

Un eccesso di ferro libero nel plasma è citotossico perché catalizza la formazione del radicale idrossile ( $\bullet\text{OH}$ ) attraverso la reazione di Fenton, che causa la perossidazione dei lipidi determinando un danno a livello mitocondriale e sintesi di collagene.

A lungo andare, l'accumulo di ferro può comportare disturbi epatici, danni cardiovascolari e diabete, pertanto si rende necessario un trattamento che prevede generalmente l'utilizzo di ferrocchelanti, i quali, attraverso il legame con il ferro, formano un complesso che permette la sua escrezione con la bile e/o le urine (Tura et al., 2018).

In condizioni fisiologiche, invece, i livelli di ferro nel corpo sono strettamente controllati attraverso l'attività di diverse proteine regolatrici, la più importante delle quali è l'epcidina, un ormone prodotto dagli epatociti, che interagisce con il suo recettore, la ferroportina, che trasporta il ferro fuori dalla cellula. Inibendo la ferroportina, l'epcidina inibisce così l'efflusso di ferro nel plasma (Nemeth et al., 2004).

Il *Luspatercept* è un farmaco che è stato approvato per il trattamento di pazienti dipendenti da trasfusione con SMD e sideroblasti ad anello che non sono responsivi agli agenti stimolanti l'eritropoietina. Si tratta di una proteina di fusione ricombinante che lega i ligandi appartenenti alla famiglia del TGF- $\beta$  (*Transforming growth factor  $\beta$* ) e,

bloccando in tal modo la trasduzione dei segnali TGF- $\beta$ -mediati, contribuisce al ripristino della maturazione delle cellule ematopoietiche e ad implementare il processo di eritropoiesi. È stato dimostrato, infatti, che nei pazienti affetti da SMD con sideroblasti ad anello il trattamento con *Luspatercept* ha determinato l'indipendenza da trasfusione nel 38% dei casi (Fenaux et al., 2020).

### **1.8. Microambiente midollare e Cellule staminali ematopoietiche**

Nel corso degli ultimi anni, l'interesse per il microambiente midollare è aumentato notevolmente, grazie anche ad una maggiore comprensione della funzione della normale nicchia staminale ematopoietica. Il midollo osseo rappresenta il microambiente che favorisce e supporta l'emopoiesi, ossia il processo di produzione delle cellule del sangue dell'individuo adulto a partire da un *pool* di cellule staminali tessuto-specifiche che prendono il nome di cellule staminali ematopoietiche.

Le diverse cellule progenitrici che discendono dalle CSE perdono progressivamente la capacità di auto-rinnovarsi nel momento in cui danno vita ai progenitori differenziati, generando dunque le cellule mature del sangue.

Le CSE sono gli unici elementi del sistema ematopoietico che in condizioni normali possiedono due proprietà fondamentali: la capacità di auto-rinnovamento (*self-renewal*) e la capacità di differenziare in tutte le linee cellulari del sangue essendo cellule pluripotenti. La capacità di auto-rinnovarsi permette alle cellule staminali di andare incontro a divisione per generare due cellule figlie con potenziale di staminalità intatto. Questa proprietà è essenziale per garantire un *reservoir* costante di CSE, necessario per tutta la durata della vita dell'individuo. In condizioni basali, le CSE si trovano in uno stato di quiescenza (fase G<sub>0</sub> del ciclo cellulare) e mostrano una ridotta capacità metabolica. In seguito a stimolazione, possono attivarsi, uscire dalla fase G<sub>0</sub> e contribuire all'emopoiesi. Perciò, in assenza della capacità di *self-renewal*, le continue stimolazioni alle quali è sottoposto un individuo durante tutta la sua vita potrebbero portare al loro esaurimento (Tura et al., 2018).

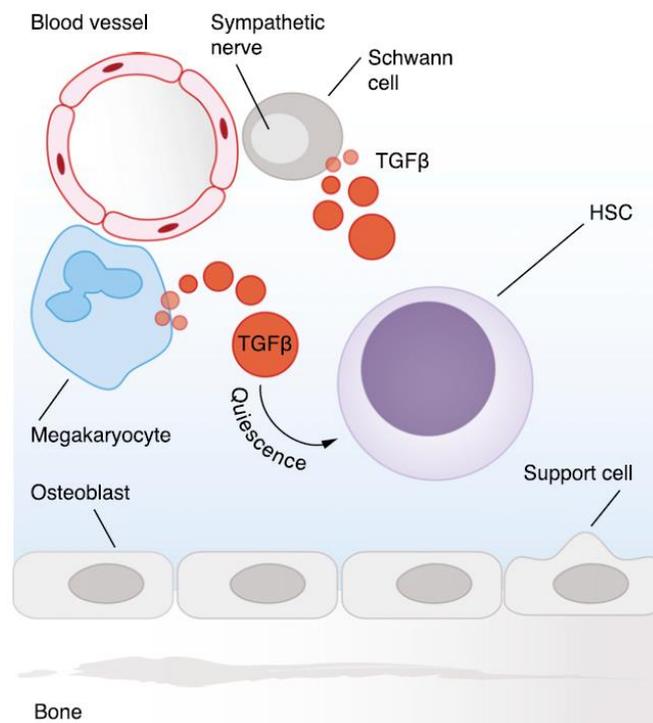
A livello midollare troviamo non solo cellule ematopoietiche, ma diverse popolazioni cellulari che sono di supporto all'emopoiesi e sono coinvolte nella sua regolazione

(Morrison and Scadden, 2014). Tra queste, cellule di origine mesenchimale (staminali mesenchimali e fibroblasti), cellule endoteliali, cellule della linea osteogenica, osteoclasti, osteociti, adipociti, cellule del sistema nervoso.

Il microambiente midollare è formato, oltre che da elementi cellulari, anche da proteine con funzione di *scaffold*, quali la fibronectina, il collagene e altre glicoproteine. Esse sono prodotte dai fibroblasti di origine mesenchimale sotto l'influenza delle cellule ematopoietiche mature e assolvono al ruolo di compartimentalizzazione del midollo osseo in aree specifiche per ogni linea di maturazione.

Il microambiente midollare rappresenta un crocevia per diversi *pathway*, quali la via di *signaling* PI3K/Akt, coinvolta nella regolazione del ciclo cellulare, nella quiescenza e nella proliferazione cellulare e nel cancro; la via di *signaling* Wnt (*Wnt family member*), una delle principali vie di segnalazione nelle cellule staminali e tumorali; il *pathway* del TGF- $\beta$  che regola la crescita e il differenziamento cellulare e l'apoptosi.

Il TGF- $\beta$  viene prodotto da una grande varietà di tipi cellulari presenti nel midollo osseo, dove si trova depositato in forma latente a livello della matrice midollare (**Figura 8**).



**Figura 8.** Il TGF- $\beta$  nel microambiente midollare (Blank and Karlsson, 2015).

Le cellule di Schwann regolano il processo di attivazione del TGF- $\beta$  attraverso il legame del suo precursore inattivo tramite una integrina presente sulla superficie cellulare. In questo modo, ne promuovono l'attivazione esponendolo a un taglio proteolitico da parte delle metalloproteasi della matrice (Yamazaki et al., 2011).

Anche i megacariociti rappresentano una fonte di TGF- $\beta$  a livello del microambiente. Infatti, le CSE che si trovano in loro prossimità mostrano l'attivazione delle proteine SMAD2/3 (*Small Mother Against Decapentaplegic*), mentre l'assenza dei megacariociti comporta una ridotta fosforilazione di SMAD2/3 e dunque la perdita di quiescenza e aumento della proliferazione delle CSE (Zhao et al., 2014).

### **1.9. TGF- $\beta$ signaling pathway**

Il TGF- $\beta$  rappresenta uno dei principali segnali di inibizione della crescita che le cellule normali, in particolare le cellule epiteliali, devono eludere per diventare cancerose. Esso esercita sulle cellule anche altri effetti del tutto diversi, costringendole a modificare il loro programma differenziativo.

#### **• Struttura e attivazione del TGF- $\beta$**

Il TGF- $\beta$  viene sintetizzato e secreto sotto forma di una proteina precursore, definita pre-pro-peptide, che è biologicamente inattiva e consta di un peptide segnale (SP, *Signal Peptide*), un peptide associato alla latenza (LAP, *Latency-associated Peptide*) e una proteina TGF- $\beta$  matura (MP, *Mature Protein*).

L'attivazione del TGF- $\beta$  avviene tramite due tagli proteolitici sequenziali. A livello del reticolo endoplasmatico rugoso, si verifica il primo taglio proteolitico della proteina precursore che porta all'eliminazione del peptide segnale idrofobico, localizzato all'estremità N-terminale, formando la proteina dimerica pro-TGF- $\beta$ . Il secondo taglio proteolitico, che avviene nell'apparato di Golgi, porta alla separazione del peptide associato alla latenza dalla proteina TGF- $\beta$  matura.

Una volta sintetizzato e processato, il TGF- $\beta$  viene rilasciato dalle cellule sotto forma di complessi latenti biologicamente inattivi. Si distinguono due tipologie di complessi latenti, uno piccolo (SLC, *Small Latent Complex*) e uno grande (LLC, *Large Latent*

Complex). Nel complesso SLC, una molecola TGF- $\beta$  matura è legata in modo non covalente al peptide LAP, mentre nel complesso LLC, il peptide associato alla latenza è legato tramite un ponte disolfuro a un membro della famiglia di proteine ad alto peso molecolare, chiamato LTBP (*Latent TGF- $\beta$  Binding Protein*), che conferisce al complesso la capacità di rimanere ancorato a componenti della matrice extracellulare, come la fibronectina (Annes et al., 2003; Fontana et al., 2005).

Il rilascio del TGF- $\beta$  latente, inoltre, è operato da enzimi in grado di tagliare la proteina LTBP (Figura 9).

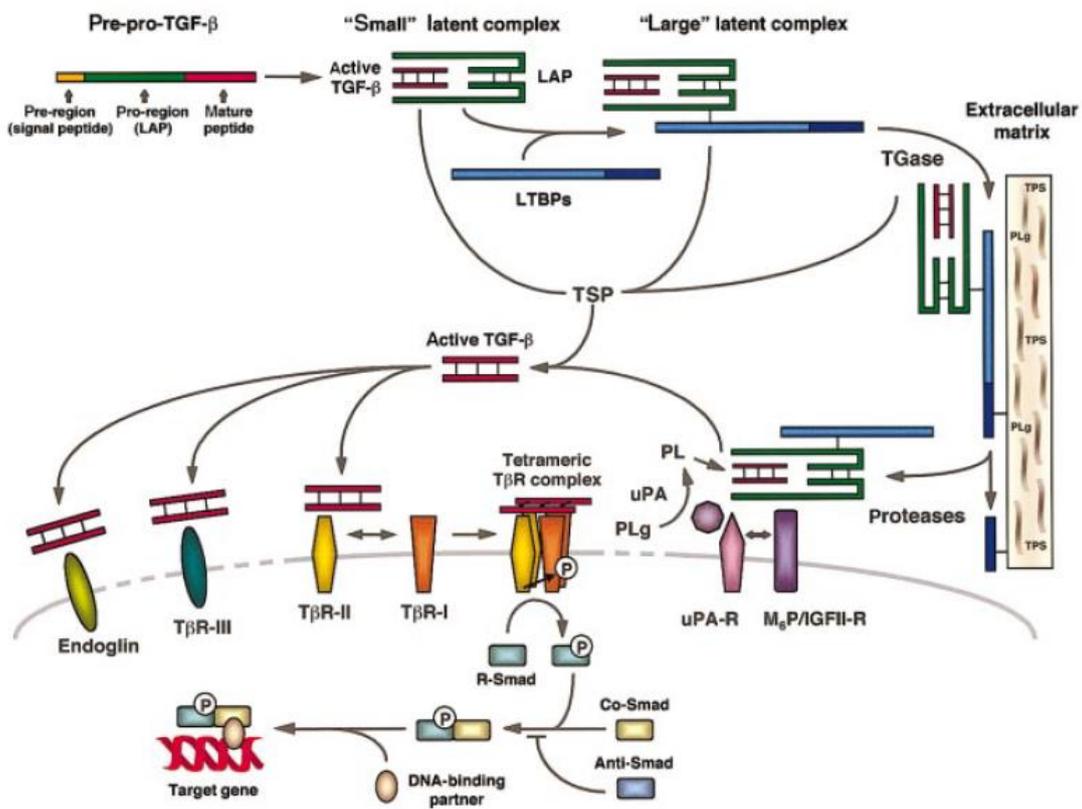


Figura 9. Processamento e attivazione del TGF- $\beta$  (Fortunel et al., 2000).

Tra gli enzimi proteolitici che promuovono l'attivazione del TGF- $\beta$  troviamo la plasmina, una serin proteasi attivata dall'attivatore urochinasi del plasminogeno (uPA), codificato dal gene *PLAU* (*Plasminogen activator, urokinase*). La plasmina attiva la forma latente del TGF- $\beta$  tramite taglio proteolitico all'estremità N-terminale del peptide LAP,

distruggendo in tal modo i legami covalenti e consentendo il rilascio della forma biodisponibile del TGF- $\beta$  (Lyons et al., 1990).

Similmente, per la metalloproteasi di matrice MMP2 (*Matrix metalloproteinase 2*) è stata dimostrata la sua capacità di attivare direttamente il TGF- $\beta$  *in vitro* (Jenkins, 2008). Al contrario, sembrerebbe che le metalloproteasi di matrice agiscano *in vivo* come tramite per le integrine, al fine di facilitare il taglio del peptide LAP e rilasciare il TGF- $\beta$  in forma attiva (Robertson and Rifkin, 2016).

- *Network di segnalazione del TGF- $\beta$*

La famiglia del TGF- $\beta$  include più di 30 proteine, secrete sotto forma di omodimeri con un motivo a nodo di cistina, suddivise in due sottofamiglie: TGF- $\beta$  e proteina morfogenetica dell'osso (BMP, *Bone Morphogenetic Protein*), con funzioni opposte. La prima include TGF- $\beta$ 1, TGF- $\beta$ 2, TGF- $\beta$ 3, attivina A e B, miostatina e diversi membri della famiglia del GDF (*Growth/Differentiation Factor*); la seconda, include più di 10 differenti proteine e l'ormone antimulleriano.

Il TGF- $\beta$  e i ligandi ad esso correlati hanno recettori simili ai recettori tirosin-chinasici, in quanto dotati di un dominio extracellulare che riconosce il ligando, un dominio trans-membrana, e un dominio citoplasmatico con attività chinasi.

Il recettore della famiglia del TGF- $\beta$  agisce come un complesso eterotetramerico costituito dall'assemblaggio di componenti con attività di serina/treonina chinasi: il recettore di tipo I, che propaga il segnale, e il recettore di tipo II.

In seguito all'interazione con il ligando, la catena del recettore di tipo II, che possiede un dominio Ser/Thr-chinasico costitutivamente attivo, entra in contatto e fosforila la subunità del recettore di tipo I, attivando il dominio chinasi. Questo dominio fosforila delle proteine citosoliche chiamate SMAD2 e SMAD3 e, alternativamente, in alcuni tipi cellulari possono essere fosforilate e attivate le proteine SMAD1, SMAD5 e SMAD8 a seconda del recettore di tipo I che viene espresso.

In particolare, una volta fosforilata nel citoplasma da un recettore del TGF- $\beta$ , una coppia di molecole SMAD attivate (R-SMAD) si assembla con una proteina SMAD4 e il complesso così generato migra nel nucleo dove recluta regolatori trascrizionali

aggiuntivi, quali fattori di legame al DNA, co-attivatori, co-repressori, rimodellatori della cromatina, che controllano l'espressione di numerosi geni *target* (Shi and Massagué, 2003).

Il TGF- $\beta$  può attivare numerosi *pathway* SMAD-indipendenti, incluso quello delle MAPK (*Mitogen-activated protein kinase*) e di PI3K/Akt, che sono frequentemente coinvolti nel cancro. Sembrerebbe, infatti, che il comportamento di *tumor promoter* del TGF- $\beta$ , quale l'induzione della transizione epitelio-mesenchima (EMT, *Epithelial-mesenchymal transition*) e la formazione di metastasi, sia dovuto a uno *switch* dalla via di segnalazione canonica SMAD-dipendente alla via di segnalazione non canonica (Massagué, 2012).

### 1.9.1. Ruolo del TGF- $\beta$ nell'emopoiesi e nelle mielodisplasie

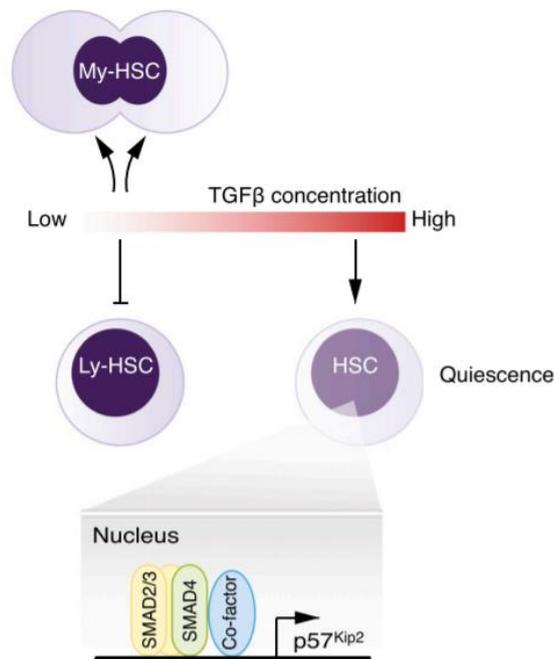
Il TGF- $\beta$  gioca un ruolo importante nel definire il comportamento delle cellule staminali ematopoietiche, in particolar modo regola la quiescenza e il *self-renewal*, impedendo il *commitment* verso una linea cellulare specifica. I differenti effetti biologici sono strettamente dipendenti dalla concentrazione di fattore a cui le cellule sono esposte.

È stato osservato che a basse concentrazioni stimola la crescita delle CSE, in particolare quella delle cellule della linea mieloide, mentre inibisce la crescita delle cellule della linea linfoide. A concentrazioni crescenti, invece, inibisce la proliferazione e l'eritropoiesi e accelera il differenziamento eritroide terminale di linee cellulari e progenitori eritroidi (Blank and Karlsson, 2015) (**Figura 10**).

Il TGF- $\beta$  è uno dei regolatori chiave dell'eritropoiesi perché è coinvolto nel controllo sia degli stadi precoci sia degli stadi tardivi dello sviluppo dei progenitori cellulari eritroidi. La prima isoforma, TGF- $\beta$ 1, è stata identificata come un inibitore del ciclo cellulare per le cellule progenitrici BFU-E (*Burst Forming Units – Erythroid*), che rispondono al TGF- $\beta$ 1 con l'arresto della proliferazione, aumentando il tempo di permanenza nella fase G<sub>1</sub> del ciclo cellulare.

Oltre al suo ruolo nel controllo della proliferazione delle cellule progenitrici eritroidi, il TGF- $\beta$ 1 è coinvolto nella regolazione del differenziamento eritroide. È stato osservato come le cellule BFU-E, sotto l'influenza del fattore di crescita, differenzino in cellule BFU-E *late* e in cellule CFU-E (*Colony Forming Units – Erythroid*).

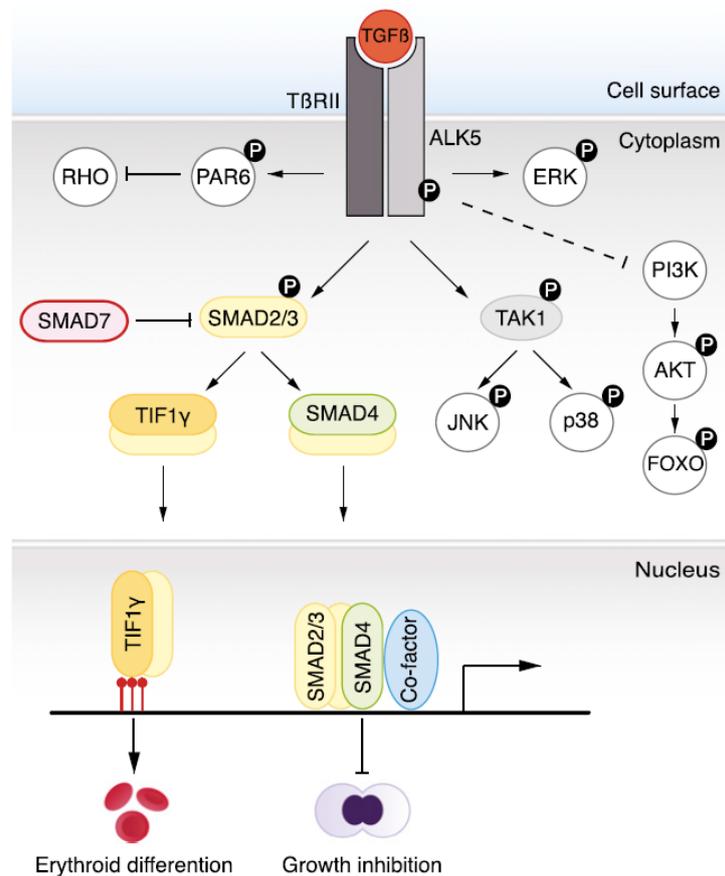
In modo del tutto simile, TGF- $\beta$ 1 riduce l'attività mitotica delle cellule CFU-E e induce il differenziamento delle stesse in tipi cellulari più maturi, grazie all'azione sinergica con l'eritropoietina (Zermati et al., 2000).



**Figura 10.** Effetto bidirezionale del TGF- $\beta$  (Blank and Karlsson, 2015).

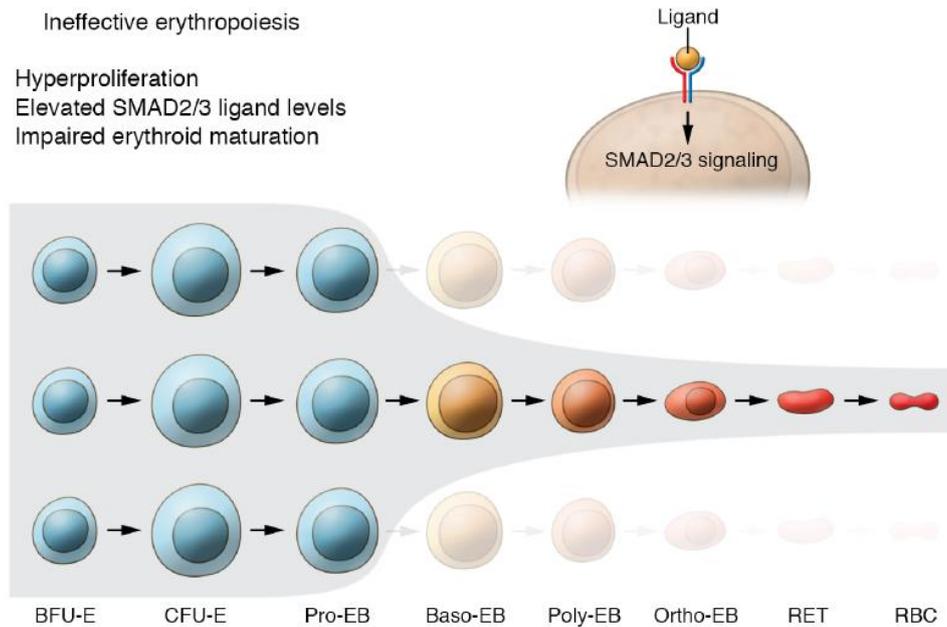
Nelle cellule staminali e nei progenitori ematopoietici, il TGF- $\beta$  si rende promotore del differenziamento eritroide o dell'inibizione della crescita attraverso un meccanismo dipendente dal legame competitivo di SMAD4 e TIF1 $\gamma$  (*Transcriptional intermediary factor 1 gamma*) a SMAD2/3. In particolare, quando le proteine SMAD2/3 si associano al cofattore TIF1 $\gamma$  viene stimolato il differenziamento eritroide, mentre la formazione del complesso SMAD2/3-SMAD4 inibisce la proliferazione delle cellule staminali ematopoietiche (He et al., 2006) (**Figura 11**).

L'effetto, dunque, dei ligandi della famiglia del TGF- $\beta$  (in particolare TGF- $\beta$ 1) sui progenitori eritroidi risulta in una ridotta eritropoiesi a causa dell'inibizione della loro attività mitotica, sebbene una quota di cellule progenitrici vada incontro a un rapido differenziamento verso forme cellulari mature.



**Figura 11.** TGF-β signaling pathway (Blank and Karlsson, 2015).

È stato dimostrato che la via di *signaling* del TGF-β riveste un ruolo chiave nella fisiopatologia delle sindromi mielodisplastiche (Bataller et al., 2019), patologie clonali della cellula staminale ematopoietica, caratterizzate da citopenie derivanti da un alterato processo di emopoiesi. Il TGF-β regola la maturazione eritroide in condizioni fisiologiche e contribuisce ad una maturazione eritroide deregolata in condizioni patologiche. Infatti, l'*up*-regolazione di questo *pathway* è stata indicata come una delle possibili cause dell'eritropoiesi inefficace, una peculiarità delle SMD, definita dalla incapacità di produrre un numero adeguato di globuli rossi nonostante la presenza di un numero aumentato di precursori eritroidi immaturi. Il risultato è una iperplasia delle cellule eritroidi immature nel midollo osseo e anemia a livello periferico (Verma et al., 2020) (**Figura 12**).



**Figura 12.** Eritropoiesi inefficace e compromessa maturazione eritroide (Verma et al., 2020).

Quanto osservato nelle sindromi mielodisplastiche è il risultato dell'azione delle proteine SMAD, regolatori fisiologici importanti dell'emopoiesi. È stato dimostrato come nei pazienti affetti da SMD sia presente a livello midollare un aumentato numero di cellule che presentano la proteina SMAD2 attivata (fosforilata), la quale risulta costitutivamente attiva e sovraespressa nei precursori ematopoietici di tali pazienti (Zhou et al., 2008).

Inoltre, è stato osservato come nelle SMD vi sia una ridotta presenza di fattori con azione inibitoria nei confronti del *pathway* del TGF- $\beta$ , quali le proteine SMAD7 e SKI, che comporta una attivazione costitutiva del *signaling* intracellulare SMAD2-mediato e la soppressione dell'emopoiesi (Zhou et al., 2011).

## 2. SCOPO DELLA TESI

Le sindromi mielodisplastiche sono un gruppo eterogeneo di emopatie clonali della cellula staminale, caratterizzate da un alterato processo di emopoiesi che risulta in un ridotto numero di cellule del sangue a livello periferico.

Per molto tempo, il criterio diagnostico principale è stato l'esame morfologico, volto a identificare segni di displasia nelle varie linee cellulari. Numerosi studi, pubblicati negli ultimi anni, hanno messo in luce un ruolo sempre più preponderante delle mutazioni somatiche sia nella patogenesi sia nell'evoluzione delle sindromi mielodisplastiche.

Lo scopo del mio progetto di tesi è stato quello di analizzare il profilo di espressione genica di pazienti affetti da sindrome mielodisplastica, prevalentemente a basso rischio, con mutazioni nei geni implicati nello *splicing* dell'RNA.

Gli obiettivi primari del mio lavoro sono stati:

- Selezione dei pazienti della coorte di studio sulla base delle caratteristiche del loro profilo mutazionale, utilizzando la tecnologia *targeted*-NGS.
- Analisi del *landscape* mutazionale nei pazienti affetti da SMD tramite WGS, al fine di identificare mutazioni in associazione con i geni dello *splicing*.
- Valutazione del profilo di espressione del trascrittoma nei pazienti affetti da SMD rispetto ai controlli.
- Stratificazione dell'analisi del trascrittoma in relazione al profilo mutazionale.

### 3. MATERIALI E METODI

#### 3.1. Popolazione di studio

La nostra coorte di studio comprende 50 pazienti (30 maschi, 20 femmine, età mediana alla diagnosi 75 anni, *range* 40-91 anni) affetti da sindrome mielodisplastica, stratificati secondo la classificazione IPSS-R (Greenberg *et al.*, 2012).

Dei 50 pazienti, 32 sono stati selezionati sulla base del *landscape* mutazionale a partire da una coorte iniziale di 300 pazienti con SMD raccolti nella nostra biobanca (GROM - Gruppo Romano Mielodisplasie, Protocollo BMT-AZA, Laboratorio di Diagnostica Avanzata Oncoematologica – Fondazione PTV, Roma). I 18 pazienti restanti, di cui non è noto il profilo mutazionale, sono stati selezionati al momento della diagnosi presso il Laboratorio di Diagnostica Avanzata Oncoematologica – Fondazione PTV, Roma.

Di tutti i pazienti oggetto di studio si dispone di campioni appaiati di DNA e RNA. Le caratteristiche cliniche dei pazienti studiati sono riassunte nella **Tabella 6**.

Caratteristiche dei pazienti	N. ( <i>range</i> )
N° di pazienti	50
Età mediana alla diagnosi	75 (40-91)
<b>Sesso</b>	
Maschile	30
Femminile	20
<b>IPSS-R</b>	
<i>Very low</i>	7
<i>Low</i>	27
<i>Intermediate</i>	7
<i>High</i>	7
<i>Very high</i>	1
NA	1
<b>Percentuale di blasti</b>	
0% - 4,5%	34
5% - 9,5%	4
10% - 15%	6
NA	6

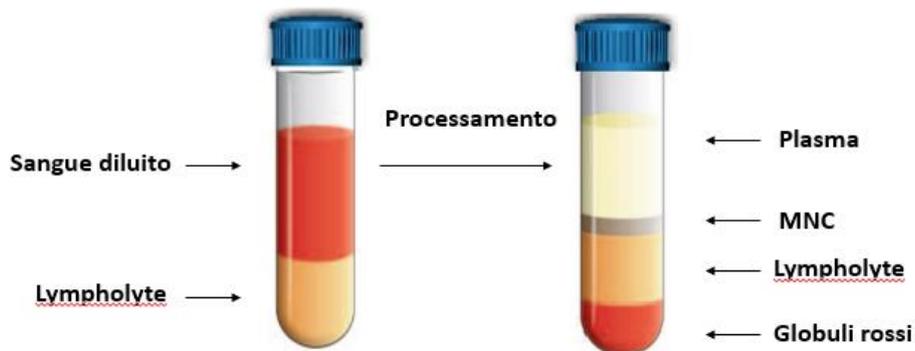
**Tabella 6.** Caratteristiche cliniche dei pazienti con SMD studiati. NA: Not Available

Per l'analisi del trascrittoma, tramite RNA-Seq, sono stati utilizzati come controlli 64 pazienti ospedalieri non affetti da patologie ematologiche, accoppiati per età e sesso con i pazienti affetti da SMD.

La popolazione di controllo utilizzata per gli studi di espressione genica mediante qRT-PCR è costituita da donatori sani di midollo osseo (n=5).

### 3.2. Separazione degli elementi mononucleati del sangue mediante gradiente di densità

La separazione delle cellule mononucleate da campioni di sangue midollare (BM-MNCs, *Bone Marrow-Mononuclear Cells*) si basa sulla centrifugazione in gradiente di densità. La frazione di cellule mononucleate da midollo osseo (MO) è stata isolata mediante separazione con *Ficoll-Hypaque*. Il campione di sangue, presente all'interno di una provetta contenente sodio citrato, viene trasferito in una provetta Falcon sterile da 15 ml. Una volta diluito con Dulbecco's PBS (*Phosphate Buffer Saline*) in una proporzione di 1:2, viene poi stratificato in una nuova provetta contenente 3,5 ml di Ficoll (*Lympholyte*® - *Cederlane*) ogni 5 ml di sangue diluito. Segue una centrifugazione a 800 g per 25 minuti che determina la formazione di un anello di cellule mononucleate (MNC, *Mononuclear Cells*) all'interfaccia tra il plasma (fase superiore) e il Ficoll (fase inferiore), mentre gli eritrociti sedimentano sul fondo della provetta come illustrato nella **Figura 13**.



**Figura 13.** Separazione degli elementi mononucleati del sangue su gradiente di densità. Dopo centrifugazione si ha la formazione di diverse fasi, le cellule mononucleate (MNC) si trovano all'interfaccia tra il plasma e il Lympholyte (Immagine adattata da Ficoll-Paque PLUS. Amersham Biosciences).

L'anello di cellule mononucleate viene recuperato e trasferito in una nuova provetta Falcon contenente 10 ml di soluzione fisiologica per lavare le cellule.

Dopo aver centrifugato a 1200 rpm per 10 minuti, viene rimosso il surnatante e il *pellet* cellulare viene risospeso in un appropriato volume di Dulbecco's PBS. Si procede quindi con la conta cellulare utilizzando il colorante Trypan Blue 0,5% (Euroclone) e una camera contaglobuli (camera di Bürker). Il Trypan Blue penetra esclusivamente nelle cellule la cui membrana plasmatica è danneggiata, quindi colora selettivamente in blu le cellule morte permettendo in tal modo di discriminarle dalle vive con una rapida analisi al microscopio ottico.

Il *pellet* ottenuto dopo la centrifugazione viene ripartito in due aliquote in provette sterili da 2 ml (a seconda delle concentrazioni cellulari ottenute), una per l'estrazione del DNA e una per l'estrazione dell'RNA.

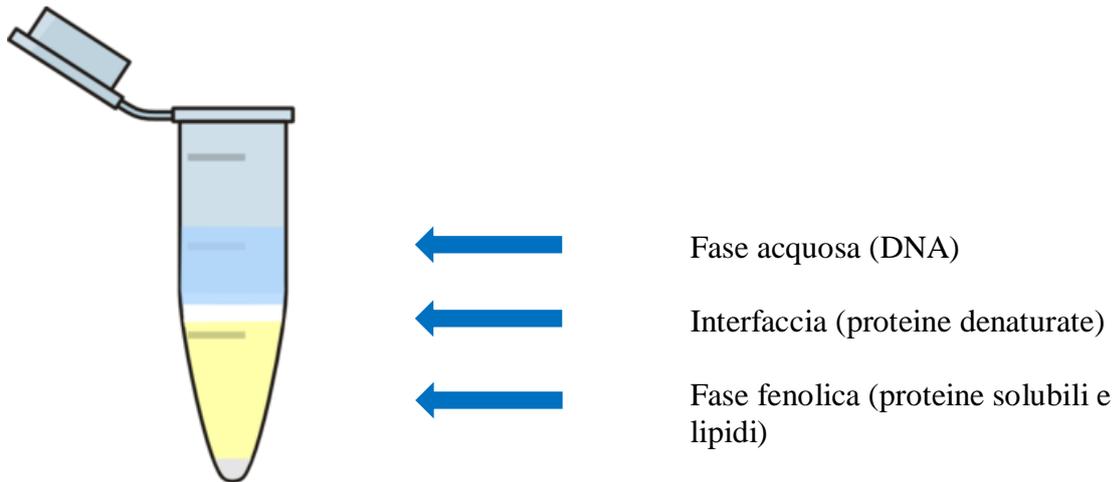
### **3.3. Analisi del genoma**

#### **3.3.1. Estrazione del DNA genomico**

Il protocollo adottato per l'estrazione del DNA prevede che tutte le fasi vengano eseguite manualmente dall'operatore. Il *pellet* cellulare viene risospeso con 200 µl di PBS, 20 µl di Proteinasi K (20 mg/ml) e 200 µl di *Buffer AL* (*Lysis Buffer*, Qiagen).

La miscela così preparata viene agitata brevemente con il vortex e lasciata in incubazione a 56°C nel Thermoblock per 20 minuti. Si aggiungono 440 µl di una soluzione 25:24:1 di Fenolo/Cloroformio/Alcool isoamilico (Acros Organics), si agita velocemente e si centrifuga a 8595 g a T=amb per 10 minuti.

Al termine della centrifugazione si potranno distinguere nella provetta una fase inferiore contenente la miscela di fenolo/cloroformio/alcool isoamilico e nella parte superiore la fase acquosa contenente il DNA (**Figura 14**).



**Figura 14.** Dopo centrifugazione si ottengono tre fasi: una fase acquosa contenente il DNA, una fase intermedia costituita da proteine denaturate, e una fase fenolica contenente lipidi, RNA e proteine ricche di amminoacidi idrofobici.

Si preleva la fase acquosa e la si trasferisce in una nuova provetta da 1,5 ml. Si aggiungono 40  $\mu$ l di acetato di sodio 3M e 800  $\mu$ l di etanolo 100%, si agita velocemente e si lascia ad incubare a  $-20^{\circ}\text{C}$  per almeno 20 minuti. Dopo una centrifugazione a 8595 g a  $4^{\circ}\text{C}$  per 20 minuti, si rimuove il surnatante per inversione, si aggiunge 1 ml di etanolo 70% e si centrifuga a 8595 g a  $4^{\circ}\text{C}$  per 20 minuti. Si elimina nuovamente il surnatante per inversione, si lascia asciugare il *pellet* rimasto sul fondo della provetta ed infine lo si risospende nel quantitativo opportuno di ddH<sub>2</sub>O.

### 3.3.2. Valutazione qualitativa e quantitativa del DNA

Il DNA ottenuto è stato valutato quantitativamente e qualitativamente. La valutazione quantitativa permette di calcolare la concentrazione esatta di DNA estratto e viene effettuata attraverso lo strumento Qubit® *dsDNA assay*.

I test di quantificazione basati su fluorescenza sono altamente sensibili e precisi. La valutazione qualitativa si esegue, invece, per avere informazioni sul grado di purezza del DNA estratto rispetto ad eventuali proteine e RNA contaminanti. Per effettuare la valutazione qualitativa, si ricorre alla lettura spettrofotometrica a 260 nm (lunghezza d'onda alla quale assorbono gli acidi nucleici) e 280 nm (lunghezza d'onda alla quale

assorbono le proteine), attraverso lo strumento NanoPhotometer N60 (Implen). La purezza è espressa dal rapporto tra i valori di assorbanza misurati a 260 nm e 280 nm ed è di norma un valore corrispondente a  $2,0 \pm 0,2$ .

Un'ulteriore valutazione quantitativa del DNA estratto viene fatta con lo strumento Bioanalyzer (Agilent Technologies), riportato nella **Figura 15**.



*Figura 15. Bioanalyzer (Agilent Technologies).*

Lo strumento permette di avere allo stesso tempo una caratterizzazione qualitativa e quantitativa del campione di DNA mediante elettroforesi capillare. L'Agilent Bioanalyzer 2100 restituisce un valore che permette di attribuire un indice qualitativo al DNA totale esaminato. Tale parametro presenta una scala che va da 10 (DNA di ottima qualità) a zero (DNA completamente degradato).

Il DNA da analizzare viene caricato su un chip che contiene una serie di microcanali strettamente interconnessi: i frammenti di acidi nucleici vengono così separati in base al loro peso molecolare come avverrebbe in una normale elettroforesi su gel di agarosio. I microcanali di ogni chip sono riempiti con una matrice ed un fluorocromo. I campioni si muovono attraverso i microcanali a partire dal pozzetto di caricamento e vengono separati mediante corsa elettroforetica.

Al termine della corsa i campioni vengono letti in base alla loro fluorescenza e tale informazione è tradotta in elettroferogrammi. Abbiamo bisogno di un rapporto tra i

valori di assorbanza misurati a 260 nm e 280 nm compreso tra 1,8 e 2,0. I frammenti di DNA genomico devono essere superiori a 300 bp.

### **3.3.3. Next Generation Sequencing (NGS)**

Il sequenziamento di nuova generazione (NGS, *Next Generation Sequencing*) consente la lettura multipla in parallelo di singoli frammenti di DNA analizzando milioni di paia di basi in poche ore.

Tutte le piattaforme NGS disponibili presentano una caratteristica tecnologica comune: il sequenziamento parallelo massivo di molecole di DNA amplificate in modo clonale e separate spazialmente in una cella a flusso (*flow cell*). Si tratta di una innovazione rispetto al metodo di sequenziamento Sanger, perché la determinazione della sequenza non richiede il passaggio limitante della separazione elettroforetica in quanto via via che i nucleotidi vengono incorporati nella reazione di sequenziamento vengono simultaneamente identificati.

Nella tecnologia NGS, il sequenziamento viene effettuato mediante cicli ripetuti di estensioni nucleotidiche ad opera di una DNA polimerasi o, in alternativa, mediante cicli iterativi di ligazione di oligonucleotidi. Poiché la procedura è parallela e massiva, tali piattaforme consentono di sequenziare fino a miliardi di paia di basi di DNA in un'unica seduta analitica.

Per il nostro studio, il sequenziamento è stato realizzato sfruttando il pannello *MYeloid Solution* di *SOPHiA GENETICS* (*SOPHiA GENETICS*, Saint-Sulpice, Svizzera) e la piattaforma di sequenziamento MiniSeq® di Illumina (Illumina, San Diego, California).

Per la preparazione della *library* abbiamo utilizzato tra i 10-200 ng di DNA genomico estratto dalle cellule mononucleate di midollo osseo al fine di analizzare, con estrema profondità e sensibilità, le regioni *target* dei geni di interesse presenti in un pannello di 30 geni.

Il pannello comprende geni noti per essere associati alla patogenesi delle neoplasie mieloidi (**Tabella 7**).

Gene	Regione target (esone)	Gene	Regione target (esone)	Gene	Regione target (esone)
ABL	4 – 9	FLT3	13 – 15 e 20	PTPN11	3, 7 – 13
ASXL1	9, 11, 12	HRAS	2, 3	RUNX1	Tutta
BRAF	15	IDH1	4	SETBP1	4
CALR	9	IDH2	4	SF3B1	10 – 16
CBL	8, 9	JAK2	Tutta	SRSF2	1
CEBPA	Tutta	KIT	2, 8 – 11, 13, 17 e 18	TET2	Tutta
CSF3R	Tutta	KRAS	2, 3	TP53	Tutta
DNMT3A	Tutta	MPL	10	U2AF1	2, 6
ETV6	Tutta	NPM1	10, 11	WT1	6 – 10
EZH2	Tutta	NRAS	2, 3	ZRSR2	Tutta

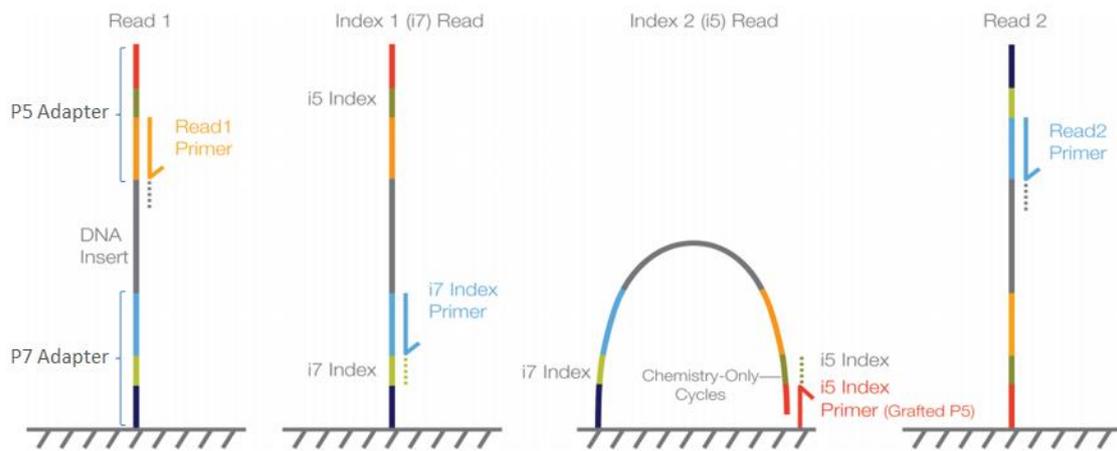
**Tabella 7.** Geni del pannello mieloide di SOPHiA GENETICS.

### 3.3.3.1. Il flusso di lavoro in NGS

Il flusso di lavoro del pannello da noi utilizzato prevede una prima fase di lavorazione da parte dell'operatore, suddivisa in due giornate, seguita da una fase che viene espletata all'interno dello strumento MiniSeq® di Illumina (Illumina, San Diego, California).

Per prima cosa, viene effettuata una valutazione quantitativa del DNA tramite l'utilizzo del fluorimetro Qubit®, a cui segue una fase di amplificazione clonale del DNA ottenuto dalla frammentazione del campione in esame. I frammenti di DNA ottenuti, di lunghezza pari a 150-600 bp, vengono opportunamente modificati dall'aggiunta di una coppia di indici (sequenze di circa 6-8 bp che permettono di identificare i frammenti associandoli in maniera univoca ad ogni paziente) e degli adattatori interni ed esterni (**Figura 16**).

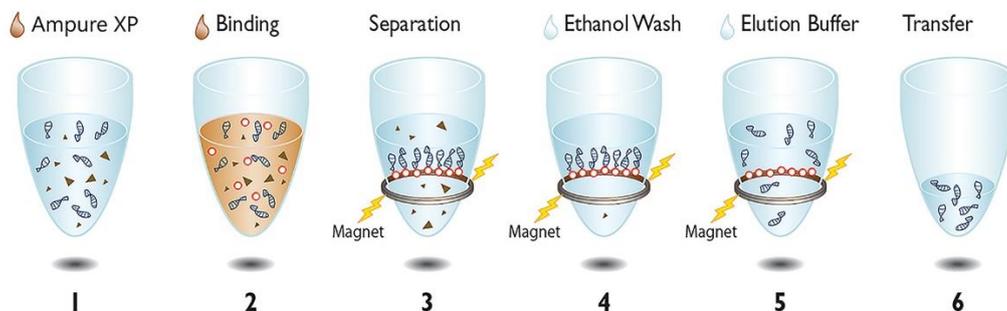
Gli adattatori interni agiscono come innesco per la reazione di polimerizzazione e sequenziamento, legandosi a delle sonde biotinilate; gli adattatori esterni, P5 e P7, sono quelli che permetteranno il legame con la *flow cell* (regione fisica dove avverrà la vera e propria reazione di sequenziamento).



**Figura 16.** Ligazione di incidi e adattatori interni ed esterni (ILLUMINA). Index: indice; P5 e P7: adattatori esterni complementari ai primer della flow cell; Rd1 e 2 SP (Read 1 e 2 sequencing primer): adattatori interni utilizzati come innesco per la reazione di polimerizzazione e sequenziamento.

Successivamente viene eseguito uno step di selezione dei frammenti (*Dual Size Selection*) nel quale si rimuovono tutti i frammenti privi degli adattatori e tutti i dimeri di adattatori (cioè i dimeri che gli adattatori in eccesso hanno formato legandosi fra di loro), che possono ridurre sensibilmente la resa della reazione di sequenziamento. La rimozione avviene mediante l'utilizzo di biglie magnetiche rivestite di streptavidina (**Figura 17**).

Più è elevato il numero di biglie, più si riescono a recuperare frammenti di piccole dimensioni. Al contrario, poche biglie si legheranno al DNA più carico e quindi ai frammenti più grandi.



**Figura 17.** Meccanismo d'azione delle biglie (ThermoScientificTM).

I frammenti di DNA ottenuti tramite l'aggiunta degli adattatori costituiscono la cosiddetta libreria di sequenziamento. Dopo aver creato la libreria, la si amplifica tramite PCR, viene effettuata nuovamente una purificazione tramite biglie magnetiche (*Post Amplification Clean Up*) e si valutano di nuovo le dimensioni dei frammenti tramite Bioanalyzer. I frammenti ottenuti hanno dimensioni di circa 300-400 bp e una concentrazione di DNA compresa tra 500 e 1000 ng.

Si prepara una *pool library*, ovvero una miscela che conterrà tutte le librerie di ogni singolo paziente precedentemente amplificate e riconoscibili tramite gli indici adattatori. A questo punto è possibile interrompere la lavorazione che verrà ripresa la giornata successiva.

Alle sequenze di DNA precedentemente ottenute sono aggiunte delle sonde biotilate (specifiche per tutte le porzioni codificanti del genoma) che permetteranno successivamente, tramite l'uso di biglie magnetiche legate a streptavidina, di catturare le regioni del DNA di interesse. Queste regioni di DNA così arricchite vengono purificate mediante lavaggi con etanolo ed amplificate ulteriormente tramite PCR.

Al termine della reazione di amplificazione, viene effettuato un ulteriore lavaggio per la rimozione dei primer in eccesso e dei sali di reazione. Gli amplificati vengono valutati mediante analisi fluorimetrica (Qubit®) e Bioanalyzer.

La concentrazione della libreria ottenuta mediante Qubit® viene successivamente adattata alla taglia media dei frammenti di libreria ottenuti tramite Bioanalyzer in accordo con la seguente formula:

$$\frac{\text{concentrazione ng}/\mu\text{l}}{660 \text{ g/mol} * \text{taglia media della libreria}} \times 10^6 = \text{concentrazione in nM}$$

Della concentrazione della libreria ottenuta tramite la precedente formula sono caricati 1,6 pM nella cartuccia di sequenziamento. Gli step successivi sono condotti nella piattaforma MiniSeq® di Illumina (Illumina, San Diego, California).

### 3.3.4. Whole Genome Sequencing (WGS)

L'analisi di WGS consiste nel sequenziamento dell'intero genoma, ossia di tutti gli esoni, tutti gli introni e tutte le regioni intergeniche. Infatti, sebbene la maggior parte delle mutazioni associate ad una qualche patologia cada all'interno dell'esoma, alcune mutazioni patogene possono cadere anche in regioni introniche profonde o in regioni a monte o a valle dei geni a volte non evidenziabili con l'analisi dell'esoma.

Per il sequenziamento del genoma, si utilizza il genoma umano di riferimento GRCh37/hg19. Il rapporto tra la quantità totale delle sequenze prodotte per il sequenziamento e la dimensione del genoma definisce il grado di copertura o *coverage*, cioè il livello di profondità di lettura. Esso non è costante, ma dipende dalle dimensioni del genoma, dal grado di ripetitività, dalla lunghezza e dalla natura delle sequenze generate.

Le librerie per il sequenziamento dell'intero genoma sono state generate a partire da 1 µg di DNA utilizzando il kit di preparazione *TruSeq PCR-Free*, seguendo le raccomandazioni del produttore (Illumina, San Diego, CA, USA) e sequenziate con lo strumento NovaSeq6000 Illumina seguendo un protocollo standard 2x150bp *paired-end reads* ad una profondità media di copertura di 141x (104x - 184x).

L'analisi bioinformatica primaria (mappatura delle *reads* al genoma umano di riferimento GRCh37/hg19, chiamata delle varianti somatiche) è stata eseguita all'interno del *BaseSpace Sequence Hub* di Illumina utilizzando la *DRAGEN Somatic Pipeline 3.4.5*. Le varianti a singolo nucleotide, le inserzioni (< 50 bp) e le delezioni (*indel*) sono state identificate usando *Ensembl Variant Effect Predictor*. Le varianti identificate sono state poi funzionalmente annotate usando i *database* pubblici disponibili (gnomAD, ClinVar, COSMIC).

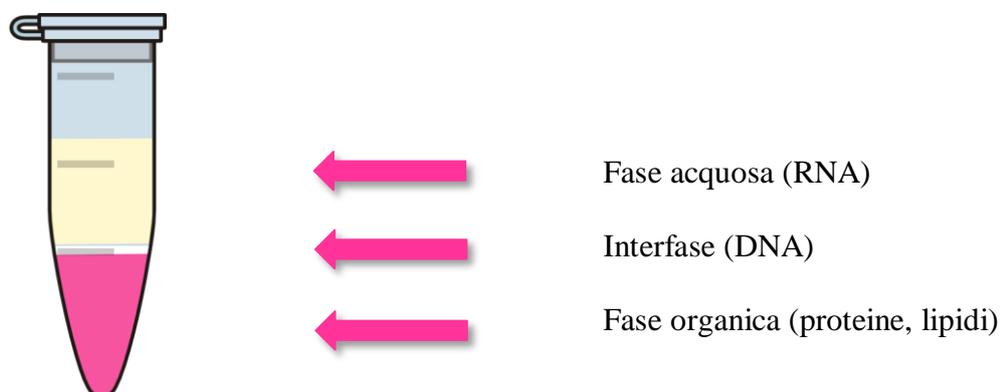
Dalla lista delle varianti identificate sono state escluse quelle appartenenti alle regioni non codificanti, quelle sinonime, quelle di natura benigna (gnomAD, ClinVar e TCGA per la predizione *in vivo* e Polyphen 2 e SIFT per la predizione *in silico*), quelle con frequenza polimorfica (COSMIC e NCBI), con VAF < 3% e con MLL Pred < 0,6 (*Munich Leukemia Laboratory Predictor scores*).

Il WGS e l'analisi statistica primaria sono state effettuate in collaborazione con il "MLL Munich Leukemia Laboratory" diretto dal Prof. Torsten Haferlach.

### 3.4. Analisi del trascrittoma

#### 3.4.1. Estrazione dell'RNA totale

L'estrazione dell'RNA viene eseguita mediante *TRIZOL*<sup>®</sup> reagent, una soluzione monofasica di fenolo e guanidinio isotiocianato in grado di dissolvere la membrana delle cellule ed inibire l'attività delle RNasi. Alla sospensione cellulare contenente 1 ml di *TRIZOL* vengono aggiunti 200 µl di cloroformio. Si lascia in incubazione per 2-3 minuti a temperatura ambiente, si agita per inversione e si centrifuga a 12000 g per 15 minuti a 4°C. In seguito ad agitazione ed ultracentrifugazione è possibile distinguere tre fasi: una fase acquosa meno densa contenente l'RNA, un'interfaccia di separazione biancastra contenente le proteine, e una fase fenolica sul fondo contenente DNA, membrane, proteine solubili e altre componenti cellulari (**Figura 18**).



**Figura 18.** In seguito alla centrifugazione si formano le tre fasi: una fase acquosa contenente l'RNA, una fase intermedia costituita da proteine denaturate, e una fase fenolica contenente lipidi, DNA e proteine ricche di amminoacidi idrofobici.

Dopo centrifugazione, la fase acquosa viene prelevata (circa 500 µl) e trasferita in una provetta sterile da 2 ml. A questa viene aggiunto un ugual volume di isopropanolo (rapporto 1:1) per consentire la precipitazione dell'acido nucleico. La sospensione di RNA in isopropanolo viene incubata per 10 minuti a temperatura ambiente. Segue una

centrifugazione a 12000 g per 10 minuti a 4°C. Il supernatante viene rimosso e il *pellet* lavato con 1 ml di etanolo 75%. Il *pellet*, ottenuto dopo un'ulteriore centrifugazione a 12000 g per 5 minuti a 4°C, viene lasciato asciugare, in quanto l'etanolo interferisce con la successiva reazione di PCR, e risospeso, a seconda della grandezza dello stesso, in 10-100 µl di acqua bidistillata sterile RNasi *free*.

### 3.4.2. Quantificazione spettrofotometrica dell'RNA

La quantificazione dell'RNA viene eseguita per determinarne la concentrazione in soluzione e verificare l'assenza di contaminanti quali proteine, polisaccaridi o fenolo. L'RNA è stato analizzato mediante uno spettrofotometro (NanoPhotometer N60 - Implen) in grado di effettuare analisi di volumi estremamente ridotti di campione (1 µl) con accuratezza e riproducibilità elevate. Per valutare la concentrazione e il grado di purezza rispetto ad eventuali contaminanti, si ricorre alla lettura spettrofotometrica a 260 nm e 280 nm, lunghezze d'onda alle quali assorbono rispettivamente gli acidi nucleici e le proteine. La purezza è data dal rapporto tra i valori di assorbanza misurati a 260 nm e 280 nm (260/280), il valore corrisponde generalmente a  $2,0 \pm 0,2$ .

### 3.4.3. RNA-Seq

L'RNA-Seq rappresenta un'analisi su larga scala del trascrittoma basata su tecnologie NGS, fondamentale per determinare a livello qualitativo e quantitativo il piano di espressione genica e la caratterizzazione delle diverse varianti di *splicing*. Il sequenziamento del trascrittoma prevede la sintesi di cDNA mediante retrotrascrizione da RNA; la generazione, mediante l'utilizzo di *primer*, di frammenti di cDNA a cui vengono aggiunti degli adattatori specifici, ossia sequenze oligonucleotidiche necessarie per il sequenziamento; il sequenziamento massivo diretto della libreria di frammenti così generata; la formazione delle *reads*, piccole sequenze di cDNA che devono essere allineate lungo il genoma, il cui numero è direttamente proporzionale al livello di espressione di un gene.

Le librerie per il sequenziamento sono state generate a partire da 250 ng di RNA totale per campione con il *TruSeq Total Stranded RNA library kit* (Illumina, San Diego, CA),

secondo le istruzioni del produttore. Le *reads paired-end* di 100 bp sono state generate con il sistema NovaSeq 6000 con un conteggio mediano di 86 milioni di *paired-reads* per campione (11 - 190 milioni). Le letture sono state allineate con *STAR aligner 2.5.0a* lungo il genoma umano di riferimento GRCh37/hg19 e i conteggi delle *reads* per gene sono stati stimati con Cufflinks 2.2.1 utilizzando RefSeqGene. La chiamata delle varianti somatiche è stata effettuata utilizzando il pacchetto GATK 4.1.6.0, seguendo le indicazioni del produttore, per identificare tali varianti a livello dell'RNA.

La normalizzazione dell'espressione genica e i test per l'espressione genica differenziale sono stati eseguiti utilizzando DeSeq2.

L'analisi di PCA (*Principal Component Analysis*) è stata realizzata utilizzando DESeq2 e la piattaforma *Perseus* (<https://maxquant.net/perseus/>).

L'RNA-Seq è stato effettuato in collaborazione con il "MLL Munich Leukemia Laboratory" diretto dal Prof. Torsten Haferlach.

#### **3.4.4. Retrotrascrizione dell'RNA in cDNA**

Le reazioni di retrotrascrizione sono state effettuate mediante *QuantiTect® Reverse Transcription Kit* (Qiagen AG, Hilden, Germany), seguendo le istruzioni della casa fornitrice. In una prima fase è stata preparata una miscela di reazione i cui componenti sono: 1 µg di RNA, 2 µl di *gDNA Wipeout Buffer 7X* e acqua RNasi *free* per un volume finale di 14 µl. La miscela di reazione è stata posta nel termociclatore a 42°C per 5 minuti. Al termine dell'incubazione, sono stati aggiunti i seguenti reagenti: 4 µl di *Quantiscript RT Buffer 5X*, 1 µl di *RT Primer Mix* e 1 µl di *Quantiscript Reverse Transcriptase*. La miscela di reazione è stata posta nel termociclatore e incubata con il seguente profilo termico: 42°C per 30 minuti; 95°C per 3 minuti.

#### **3.4.5. Real-Time PCR (qRT-PCR)**

A differenza di una PCR *standard* in cui si va ad analizzare il prodotto amplificato alla fine dei cicli sottomettendo a migrazione elettroforetica la miscela di reazione, la qRT-PCR consente di seguire il corso dell'amplificazione del template, ottenendo un profilo che descrive l'andamento della reazione stessa e utilizzando metodiche di rilevazione basate

sulla fluorescenza. Per la valutazione semiquantitativa del cDNA è stata eseguita una qRT-PCR basata sull'uso del SYBR *Green*, un composto fluorescente che lega specificatamente il DNA a doppio filamento ma non a singolo filamento e, quando intercalato, emette fluorescenza molto intensamente. Una volta eccitato all'appropriata lunghezza d'onda (497 nm), il SYBR *Green* intercalato al DNA, emette energia luminosa a 520 nm, con un'intensità fino a 200 volte superiore rispetto alla sua forma solubile. Seguendo spettroscopicamente la fluorescenza della soluzione dove avviene l'amplificazione, si otterrà una curva che descrive l'andamento dell'amplificazione. La rilevazione della fluorescenza emessa dal SYBR *Green* viene effettuata al termine della fase di allungamento di ciascun ciclo di amplificazione, consentendo di monitorare l'andamento dell'intera reazione in tempo reale.

La metodica si basa sulla valutazione del Ct (*Threshold cycle*), ossia del ciclo soglia al quale la fluorescenza emessa dalla reazione diventa significativamente superiore rispetto al *background*. Tale valore corrisponde all'inizio della fase esponenziale di amplificazione ( $n$  copie prodotte =  $2^n$  cicli) della reazione ed è direttamente correlato alla quantità iniziale di template. Graficamente, il ciclo soglia corrisponde al punto in cui la curva di amplificazione interseca la linea di base (*Threshold line*).

Più basso è il numero dei Ct, minore è il numero di cicli affinché la fluorescenza superi la *baseline*, tanto più mRNA sarà presente nei campioni di interesse.

I livelli di espressione genica vengono riportati tramite quantificazione relativa che definisce per l'appunto una misura relativa dei livelli di espressione rispetto a una specifica condizione sperimentale.

La contemporanea amplificazione di un gene costitutivo (*Housekeeping gene*) consente di quantificare in base al  $\Delta Ct$  la differenza tra il gene *test* e il gene di riferimento. Mediante il calcolo del  $2^{-\Delta Ct}$  sarà possibile individuare il *Fold Change*, ossia l'incremento o il decremento del livello di espressione del gene testato rispetto al gene di riferimento. Poiché la specificità della qRT-PCR in SYBR *Green* dipende esclusivamente dalla quantità del prodotto di amplificazione (sequenza a doppio filamento), è importante verificare la specificità del prodotto di amplificazione. Ciò è reso possibile effettuando una curva di *melting* al termine della reazione, costruita rappresentando in grafico i valori di

emissione della molecola fluorescente in funzione della temperatura. Le curve di *melting* vengono costruite sottoponendo il prodotto di amplificazione a graduali incrementi di temperatura e misurando ad ogni step la fluorescenza emessa. Poiché la temperatura di *melting* di un amplicone (caratterizzata in questo caso dall'improvvisa interruzione della fluorescenza) è fortemente dipendente dalla sua composizione in nucleotidi, sarà possibile verificare la specificità del prodotto ottenuto.

Le sequenze dei *primer* impiegati sono riportate in **Tabella 8**.

Tutte le amplificazioni sono state eseguite nel termociclatore *QuantStudio 1 Real-Time PCR System* (Applied Biosystems, Waltham, USA), utilizzando un programma costituito da 40 cicli di due fasi ciascuno (95°C per 10 minuti e 60°C per 1 minuto).

Gene	Sequenza <i>Forward</i> 5' → 3'	Sequenza <i>Reverse</i> 5' → 3'
GAPDH	GAAGGTGAAGGTCGGAGTCAAC	TGGAAGATGGTGATGGGATTC
DCN	CCTCCGCTGTCAATGCCATCTTC	TGCAGGTCTAGCAGAGTTGTGTCAG
FN1	AGGCAGGCTCAGCAAATGGTTCAG	CTCTCGCAGTTAAAACCTCGGCTTC
FST	ACCGAGGAGGACGTGAATGACAAC	CGGCATTTTTTCCCAGGTCCACAG
IGF1	CATGTCCTCCTCGCATCTCTTCTACC	TCTCCAGCCTCCTTAGATCACAGCTC
MMP2	ATGTCGCCCCAAAACGGACAAAG	CGCATGGTCTCGATGGTATTCTGGTCAAG
PLAU	CCAAAATGCTGTGTGCTGCT	TGTCCTTCAGGGCACATCC
SMAD9	CTGGAGTGCTGTGAGTCCCATTG	AGGCTGCTGGAAAGAGTCAGGATAGG

**Tabella 8.** *Primer utilizzati nella reazione di Real-Time PCR e relative sequenze.*

### 3.5. Analisi bioinformatica

Una volta completato il sequenziamento in NGS, i dati vengono analizzati nella fase computerizzata di analisi bioinformatica *home-made*. I file di sequenziamento FASTQ generati sono stati caricati sulla piattaforma SOPHiA DDM®, versione 4 (SOPHiA GENETICS, Saint-Sulpice, Svizzera), piattaforma che permette di individuare, annotare e classificare le alterazioni genetiche (SNV, *Single Nucleotide Variant* e inserzioni/delezioni) attraverso la sua intelligenza artificiale.

Per ogni variante, la piattaforma specifica il nome, il tipo di variante somatica individuato (INDEL e SNV), il genoma di riferimento utilizzato per l'allineamento, il

numero di *reads*, la posizione cromosomica e la sua classificazione (altamente patogena - probabilmente patogena – di incerto significato e probabilmente benigna). Dopo la selezione delle varianti tramite i filtri della piattaforma SOPHiA DDM® (*coverage* > 1000X), sono stati applicati dei filtri manuali aggiuntivi per escludere dall'analisi gli SNP (*Single Nucleotide Polymorphism*) noti, le varianti localizzate in regioni introniche o non tradotte (UTR, *Untranslated Region*) e le varianti sinonime.

In questo studio, sono state prese in considerazione solo:

- 1) le varianti localizzate nelle regioni esoniche o nei siti di *splicing*;
- 2) le varianti con una *variant allele frequency* (VAF) superiore all'1%. (Per *variant allele frequency* si intende la misura della frequenza relativa di un allele su un *locus* genetico nella popolazione. Di solito si esprime come proporzione o percentuale);
- 3) le varianti considerate dai *tools* di predizione come altamente o probabilmente patogeniche.

### **3.5.1. Analisi differenziale dell'espressione genica e di arricchimento dei geni**

I dati di espressione genica presenti nel file DESeq2 sono stati utilizzati per eseguire l'analisi differenziale dei geni con Excel. L'analisi di espressione differenziale permette di definire i geni che hanno un valore di espressione significativamente diverso in due o più condizioni. Dal punto di vista della statistica, si sfrutta un test di verifica di ipotesi e, in questo caso, è stato adottato un *p-value adjusted* per ridurre la proporzione di falsi positivi. Inizialmente, i geni sono stati filtrati scegliendo un *p-value adjusted* < 0,05. In seguito, per individuare i geni *up-regolati* e *down-regolati* è stato adottato il logaritmo in base 2 del *Fold Change* ( $\log_2FC \geq 2$  e  $\leq -2$ , rispettivamente).

Successivamente, i geni identificati come differenzialmente espressi (DEG, *Differentially expressed genes*) sono stati raggruppati in una lista per valutare quali annotazioni funzionali fossero significativamente associate a tale gruppo di geni correlati e con livelli differenti di espressione (*Gene set enrichment analysis*). L'analisi di "arricchimento" è stata condotta usando il *tool* bioinformatico DAVID (<https://david.ncifcrf.gov/>). Dopo aver caricato sul sito la lista dei geni *up-* e *down-regolati*, è stata utilizzata la funzione

*Gene Functional Annotation Chart*, che raggruppa i termini più significativi all'interno della lista data come input per visualizzare i KEGG *pathway*.

### **3.6. Analisi statistica**

L'analisi statistica è stata condotta utilizzando il programma *GraphPad Prism Statistical PC* (GraphPad Software, San Diego, CA), un software di creazione di grafici scientifici 2D/3D polivalente che integra funzioni statistiche e permette di organizzare, analizzare e visualizzare tutti i tipi di dati scientifici provenienti da esperimenti. I dati sono presentati come media  $\pm$  deviazione standard (DS). I dati raggruppati sono stati confrontati con il test non parametrico Mann-Whitney U-test. Valori di  $p < 0.05$  sono stati considerati come soglia di significatività.

## 4. RISULTATI

### 4.1. Analisi del profilo mutazionale della popolazione di studio tramite *targeted-NGS*

La prima parte del progetto di tesi ha previsto la selezione dei pazienti da arruolare per lo studio. Attraverso la metodica di *targeted-NGS* (pannello *MYeloid Solution* di *SOPHiA GENETICS*), che permette di studiare i 30 geni più frequentemente mutati nelle neoplasie mieloidi, è stato caratterizzato il *pattern* mutazionale di 300 pazienti con SMD. Di questi 300 pazienti, di cui si disponeva di campioni appaiati di DNA e RNA, estratti al tempo della diagnosi, ne sono stati selezionati 32 in base alla presenza o meno di mutazioni nei geni implicati nel meccanismo di *splicing* e in base alla categoria di rischio IPSS-R di appartenenza (prevalentemente pazienti a basso rischio) (22 maschi, 10 femmine, età mediana alla diagnosi 71 anni, *range* 42-87 anni).

La scelta dei pazienti è stata effettuata in modo da consentire un'analisi statistica di potenza sufficiente a garantire una significativa caratterizzazione del profilo di espressione genica associato alla presenza di specifiche mutazioni nei geni dello *splicing*. Dei 32 pazienti selezionati, un totale di 25 pazienti mostravano almeno una mutazione in un gene dello *splicing*, mentre 7 pazienti erano *wild-type* per tutti i geni dello *splicing* del pannello utilizzato.

In particolare, 11 pazienti presentavano mutazioni nel gene *SRSF2* (34,4%), 9 nel gene *SF3B1* (28,1%), 7 nel gene *U2AF1* (21,9%) e 3 nel gene *ZRSR2* (9,4%), mentre 5 pazienti (15,6%) presentavano mutazioni in più geni di *splicing* (2 pazienti, UPN2-UPN50, con mutazioni nei geni *SRSF2* e *SF3B1*; 2 pazienti, UPN16-UPN17, con mutazioni nei geni *ZRSR2* e *U2AF1*; un paziente, UPN13, con mutazione nei geni *SRSF2* e *ZRSR2*).

Nella suddetta coorte di studio abbiamo identificato un totale di 82 mutazioni, con una mediana di 2 mutazioni per paziente (*range* 0-9), in 27 dei 32 pazienti studiati (84,4%).

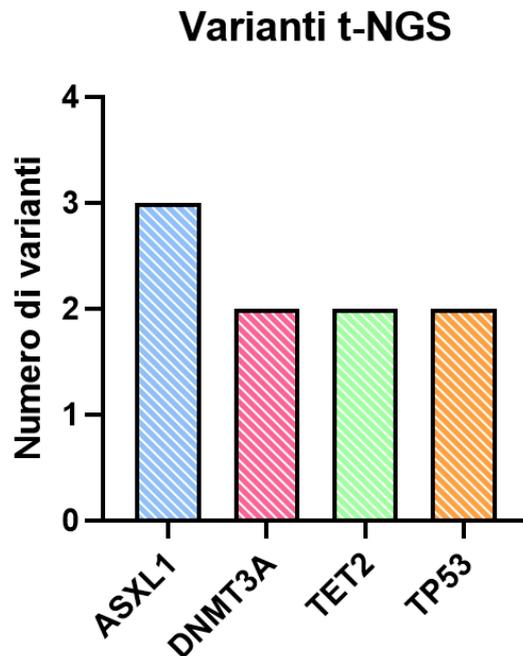
Come mostrato in **Tabella 9**, i geni più frequentemente mutati sono risultati essere *TET2* (9 pazienti su 32, 28,1%), *ASXL1* (7 pazienti su 32, 21,9%), *TP53* (5 pazienti su 32, 15,6%), *CBL* (4 pazienti su 32, 12,5%), *CSF3R* (4 pazienti su 32, 12,5%), *DNMT3A* (4 pazienti su 32, 12,5%) e *JAK2* (3 pazienti su 32, 9,4%).

La *variant allele frequency* delle mutazioni identificate va da un minimo di 2,9% per il gene *WT1* ad un massimo di 74,2% per il gene *ZRSR2*, con una VAF mediana del 37,95%.

Gene	UPN9	UPN10	UPN11	UPN12	UPN13	UPN42	UPN44	UPN46	UPN49	UPN50	UPN2	UPN3	UPN4	UPN6	UPN8	UPN1	UPN5	UPN7	UPN16	UPN15	UPN37	UPN43	UPN17	UPN14	UPN47	UPN41	UPN48
SRSF2																											
SF3B1																											
TET2			2	2												2											
ASXL1	2													2				3									
UZAF1																											
TP53					2																						
CBL																											
CSF3R																											
DNMT3A					2																						
JAK2																											
ZRSR2																											
IDH1																											
PTPN11																											
SETBP1																											
EZH2																											
IDH2																											
RUNX1																											
WT1																											
ABL																											
BRAF																											
CALR																											
CEBPA																											
ETV6																											
FLT3																											
HRAS																											
KIT																											
KRAS																											
MPL																											
NPM1																											
NRAS																											

**Tabella 9.** Distribuzione delle mutazioni identificate in 27 dei 32 pazienti affetti da sindrome mielodisplastica selezionati mediante t-NGS.

In 7 dei 32 pazienti abbiamo identificato più di una variante nello stesso gene. Nello specifico, 2 varianti nel gene *ASXL1* per il paziente UPN6 e UPN9 e 3 varianti per il paziente UPN7; 2 varianti nel gene *DNMT3A* per il paziente UPN13; 2 varianti nel gene *TET2* per i pazienti UPN1, UPN11 e UPN12; e, infine, nel paziente UPN13 sono state identificate 2 varianti nel gene *TP53* (**Figura 19**).



**Figura 19.** Geni con il maggior numero di varianti identificate nel singolo paziente.

#### 4.2. Analisi di WGS

La seconda parte del progetto ha previsto lo studio del genoma nella nostra casistica di pazienti, con lo scopo di identificare mutazioni di potenziale interesse ematologico associate ai geni dello *splicing*. Sono state prese in considerazione solo le varianti esoniche, non sinonime e nei siti di *splicing*. Sono state escluse dall'analisi le varianti nelle regioni introniche e le varianti esoniche sinonime.

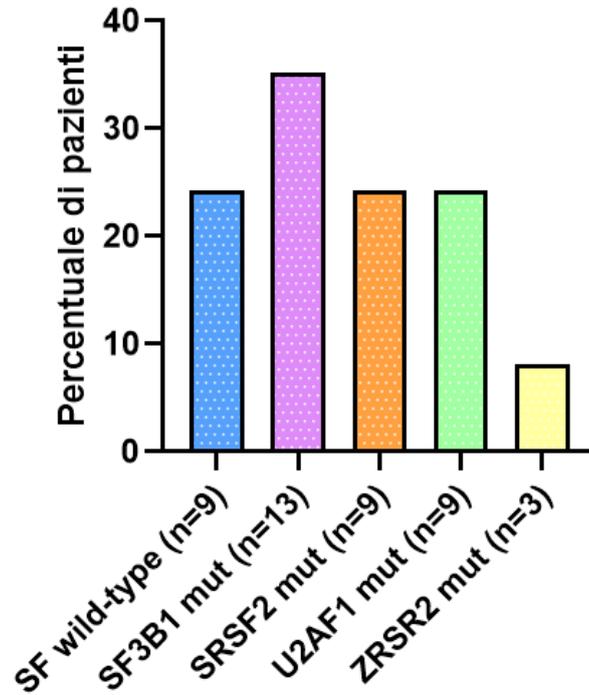
Il WGS è stato eseguito su un totale di 50 pazienti (32 pazienti appartenenti alla coorte di studio selezionata con t-NGS e 18 pazienti con profilo mutazionale non noto al fine di ampliare la coorte di studio).

L'analisi è stata effettuata con successo in 37 dei 50 pazienti selezionati, mentre per i restanti 13 (5 del primo campione e 8 del campione aggiuntivo) il DNA è risultato quantitativamente o qualitativamente inadeguato per la riuscita della stessa.

Il WGS ha permesso di identificare varianti non solo nei geni non compresi nel pannello *MYeloid Solution*, ma anche in esoni che non ricadono nelle regioni *target* dei geni del pannello stesso (vedi materiali e metodi).

A seguito dell'analisi di WGS, i pazienti che presentavano mutazioni in almeno uno dei geni di *splicing* sono risultati 28, mentre i restanti 9 pazienti sono risultati *wild-type* per

i tutti i fattori di *splicing*. I pazienti mutati, di cui 4 con 2 mutazioni e un paziente con 3 mutazioni, sono stati ulteriormente stratificati in pazienti *SF3B1* mutati (n=13), *SRSF2* mutati (n=9), *U2AF1* mutati (n=9) e *ZRSR2* mutati (n=3) (**Figura 20**).



**Figura 20.** Numero di pazienti mutati per categoria.

*SF: Splicing Factor*

Tra i 5 pazienti con più di una mutazione sono stati riscontrati 2 doppi mutanti per i geni *SF3B1* e *SRSF2*, 2 doppi mutanti per i geni *U2AF1* e *ZRSR2*, mentre 1 paziente presentava mutazioni in ben tre geni, *SF3B1*, *SRSF2* e *ZRSR2*.

Nonostante le mutazioni nei geni dello *splicing* vengano riportate in letteratura prevalentemente come mutualmente esclusive, studi di recente pubblicazione (Taylor et al., 2020) mostrano come sia possibile la coesistenza di mutazioni in più di uno di questi geni qualora esse coinvolgano codoni non frequentemente mutati (**Tabella 10**).

Codice paziente	Geni mutati e relativa mutazione		
UPN 2	SF3B1 E783K	SRSF2 P95H <sup>†</sup>	
UPN 13	SF3B1 P370T	SRSF2 P95H <sup>†</sup>	ZRSR2 R452C
UPN 16	U2AF1 Q157P <sup>†</sup>	ZRSR2 K405Rfs*Ter	
UPN 17	U2AF1 Q157P <sup>†</sup>	ZRSR2 W340Ter	
UPN 27	SF3B1 K700E <sup>†</sup>	SRSF2 R94H100del	

**Tabella 10.** Tabella che mostra la coesistenza di mutazioni nei geni dello *splicing*. Sono indicate con <sup>†</sup> le mutazioni canoniche per il gene considerato.

Dall'analisi è stato scelto di escludere i 3 pazienti con mutazione del gene *ZRSR2* in quanto di dubbia interpretazione, essendo le mutazioni localizzate lungo tutta la porzione codificante del gene.

Sono stati inoltre esclusi i 5 pazienti che presentavano mutazioni in più di uno dei geni di *splicing*, al fine di associare in maniera univoca il *pattern* mutazionale alla mutazione in quel particolare gene di *splicing*. Pertanto, il campione finale risulta costituito da 10 pazienti *SF3B1* mutati, 6 pazienti *SRSF2* mutati e 7 pazienti *U2AF1* mutati.

Come mostrato in **Tabella 11**, i geni più frequentemente mutati nei pazienti con mutazione del gene *SF3B1* (n=10) sono risultati essere *ACLY* (*ATP-citrate lyase*) (9 su 10 pazienti, 90%), *TTN* (*Titin*) (7 pazienti su 10, 70%), *FAM20C* (*FAM20C golgi associated secretory pathway kinase*) (6 pazienti su 10, 60%), *SH3TC1* (*SH3 domain and tetratricopeptide repeats 1*) e *SLC7A8* (*solute carrier family 7 member 8*) (5 pazienti su 10, 50%).

Per quanto riguarda i pazienti con mutazione *SRSF2* (n=6), sono risultati più frequentemente mutati *ARSD* (*Arylsulfatase D*) e *NLGN2* (*Neurologin-2*) (4 pazienti su 6, 67% circa), *ACLY*, *ASXL1*, *CACNA1H* (*Calcium voltage-gated channel subunit alpha1 H*), *FAM20C*, *ITGA10* (*integrin subunit alpha 10*), *KDM4B* (*Lysine demethylase 4B*), *PIEZO1* (*Piezo-type mechanosensitive ion channel component 1*) e *TET2* (3 pazienti su 6, 50%).

Nei pazienti con mutazione *U2AF1* (n=7) sono stati trovati più frequentemente mutati *ARSD*, *FAM20C*, *KDM4B* (5 pazienti su 7, 71% circa), *ACTN3* (*Actinin alpha 3*), *ATXN3* (*Ataxin-3*), *SLC7A8* (4 pazienti su 7, 57% circa).

Gene	UPN1	UPN3	UPN4	UPN5	UPN6	UPN7	UPN8	UPN22	UPN24	UPN33	UPN9	UPN10	UPN11	UPN12	UPN46	UPN49	UPN14	UPN15	UPN21	UPN29	UPN37	UPN43	UPN45	
ACLY																								
ACTN3																								
ARSD																								
ASXL1																								
ATXN3																								
CACNA1H																								
CUL7																								
DNAH2																								
DNAH3																								
DNAH7																								
FAM20C																								
ITGA10																								
KDM4B																								
LNP1																								
NLGN2																								
PIEZO1																								
SLC7A8																								
SH3TC1																								
TET2																								
TTN																								

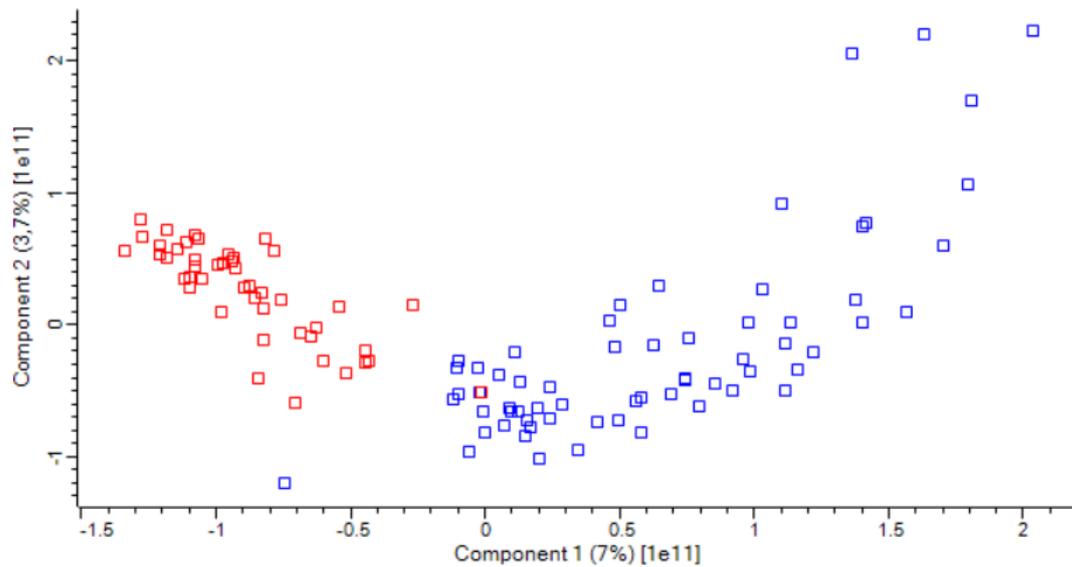
**Tabella 11.** Distribuzione delle mutazioni identificate nei pazienti *SF3B1* mutati (in viola), *SRSF2* mutati (in arancione), *U2AF1* mutati (in verde).

È interessante notare come i geni *FAM20C* e *KDM4B* siano frequentemente mutati in tutte e tre le categorie, mentre *ACLY* sia mutato ad alta frequenza nei pazienti *SF3B1* mutati (90%) e *SRSF2* mutati (50%), ma non nei pazienti con mutazione del gene *U2AF1*. In particolare, evidenze scientifiche recenti mostrano come le mutazioni in questo ultimo gene siano correlate a una deregolazione del metabolismo energetico cellulare nelle sindromi mielodisplastiche (Coelho-Silva et al., 2021). La validazione mediante metodiche di sequenziamento di Sanger e PCR allele-specifica è tuttora in corso.

### 4.3. Identificazione e analisi dei geni differenzialmente espressi

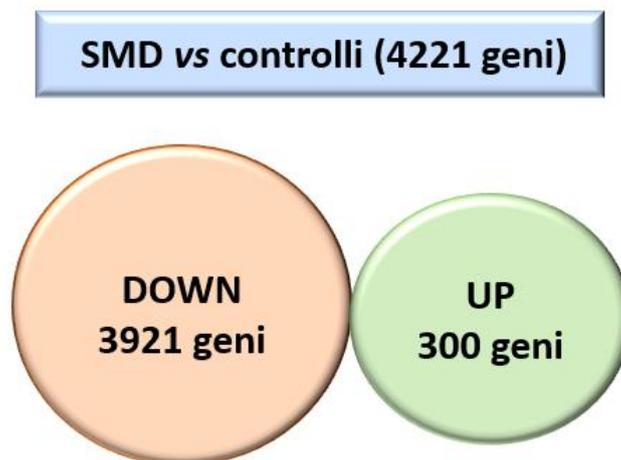
L'analisi di espressione genica, tramite *RNA-Seq*, è stata effettuata sul totale dei 50 pazienti con SMD arruolati per lo studio e su 64 controlli selezionati tra pazienti ospedalieri non affetti da patologie di carattere ematologico. È stato possibile ricavare i dati di espressione di 46 dei 50 pazienti con SMD.

Ottenuto dall'analisi di *RNA-Seq* il numero delle *reads* dei circa 26000 geni ad oggi annotati, queste sono state esportate per effettuare l'analisi delle componenti principali (*PCA, Principal Component Analysis*). I risultati della *PCA* mostrano come i pazienti affetti da sindrome mielodisplastica e i controlli si dispongano in *cluster* ben distinti (**Figura 21**). Un solo paziente (UP50) mostrava un profilo di espressione sovrapponibile a quello dei controlli ed è stato pertanto escluso dall'analisi statistica. Non sono state identificate caratteristiche cliniche specifiche che potessero giustificare un tale comportamento rispetto al gruppo di appartenenza.



**Figura 21.** Analisi delle componenti principali (PCA). In rosso sono mostrati i pazienti affetti da SMD ( $n=46$ ); in blu sono mostrati i campioni controllo ( $n=64$ ).

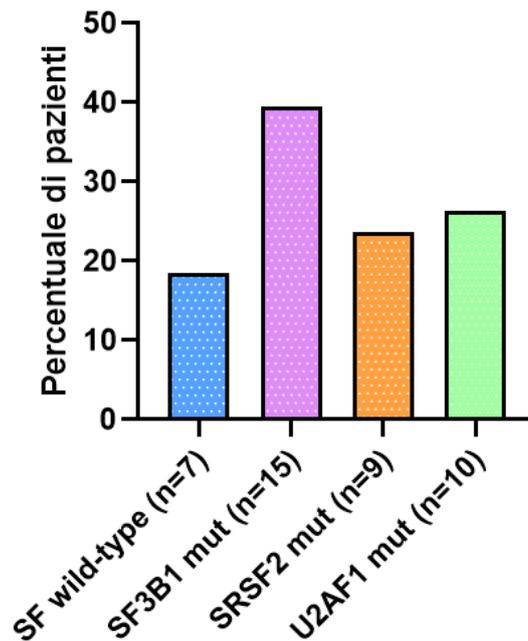
Il confronto tra pazienti con SMD e controlli ha permesso di identificare più di 4000 geni differenzialmente espressi (DEG). In particolare, nella coorte di pazienti affetti da sindrome mielodisplastica 300 geni sono risultati *up*-regolati, mentre 3921 geni sono risultati *down*-regolati (**Figura 22**).



**Figura 22.** Geni differenzialmente espressi nei pazienti affetti da sindrome mielodisplastica rispetto ai controlli.

L'analisi del trascrittoma, tramite il sequenziamento dell'RNA, e la successiva analisi differenziale dell'espressione genica hanno consentito l'identificazione di geni DE per le

tre sottocategorie di pazienti mutati per i geni di *splicing* considerati (*SF3B1*, *SRSF2* e *U2AF1*). I pazienti arruolati per l'analisi di espressione differenziale (n=38) sono stati stratificati in accordo con la presenza o meno di mutazioni nei geni dello *splicing*, identificate tramite sequenziamento del genoma (WGS) o dell'RNA (RNA-Seq) (**Figura 23**).

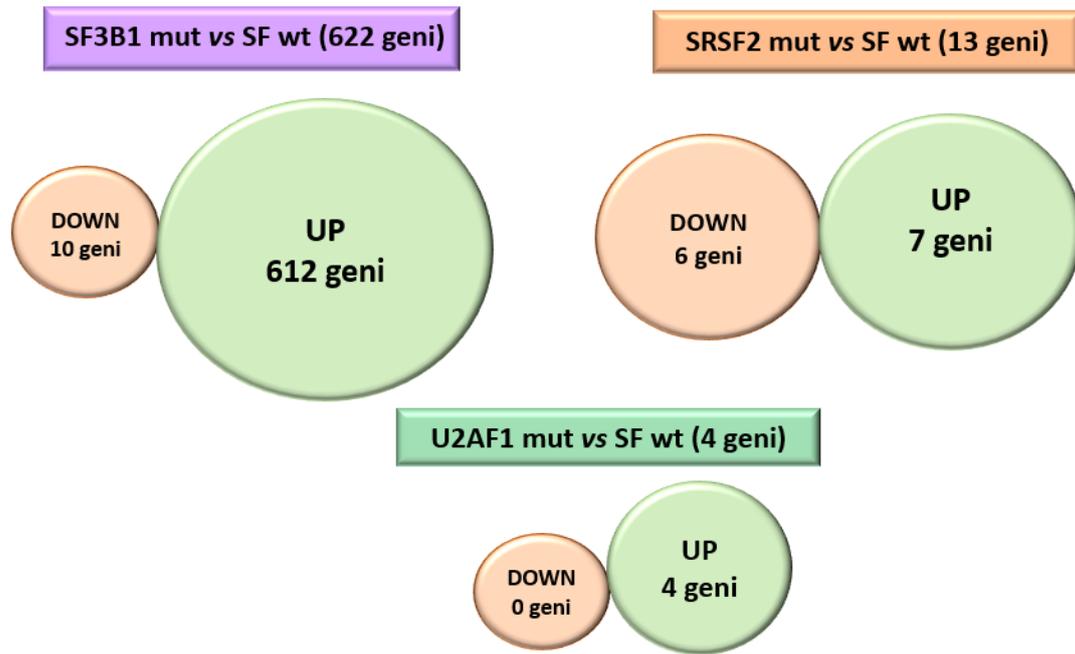


**Figura 23.** Distribuzione delle mutazioni nei pazienti affetti da sindrome mielodisplastica esaminati con successo per l'analisi di espressione differenziale. In grafico sono riportati anche i pazienti con più di una mutazione nei geni di *splicing* che sono stati categorizzati in base al valore di VAF.

SF: Splicing Factor

L'analisi di espressione differenziale è stata condotta ponendo a confronto i pazienti con mutazione in uno specifico gene dello *splicing* rispetto ai pazienti *wild-type* per tutti i geni dello *splicing* presi in considerazione (n=7).

Per i pazienti con mutazione del gene *SF3B1* (n=15) sono risultati differenzialmente espressi 622 geni e, di questi, 612 sono *up*-regolati nei pazienti mutati e 10 *down*-regolati. Per i pazienti con mutazione del gene *SRSF2* (n=9) sono risultati differenzialmente espressi 13 geni e, di questi, 7 sono *up*-regolati nei pazienti mutati e 6 *down*-regolati. Infine, nei pazienti con mutazione del gene *U2AF1* (n=10) sono risultati differenzialmente espressi 4 geni, tutti *up*-regolati nei pazienti mutati (**Figura 24**).



**Figura 24.** Geni deregolati in accordo con lo stato mutazionale.

SF wt: Splicing Factor wild-type

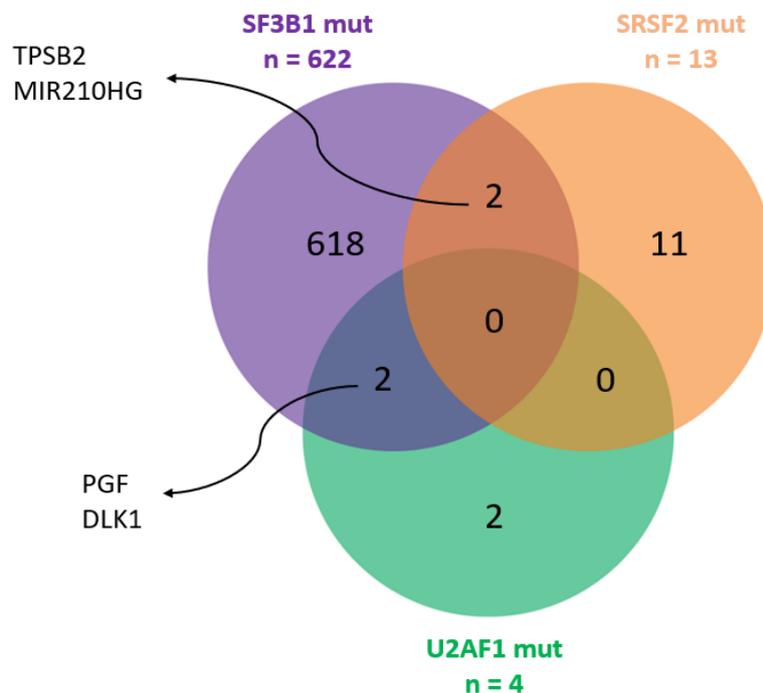
I geni differenzialmente espressi identificati in ciascuna delle tre diverse classi di pazienti mutati nei geni di *splicing* (*SF3B1*, *SRSF2* e *U2AF1*) presentano una modesta sovrapposizione.

Il diagramma di Venn (**Figura 25**) mostra come dei 622 geni DE nei pazienti con mutazione del gene *SF3B1*, solo 4 siano in comune con le altre due classi di pazienti. Nel dettaglio, i geni *TPSB2* e *MIR210HG* sono risultati differenzialmente espressi sia nei pazienti con mutazione del gene *SF3B1* che in quelli con mutazione di *SRSF2* e, allo stesso modo, i geni *PGF* e *DLK1* sono risultati DE nei pazienti mutati per *SF3B1* e in quelli con mutazione di *U2AF1*. I rimanenti 618 geni DE sono stati identificati esclusivamente nel gruppo dei pazienti *SF3B1* mutati.

Nei pazienti con mutazione del gene *SRSF2*, 2 dei 13 geni DE sono in comune con quelli dei pazienti *SF3B1* mutati, mentre non vi sono geni in comuni con i pazienti *U2AF1* mutati.

Nei pazienti con mutazione del gene *U2AF1*, 2 dei 4 geni DE sono in comune con quelli dei pazienti *SF3B1* mutati, mentre non vi sono geni in comune con i pazienti *SRSF2* mutati.

Nessun gene è risultato deregolato in tutte e tre le classi di pazienti con geni di *splicing* mutati.



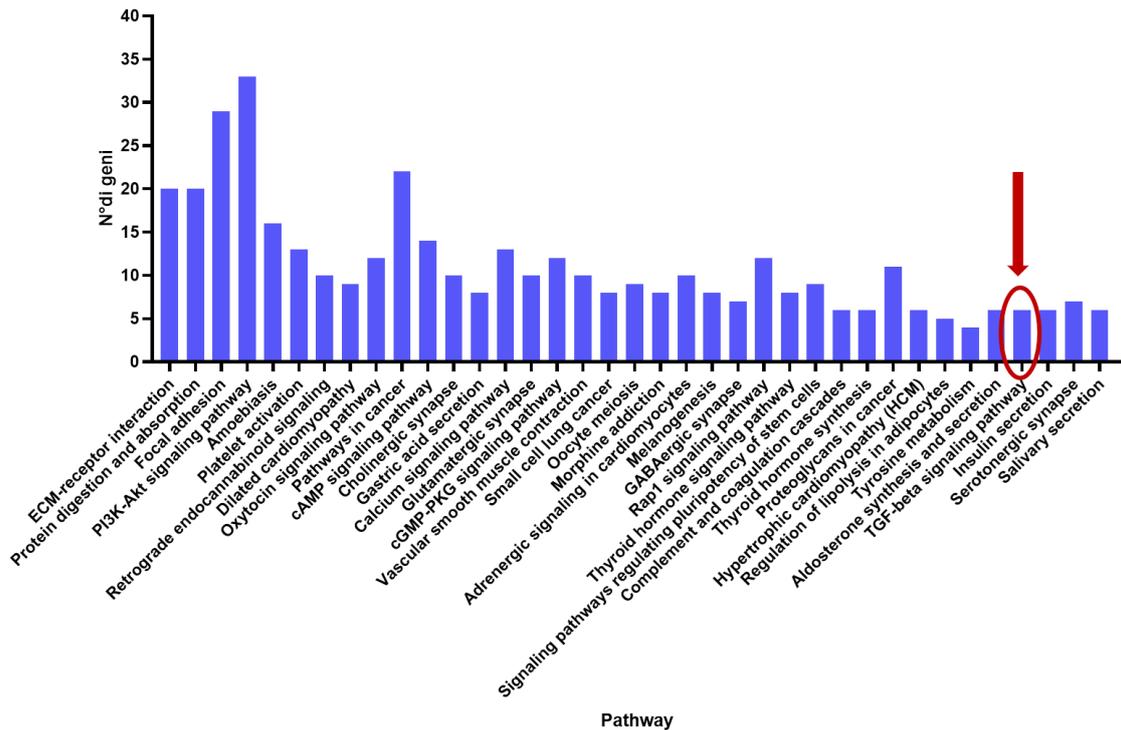
**Figura 25.** Diagramma di Venn in cui è mostrata la sovrapposizione dei geni deregolati identificati nei pazienti *SF3B1* mutati, *SRSF2* mutati, *U2AF1* mutati.

Alla fase di identificazione dei geni differenzialmente espressi ha fatto seguito l'analisi di "arricchimento", che ha permesso di raggruppare in *pathway* i geni identificati come differenzialmente espressi.

L'analisi è stata portata a termine sfruttando la funzione *Gene Functional Annotation* del tool DAVID e fornendo in input al sistema la lista dei geni DE per ognuno dei tre sottogruppi di pazienti considerati.

Per quanto riguarda i pazienti con mutazione del gene *SRSF2* e del gene *U2AF1*, il sistema non è stato in grado di raccogliere in *cluster* i geni presenti nella lista in base ai *cut-off* prescelti ( $padj < 0,05$  e  $\log_2FC \geq 2$  e  $\leq -2$ ). Mentre, per quanto riguarda i pazienti mutati per *SF3B1*, DAVID ha restituito 37 *pathway* (**Figura 26**).

Nella lista di geni DE sono stati individuati principalmente geni coinvolti nell'interazione con la matrice extracellulare (n=20), nella digestione delle proteine e assorbimento (n=20), nell'adesione focale (n=29) e nella via di trasduzione del segnale PI3K/Akt (n=33).



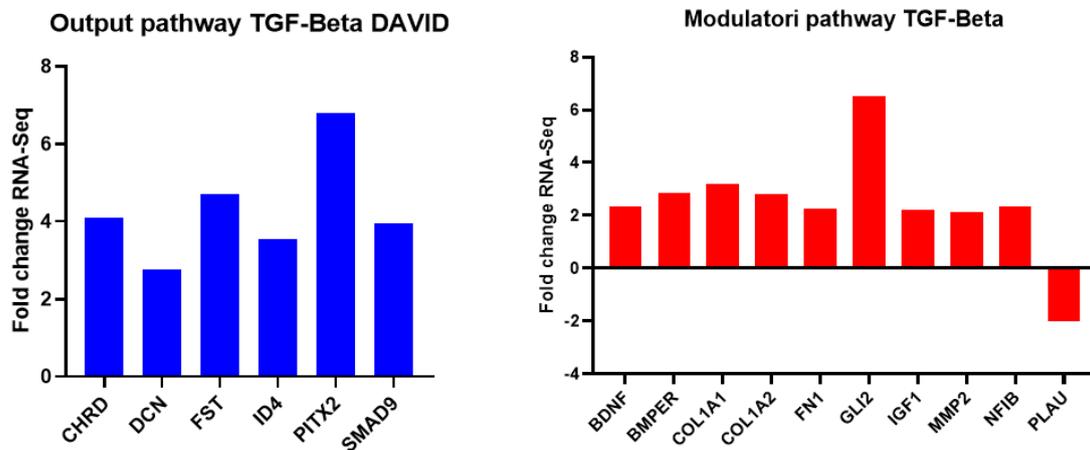
**Figura 26.** Pathway per cui sono arricchiti i pazienti con mutazione del gene *SF3B1* (DAVID).

Nonostante i *pathway* sopracitati siano indicati da DAVID come i più statisticamente significativi in base al valore di *p-value* (2,3E-12; 2,9E-12; 3,0E-12; 2,0E-9 rispettivamente) ed includano un numero cospicuo di geni DE, in questo studio abbiamo scelto di approfondire gli aspetti relativi alla via di *signaling* del TGF- $\beta$ . Il motivo di tale scelta risiede nel fatto che, ad oggi, i pazienti affetti da sindrome mielodisplastica, con mutazione del gene *SF3B1* e non responsivi al trattamento con eritropoietina, la terapia *gold standard* per questa categoria di pazienti, vengono sottoposti a trattamento con il *Luspatercept*, una proteina di fusione ricombinante utilizzata nella pratica clinica che agisce sulla trasduzione del segnale del TGF- $\beta$ .

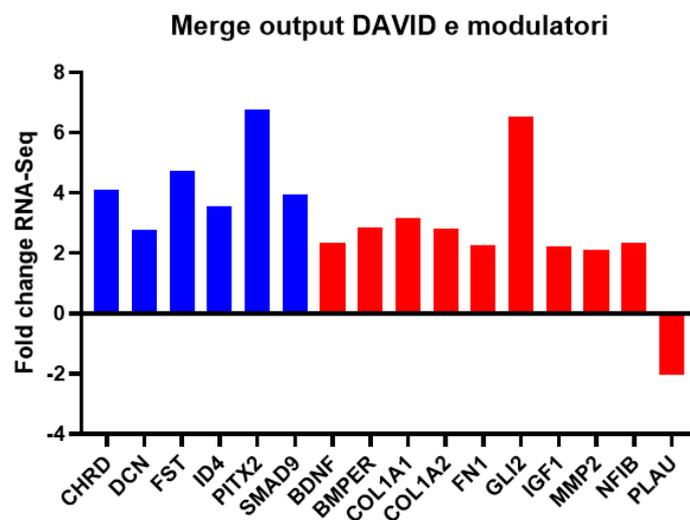
DAVID ha restituito come arricchiti per questo *pathway* 6 geni, *CHRD* (*Chordin*), *DCN* (*Decorin*), *SMAD9* (*SMAD family member 9*), *FST* (*Follistatin*), *ID4* (*Inhibitor of DNA binding 4*, *HLH protein*) e *PITX2* (*Paired like homeodomain 2*).

Sulla base dei dati annotati in letteratura, abbiamo individuato tra i nostri geni DE altri 10 geni riconducibili al *pathway* del TGF- $\beta$ , *BDNF* (*Brain-derived neurotrophic factor*), *BMPER* (*BMP-binding endothelial regulator*), *COL1A1* (*Collagen alpha-1(I) chain*), *COL1A2* (*Collagen alpha-2(I) chain*), *FN1* (*Fibronectin*), *GLI2* (*GLI family zinc finger 2*),

*IGF1* (Insulin-like growth factor 1), *MMP2* (Matrix metalloproteinase 2), *NFIB* (Nuclear factor 1 B) e *PLAU* (Plasminogen activator, urokinase) (**Figura 27**).



**Figura 27.** Geni deregolati appartenenti al pathway del TGF- $\beta$  ottenuti tramite il tool bioinformatico DAVID (A sinistra). Modulatori del pathway del TGF- $\beta$  annotati in letteratura (A destra).



**Figura 28.** Merge dei geni deregolati.

Dalla lista di geni così ottenuta (**Figura 28**), abbiamo scelto di studiare 7 geni differenzialmente espressi nella coorte di pazienti mutati per *SF3B1*, ovvero *DCN*, *FST*, *SMAD9*, *FN1*, *IGF1*, *MMP2* e *PLAU*.

#### 4.4. Validazione dei risultati di RNA-Seq

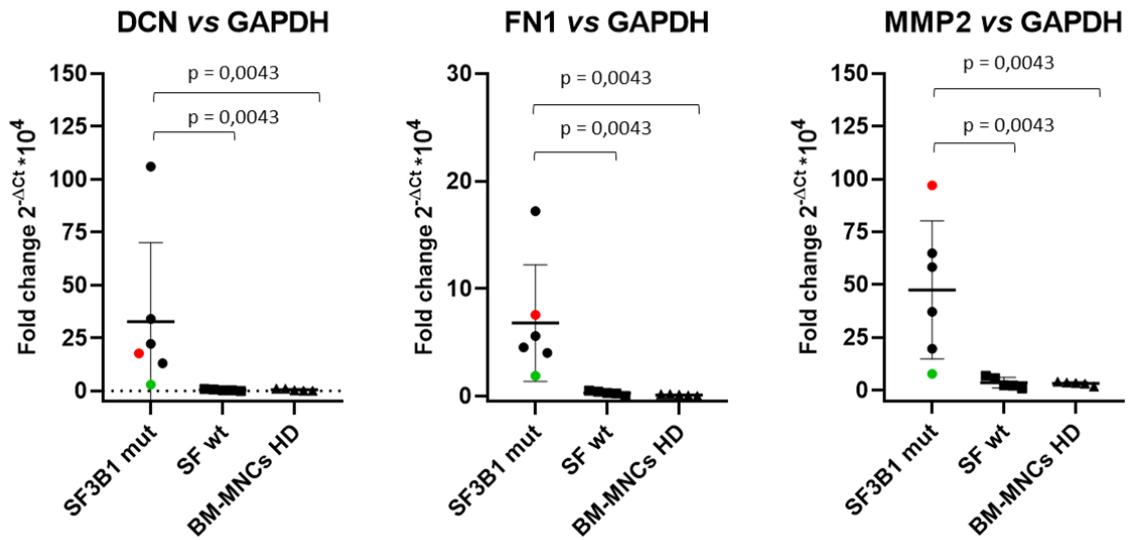
Per validare i risultati dell'RNA-seq, i 7 geni differenzialmente espressi selezionati sono stati studiati mediante Real-Time PCR, al fine di valutarne l'espressione attraverso un secondo metodo sperimentale.

L'analisi di qRT-PCR ha confermato i dati di espressione ottenuti tramite l'analisi del trascrittoma per tutti i geni analizzati.

I livelli di espressione dei geni considerati sono risultati alterati nelle BM-MNCs isolate dai pazienti con mutazione del gene *SF3B1* (n=6) rispetto ai pazienti SF *wild-type* (n=5) e ai controlli sani (n=5). In particolare, nei pazienti *SF3B1* mutati è stato osservato un aumento significativo dei livelli di espressione di *DCN*, *FST*, *SMAD9*, *FN1*, *IGF1* e *MMP2*. Al contrario, è stata apprezzata una diminuzione dei livelli di espressione di *PLAU*.

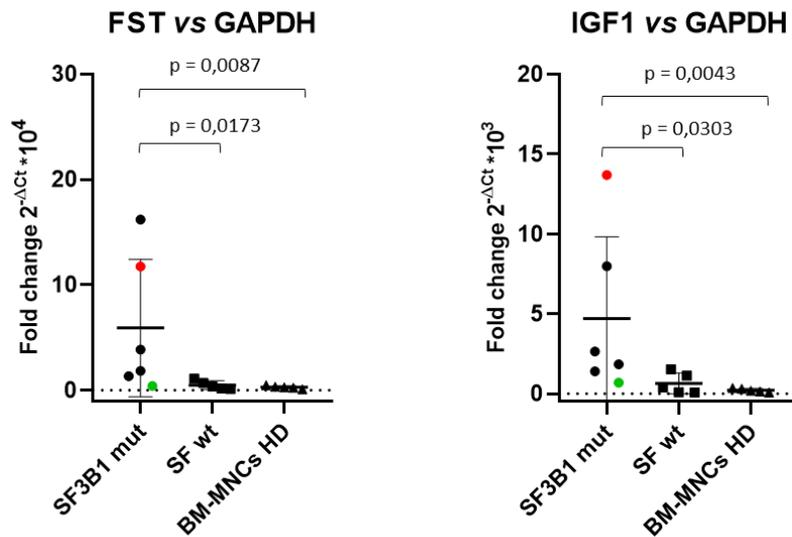
*DCN*, *FN1* e *MMP2* risultano *up-regolati* nei pazienti *SF3B1* mutati sia rispetto ai pazienti SF *wild-type* [FC (*fold change*) = 59,55, FC = 22,17, FC = 13,07 rispettivamente; *p-value* = 0,0043 per tutti i test] sia rispetto ai donatori sani di midollo osseo (FC = 53,92, FC = 51,67, FC = 14,55 rispettivamente; *p-value* = 0,0043 per tutti i test) (**Figura 29**).

*FST* risulta *up-regolato* nei pazienti *SF3B1* mutati sia rispetto ai pazienti SF *wild-type* (FC = 11,93; *p-value* = 0,0173) sia rispetto ai donatori sani di midollo osseo (FC = 19,75; *p-value* = 0,0087). *IGF1* risulta *up-regolato* nei pazienti *SF3B1* mutati sia rispetto ai pazienti SF *wild-type* (FC = 7,86; *p-value* = 0,0303) sia rispetto ai donatori sani di midollo osseo (FC = 22,08; *p-value* = 0,0043) (**Figura 30**). *SMAD9* risulta *up-regolato* nei pazienti *SF3B1* mutati sia rispetto ai pazienti SF *wild-type* (FC = 9,79; *p-value* = 0,0303) sia rispetto ai donatori sani di midollo osseo (FC = 19,55; *p-value* = 0,0087). *PLAU* risulta *down-regolato* nei pazienti *SF3B1* mutati rispetto ai pazienti SF *wild-type* (FC = -4,96; *p-value* = 0,0087), mentre non si apprezza una differenza statisticamente significativa rispetto ai donatori sani di midollo osseo (FC = -1,13; *p-value* = 0,6623) (**Figura 31**).



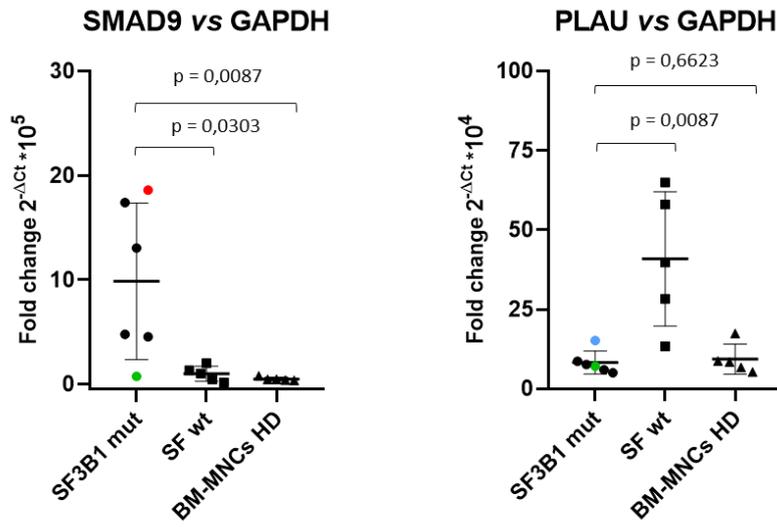
**Figura 29.** Profilo di espressione dei geni DCN, FN1 e MMP2 nelle BM-MNCs dei pazienti SF3B1 mutati, dei pazienti SF wild-type e dei donatori sani di midollo osseo.

In colori diversi, la mutazione SF3B1 K700E (in nero); SF3B1 P370T (in rosso); SF3B1 G740E (in verde).



**Figura 30.** Profilo di espressione dei geni FST e IGF1 nelle BM-MNCs dei pazienti SF3B1 mutati, dei pazienti SF wild-type e dei donatori sani di midollo osseo.

In colori diversi, la mutazione SF3B1 K700E (in nero); SF3B1 P370T (in rosso); SF3B1 G740E (in verde).



**Figura 31.** Profilo di espressione dei geni SMAD9 e PLAU nelle BM-MNCs dei pazienti SF3B1 mutati, dei pazienti SF wild-type e dei donatori sani di midollo osseo.  
 In colori diversi, la mutazione SF3B1 K700E (in nero); SF3B1 P370T (in rosso); SF3B1 G740E (in verde); SF3B1 H662Q (in blu).

## 5. DISCUSSIONE

Il termine “sindrome mielodisplastica” identifica un gruppo di disordini mieloidi clonali che contraddistinguono la cellula staminale ematopoietica, risultando in un compromesso processo di emopoiesi, citopenie periferiche e un aumentato rischio di progressione leucemica (Shallis et al., 2018).

Le sindromi mielodisplastiche presentano alterazioni citogenetiche in oltre il 50% dei casi. Tali alterazioni hanno una forte valenza prognostica, in quanto, insieme alle citopenie e alla quota di blasti midollari, consentono di stratificare le SMD in varie categorie di rischio, in accordo con lo schema di punteggio IPSS-R (Greenberg et al., 2012).

Grazie allo sviluppo e al perfezionamento delle metodiche di sequenziamento massivo, l'identificazione delle mutazioni somatiche sta assumendo un ruolo sempre più preponderante nell'inquadramento diagnostico e prognostico dei pazienti affetti da SMD. Si tratta per lo più di mutazioni che interessano modificatori epigenetici, quali *ASXL1*, *DNMT3A* e *TET2*, e geni coinvolti nello *splicing* dell'RNA, quali *SF3B1*, *SRSRF2*, *U2AF1* (Papaemmanuil et al., 2013; Voso et al., 2013).

Una percentuale significativa delle mutazioni identificate, infatti, riguarda geni che codificano per proteine che regolano l'espressione genica, mutazioni spesso associate ad una prognosi sfavorevole. Ciò trova conferma nel fatto che si tratta di geni che determinano una aumentata metilazione del DNA, che comporta un maggior grado di compattamento della cromatina ed è associata a silenziamento genico.

Quanto al processo di *splicing*, esso risulta implicato nella patogenesi delle sindromi mielodisplastiche e assolve al ruolo di potenziale bersaglio terapeutico in patologie oncoematologiche (Inoue et al., 2016). Lo *splicing* prevede, per definizione, la rimozione degli introni e il ricongiungimento degli esoni nei precursori dei trascritti di RNA messaggero (pre-mRNA).

Le mutazioni identificate nei pazienti affetti da SMD alterano la funzione dei fattori di *splicing*, determinando difetti nel processo e portando alla produzione di varianti con una diversa o compromessa attività biologica, con conseguente alterazione generalizzata del profilo di espressione genica (Matera and Wang, 2014).

Il profilo mutazionale dei pazienti *SF3B1* mutati, *SRSF2* mutati e *U2AF1* mutati è estremamente eterogeneo sia per quanto riguarda le mutazioni identificate sia per quanto riguarda la loro frequenza.

Grazie all'analisi di WGS, abbiamo identificato una stretta associazione tra le mutazioni nei geni dello *splicing* e i geni *ACLY*, *FAM20C* e *KDM4B* su cui abbiamo focalizzato la nostra attenzione.

*ACLY* è risultato mutato nel 90% dei pazienti con mutazione del gene *SF3B1* e nel 50% dei pazienti con mutazione del gene *SRSF2*, mentre non sono state identificate mutazioni nei pazienti *U2AF1* mutati.

Si tratta di un gene che non è mai stato identificato come mutato ad alta frequenza nelle SMD e normalmente non fa parte dei pannelli impiegati per lo studio delle neoplasie mieloidi. La forza del nostro progetto risiede proprio nell'utilizzo di metodiche, come il WGS, che ci hanno permesso di individuare nuovi geni potenzialmente implicati nella patogenesi delle mielodisplasie.

*ACLY* codifica per un enzima che catalizza la sintesi dell'acetil-CoA citosolico, un catabolita del glucosio, che viene processato nel ciclo dell'acido citrico. È noto che le cellule tumorali presentano un metabolismo energetico alterato, che risulta nell'utilizzo preferenziale della glicolisi aerobica, il cosiddetto "effetto Warburg", a scapito del ciclo di Krebs e della fosforilazione ossidativa (Coelho-Silva et al., 2021).

Avendo constatato un'elevata frequenza di mutazione del gene *ACLY* in due delle categorie di pazienti mutati per i geni dello *splicing*, stiamo mettendo a punto un saggio di PCR allele-specifica al fine di verificare l'effettiva presenza della mutazione.

La scelta di tale metodica risiede nel fatto che il sequenziamento di Sanger, *gold standard* per la validazione di mutazioni ad alta frequenza allelica, potrebbe non riuscire a rilevare, a causa del suo limite di sensibilità, la mutazione del gene *ACLY*, identificata tramite WGS con una VAF mediana del 12,4%. Al contrario, abbiamo già confermato la presenza dell'unica mutazione missenso identificata, che presentava una VAF del 45,6%. *FAM20C* e *KDM4B*, che sono risultati mutati in tutte e tre le categorie di pazienti considerati, sono geni coinvolti rispettivamente nel differenziamento delle cellule ematopoietiche e nella metilazione del DNA. Un alterato *pattern* di metilazione del DNA

è noto per essere associato con le sindromi mielodisplastiche, soprattutto con le SMD ad alto rischio, che richiedono un trattamento con agenti ipometilanti, quali azacitidina e decitabina (Falconi et al., 2019; Fenaux et al., 2009).

Grazie all'analisi di espressione genica differenziale e all'analisi di "arricchimento", abbiamo ottenuto una lista di *pathway* deregolati nelle SMD per i pazienti *SF3B1* mutati. La maggior parte dei *pathway* identificati sono responsabili di processi ampiamente studiati e collegati a diversi aspetti delle mielodisplasie (Pellagatti et al., 2010).

In questo studio abbiamo scelto di approfondire gli aspetti relativi al *pathway* del TGF- $\beta$  in considerazione della recente approvazione di una nuova terapia a bersaglio molecolare basata sull'utilizzo di una proteina ricombinante (*Luspatercept*), capace di modularne la trasduzione del segnale, ripristinando una normale eritropoiesi.

Il *Luspatercept* agisce su vari membri appartenenti alla famiglia del TGF- $\beta$ , quali GDF11 (*Growth differentiation factor 11*), GDF8 e activina B, determinandone l'inibizione e il blocco del segnale mediato dalle proteine SMAD (Fenaux et al., 2020).

In considerazione di ciò, abbiamo analizzato l'espressione di geni modulatori del TGF- $\beta$ , al fine di confermare i risultati dell'analisi del trascrittoma.

I geni *DCN* e *FST*, noti inibitori di fattori attivatori della via di *signaling* del TGF- $\beta$  (GDF11, activine e TGF- $\beta$ ) (Jamaiyar et al., 2017), sono risultati *up*-regolati nei pazienti *SF3B1* mutati rispetto ai pazienti *wild-type* per i geni dello *splicing* e rispetto ai controlli. L'aumentata espressione di questi geni in combinazione con l'effetto inibitorio del *Luspatercept* stesso potrebbe esercitare un effetto additivo e sinergico nei confronti del TGF- $\beta$ . L'effetto inibitorio dei geni *DCN* e *FST* su molecole in grado di determinare una maturazione eritroide alterata (GDF11, GDF8 e activine), tramite la trasduzione del segnale del TGF- $\beta$ , potrebbe giustificare la riduzione o lo spegnimento del *pathway* SMAD-mediato e il ripristino di un normale processo di differenziamento eritroide.

L'analisi di espressione ha evidenziato una deregolazione dei geni *MMP2* e *PLAU* nei pazienti con mutazione del gene *SF3B1* rispetto ai pazienti *SF wild-type* e ai controlli (*up*-regolazione e *down*-regolazione, rispettivamente). È noto che *MMP2*, come alcuni altri membri della famiglia delle metalloproteasi della matrice, funge da tramite per le

integrine per l'attivazione *in vivo* del TGF- $\beta$ . Tuttavia, i nostri dati mostrano come il profilo di espressione genica relativo alle integrine non sia deregolato nei pazienti *SF3B1* mutati (dati non mostrati). Anche nel caso del gene *PLAU*, la *down*-regolazione osservata potrebbe rendersi responsabile della mancata attivazione della plasmina, una idrolasi che promuove l'attivazione del TGF- $\beta$ . La deregolazione osservata nei geni *MMP2* e *PLAU* non ci permette, ad oggi, di confermare, in accordo con i dati presenti in letteratura, un loro ruolo nel mediare l'attivazione della proteina TGF- $\beta$  matura e, di conseguenza, del *pathway* SMAD-mediato. Inoltre, studi di recente pubblicazione (Fenaux et al., 2020; Parisi et al., 2021) mostrano come il *Luspatercept* non sia in grado di spegnere il *signaling* promosso dai ligandi TGF- $\beta$ 1, TGF- $\beta$ 2 e TGF- $\beta$ 3 nonostante questi, e in particolar modo il TGF- $\beta$ 1, siano indicati come regolatori negativi della maturazione eritroide nella sua fase terminale.

Un altro gene deregolato oggetto di studio è *SMAD9*, fattore trascrizionale attivato dalla proteina morfogenetica dell'osso (BMP), noto per essere uno dei regolatori della produzione di epcidina, un ormone responsabile del riciclo del ferro. Un tratto clinico di rilievo delle sindromi mielodisplastiche, in particolar modo delle SMD con sideroblasti ad anello e mutazione del gene *SF3B1*, è l'anemia e la conseguente dipendenza da trasfusione che, a lungo andare, comporta un sovraccarico di ferro secondario. Un aumento dei livelli di ferro potrebbe determinare, per un meccanismo di compensazione omeostatica, un aumento della sintesi di epcidina (soprattutto nei pazienti trasfusi), in linea con l'aumentata espressione del gene *SMAD9* osservata nei nostri pazienti.

*IGF1* regola la capacità del TGF- $\beta$  di indurre l'apoptosi, bloccando questo processo attraverso il *signaling* PI3K-Akt e determinando in tal modo la sopravvivenza di cellule tumorali di origine ematopoietica. L'*up*-regolazione di questo gene, dunque, giustificerebbe una ridotta apoptosi, che assolve a un ruolo chiave nella patogenesi delle sindromi mielodisplastiche e nella progressione a leucemia acuta mieloide. Infatti, un blocco del processo apoptotico determinerebbe un aumento del grado di sopravvivenza delle cellule neoplastiche, portando alla successiva espansione di cellule precursori anormali (Vishwamitra et al., 2017).

Infine, il gene *FN1*, che codifica per una proteina della matrice extracellulare con funzione di *scaffold* insieme al collagene e ad altre glicoproteine, è risultato overespresso nei pazienti *SF3B1* mutati. Partendo dalle evidenze scientifiche che suggeriscono come la deposizione di fibre e di collagene nel midollo osseo ematopoietico possa determinare fibrosi, l'*up*-regolazione di *FN1* sembrerebbe indirizzarci verso questo aspetto della patologia. Tuttavia, un'eccessiva deposizione di fibre a livello midollare risulterebbe in contrasto con l'iperproliferazione dei precursori eritroidi immaturi e un midollo osseo a cellularità aumentata, una peculiarità delle SMD. Sebbene i dati siano ancora preliminari e necessitino di una validazione in una coorte prospettica indipendente, il nostro studio ha permesso di dare risalto alla deregolazione di alcuni dei geni implicati nel *signaling* del TGF- $\beta$  e ha mostrato un significativo aumento dei livelli di espressione di *DCN*, *FST*, *SMAD9*, *FN1*, *IGF1*, *MMP2* e una significativa diminuzione del livello di espressione di *PLAU* nei pazienti con mutazione del gene *SF3B1* rispetto ai pazienti *wild-type* per i fattori di *splicing*, in accordo con i dati del trascrittoma.

In conclusione, il mio progetto di tesi mostra come i pazienti affetti da SMD presentino un profilo di espressione alterato ed estremamente eterogeneo. Tale eterogeneità appare riconducibile alla presenza di specifiche mutazioni nei geni implicati nel processo di *splicing*. In particolare, i pazienti con mutazioni del gene *SF3B1* mostrano una deregolazione dei geni appartenenti al *pathway* del TGF- $\beta$ , suggerendo che alcuni di essi possano essere utilizzati, in futuro, come bersagli molecolari con l'obiettivo del ripristino della corretta eritropoiesi e del raggiungimento dell'indipendenza da trasfusione.

## 6. BIBLIOGRAFIA

1. Annes, J.P., Munger, J.S., and Rifkin, D.B. (2003). Making sense of latent TGF $\beta$  activation. *J. Cell Sci.* *116*, 217–224.
2. Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* *127*, 2391–2405.
3. Avvisati, G. (2019). *Ematologia di Mandelli (Piccin)*.
4. Bataller, A., Montalban-Bravo, G., Soltysiak, K.A., and Garcia-Manero, G. (2019). The role of TGF $\beta$  in hematopoiesis and myeloid disorders. *Leukemia* *33*, 1076–1089.
5. Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R.L., Neuberg, D., et al. (2011). Clinical Effect of Point Mutations in Myelodysplastic Syndromes. *N. Engl. J. Med.* *364*, 2496–2506.
6. Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A.G., Gralnick, H.R., and Sultan, C. (1982). Proposals for the classification of the myelodysplastic syndromes. *Br. J. Haematol.* *51*, 189–199.
7. Blank, U., and Karlsson, S. (2015). TGF- $\beta$  signaling in the control of hematopoietic stem cells. *Blood* *125*, 3542–3550.
8. Boulwood, J., Pellagatti, A., Cattani, H., Lawrie, C.H., Giagounidis, A., Malcovati, L., Porta, M.G.D., Jädersten, M., Killick, S., Fidler, C., et al. (2007). Gene expression profiling of CD34 + cells in patients with the 5q- syndrome. *Br. J. Haematol.* *139*, 578–589.
9. Buscarlet, M., Provost, S., Zada, Y.F., Barhdadi, A., Bourgoin, V., Lépine, G., Mollica, L., Szuber, N., Dubé, M.P., and Busque, L. (2017). DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* *130*, 753–762.
10. Cazzola, M. (2020). Myelodysplastic Syndromes. *N. Engl. J. Med.* *383*, 1358–1374.

11. Cazzola, M., Della Porta, M.G., and Malcovati, L. (2013). The genetic basis of myelodysplasia and its clinical relevance. *Blood* 122, 4021–4034.
12. Cedena, M.T., Rapado, I., Santos-Lozano, A., Ayala, R., Onecha, E., Abaigar, M., Such, E., Ramos, F., Cervera, J., Díez-Campelo, M., et al. (2017). Mutations in the DNA methylation pathway and number of driver mutations predict response to azacitidine in myelodysplastic syndromes. *Oncotarget* 8, 106948–106961.
13. Chen, J., Kao, Y.R., Sun, D., Todorova, T.I., Reynolds, D., Narayanagari, S.R., Montagna, C., Will, B., Verma, A., and Steidl, U. (2019). Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nat. Med.* 25, 103–110.
14. Clarke, M.F. (2019). Clinical and Therapeutic Implications of Cancer Stem Cells. *N. Engl. J. Med.* 380, 2237–2245.
15. Coelho-Silva, J.L., Silveira, D.R.A., Pereira-Martins, D.A., Rojas, C.A.O., Lucena-Araujo, A.R., Rego, E.M., Machado-Neto, J.A., Bendit, I., Rocha, V., and Traina, F. (2021). Molecular-Based Score inspired on metabolic signature improves prognostic stratification for myelodysplastic syndrome. *Sci. Rep.* 11, 1–11.
16. Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della Valle, V., Couronné, L., Scourzic, L., Chesnais, V., et al. (2012). Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* 119, 3211–3218.
17. Dolatshad, H., Pellagatti, A., Liberante, F.G., Llorian, M., Repapi, E., Steeples, V., Roy, S., Scifo, L., Armstrong, R.N., Shaw, J., et al. (2016). Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia* 30, 2322–2331.
18. Falconi, G., Fabiani, E., Piciocchi, A., Criscuolo, M., Fianchi, L., Lindfors Rossi, E.L., Finelli, C., Cerqui, E., Ottone, T., Molteni, A., et al. (2019). Somatic mutations as markers of outcome after azacitidine and allogeneic stem cell transplantation in higher-risk myelodysplastic syndromes. *Leukemia* 33, 785–790.
19. Fenaux, P., Mufti, G.J., Hellstrom-Lindberg, E., Santini, V., Finelli, C., Giagounidis,

- A., Schoch, R., Gattermann, N., Sanz, G., List, A., et al. (2009). Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol.* *10*, 223–232.
20. Fenaux, P., Platzbecker, U., Mufti, G.J., Garcia-Manero, G., Buckstein, R., Santini, V., Díez-Campelo, M., Finelli, C., Cazzola, M., Ilhan, O., et al. (2020). Luspatercept in Patients with Lower-Risk Myelodysplastic Syndromes. *N. Engl. J. Med.* *382*, 140–151.
21. Fontana, L., Chen, Y., Prijatelj, P., Sakai, T., Fässler, R., Sakai, L.Y., and Rifkin, D.B. (2005). Fibronectin is required for integrin  $\alpha$ v $\beta$ 6-mediated activation of latent TGF- $\beta$  complexes containing LTBP-1. *FASEB J.* *19*, 1798–1808.
22. Fortunel, N.O., Hatzfeld, A., and Hatzfeld, J.A. (2000). Transforming growth factor- $\beta$  pleiotropic role in the regulation of hematopoiesis. *Blood* *96*, 2022–2036.
23. Fu, X.D., and Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* *15*, 689–701.
24. Giuliani, N., and Olivieri, A. (2020). *Ematologia per Medicina – Scienze Biologiche – Biotecnologie Mediche (Idelson-Gnocchi)*.
25. Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* *44*, 53–57.
26. Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F., et al. (2012). Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* *120*, 2454–2465.
27. Haase, D., Germing, U., Schanz, J., Pfeilstöcker, M., Nösslinger, T., Hildebrandt, B., Kundgen, A., Lübbert, M., Kunzmann, R., Giagounidis, A.A.N., et al. (2007). New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: Evidence from a core dataset of 2124 patients. *Blood* *110*, 4385–4395.

28. Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28, 241–247.
29. He, W., Dorn, D.C., Erdjument-Bromage, H., Tempst, P., Moore, M.A.S., and Massagué, J. (2006). Hematopoiesis Controlled by Distinct TIF1 $\gamma$  and Smad4 Branches of the TGF $\beta$  Pathway. *Cell* 125, 929–941.
30. Holliday, R., and Grigg, G.W. (1993). DNA methylation and mutation. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* 285, 61–67.
31. Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res.* 25, 14–26.
32. Inoue, D., Bradley, R.K., and Abdel-Wahab, O. (2016). Spliceosomal gene mutations in myelodysplasia: Molecular links to clonal abnormalities of hematopoiesis. *Genes Dev.* 30, 989–1001.
33. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burt, N., Chavez, A., et al. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 371, 2488–2498.
34. Jamaiyar, A., Wan, W., Janota, D.M., Enrick, M.K., Chilian, W.M., and Yin, L. (2017). The versatility and paradox of GDF 11. *Pharmacol. Ther.* 175, 28–34.
35. Jenkins, G. (2008). The role of proteases in transforming growth factor- $\beta$  activation. *Int. J. Biochem. Cell Biol.* 40, 1068–1078.
36. Kaneko, H., Misawa, S., Horiike, S., Nakai, H., and Kashima, K. (1995). TP53 mutations emerge at early phase of myelodysplastic syndrome and are associated with complex chromosomal abnormalities. *Blood* 85, 2189–2193.
37. Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C.W., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* 27, 617–630.

38. Kon, A., Yamazaki, S., Nannya, Y., Kataoka, K., Ota, Y., Nakagawa, M.M., Yoshida, K., Shiozawa, Y., Morita, M., Yoshizato, T., et al. (2018). Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood* *131*, 621–635.
39. Lin, M.E., Hou, H.A., Tsai, C.H., Wu, S.J., Kuo, Y.Y., Tseng, M.H., Liu, M.C., Liu, C.W., Chou, W.C., Chen, C.Y., et al. (2018). Dynamics of DNMT3A mutation and prognostic relevance in patients with primary myelodysplastic syndrome. *Clin. Epigenetics* *10*.
40. Liu, X.-L., Yuan, J.-Y., Zhang, J.-W., Zhang, X.-H., and Wang, R.-X. (2007). Differential gene expression in human hematopoietic stem cells specified toward erythroid, megakaryocytic, and granulocytic lineage. *J. Leukoc. Biol.* *82*, 986–1002.
41. Lowe, S.W., Cepero, E., and Evan, G. (2004). Intrinsic tumour suppression. *Nature* *432*, 307–315.
42. Lyons, R.M., Gentry, L.E., Purchio, A.F., and Moses, H.L. (1990). Mechanism of activation of latent recombinant transforming growth factor  $\beta$ 1 by plasmin. *J. Cell Biol.* *110*, 1361–1367.
43. Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., et al. (2015). Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* *6*, 1–14.
44. Malcovati, L., and Cazzola, M. (2016). Recent advances in the understanding of myelodysplastic syndromes with ring sideroblasts. *Br. J. Haematol.* *174*, 847–858.
45. Malcovati, L., Papaemmanuil, E., Bowen, D.T., Boultonwood, J., Della Porta, M.G., Pascutto, C., Travaglino, E., Groves, M.J., Godfrey, A.L., Ambaglio, I., et al. (2011). Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* *118*, 6239–6246.
46. Malcovati, L., Hellström-Lindberg, E., Bowen, D., Adès, L., Cermak, J., Del Cañizo, C., Della Porta, M.G., Fenaux, P., Gattermann, N., Germing, U., et al. (2013).

- Diagnosis and treatment of primary myelodysplastic syndromes in adults: Recommendations from the European LeukemiaNet. *Blood* 122, 2943–2964.
47. Malcovati, L., Karimi, M., Papaemmanuil, E., Ambaglio, I., Jädersten, M., Jansson, M., Elena, C., Galli, A., Walldin, G., Porta, M.G.D., et al. (2015). SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood* 126, 233–241.
  48. Malcovati, L., Galli, A., Travaglino, E., Ambaglio, I., Rizzo, E., Molteni, E., Elena, C., Ferretti, V.V., Catricalà, S., Bono, E., et al. (2017). Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood* 129, 3371–3378.
  49. Massagué, J. (2012). TGF $\beta$  signalling in context. *Nat Rev Mol Cell Biol* 13, 616–630.
  50. Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121.
  51. Morrison, S.J., and Scadden, D.T. (2014). The bone marrow niche for haematopoietic stem cells. *Nature* 505, 327–334.
  52. Mortera-Blanco, T., Dimitriou, M., Woll, P.S., Karimi, M., Elvarsdottir, E., Conte, S., Tobiasson, M., Jansson, M., Douagi, I., Moarii, M., et al. (2017). SF3B1-initiating mutations in MDS-RSs target lymphomyeloid hematopoietic stem cells. *Blood* 130, 881–890.
  53. Nagata, Y., and Maciejewski, J.P. (2019). The functional mechanisms of mutations in myelodysplastic syndrome. *Leukemia* 33, 2779–2794.
  54. Nemeth, E., Tuttle, M.S., Powelson, J., Vaughn, M.D., Donovan, A., Ward, D.M.V., Ganz, T., and Kaplan, J. (2004). Heparin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* (80-. ). 306, 2090–2093.
  55. Papaemmanuil, E., Cazzola, M., Boulwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al. (2011). Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts . *N. Engl. J. Med.* 365, 1384–1395.
  56. Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and

- biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616–3627.
57. Parisi, S., Finelli, C., Fazio, A., De Stefano, A., Mongiorgi, S., Ratti, S., Cappellini, A., Billi, A.M., Cocco, L., Follo, M.Y., et al. (2021). Clinical and molecular insights in erythropoiesis regulation of signal transduction pathways in myelodysplastic syndromes and  $\beta$ -thalassemia. *Int. J. Mol. Sci.* 22, 1–15.
58. Pellagatti, A., Cazzola, M., Giagounidis, A., Perry, J., Malcovati, L., Della Porta, M.G., Jädersten, M., Killick, S., Verma, A., Norbury, C.J., et al. (2010). Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* 24, 756–764.
59. Prensner, J.R., and Chinnaiyan, A.M. (2011). Metabolism unhinged: IDH mutations in cancer. *Nat. Med.* 17, 291–293.
60. Robertson, I.B., and Rifkin, D.B. (2016). Regulation of the bioavailability of TGF- $\beta$  and TGF- $\beta$ -related proteins. *Cold Spring Harb. Perspect. Biol.* 8, 21907–21908.
61. Schanz, J., Tüchler, H., Solé, F., Mallo, M., Luño, E., Cervera, J., Granada, I., Hildebrandt, B., Slovak, M.L., Ohyashiki, K., et al. (2012). New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. *J. Clin. Oncol.* 30, 820–829.
62. Schneider, R.K., Ademà, V., Heckl, D., Järås, M., Mallo, M., Lord, A.M., Chu, L.P., McConkey, M.E., Kramann, R., Mullally, A., et al. (2014). Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell* 26, 509–520.
63. Shallis, R.M., Ahmad, R., and Zeidan, A.M. (2018). The genetic and molecular pathogenesis of myelodysplastic syndromes. *Eur. J. Haematol.* 101, 260–271.
64. Shi, Y., and Massagué, J. (2003). Mechanisms of TGF- $\beta$  signaling from cell membrane to the nucleus. *Cell* 113, 685–700.
65. Shirai, C.L., Ley, J.N., Graubert, T.A., and Correspondence, M.J.W. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In&nbsp;Vivo.
66. Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2017). The genetics of myelodysplastic

- syndrome: From clonal haematopoiesis to secondary leukaemia. *Nat. Rev. Cancer* *17*, 5–19.
67. Taylor, J., Mi, X., North, K., Binder, M., Penson, A., Lasho, T., Knorr, K., Haddadin, M., Liu, B., Pangallo, J., et al. (2020). Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms. *Blood* *136*, 1477–1486.
68. Tura, S., Cavo, M., and Zinzani, P.L. (2018). Hematology. Pathophysiology, diagnosis and treatment (Esculapio).
69. Vardiman, J. (2012). The classification of MDS: From FAB to WHO and beyond. *Leuk. Res.* *36*, 1453–1458.
70. Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellström-Lindberg, E., Tefferi, A., et al. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: Rationale and important changes. *Blood* *114*, 937–951.
71. Verma, A., Suragani, R.N.V.S., Aluri, S., Shah, N., Bhagat, T.D., Alexander, M.J., Komrokji, R., and Kumar, R. (2020). Biological basis for efficacy of activin receptor ligand traps in myelodysplastic syndromes. *J. Clin. Invest.* *130*, 582–589.
72. Vishwamitra, D., George, S.K., Shi, P., Kaseb, A.O., and Amin, H.M. (2017). Type I insulin-like growth factor receptor signaling in hematological malignancies. *Oncotarget* *8*, 1814–1844.
73. Voso, M.T., Fabiani, E., Fianchi, L., Falconi, G., Criscuolo, M., Santangelo, R., Chiusolo, P., Betti, S., D'Alo', F., Hohaus, S., et al. (2013a). Mutations of epigenetic regulators and of the spliceosome machinery in therapy-related myeloid neoplasms and in acute leukemias evolved from chronic myeloproliferative diseases. *Leukemia* *27*, 982–985.
74. Voso, M.T., Fenu, S., Latagliata, R., Buccisano, F., Piciocchi, A., Aloe-Spiriti, M.A., Breccia, M., Criscuolo, M., Andriani, A., Mancini, S., et al. (2013b). Revised International Prognostic Scoring System (IPSS) predicts survival and leukemic evolution of myelodysplastic syndromes significantly better than IPSS and WHO

- prognostic scoring system: Validation by the Gruppo Romano Mielodisplasie Italian regional database. *J. Clin. Oncol.* *31*, 2671–2677.
75. Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* *136*, 701–718.
76. Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* *456*, 470–476.
77. Weinberg, R. (2016). *La biologia del cancro* (Zanichelli).
78. Yamazaki, S., Ema, H., Karlsson, G., Yamaguchi, T., Miyoshi, H., Shioda, S., Taketo, M.M., Karlsson, S., Iwama, A., and Nakauchi, H. (2011). Nonmyelinating schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* *147*, 1146–1158.
79. Yang, H., Ye, D., Guan, K.L., and Xiong, Y. (2012). IDH1 and IDH2 mutations in tumorigenesis: Mechanistic insights and clinical perspectives. *Clin. Cancer Res.* *18*, 5562–5571.
80. Zermati, Y., Fichelson, S., Valensi, F., Freyssinier, J.M., Rouyer-Fessard, P., Cramer, E., Guichard, J., Varet, B., and Hermine, O. (2000). Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors. *Exp. Hematol.* *28*, 885–894.
81. Zhao, M., Perry, J.M., Marshall, H., Venkatraman, A., Qian, P., He, X.C., Ahamed, J., and Li, L. (2014). Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat. Med.* *20*, 1321–1326.
82. Zhou, L., Nguyen, A.N., Sohal, D., Ma, J.Y., Pahanish, P., Gundabolu, K., Hayman, J., Chubak, A., Mo, Y., Bhagat, T.D., et al. (2008). Inhibition of the TGF- $\beta$  receptor I kinase promotes hematopoiesis in MDS. *Blood* *112*, 3434–3443.
83. Zhou, L., McMahon, C., Bhagat, T., Alencar, C., Yu, Y., Fazzari, M., Sohal, D., Heuck, C., Gundabolu, K., Ng, C., et al. (2011). Reduced SMAD7 leads to overactivation of TGF- $\beta$  signaling in MDS that can be reversed by a specific inhibitor of TGF- $\beta$  receptor I kinase. *Cancer Res.* *71*, 955–963.

## 7. RINGRAZIAMENTI

A conclusione di questo percorso, desidero ringraziare tutti coloro che mi hanno accompagnata in questa esperienza, fatta di duro lavoro, ma ricca di soddisfazioni.

*In primis*, un ringraziamento speciale alla Professoressa Voso, che mi ha permesso di svolgere la mia attività di tirocinio presso il Laboratorio di Oncoematologia da lei diretto e di avermi fatta sentire sin da subito parte integrante del suo gruppo.

Ringrazio la mia relatrice, la Professoressa Ciminelli, che mi ha seguita passo dopo passo con premura e attenzione, rendendo possibile il raggiungimento di questo traguardo.

Ringrazio Emiliano, per avermi dato fiducia sin da principio, farò tesoro dei suoi insegnamenti.

Ringrazio Giulia, per esserci stata sempre, anche da dietro le quinte.

Ringrazio Antonio, che ha guidato i miei primi passi in questo laboratorio e che mi ha supportata e sopportata quando ormai il mio percorso qui stava volgendo al termine.

Ringrazio Elisa, la mia compagna “di banco” e di bancone. Grazie per avermi resa partecipe dei tuoi progetti (e dei tuoi protocolli!), per avermi coinvolta sempre con entusiasmo in tutti i tuoi esperimenti. Grazie per avermi aspettata e per aver rispettato i miei tempi. In te ho trovato non solo una tutor, ma anche un’amica.

Ringrazio Hajro, per avermi permesso di mettermi alla prova come “insegnante”.

Ringrazio Federica e Silvia, le mie compagne di viaggio, per aver arricchito di allegria e spensieratezza le giornate in laboratorio.

Un ringraziamento a Tiziana, Serena e Anna Maria. Da ognuna di voi ho imparato qualcosa, sia in termini umani che in termini di conoscenza.