

***Identification of New Therapeutic Targets and Drug
Repositioning in Myelodysplastic Syndromes Using
Transcriptional Profiling and Connectivity Mapping***



Tianyu Sun

St Anne's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Dr Andrea Pellagatti and Prof Jacqueline Boulton, who expertly guided me through my DPhil research and who shared the excitement of four years of discovery. Their patience, knowledge, and continuous support helped me in all the time of research and writing of this thesis.

Dozens of people have helped and taught me immensely in our laboratory, I wish to express my sincere appreciation to my lab mates: Dr. Hamid Dolatshad, Dr Shalini Singh, Dr Richard Armstrong, Dr Violetta Steeples, Dr Dharamveer Tatwavedi, Dr Doaa Ahmed, Juseong Lee and Vickram Tittrea, for giving the encouragement and sharing insightful suggestions. Without their precious support, it would not be possible to conduct this research.

A special thanks to my family. Words cannot express how grateful I am to my parents for all of the sacrifices that they've made on my behalf. I consider myself nothing without them. They gave me enough moral support, encouragement and motivation to accomplish the personal goals.

Abstract

The myelodysplastic syndromes (MDS) are a heterogeneous group of clonal haematopoietic stem cell (HSC) malignancies that are characterised by ineffective haematopoiesis leading to peripheral blood cytopenias at one or more lineages. There are few effective treatments for this disorder, and the identification of new therapeutic agents in MDS is urgently needed. Drug repurposing is a potential alternative to novel drug discovery by redirecting existing drugs for treating another disease, which dramatically reduces drug development time, cost and safety risk. In the last few years, gene expression signatures have been identified and compared based on genome-wide similarity metrics, unveiling ‘connections’ among drugs, pathways and disease phenotype by matching corresponding signatures of disease and drug responses. Some bioinformatic database and tools, such as connectivity map (CMap), containing transcriptomics of cell lines treated with small molecules were developed, greatly facilitating the discovery of new therapeutic applications of existing compounds.

In this study, we performed CMap on lists of differentially expressed genes (DEG) obtained from the analysis of RNA-seq or microarray-based gene expression profiling (GEP) data generated from bone marrow CD34⁺ cells of MDS patients with splicing factor mutations or with advanced MDS (Refractory anaemia with excess blasts-2 subtype, RAEB-2 subtype). Compounds inducing a transcriptional response opposite to that of MDS cases were identified, which might be capable of reverting the gene

expression signature, and hence may change the disease phenotype. The compounds were ranked by the absolute value of connectivity scores and were primarily screened by comprehensive literature review based on their Mode of Action (MoA) and toxicity profiles in the context of leukaemia, cancer and other diseases. Several top-ranking candidate drugs for repositioning to MDS were selected for in vitro experimental validation in leukaemia cell lines, patient sample derived induced pluripotent stem cells (iPSCs), and patient primary bone marrow cells. Celastrol and Withaferin A were identified from the CMap analysis of RNA-seq data set of splicing factor mutant MDS patients, and Parbendazole and Fenbendazole were identified by mapping a microarray-based MDS-RAEB-2 gene signature to the CMap database. These drugs markedly inhibited proliferation, arrested cell cycle, and induced apoptosis of leukaemia cells. Notably, our study provided evidence of the involvement of splicing pathway modulation in the anti-proliferative activity of these drugs in MDS and leukaemia cells. Higher cytotoxicity induced by Celastrol, Withaferin A, Parbendazole, and Fenbendazole treatment was observed in splicing factor mutant cells, including HUDEP2-U2AF1^{Q157} cells, K562-SF3B1^{K700E} cells and K562-U2AF1^{S34F} cells. Taken together, our data show that these compounds are promising candidates for repurposing to MDS and other related myeloid malignancies.

Impact of Covid-19

In response to the Covid-19 pandemic and as instructed by our University and Department, our laboratory had to cease all experimental work on the 23rd March 2020. As our laboratory is located within the John Radcliffe Hospital, our University/Department had to liaise and work closely with Oxford University Hospitals to establish operational guidance to enable a safe return to on-site working. Our laboratory was allowed to re-open on the 26th August 2020. This 5-month closure has impacted some of my experimental work. Further Covid-19 related disruption includes difficulties in obtaining some reagents and consumables because of stock shortages and backlog of orders. The work I was planning to complete before the closure of University facilities to non-Covid-19 research includes some drug efficacy and MoA studies. These planned experiments aimed at evaluating drug efficacy in bone marrow HSCs from additional MDS patients (and also from healthy individuals as controls) by assessing their effects on cell growth and proliferation, apoptosis, cell cycle, and differentiation. It would have been interesting to evaluate the expression and phosphorylation of cell cycle and apoptosis relevant protein and signalling, and changes of intracellular reactive oxygen species (ROS) levels after drug treatment. To further reveal the MoA of candidate drugs, I planned to perform NF- κ B luciferase reporter assays or western blotting for I κ B α degradation together with phospho-p65 accumulation to detect NF- κ B activation, and qPCR to measure the expression level of IRAK4-Long/IRAK4-Short in drug treated splicing factor mutant cells. A detailed

description of the studies that I was intending to conduct is included in the *Chapter VI*
Discussion-Section B Future Directions.

Contents

<i>List of Tables</i>	<i>11</i>
<i>List of Figures</i>	<i>12</i>
<i>Abbreviations</i>	<i>15</i>
<i>Chapter I Introduction</i>	<i>18</i>
A. Myelodysplastic syndromes (MDS).....	18
1. Epidemiology	18
2. Diagnosis and classification.....	19
3. Prognosis	22
4. Treatment	27
5. Somatic mutations in MDS and their clinical value.....	28
6. Gene expression profiling and gene signatures in MDS	37
B. Splicing factor mutations and spliceosome inhibitors	45
1. Splicing factors mutations in MDS	45
2. Spliceosomes and pre-mRNA splicing: normal and aberrant splicing	48
3. Clinical and biological significance of splicing factor mutations in myeloid malignancies	52
3.1 <i>SF3B1</i> mutation.....	55
3.2 <i>SRSF2</i> mutation.....	60
3.3 <i>U2AF1</i> mutation.....	63
4. Spliceosome mutations and hyperactivation of NF- κ B signalling.....	66
5. R-loop formation is associated with splicing factor mutations.....	70
6. Spliceosome Inhibitors.....	71
6.1 Categories of spliceosome inhibitors	71
6.2 Spliceosome inhibitors in pre-clinical models of MDS	74
6.3 Clinical Trials of Splicing Inhibitors in MDS.....	77
C. Drug repurposing and connectivity mapping.....	79

1. Defining Drug repurposing	79
2. Application of transcriptome signature in drug repositioning	84
3. Connectivity mapping methods and platforms	86
4. Connectivity Mapping Resource, variations and extensions	90
D. Objectives of the research.....	94
Chapter II Materials and Methods	95
1. Patient samples for RNA-sequencing	95
1.1 Patient samples	95
2. Identification of differentially expressed genes	96
3. Identifying candidate drugs using connectivity map (CMap).....	96
3.1 CMap analysis	96
3.2 MANTRA drug network	97
4. Cell lines and cell culture.....	97
5. Genetically modified cell lines.....	99
6. Isolation of CD34 ⁺ cells from bone marrow mononuclear cells.....	100
7. Generation and characterization of induced pluripotent stem cells (iPSCs).....	100
7.1 Generation of iPSCs	100
7.2 Confirmation of clearance of Sendai virus reprogramming factors	101
7.3 Characterization of iPSCs	101
7.4 Immunocytochemistry for expression of immature ES cell markers.....	102
7.5 Flow Cytometry for expression of immature ES cell markers.....	102
7.6 Trilineage differentiation of iPSCs analysed by flow cytometry	103
8. Compounds	103
9. Cell viability assay	105
10. Cell cycle analysis.....	105
11. Cell Apoptosis analysis	106
12. Cell differentiation assays	106
13. Colony-Forming Cell Assay	107

14. Real-Time Quantitative Polymerase Chain Reaction	107
15. Molecular docking	108
16. Phospho-kinase array assay.....	109
17. Kinase inhibition assay	110
18. Statistics	110

Chapter III Gene expression profiling of MDS patient CD34 cells and computational drug screening via connectivity mapping 112

A. Introduction and aims.....	112
B. Results	114
1. RNA-seq based transcriptome analysis of MDS patients with splicing factor mutations .	114
2. Generation of RNA-Seq based DEG lists used as input to the connectivity mapping analysis	117
3. Candidate compounds identified by Connectivity map analysis of RNA-Seq based DEG lists	118
3.1 Drug screening by Connectivity map.....	118
3.2 Gene expression profile based ‘Drug network’ of Celastrol and WA	120
4. Generation of microarray based DEG lists used as input to the connectivity mapping analysis (MDS-RAEB-2 patients)	124
5. Candidate compounds identified by Connectivity map analysis of Microarray based DEG lists	126
5.1 Drug screening by Connectivity map.....	126
5.2 Gene expression profile based ‘Drug network’ of Parbendazole.....	128
C. Discussion	130

Chapter IV Investigation of the effects of drug candidates on cell viability, cell apoptosis, cell cycle, and differentiation..... 135

A. Introduction and aims.....	135
B. Results	138

1. Celastrol and Withaferin A selectively inhibited the proliferation of leukaemia cell lines	138
2. Celastrol and Withaferin A induced cell cycle arrest at G2/M phase and apoptosis in leukaemia cell lines.....	140
3. Drug combinations showed synergistic effects in inhibiting leukaemia cell proliferation.....	144
4. Celastrol and Withaferin A preferentially inhibit the proliferation of MDS patient CD34+ cells.....	147
5. Piperlogumine and Parthenolide inhibited the proliferation of leukaemia cell lines	148
6. PBD and FBD inhibit the proliferation of leukaemia cell lines	150
7. PBD and FBD induce cell cycle arrest at G2/M phase and apoptosis in leukaemia cell lines.....	152
8. PBD and FBD inhibit the proliferation of the mononuclear cells and CD34+ cells from MDS patients	155
9. PBD and FBD induce myeloid differentiation in MDS/AML cell lines.....	156
C. Discussion	160

Chapter V Investigation of the molecular mechanisms of drug candidates... 163

A. Introduction and aims.....	163
B. Results	166
1. Celastrol and WA preferentially inhibit the proliferation of splicing factor mutant cells.	166
2. Effects of Celastrol and WA on NFκB signalling and inflammatory cytokine production	169
3. PBD and FBD preferentially inhibit the proliferation of splicing factor mutant cells.....	173
4. PBD and FBD preferentially inhibit iPSC derived from MDS patient cells with splicing factor mutations	179
5. Investigation of the MoA of PBD and FBD in leukaemia cells.....	181
C. Discussion	189
1. MoA of Celastrol and WA	189

2. MoA of Benzimidazoles	191
<i>Chapter VI Discussion and Future directions</i>	<i>195</i>
A. Discussion	195
1. Drugs identified by CMap analysis of RNA-seq GEP	196
2. Drugs identified by CMap analysis of microarray-based GEP	205
B. Future Directions.....	208
1. Drug efficacy and MoA studies	208
2. Single-cell analysis of haematopoietic stem cell in SF mutant MDS	209
<i>Reference... ..</i>	<i>211</i>

List of Tables

Table 1.1	FAB classification of MDS
Table 1.2	Comparison between the editions 2008 and 2016 WHO classification of MDS
Table 1.3	Revised International Prognostic Scoring System (IPSS-R) for MDS
Table 1.4	Most frequent somatic mutations observed in MDS
Table 1.5	Computational methods used in drug repurposing
Table 2.1	The cell culture conditions of leukaemia cell lines
Table 2.2	The names and suppliers of the compounds
Table 2.3	Primers used for RT-PCR
Table 3.1	Distance table of Celastrol and WA generated by MANTRA
Table 3.2	Top 30 CMap drugs identified by using microarray based DEG in MDS-RAEB-2 patients
Table 3.3	Distance table of Parbendazole generated by Mantra
Table 4.1	Clinical and biological information of leukaemia cell lines
Table 4.2	The IC ₅₀ of 72h drug treatment in cell lines (nM)
Table 4.3	The IC ₅₀ of Parthenolide and Piperlongumin treatment in cell lines (μM)
Table 4.4	The IC ₅₀ of PBD and FBD treatment in cell lines (nM)
Table 4.5	Drug selected for experimental validation
Table 5.1	Top 30 compounds generated from the CMap analysis of Top200 gene list
Table 5.2	Top 30 compounds generated from the CMap analysis of Up-down 100 gene list
Table 6.1	Molecular mechanisms of Candidate compounds in malignancies

List of Figures

- Figure 1.1 Mutated genes with independent prognostic significance by MDS bone marrow blast proportion
- Figure 1.2 Enrichment of mutations in secondary AML and high risk MDS versus high-risk and low-risk MDS respectively
- Figure 1.3 Schematic Representation of hotspot residue mutations in *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*
- Figure 1.4 Schematics of the stepwise pre-mRNA splicing process
- Figure 1.5 Schematic of alternative splicing events: normal and abnormal (aberrant) splicing
- Figure 1.6 Functional consequences of SF3B1, SRSF2, and U2AF1 mutations on RNA splicing
- Figure 1.7 Rationale for the use of spliceosome inhibitors in splicing factor mutant cells
- Figure 1.8 Workflow of GEP-based drug screening through connectivity mapping
- Figure 3.1 Disease subtypes and mutations status of MDS patients
- Figure 3.2 Differentially expressed genes in MDS with splicing factor gene mutations
- Figure 3.3 Venn diagrams show the overlapping genes of the three DEG lists for CMAP queries in three comparisons
- Figure 3.4 Celastrol and Withaferin A were identified as top-ranking candidates by CMAP
- Figure 3.5 Drug networks of Celastrol and WA (MANTRA)
- Figure 3.6 Top 500 up-regulated and down-regulated genes induced by four drugs from CMAP database
- Figure 3.7 The overlapping genes of the three microarray-based DEG lists from MDS-RAEB-2 patients for CMAP queries
- Figure 3.8 Drugs identified from the CMap analysis in RAEB-2

- Figure 3.9 Drug networks of Parbendazole (MANTRA)
- Figure 4.1 Drug dose-response curves in cells by Celastrol or WA treatment.
- Figure 4.2 Cell cycle analyses by PI staining
- Figure 4.3 Cell apoptosis induced by Celastrol and WA
- Figure 4.4 Visualization of 3D synergy maps showing overall synergy scores (δ -score) for combination treatment of Celastrol or WA and Azacitidine
- Figure 4.5 Visualization of 3D synergy maps showing overall synergy scores (δ -score) for combination treatment of Celastrol and WA
- Figure 4.6 Cell viability of CD34⁺ cells from MDS patients and healthy donors after 72h Celastrol or WA treatment with indicated concentrations
- Figure 4.7 Drug dose-response curves in cells treated by Piperlogumine or Parthenolide
- Figure 4.8 Drug dose-response curves in cells treated with PBD or FBD
- Figure 4.9 Cell cycle analyses by PI staining
- Figure 4.10 Cell apoptosis induced by PBD and FBD
- Figure 4.11 Cell viability of MDS patient cells after treatment with PBD and FBD
- Figure 4.12 FBD and PBD induced myeloid differentiation in leukaemia cell lines
- Figure 4.13 Colony-forming assay of CD34⁺ cells treated with PBD and FBD
- Figure 5.1 Responses of HUDEP2 cells to drug treatments
- Figure 5.2 Responses of HUDEP2 cell and K562 cell to drug treatments
- Figure 5.3 Responses of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} cells to drugs
- Figure 5.4 IL-6 mRNA expression in spliceosome mutant cells after drug treatment
- Figure 5.5 Molecular docking of Celastrol or WA with IRAK4
- Figure 5.6 Differentially expressed genes in K562 cells with SF3B1-K700E mutation
- Figure 5.7 Responses of HUDEP2 cell and K562 cell to drug treatments

- Figure 5.8 Responses of iPSCs to drug treatments
- Figure 5.9 Responses of iPSCs to drug treatments
- Figure 5.10 Phospho-kinase array assays of SKM1 cells
- Figure 5.11 Expression levels of STAT5a/b target genes in SKM1 cells after treatment with PBD or FBD
- Figure 5.12 Molecular docking of PBD or FBD with TNK1
- Figure 5.13 TNK1 kinase activity inhibited by PBD and FBD
- Figure 5.14 Expression levels of Wnt target genes in SKM1 cells after treatment with PBD or FBD
- Figure 5.15 Molecular docking of PBD with IRAK1/4
- Figure 5.16 Molecular docking of FBD with IRAK1/4
- Figure 5.17 N-Acyl-2-aminobenzimidazole inhibitors of IRAK4 discovered via high-throughput screening by Amgen
- Figure 6.1 NF- κ B signalling in cancer and inflammatory disease

Abbreviations

AML	Acute myeloid leukaemia
AS	Alternative splicing
ASCT	Allogeneic stem cell transplantation
BP	Branch point
CFU	Colon-forming unit
CMap	Connectivity map
Coxnet	Cox proportional hazards regression
CP	Cleavage and polyadenylation
DEG	Differentially expressed genes
DMSO	Dimethylsulphoxide
ECOG	Eastern Cooperative Oncology Group
ESE	Exon splicing enhancer
FAB	French-American-British
FACTS	Fast analytical continuum treatment of solvation
FBD	Fenbendazole
GEO	Gene Expression Omnibus
GEP	Gene expression profiling
GWAS	Genome-wide association studies
HAS	Highest single agent
hESC	Human embryonic stem cell
HMA	Hypomethylating agent
HPCs	Haematopoietic progenitor cells
HSCs	Haematopoietic stem cells
HSPCs	Haematopoietic stem and progenitor cells
HTS	High throughput screening
HUDEP2	Erythroid progenitor cell line 2
IC50	Half maximal inhibitory concentration

IKK β	I κ B kinase β
IL-	Interleukin-
iPSCs	Induced pluripotent stem cells
IPSS	International Prognostic Scoring System
IPSS-R	Revised International Prognostic Scoring system
IRAK	Interleukin-1 receptor-associated kinase
LGA	Lamarckian genetic algorithm
LINCS	Library of Integrated Network-Based Cellular Signatures
MANTRA	Mode of Action by NeTwoRk Analysis
MDS	Myelodysplastic syndromes
MDS-EB	MDS with excess blasts
MDS-MLD	MDS with multilineage dysplasia
MDS-RS	MDS with ring sideroblasts
MDS-SLD	MDS with single lineage dysplasia
MDS-U	MDS, unclassifiable
MoA	Mode of Action
NIH	National Institutes of Health
NMD	Nonsense-mediated decay
PBD	Parbendazole
PI	Propidium iodide
RA	Refractory anaemia
RAEB	Refractory anaemia with excess of blasts
RAEB-2	Refractory anaemia with excess blasts-2
RAEB-T	Refractory anaemia with excess blast in transformation
RARS	Refractory anaemia with ringed sideroblasts
RCMD-RS	Refractory cytopenia with multilineage dysplasia and ring sideroblasts
RCUD	Refractory cytopenia with unilineage dysplasia
RN	Refractory neutropenia

ROS	Reactive oxygen species
RRM	RNA recognition motif
SAR	Structure-activity relationship
SF	Splicing factor
SR	Serine/arginine
ssODN	Single-stranded oligodeoxynucleotide
SuperPC	Supervised principal components
TGF- β	Transforming growth factor β
TNF α	Tumor necrosis factor α
TNIK	TRAF2 and NCK-interacting kinase
WA	Withaferin A
WHO	World Health Organization
WPSS	WHO classification–based prognostic scoring system

Chapter I Introduction

A. Myelodysplastic syndromes (MDS)

1. Epidemiology

The myelodysplastic syndromes (MDS) represent a heterogeneous group of clonal haematopoietic stem cell malignancies that are characterized by ineffective haematopoiesis leading to peripheral blood cytopenias at one or more lineages [1]. The MDS are pre-leukaemic conditions showing frequent evolution to acute myeloid leukaemia (AML), occurring in approximately one third of MDS cases over months to years after diagnosis. MDS occurs most commonly in older adults, and the risk of developing MDS increases with age. The median age at presentation is 70 years and disease onset before age 50 is unusual (except for therapy-related MDS), with rare cases having been reported in children [2-4]. There is a male predominance in most categories of MDS, apart from MDS with isolated del(5q), which is more common in women [5, 6]. The annual incidence of MDS has been estimated to be 1 to 5 cases per 100,000 persons per year in the general population; however, for individuals aged >70 years the incidence is at least 20 cases/100,000 per year [2, 7-9]. There are no known ethnic differences in the incidence of MDS, but in Asian populations, MDS tend to occur at an earlier age, patients more often have a hypocellular marrow and present less often with isolated 5q deletion (5q- syndrome), while trisomy 8 seems to be more frequent than in Western populations [10, 11].

2. Diagnosis and classification

Survival following a diagnosis of MDS varies from a few months to more than ten years [12]. The highly variable prognosis underscores the importance of a classification system, supplemented by a prognostic index, to predict the survival of patients with MDS and the likelihood of transformation to AML. With the recent development and introduction of some effective treatment options for MDS, the need for a classification system to predict responsiveness to treatment and clinical outcomes for individual patients has become even more important. In the last decades, different MDS classification and prognostic scoring systems have been proposed. The most common approaches for diagnosis and classification of MDS are the French-American-British (FAB) classification system [13] and 2008 World Health Organization (WHO) classification system [14], which are primarily based on morphologic anomalies of blood and bone marrow cells. The FAB system was introduced in 1982 and is based on the percentage of blasts and morphological dysplastic features of blood and bone marrow. According to this system, patients are diagnosed with MDS when dysplastic bone marrow haematopoiesis is present and/or myeloblasts are between 5 and 30% of all bone marrow cells. Five subgroups with different prognoses were established: refractory anaemia (RA) with blasts <5% in BM, refractory anaemia with ringed sideroblasts (RARS) with blasts <5% and ring sideroblasts >15%, refractory anaemia with excess of blasts between 5 and 20% (RAEB), RAEB in transformation to acute leukaemia and blast cells ranged between 20 and 30% (RAEB-T), and chronic

myelomonocytic leukaemia (CMML) characterized by increase of peripheral blood monocytes (Table 1.1). The FAB system has served as the standard for MDS classification for more than 20 years and provided considerable prognostic information [13, 15]. Nonetheless, the clinical outcomes of patients assigned to the same MDS subgroup remain too variable to accurately predict survival or transformation to AML in individual patients.

Table 1.1 FAB classification of MDS

Type	Blasts in blood	Blasts in bone marrow
Refractory anaemia (RA)	<1%	Blasts <5%, ring sideroblasts <15%
Refractory anaemia with ring sideroblastic (RARS)	<1%	Blasts <5%, ring sideroblasts >15%
Refractory anaemia with excess of blast (RAEB)	<5%	Blasts 5–20%
Refractory anaemia with excess blast in transformation (RAEB-T)	<30%	Blasts 21–30%
Chronic myelomonocytic leukaemia (CMML)	<5% with increase monocytes	Blasts 0–20%
Acute myeloid leukaemia (AML)	>30%	>30%

In 2001, the WHO proposed an alternative classification for MDS that was modified from the FAB approach [16, 17]. The revisions included lowering the threshold for the percentage of blasts required to make the diagnosis of AML from 30% to 20%, thus eliminating the MDS subcategory of RAEB-T. In addition, CMML was reclassified from a subcategory of MDS to a subcategory of myelodysplastic/myeloproliferative disorder. Since then, the WHO classification has been updated twice (2008 and 2016) (Table 1.2) [14, 18-22]. The classification of 2016 WHO of MDS was according to factors that differ from those of the FAB system and defined by precise criteria including: (1) dysplastic changes (2) number of cytopenia in peripheral blood (3) percentage of ring sideroblasts [22, 23].

Table 1.2 Comparison between the editions 2008 and 2016 WHO classification of MDS

2008	2016
Refractory cytopenia with unilineage dysplasia (RCUD)	MDS with single lineage dysplasia (MDS-SLD)
Refractory anaemia (RA)	
Refractory neutropenia (RN)	
Refractory thrombocytopenia (RT)	
Refractory anaemia with ring sideroblasts (RARS)	MDS with ring sideroblasts (MDS-RS)
	MDS-RS-SLD
	MDS-RS-MLD
Refractory cytopenias with multilineage dysplasia	MDS with multilineage dysplasia (MDS-MLD)
Refractory anaemia with excess blasts (RAEB)	MDS with excess blasts (MDS-EB)
RAEB-1	MDS-EB-1
RAEB-2	MDS-EB-2
MDS with isolated del(5q)	MDS with isolated del(5q)
MDS, unclassifiable (MDS-U)	MDS, unclassifiable (MDS-U)
Refractory cytopenia of childhood (provisional)	Refractory cytopenia of childhood (provisional)

3. Prognosis

Disease course and prognosis are highly variable in MDS. The main prognostic factors of MDS, for progression to AML and survival, include the marrow blast percentage, extent of cytopenias and cytogenetic abnormalities. The current clinical management

of MDS focuses on assessing disease risks at diagnosis by assaying clinical and patient characteristics and using established prognostic scoring systems to estimate survival and risk of evolution to AML [2, 3, 12]. Since the publication of FAB classifications, several prognostic scoring systems and risk models have been proposed which were generated using several validated prognostic features, acting as surrogates for the underlying disease biology and patient host characteristics, to generate weighted scoring systems that divide patients into defined risk categories [7]. Predicting severity of disease progression and likelihood of AML transformation is useful to guide therapeutic recommendations for patients. Some patients belong to a low-risk cohort best managed with conservative supportive care, whereas others are included in a high-risk cohort that requires decisive therapy with haematopoietic stem cell transplantation or hypomethylating agent (HMA) administration. The International Prognostic Scoring System (IPSS), the recently Revised International Prognostic Scoring system (IPSS-R) and the WHO classification-based prognostic scoring system (WPSS) provide clinicians with useful risk stratification and prognostic models, and can, therefore, be used for implementing risk-adapted treatment strategies in MDS patients [24-28]. The IPSS score was generated by assessing three prognostic indicators to predict the course of the patient's disease: 1) the percentage of blast cells in the marrow; 2) the type of chromosomal changes, if any, in the marrow cells (cytogenetics); 3) the presence of one or more low blood cell counts (cytopenias). It ranked patients into four cohorts of risk: low, intermediate-1, intermediate-2, and high [28].

A revision of the IPSS released in 2012 improved risk prediction by redefining cutoffs for bone marrow blast percentages and cytopenias, adding numerous additional cytogenetic features, and establishing five risk cohorts instead of four: very low, low, intermediate, high, and very high [24]. This revised IPSS-R covers the same disease factors as the IPSS, but the parameters are identified in a more detailed way. The IPSS-R is based on five disease factors: blasts, cytogenetics (includes new classifications of a number of less common cytogenetic subsets), hemoglobin, platelet count, absolute neutrophil count (Table 1.3) [11, 24]. Currently, the IPSS-R is the most widely used prognostic scoring system for MDS. The WPSS is used less commonly than the IPSS and IPSS-R. It differs from the other two systems in which it includes WHO-defined MDS subtypes as a prognostic factor. It also assigns a score based on the presence or absence of severe anaemia. WPSS stratifies MDS patients into 5 risk groups based on 3 factors: 1) The type of MDS based on the WHO classification (e.g. certain types of MDS-SLD and MDS-del(5q) tend to have the best outlook, whereas MDS-EB tends to have the worst); 2) Chromosome abnormalities (grouped as good, intermediate, or poor); 3) Whether or not the patient needs regular blood transfusions [24, 26]. Patient-related characteristics such as age, Eastern Cooperative Oncology Group (ECOG) performance status and comorbidities are also relevant for establishing prognosis and treatment choice, particularly in lower risk MDS. Other disease-related factors include such as albumin, bone marrow fibrosis, ferritin, and LDH levels have been shown to

have prognostic significance. Ferritin and LDH levels can add to the IPSS and were considered for inclusion in the IPSS-R [2, 24, 25, 29].

It is important to note that prognostic systems and risk groups do not predict how MDS will respond to treatment but instead how MDS is likely to behave over time without treatment. Lower-risk MDS tends to progress slowly, and it may not cause severe symptoms in the long-term. Hence, less intensive treatment is frequently used. In contrast, higher-risk MDS is likely to progress more quickly or become AML without treatment. It may cause more symptoms and health complications in a shorter time. Thus, more intensive treatment is required.

Table 1.3 Revised International Prognostic Scoring System (IPSS-R) for MDS

Prognostic characteristics	Points						
	0	0.5	1	1.5	2	3	4
Cytogenetic risk category	Very good		Good		Intermediate	Poor	Very poor
Blasts in bone marrow, %	≤ 2		$>2\% - 5\%$		$5\% - 10\%$	$>10\%$	
Haemoglobin, g/dl	≥ 10		$8 - <10$	<8			
Platelet count, $\times 10^9/l$	≥ 100	$50 - <100$	<50				
Absolute neutrophil count, $\times 10^9/l$	≥ 0.8	<0.8					
IPSS-R risk group	Score	Median overall survival, years			Median time to 25% AML evolution, years		
Very low	≤ 1.5	8.8			NR		
Low	$>1.5 - 3$	5.3			9.4		
Intermediate	$>3 - 4.5$	3			2.5		
High	$>4.5 - 6$	1.6			1.7		
Very high	> 6	0.8			0.7		

Very good: -Y and del(11q) as single abnormalities; good: normal, del(5q), del(12p) and del(20q) as single abnormalities, double abnormalities including del(5q); intermediate: del(7q), +8, +19, i(17q) and any other single abnormalities, any other double abnormalities; poor: -7 and inv(3)/t(3q)/del(3q) as single abnormalities, double abnormalities including -7/del(7q), complex (3 abnormalities); very poor; >3 abnormalities.

4. Treatment

The majority of MDS patients will succumb to the disease over time, because of infection as a result of neutropenia and neutrophilic dysfunction or the development of AML [1]. Treatment options vary from the symptomatic therapy of cytopenias to bone marrow transplantation [2]. The current standard of care for MDS patients depends upon early risk stratification according to predicted overall survival, given that the choice of intervention is guided by prognosis [1, 2]. For lower risk patients with symptomatic anaemia, treatment with erythropoiesis-stimulating agents or lenalidomide (especially for patients with del(5q)) can ameliorate symptoms. Some lower risk patients may be candidates for immunosuppressive therapy, thrombopoiesis-stimulating agents, or a HMA (azacitidine or decitabine) [30]. Among higher risk patients, transplant candidates should undergo allogeneic stem cell transplantation (ASCT) as soon as possible, with HMAs useful as a bridge to transplant. Non-transplant candidates should initiate HMA therapy and continue if tolerated until disease progression [30]. Supportive care with blood transfusions and antimicrobial drugs as needed remains important in all groups [3, 31, 32]. With the exceptions of the immunomodulatory drug lenalidomide, which is particularly effective in patients with del(5q), and azacitidine and decitabine, used to treat advanced MDS, there are few effective drug treatments for this disorder. In addition, there is extensive off-label use of the erythropoiesis-stimulating agents epoetin and darbepoetin in MDS. ASCT is the only curative treatment for MDS but is only applicable in a small proportion of patients

[7, 32, 33]. The major obstacle to ASCT is that most MDS patients are above the age of 70 years. Co-morbidity, age, IPSS and IPSS-R score, cytogenetics, conditioning regimen and donor selection are predictors of post-transplant outcome, and should be taken into account carefully during the decision process [34, 35]. Only one new drug has been approved for MDS since decitabine's FDA approval in 2006. In April 2020, FDA approved luspatercept for the treatment of anaemia failing an erythropoiesis stimulating agent and requiring 2 or more red blood cell units over 8 weeks in adult patients with very low- to intermediate-risk MDS-RS or with MDS/MPN-RS-T. Luspatercept is a recombinant fusion protein that binds transforming growth factor β (TGF- β) superfamily ligands to reduce aberrant Smad2/3 signalling and enhance late-stage erythropoiesis [36]. Efficacy was demonstrated in a Phase 3 trial (NCT02631070), in which luspatercept reduced the severity of the anaemia in patients with lower-risk MDS-RS who had been receiving regular red-cell transfusions and who had disease that was refractory to or unlikely to respond to erythropoiesis-stimulating agents or who had discontinued such agents owing to an adverse event [36, 37].

5. Somatic mutations in MDS and their clinical value

During the past decade, our understanding of the genetics underlying MDS/AML has improved tremendously. Early genetic studies on MDS mostly focused on cytogenetic abnormalities detected by conventional karyotyping, which are found in 30% to >80% of patients, depending on the MDS subtype and whether the disease is de novo or

therapy-related [38]. In contrast to AML, most abnormalities in MDS are unbalanced changes, resulting in gains or losses of chromosomal materials. The most frequent among these are $-7/\text{del}(7q)$ and $-5/\text{del}(5q)$, followed by $+8$, $\text{dup}(1q)$, $\text{del}(20q)$, $\text{del}(11q)$, $\text{del}(12p)/\text{t}(12p)$, $\text{del}(17p)/\text{iso}(17q)$, $\text{del}(18q)$, $+21q$ gains, $\text{del}(13q)$, and $+\text{der}(1;7)(q10;p10)$ [39]. Most cytogenetic abnormalities have been incorporated the IPSS/IPSS-R prognosis system for risk stratification [39]. Karyotype represents one of the most powerful prognostic factors among all parameters in the IPSS-R [1, 11, 24].

Next generation sequencing based studies have illuminated the mutational landscape of MDS [40-42]. It has been shown that more than 90% of MDS patients have a somatic mutation in at least one gene [41, 43]. While a typical MDS patient harbors a median of 2 or 3 driver mutations, high-risk MDS tend to show higher numbers of driver mutations than lower-risk MDS [39]. Though the number of recurrent gene mutations in MDS is large, the affected genes can be grouped into fewer categories, based on the implicated cellular process: pre-mRNA splicing factors, epigenetic regulators and chromatin-remodelling factors, transcription factors, and signal transduction molecules (Table 1.4) [43]. These genetic lesions are presumed to have a common clonal origin, and allow for expansion of the initiating clone to compromise normal haematopoiesis, ultimately giving rise to MDS and subsequent transformation to AML in many patients [39, 44]. The driver alterations in MDS are largely shared by primary AML, suggesting common abnormalities in these disorders, but some differences exist. For example,

there is an overrepresentation of mutations in receptor tyrosine kinase (*FLT3* and *KIT*) and *RAS* pathway genes, as well as *CEBPA* and *IDH1/IDH2* mutations in AML, while mutations in splicing factors and epigenetic regulators are more prevalent in MDS [45].

Table 1.4 Most frequent somatic mutations observed in MDS (other mutations are seen in <5% of the cases)

Gene function	Gene	Mutation frequency
Epigenetic regulators and chromatin-remodelling factors	<i>TET2</i>	15%–25%
	<i>ASXL1</i>	10%–20%
	<i>DNMT3A</i>	10%
	<i>IDH1/2</i>	5%–10%
Pre-mRNA splicing factors	<i>SF3B1</i>	15%–30%
	<i>SRSF2</i>	10%–15%
	<i>U2AF1</i>	5%–10%
Transcription factors	<i>RUNX1</i>	10%–15%
	<i>TP53</i>	5%–10%
Signal transduction molecules	<i>NRAS/KRAS</i>	10%

The clinical effects of driver genetic lesions have been investigated in large cohorts of MDS patients. Several somatic mutations can serve as clinically relevant biomarkers, which are associated with clinical features such as bone marrow blast percentage, severity of certain cytopenia, and presence of dysplasia [46, 47]. Moreover, mutations in particular drivers are now beginning to play similar roles as karyotype to ascertain diagnosis and prognosis that is independent of common risk stratification tools, and to a lesser extent, can help predict response to certain therapies [41, 43, 46, 48-52]. A

landmark paper by Bejar et al. suggested that mutations in *ASXL1*, *TP53*, *EZH2*, *ETV6* and *RUNX1* are associated with poor prognosis in MDS patients, independently of established risk factors [46]. Mutation of *RUNX1*, *TP53*, *NRAS*, *ASXL1* and *U2AF1* have been associated with thrombocytopenia, while *SF3B1* mutations have been associated with a normal or increased platelet count [46, 52, 53]. Mutations in *ASXL1*, *SRSF2*, *CBL*, *RUNX1*, *WT1*, *IDH2*, *STAG2*, and *NRAS* have all been shown to correlate with a higher bone marrow blast percentage [41, 43, 52].

Specifically, *TP53* mutations in MDS have been associated with a particularly poor prognosis, rapid transformation to AML, resistance to conventional therapies and dismal outcomes[41, 46, 54-57]. *TP53* mutations occur in approximately one-fourth of high-risk MDS patients, and are especially prevalent among patients with therapy-related MDS/AML, in whom their frequency approaches 40%-50% [58-60]. Complex karyotype MDS patients have fewer mutations in other genes, but roughly half carry a *TP53* mutation and have the worst overall outcomes, even after treatment [50, 51, 61, 62]. However, for individuals with complex karyotypes, the absence of *TP53* mutations is associated with significantly improved outcomes, comparable to patients with noncomplex karyotypes, highlighting the importance of *TP53* mutation status in this group [63]. It appears that the strong negative impact of a *TP53* mutation on prognosis cannot be overcome, even with allogenic stem-cell transplantation [51, 55, 56]. A recent large cohort study reported by Bernard et al showing the clinical implications of

TP53 allelic state in MDS patients. Established correlation of *TP53* mutations with complex karyotype, few co-occurring mutations, high-risk presentation and poor outcomes were specific to multiple-hit patients only. *TP53* multiple-hit state could predict risk of death and leukaemic transformation independently of the IPSS-R, while monoallelic patients did not differ from *TP53* wild-type patients with regard to response to therapy, overall survival and AML progression [64].

SF3B1 mutations are strongly associated with increased bone marrow ring sideroblasts and a trend towards longer patient survival [65, 66]. *SF3B1* mutations are mutually exclusive not only with other splicing factor (SF) mutations but also with most other mutations, except for *DNMT3A*, *TET2*, *JAK2*, and *del(5q)* [41, 43]. On the basis of these unique genetic, morphological, and clinical features, it is suggested that *SF3B1* mutations define a unique subset of myeloid neoplasms [67]. In a recent report published by Malcovati et al., an international working group of experts in MDS has proposed the recognition of a distinct subtype of MDS based on the presence of *SF3B1* mutation [68]. In contrast to generally favorable prognosis in *SF3B1*-mutated cases, patients with other SF mutations, including *U2AF1* and *SRSF2*, have poor overall survival [69-71]. *SRSF2* mutations are among the most predominant mutations in CMML, in which mutated *SRSF2* was found in approximately half of these patients, frequently cooccurring with *TET2* and *ASXL1* mutations [72-74]. The biological mechanism and clinical relevance of SF mutations in MDS will be discussed in the

following sections. Some mutations in genes such as *CBL*, *EZH2*, *PTPN11*, and *PRPF8* are not strongly linked to clinical risk factors yet still predict a shorter overall survival (Figure 1.1) [75, 76]. This may be because mutations associated with an adverse prognosis often arise in small subclones that can be detected by sensitive sequencing techniques. These small clones may be harbingers of disease progression, but not yet large enough to alter clinical risk factors like blast proportion or cytopaenia. For adverse mutations, it has been shown that their presence in a small subclone carries the same risk as when they are present in the dominant clone [43]. Additionally, the number of pathogenic variants in a patient has been found to be prognostically significant. As the number of oncogenic mutations increases, patient outcomes progressively worsen [43, 46]. One or more mutations with adverse prognostic associations can be found in over one-third of patients, indicating that we routinely underestimate disease risk using conventional assessment tools [43, 46, 77, 78].

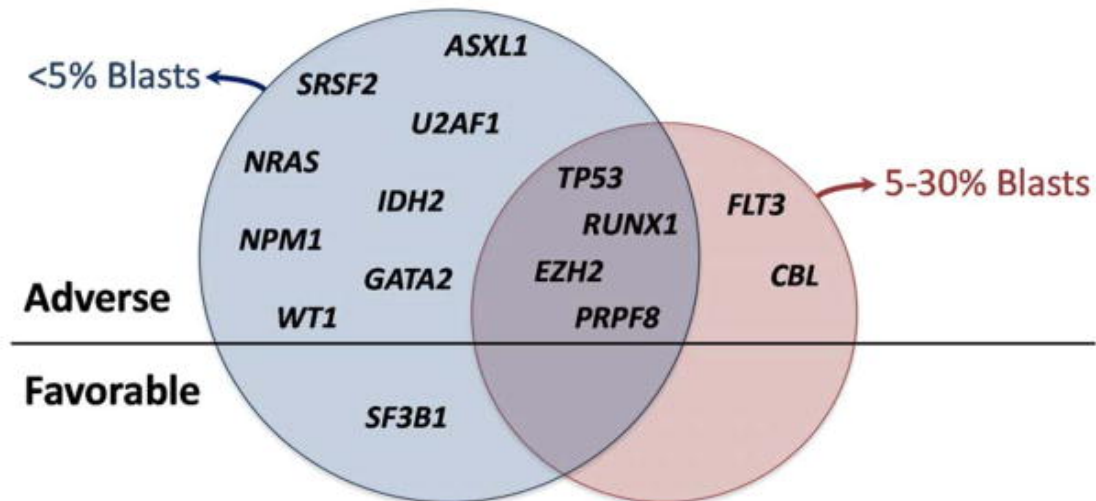


Figure 1.1 Mutated genes with independent prognostic significance by MDS bone marrow blast proportion. Genes in the figure are associated with overall survival after adjustment for IPSS-R risk groups. Genes in the blue circle are significant in patients with less than 5% blasts in the bone marrow. Genes in the red circle remain significant in patients with higher blast counts. *SF3B1* mutations are independently prognostically favorable (taken from Bejar et al [76]).

Although differences in cell biology and maturation exist between MDS and AML secondary to MDS, these two diseases are genetically related. In a clonal haematopoiesis model of disease evolution, the sequential acquisition of mutations during MDS pathogenesis implies a series of genetic states that correspond to distinct clinical phenotypes. In an initiating stage, founder mutations drive asymptomatic clonal expansion within the haematopoietic compartment. During progression, further mutations are acquired within this clone, which impair normal haematopoiesis and alter blood counts, ultimately resulting in overt MDS and in some instances, eventual secondary AML [79]. Studies suggested that the final transition to secondary AML involves the acquisition of characteristic AML-associated genetic changes, such as activating mutations in signalling molecules such as *FLT3* and *N-RAS*, as well as

inactivating mutations in *CEBPA*. At the time of leukaemic transformation, the antecedent MDS clone persists but is outcompeted by aggressive subclones that drive the development of AML [80, 81]. Substantial differences in mutational profiles are seen across different disease stages by mutation enrichment studies, which provide insight into clonal evolution of MDS and how driver mutations shape MDS development (Figure 1.2). Mutations in some genes are significantly enriched in secondary AML compared with high-risk MDS, including *FLT3*, *NPM1*, *NRAS*, *PTPN11*, *WT1*, *IDH1*, and *IDH2* (Type I), while other genes are more commonly mutated in high-risk MDS compared with low-risk MDS, such as *GATA2*, *RUNX1*, *TET2*, *ZRSR2*, *TP53*, *STAG2*, and *ASXL1* (Type II). Differences in the frequency of mutated genes in MDS and secondary AML suggest that the order of mutation acquisition is not random during progression. It is of note that all type I genes are among the most frequently mutated genes in primary AML and highly enriched in primary AML compared with MDS. Interestingly, patients with type I mutations show a significantly faster leukaemic transformation than those without type I mutations, indicating the critical roles of type I mutations in AML phenotypes and transformation to secondary AML [45, 82].

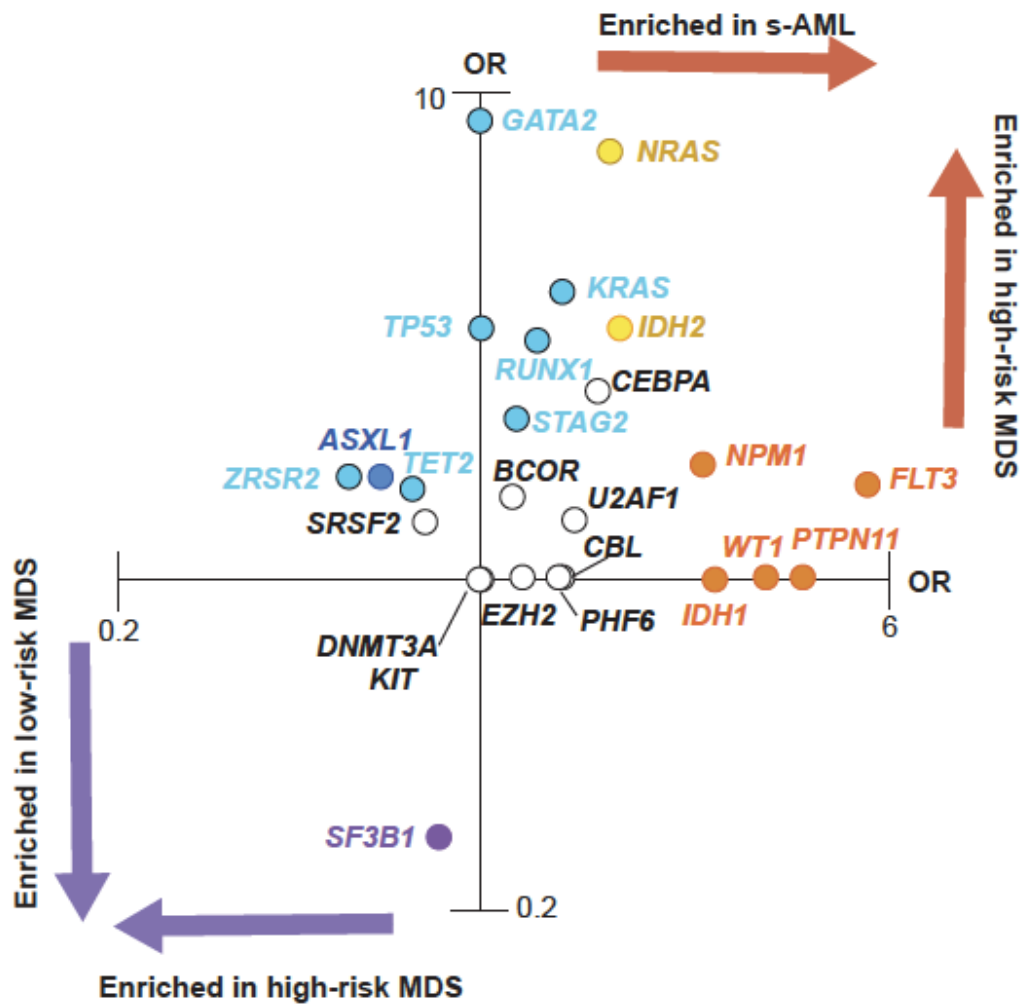


Figure 1.2. Enrichment of mutations in secondary AML and high risk MDS versus high-risk and low-risk MDS respectively. Enrichment of mutations expressed as odds ratio (OR) of mutation rates in secondary AML vs high risk MDS (x-axis) and in high risk MDS vs low risk MDS (y-axis). Non-significant OR are represented by black circles (taken from Makishima et al [82]).

Given the emerging evidence has been supporting the role of somatic gene mutations in improving risk stratification, attempts to build molecular prognostic models that incorporate molecular data into current prognostic models have been made [41, 83, 84]. Haferlach et al. and Nazha et al. demonstrated that integrating mutational status into the IPSS-R could refine risk stratification and improve the power of prognostic

prediction. However, the number of incorporated genes and patient clinical features were highly variable in these two studies [41, 84]. More recently, Hou et al. proposed a scoring system incorporating age, IPSS-R and five poor-risk mutations (*CBL*, *IDH2*, *ASXL1*, *DNMT3A*, and *TP53*) which dividing the MDS patients into four risk groups ($P < 0.001$ for both OS and leukaemia-free survival). It has been suggested that integration of gene mutations in IPSS-R improves the prognostication of MDS patients and may help identify high-risk patients in IPSS-R lower risk group for more aggressive treatment [83]. Consensus on how to best incorporate mutational information into risk assessment remains a work in progress [85].

6. Gene expression profiling and gene signatures in MDS

Gene expression profiling (GEP), based on DNA microarray and more recently RNA-seq technology [86], is a powerful technique for the measurement of patterns of gene expression on a genomic scale, which has been used to identify pathways and genes potentially involved in the pathogenesis of disease [87-90]. GEP can be used to compare gene transcription in cancer cells versus normal tissues in thousands of genes. A gene is considered to be differentially expressed if an observed difference in expression levels between two biological conditions is statistically significant [91, 92]. Recent studies suggested that any perturbations to the transcriptional profile of the cell can be summarised by a 'gene signature' [93]. A gene signature is a combined pattern of gene expression alterations, which typically contains a small number of differentially

expressed genes (DEG) with validated association in terms of diagnosis, prognosis or prediction of therapeutic response [87-90, 94]. As seen in recent GEP studies of myeloid malignancies, progress has been made in the molecular classification, prediction of clinical outcomes and treatment stratification through the application of the gene signatures [42, 94-100]. In addition, as microarray technology and next-generation sequencing has become more technically feasible and widely accessible, GEP will continue providing new insights into the molecular complexity underlying MDS and AML and lead to the dissection of pathogenic mechanisms, thereby contributing to discover novel targets for therapeutic intervention. The heterogeneity of MDS has been seen as a potential issue for GEP-based classifications, but several studies compared gene signatures between MDS patients and healthy control, between different risk groups of MDS, or between MDS-derived AML and de novo AML [101-105]. Most of these studies have been based on CD34⁺ or CD133⁺ cell populations isolated from normal/healthy and MDS bone marrow, providing valuable insights into the molecular pathophysiology of MDS [87]. However, other studies have shown that it is also feasible to use unfractionated mononuclear bone marrow cells to identify diagnostic or prognostic molecular signatures [94, 98, 99, 106].

Early GEP studies of MDS bone marrow haematopoietic stem cells (HSCs) are relatively small, and primarily focused on the pathway analysis to understand the dysregulated biologic processes and functions in different MDS subgroups. By using a

comprehensive array platform, the expression profiles of MDS CD34⁺ cells showed many similarities to interferon- γ -induced gene expression in normal CD34⁺ cells. Distinct gene expression profiles were associated with specific FAB and cytogenetic subclasses and CD34⁺ cells from RARS patients were characterized by deregulation of mitochondrial-related genes [103]. Subsequent studies suggested a strong relationship between increasing percentage of bone marrow ring sideroblasts and decreasing *ABCB7* gene expression levels in RARS subtype [107]. The *ABCB7* gene is critical for mitochondrial iron homeostasis and has been implicated in the transport of a component required for the maturation of iron-sulfur cluster proteins out of mitochondria [108, 109]. RARS erythroblasts also showed a marked deregulation of mitochondrial function pathway and severely suppressed *ABCB7* throughout erythroid maturation in RARS, in contrast to the increasing expression over time in normal erythroblast cultures [110]. A more recent RNA-Seq analysis of HSCs from *SF3B1* mutant MDS samples unveiled aberrant/cryptic splicing events in some downstream target genes including *ABCB7* induced by *SF3B1* mutation. Aberrant *ABCB7* splicing, due to usage of an alternative 3' splice site in MDS patient samples, giving rise to a premature termination codon in the *ABCB7* mRNA. Nonsense-mediated decay (NMD)-induced downregulation of *ABCB7* mRNA transcript resulting from aberrant splicing underlies the increased mitochondrial iron accumulation found in MDS patients with ring sideroblasts [111].

The del(5q) is the most commonly reported deletion in de novo MDS and is found in 10–15% of all patients [104]. Defined by isolated del(5q) and no excess of marrow blasts, 5q- syndrome is a specific type of MDS with particular characteristics, including severe anemia, frequent thrombocytosis, typical dysmegakaryopoiesis and favorable outcome[112]. Several small cohort GEP studies of 5q- syndrome also provided novel insights into the molecular pathogenesis of the disease. For example, a GEP analysis of highly purified 5q-deleted HSCs (CD34+CD38–Thy1+) in 5q- MDS supported that they might originate from and outcompete the normal counterparts. Few but distinct differences in gene expression distinguished 5q- MDS stem cells and normal HSCs, including upregulated expression of *BMII*, *ID1*, *DNMT3A*, *MYC*. Expression of *CEBPA*, encoding a myeloid transcription factor, was up-regulated in the 5q- stem cell fraction (CD34+CD38–Thy1+), but consistently downregulated in 5q- progenitor cells (CD34+CD38+Thy1–) compared with normal HSCs and haematopoietic progenitor cells (HPCs), respectively [113]. Other studies reported that candidate genes showing haploinsufficiency in the 5q- syndrome, including the tumour suppressor gene *SPARC* and *RPS14*, a component of the 40S ribosomal subunit, may play a role in the pathogenesis of the 5q- syndrome. Two genes mapping to the commonly deleted region of the 5q- syndrome, *RBM22* and *CSNK1A1*, showed a >50% reduction in gene expression, consistent with the downregulation of the remaining allele [114, 115].

In 2010, Pellagatti et al. performed the first GEP study of CD34⁺ cells from a large cohort of MDS patients (n=183) to determine deregulated gene pathways and ontology groups. GEP and pathway analysis revealed that the most deregulated pathways in MDS were interferon signalling, thrombopoietin signalling and the Wnt pathways, as compared with the HSC of healthy controls. Early MDS stages showed deregulation of immunodeficiency, apoptosis and chemokine signalling pathways, whereas advanced MDS was characterized by deregulation of DNA damage response and checkpoint pathways. Distinct gene expression profiles and deregulated gene pathways in patients with del(5q), trisomy 8 or -7/del(7q) were identified. Patients with trisomy 8 were characterized by deregulation of pathways involved in the immune response, patients with -7/del(7q) by pathways involved in cell survival, whereas patients with del(5q) showed deregulation of integrin signalling and cell cycle regulation pathways [116].

These studies have provided important molecular insights into the pathophysiology of MDS, but none of these were designed to assess the diagnostic or prognostic capabilities of GEP in this group of diseases. The predictive value of gene expression signatures was confirmed by using marrow CD34⁺ cells from MDS patients. In a small cohort study with 35 MDS patients and 6 age-matched healthy persons, a 6-gene ‘poor risk’ signature associated with acute myeloid leukaemia transformation was defined, providing additive prognostic information for IPSS Intermediate-1 patients. The ‘poor

risk' signature associated with AML transformation was shown to be enriched for 'ribosomal', MYC and Wnt signalling targets [117].

The International Microarray Innovations in Leukaemia Study Group performed a multi-center clinical study regarding the utility of microarray-based GEP in the diagnosis and sub-classification of leukaemia (MILE study). The MILE study was designed to assess the clinical accuracy of GEP (compared with current routine diagnostic work-up) of 16 acute and chronic leukaemia subclasses, MDS, and a so-called "none of the target classes" (neither leukaemia nor MDS) control group that included nonmalignant disorders and normal bone marrow [106]. Overall, the accuracy of the diagnostic classification model for subtyping leukaemia was approximately 93%, but the diagnostic classification model failed to confirm the clinical diagnosis of MDS in approximate half of the MDS specimens submitted to the study. Discordant samples of MDS were classified either into acute myeloid leukaemia (AML) or 'none-of-the-targets' categories. To clarify the discordant results, all submitted 174 MDS samples were externally reviewed, although this did not improve the molecular classification results. However, a significant correlation was noted between the AML and 'none-of-the-targets' categories and prognosis, leading to a prognostic classification model to predict for time-dependent probability of leukaemic transformation. The prognostic classification model accurately discriminated patients with a rapid transformation to AML within 18 months from those with more indolent disease [94].

More recently, Pellagatti et al. performed a GEP analysis on bone marrow CD34+ cells from 125 MDS patients to determine the relationship between gene expression levels and prognosis. Supervised principal components (SuperPC) and lasso penalized Cox proportional hazards regression (Coxnet) were used for the analysis [95]. The SuperPC method identified a total of 139 genes associated with MDS patient survival, including *LEF1*, *PTH2R*, *CDH1*, *WT1* and *MNI*. The Coxnet method was able to identify a smaller set of 20 genes associated with patient survival. The Coxnet predictor outperformed other predictors including one which additionally used clinical information. The Coxnet gene signature based on CD34+ cells significantly identified a separation of patients with a good or bad prognosis in an independent GEP dataset which is based on unsorted bone marrow mononuclear cells, demonstrating that the gene signature is robust and may be applicable to bone marrow cells without isolating CD34+ cells. Eight genes were identified as significantly associated with MDS patient survival by both the SuperPC and Coxnet methods: low expression levels of *LEF1*, *BTBD6*, *FRMD6* and *C7orf58* and high expression levels of *ADHFE1*, *CPT1B*, *GPR114* and *LOC100505956*, were associated with a poor prognosis in MDS patients. This GEP-based signature correlating with clinical outcome may significantly contribute to a refined risk classification of MDS [118].

Taken together, GEP based gene signature could enhance the MDS classifications and prognostic scoring systems by providing standardized, objective gene signatures.

However, these GEP cohorts had not taken into account the mutation status of the MDS patients studied, which highlighted the need for a comprehensive study to investigate the interconnections between mutations, gene expression profiles, clinical variables and patient outcome thereby contributing to build a more powerful and integrative prediction model [96].

B. Splicing factor mutations and spliceosome inhibitors

1. Splicing factors mutations in MDS

In the past decade, advanced high-throughput sequencing technologies have led to the identification of acquired mutations in several genes in MDS [43]. Mutations in RNA splicing regulators (*SRSF2*, *U2AF1*, *SF3B1*, and *ZRSR2*) have been shown to be the most common genomic lesions found in this disorder, and occur in over 50% of cases [119-121]. Figure 1.3 shows the hotspot residue mutations in splicing factor genes in MDS. The splicing factor genes mutated in MDS encode proteins that are involved in 3'-splice site recognition during precursor messenger RNA (pre-mRNA) processing. Mutations in these genes lead to perturbed normal splice site and exon recognition, which affects the expression of many downstream target genes and pathways, implicating aberrant spliceosome function in MDS disease pathogenesis [120, 122, 123]. Notably, mutations in *SF3B1*, *U2AF1* and *SRSF2* are predominantly found as mutually exclusive heterozygous point mutations at restricted locations, suggesting that they have a change-of-function effect, whereas *ZRSR2* mutations are scattered across the gene and are predicted to be loss-of-function mutations (Figure 1.3) [124]. Mutations in splicing factor genes are far more frequent in MDS than de novo AML, suggesting that splicing factor gene mutations may be key factors driving the different clinical features between these related disorders. These mutations are presumed to contribute to oncogenic transformation, but the underlying mechanisms remain elusive

and are currently an area of intense research. Recent studies have highlighted that the presence of splicing factor mutations can lead to impaired haematopoiesis and dysplastic differentiation that are hallmarks of MDS pathogenesis [125-127]. Given the high frequency of spliceosomal mutations and their early occurrence in MDS and other myeloid malignancies, emerging data has suggested that strategies to inhibit or modulate the spliceosome may have therapeutic value in the treatment of spliceosome-mutant disorders [128, 129].

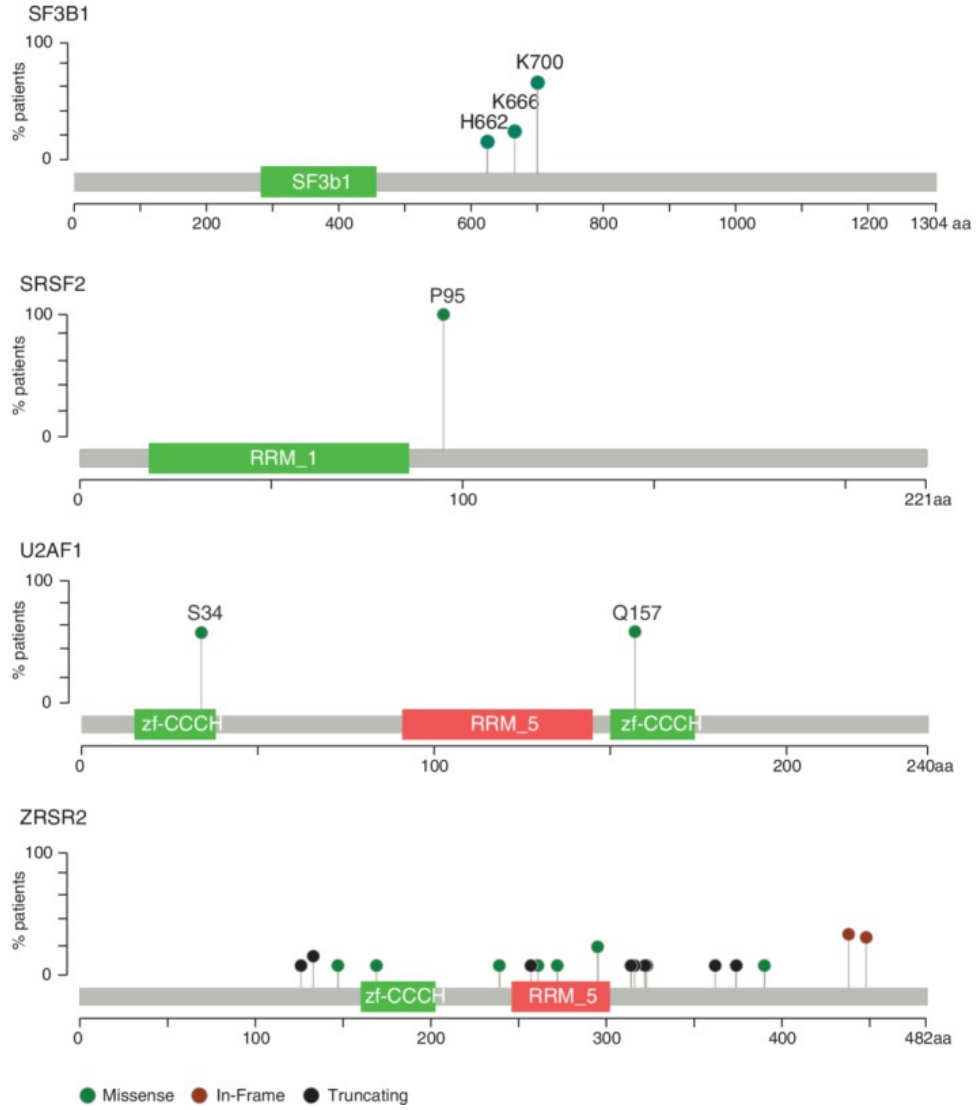


Figure 1.3. Schematic Representation of hotspot residue mutations in *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*. Mutations in *SF3B1*, *SRSF2*, and *U2AF1* occur exclusively as heterozygous missense mutations located to very restricted regions, whereas *ZRSR2* mutations are scattered across the gene and are predicted to be loss-of-function mutations. Lollipop plots showing the frequency of mutations occurring in *SF3B1*, *SRSF2*, *U2AF1*. The height of the lollipop represents the frequency of occurrence of mutations at the specific residue in that gene as a percentage of all mutations occurring in that gene across myeloid malignancy patients. Green dots represent missense mutations. Red dots represent in-frame insertions/deletions and black dots represent truncating frameshift or nonsense mutations (Taken from Taylor et al [130]).

2. Spliceosomes and pre-mRNA splicing: normal and aberrant splicing

Pre-mRNA splicing, the removal of introns and the joining of flanking exons, is an essential process to generate mature mRNAs which enables the production of multiple protein isoforms transcribed from limited eukaryotic genes [131]. For the vast majority of human genes, excision of intronic sequences in pre-mRNA is carried out by repeated assembly of a large and highly dynamic ribonucleoprotein machine called the spliceosome, a complex of five small nuclear ribonucleoproteins (snRNPs), U1, U2, U4/U6, and U5, with a myriad of associated proteins [132, 133]. The precision of splicing is achieved by stepwise assembly of spliceosome on the pre-mRNA substrate to recognize the splice junction and catalyze two trans-esterification reactions that excise the intron, while spliceosome components are recycled for additional rounds of splicing [131, 132, 134, 135]. Figure 1.4 shows the cycle of ordered assembly and disassembly of the spliceosome throughout the splicing reaction [119].

As splicing initiates, the early complex E is formed by the binding of U1 snRNP to the 5' splice site. Meanwhile, multiple factors are involved in the 3' splice site recognition process: SF1 binds to the branch point sequence upstream of the 3' splice site, and the serine/arginine (SR) rich proteins SRSF2 and ZRSR2 bind to the exon splicing enhancer (ESE) site via its RNA recognition motif (RRM) domain to aid the binding and stability of U2AF1 (U2AF35) and U2AF2 (U2AF65). In the subsequent step, U2 snRNP complex displaces bound to the branchpoint sequence and forms the A complex.

The SF3B1 protein is a core subcomponent of the U2 snRNP, that is involved in the recognition of the branch point, targeting the U2 to the 3' splice site and prevents the premature nucleophilic attack at the site by the spliceosome. The B complex is then formed by the recruitment of U4/U6–U5 tri-snRNP, followed by remodeling of the complex leading to removal of the U1 and U4 and formation of the catalytic structure (complex C). This then activates the spliceosome machinery for catalyzing intron release and exon ligation to form a functional mRNA. The spliceosome components are recycled for subsequent rounds of splicing (Figure 1.4) [119, 131, 136].

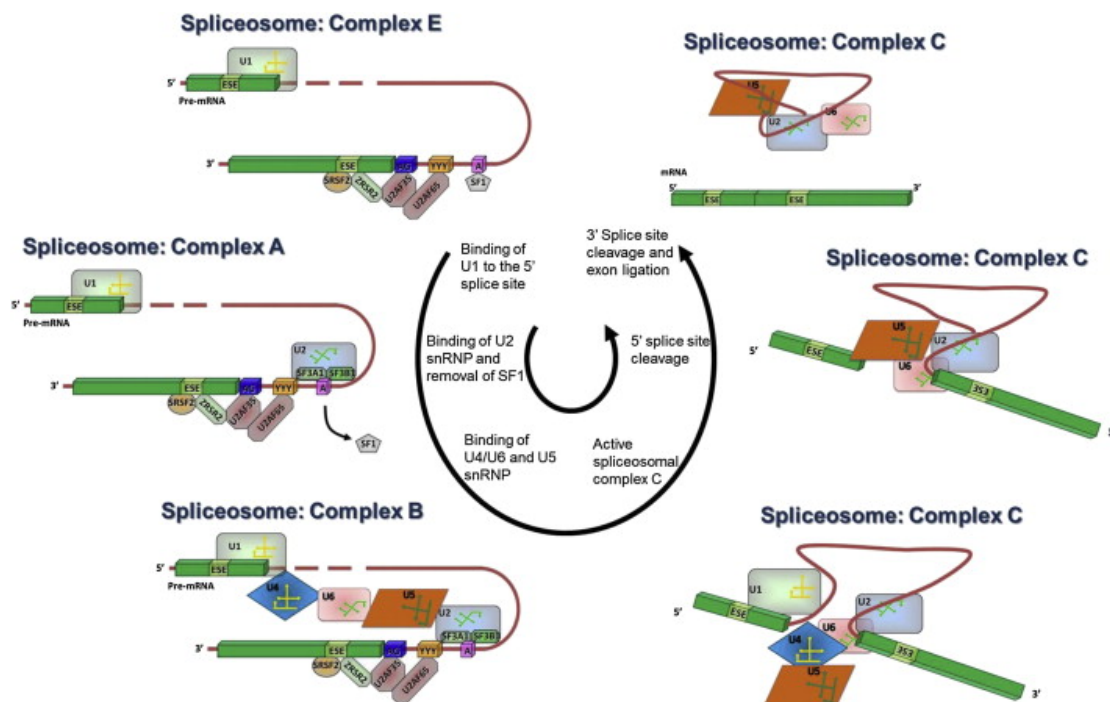


Figure 1.4. Schematics of the stepwise pre-mRNA splicing process. U1 snRNP binds to the 5' splice site and SF1, SRSF2, ZRSR2, U2AF35 and U2AF65 bind to the 3' splice site forming the Early complex (E). U2 snRNP then replaces SF1 and binds to the branch site (complex A) and aids the binding of U4/U6–U5 tri-snRNP to form complex B. This brings the two splice sites together and an intron loop is formed (complex C). Subsequently, the 5' site is spliced followed by the 3' splice site and the two exons are ligated together (Taken from Boulton et al [119]).

In the process of pre-mRNA splicing, some exons, parts of exons, or introns can be retained or skipped in the mature transcript by a process which is called alternative splicing (AS). Deep sequencing has now revealed that >95% of human genes are alternatively spliced, often in a developmental, cell-/tissue-specific, or signal transduction-dependent manner. AS arises from stimulatory and inhibitory signals coming from multiple splicing factors near weak splice sites, either helping the spliceosome to assemble at this location, or destabilizing it and giving rise to a mixed population of mature mRNAs [137, 138]. Many different types of AS events exist and are summarized in Figure 1.5 [119]. Exon skipping, whereby an exon and its flanking introns are excised together from a transcript, is most common and accounts for 30–40% of alternative splicing events. Alternative 5' and 3' splice site selection can also lead to the removal of specific portions of exons. An intron might also not be properly recognized by the spliceosome and be kept in the mature RNA, leading to intron retention. Furthermore, different transcription start sites might be used to initiate transcription, leading to alternative 5' exons. Regulation of AS is tightly controlled during normal tissue differentiation. Misregulation of AS can lead to production of aberrant protein isoforms, which may contribute to diseases including cancer [139, 140].

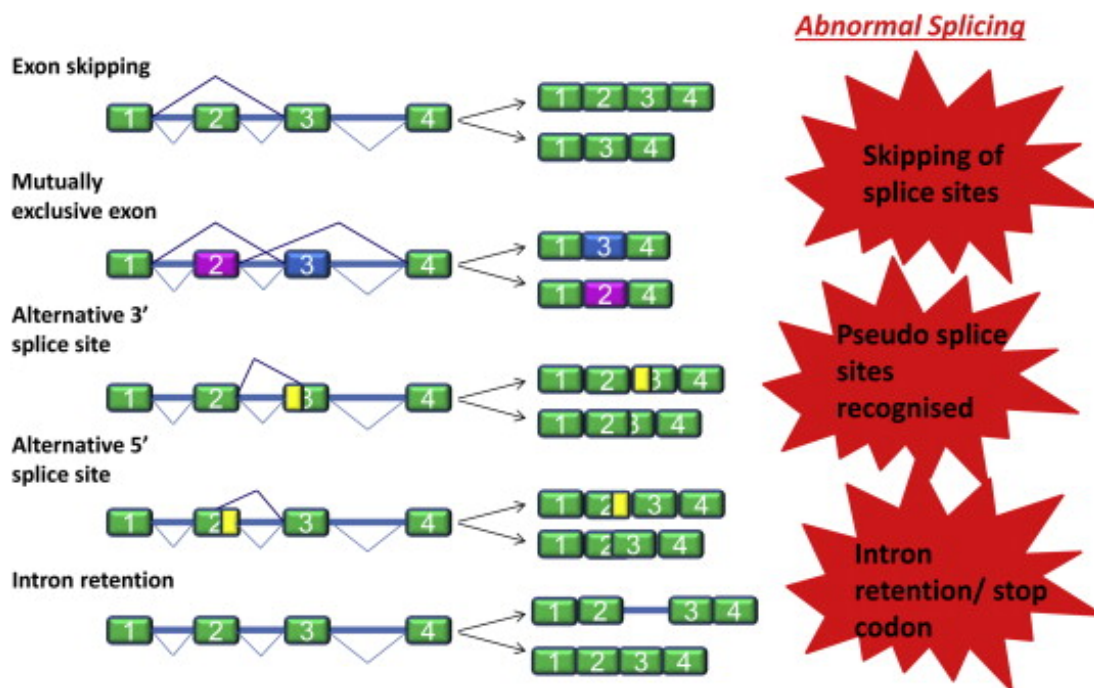


Figure 1.5. Schematic of alternative splicing events: normal and abnormal (aberrant) splicing. Normal splicing can involve exon skipping, mutually exclusive exons, alternative 5' and 3' splice sites and intron retention (in frame). Aberrant splicing can involve skipping of known splice sites, recognition of pseudo splice sites, and intron retention that causes pre-mature stop codon which subsequently initiates nonsense-mediated decay pathway (Taken from Boulton et al [119]).

As the most frequently mutated splicing factor genes in MDS, *SF3B1*, *U2AF1*, *SRSF2* and *ZRSR2* are components of the E/A splicing complex that engaging in the initial steps of RNA splicing and orchestrating 3' splice site recognition. Even though the spliceosome machinery contains more than 150 proteins, exactly why only four of these proteins are frequent targets of somatic mutations in haematological malignancies remains an open question [130]. It has been demonstrated that the mutations in splicing factor genes reflect defects in 3' splice site recognition leading to the generation of aberrantly spliced mRNA transcripts in the bone marrow cells of patients with myeloid malignancies and in mouse models expressing these mutations [119, 130, 141]. Given

the mutation frequencies and critical functions of the E/A splicing complex on 3' splice site recognition, a major molecular feature of MDS would be aberrant splicing events and diverse mRNA species caused by impaired spliceosome machinery [70, 122].

In some cases, the use of alternative 3' splice sites can result in a frameshift leading to the generation of a premature stop codon that activate NMD pathway, and therefore lead to degradation and down-regulated expression of these transcripts [122]; aberrantly spliced transcripts with retained introns can similarly lead to NMD or to increased mRNA stability [142]; and cassette exons result in alternative isoform generation, often not expected to alter gene expression levels [143-145]. In addition, it is also possible that NMD insensitive aberrant transcripts are translated into novel protein isoforms with altered function [146]. However, the link between the splicing gene mutations and clonal expansion of haematopoietic stem cells remains unclear. Emerging evidence from both in vivo and in vitro studies by using isogenic models and human cells has primarily revealed the functional consequences of splicing factors alterations in MDS.

3. Clinical and biological significance of splicing factor mutations in myeloid malignancies

As opposed to other classes of mutations occurring in haematological malignancies, splicing factor mutations are frequent MDS but less common in other blood cancers.

These mutations are heterozygous and occur in a mutually exclusive manner, suggesting that cells may tolerate only a partial deviation from the physiological splicing process [119, 130]. Splicing factor mutations define clinical phenotypes in MDS to some extent, and have differing prognostic impacts association with subtypes of MDS [119, 123]. While splicing factor genes can be found as single mutation, several studies have shown that splicing factor mutations show specific associations with epigenetic regulators and also oncogenes mutated in MDS [41, 43, 147]. Co-mutation of *ASXL1* and *U2AF1*, and *SF3B1* and *DNMT3A*, for example are frequently reported in MDS, whilst co-mutation of *TET2* with *SRSF2* or *ZRSR2* is highly predictive of a CMML disease phenotype [41, 43, 77]. Thus, co-operation between splicing factor mutations and mutations in other genes is likely required for the development of the MDS phenotype. A wealth of transcriptome-based studies have shown that the splicing factor gene mutations found in MDS result in aberrant expression and pre-mRNA splicing of many downstream target genes. Cassette exons (i.e. exons that are either included or skipped from transcripts) were found to be predominant in *SRSF2* mutant and *U2AF1* mutant MDS cases, whilst *SF3B1* mutations were mainly associated with intron retention events and the use of cryptic 3' splice sites [148, 149].

Although recent large-scale transcriptomic analysis of mutations affecting *SF3B1*, *SRSF2*, and *U2AF1* revealed global splicing dysregulation consistent with their predicted effects, there was minimal overlap between mis-spliced transcripts altered by

each of the mutant proteins. Deregulated cellular processes/pathways include genome stability and the DNA damage response, epigenetic regulation and RNA processing and splicing [111, 126, 127, 150]. Furthermore, evidence from human and mouse modeling studies suggests that splicing factor gene mutations can drive haematopoietic stem/myeloid progenitor cell expansion in the bone marrow and aspects of the dysplastic differentiation and ineffective haematopoiesis characteristic of MDS [111, 125, 126]. Pellagatti et al. showed that the aberrantly spliced genes identified in *SF3B1*, *SRSF2* and *U2AF1* mutant MDS converged in common dysregulated cellular processes and pathways, focused on RNA splicing, protein synthesis and mitochondrial dysfunction, suggesting common mechanisms of action [149]. To identify key splicing abnormalities induced by mutant splicing factors, several studies provided functional evidence that expression of the abnormal splice isoforms in normal cells, or reintroduction of the canonical isoforms in spliceosomal mutant cells, can only partially phenocopy or rescue the effects of mutant spliceosome proteins [111, 126, 151, 152]. These findings suggest that the pathological effects of spliceosome mutations are imparted through multiple mis-spliced products simultaneously. Figure 1.6 summarises the functional consequences of *SF3B1*, *SRSF2*, and *U2AF1* mutations in on RNA splicing. In the following sections, we will review the clinical and biological impact of mutations in *SF3B1*, *SRSF2* and *U2AF1*.

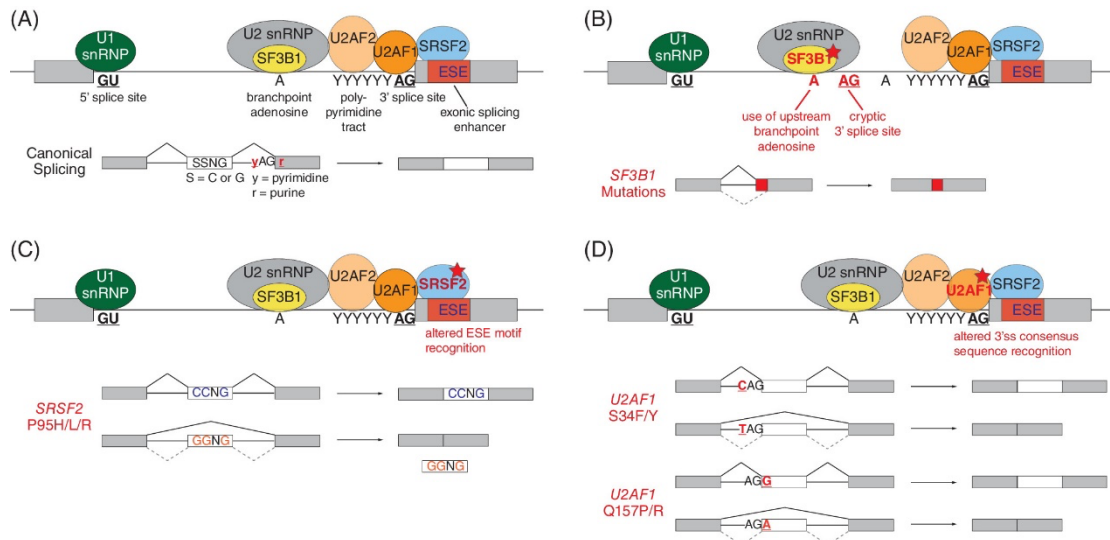


Figure 1.6. Functional consequences of SF3B1, SRSF2, and U2AF1 mutations on RNA splicing. A. Schematic representation of canonical splicing. B. SF3B1 mutations lead to preferential usage of cryptic 3' splice sites approximately 10-30 nucleotides upstream of the canonical 3' splice site (AG dinucleotide), and an upstream branch point adenosine residue is recognized by the U2 snRNP/SF3B1 complex. C. SRSF2 mutations alter the preference of ESE motif recognition. Canonical SRSF2 have equal preference for C-rich CCNG and G-rich GGNG ESE motifs in cassette exon splicing. Mutant SRSF2 show preferential inclusion of exons with CCNG motif over GGNG motif. D. Mutations in U2AF1 alter the 3' splice site usage by recognizing consensus nucleotides at the -3 and +1 positions flanking the AG dinucleotide (intron-exon junction). In the presence of U2AF1 S34 mutation, exons are preferentially included if the pyrimidine at the -3 position of the AG dinucleotide is a C instead of a T, in U2AF1 Q157P mutant cells, this preference is determined by the purine nucleotide at the +1 position relative to the AG dinucleotide, such that a G instead of A results in preferential exon inclusion (Adapted from Taylor et al[130]).

3.1 *SF3B1* mutation

3.1.1 Clinical significance

SF3B1 encodes a subunit of the U2 snRNP responsible for 3' branch site recognition and is the most frequently mutated splicing factor gene found in MDS. In many studies it has been shown that approximately 20%-28% of MDS patients carry *SF3B1* mutations, with a higher frequency (>80%) observed in patients with refractory

anaemia with ring sideroblasts (RARS) and refractory cytopenia with multilineage dysplasia and ring sideroblasts (RCMD-RS) [53, 146]. In MDS, mutated *SF3B1* clinically featured by normal or elevated platelet counts, increased bone marrow cellularity, red cell transfusion dependency and increased incidence of thrombotic events. It negatively correlates with presence of multilineage dysplasia and high-risk karyotype, as well as a lower risk of progression to secondary AML [65, 146]. *SF3B1* mutations have been shown to co-occur with mutations of genes involved in the regulation of DNA methylation, including *DNMT3A* and *TET2*. *SF3B1* mutations have also been associated with the presence of mutation in the tyrosine kinase gene *JAK2* [41, 43, 147].

It has been noted in several studies that *SF3B1* mutations tend to be heterozygous and imply a gain of function, and interestingly *SF3B1* mutations have been recently shown to be mutually exclusive with mutations in other genes that are associated with leukaemic transformation. Most of the mutations in *SF3B1* occur near a cluster of C-terminal HEAT repeat domains that are thought to be involved in binding of *SF3B1* to other members of the U2 snRNP, with the most frequently mutated residue being K700 in MDS. Patients carrying the *SF3B1* mutation show a homogeneous disease phenotype characterized by isolated erythroid dysplasia and mild dysplasia in granulocytic or megakaryocytic lineages, supporting the notion that the *SF3B1* mutation identifies a distinct entity within MDS [146]. However, how exactly the mutations affect MDS pathogenesis is not completely clear, but they are often acquired early in disease

development, tend not to be associated with complex karyotypes, and tend to be associated with a favorable clinical course.

3.1.2 Biological mechanisms

Mian et al. reported that *SF3B1* mutations in MDS patients with ring sideroblasts can arise in a rare population of HSCs (CD34⁺CD38⁻CD45RA⁻CD90⁺CD49f⁺) and are an initiating event in MDS pathogenesis. Using xenotransplantation assays, it was demonstrated that the *SF3B1* mutant clone alone or in combination with other molecular abnormalities results in a clonal growth advantage over ‘normal’ cohabitating cells in NSG mice [153]. Transcriptome analysis is a powerful tool for inferring direct targets of aberrant splicing, and several studies have investigated the functionally relevant and causative mis-splicing events associated with *SF3B1* mutations that drive MDS phenotypes. RNA sequencing studies in *SF3B1* mutated cells provided evidence that the major splicing defect associated with *SF3B1* mutations, regardless of cellular or disease origin, is the preferential usage of cryptic 3′ splice sites approximately 10-30 nucleotides upstream of the canonical 3′ splice site. This is distinct from loss-of-function of *SF3B1* that causes inefficient splicing catalysis [154-156]. Darman et al showed that SF3B1 hotspot mutation induces aberrant 3′ splice sites selection through the use of a branch point (BP) different from the BP utilized by wild-type SF3B1. It is predicted that approximately half of the aberrantly spliced transcripts induced by SF3B1 mutations undergo degradation by a NMD pathway, resulting in

down-regulation of canonical transcripts and protein expression [155]. In an animal study, SF3B1-K700E mice developed macrocytic anaemia due to a terminal erythroid maturation defect, erythroid dysplasia, and long-term haematopoietic stem cell expansion. Aberrant 3' splice-site selection associated with increased NMD was observed in SF3B1-K700E myeloid progenitors. In addition, SF3B1 mutations cooperate with TET2 loss can cause an accelerated MDS phenotype [157].

Some of the defined targets of this NMD such as *ABCB7*, *SLC25A37*, *TMEM14C*, and *ALAS2* have provided some insights into a potential causal link between SF3B1 mutations and MDS-RS. SF3B1 mutation induces mis-splicing of *ABCB7* by aberrant 3' splice site usage that results in NMD, and subsequent downregulation of *ABCB7* mRNA and protein. *ABCB7* is a mitochondrial iron exporter responsible for maintaining iron homeostasis, and there is a strong correlation between MDS-RS and dysregulation of *ABCB7*. *SLC25A37* is also another iron transporter mis-spliced by mutant SF3B1 [130]. However, specific alternative splicing events observed in human cells were often absent in mouse models. In SF3B1-K700E knock-in mice, Mupo et al. found significant differences between aberrantly spliced transcripts in mouse versus human, despite the closely similar properties of human and mouse SF3B1 mutants. The murine orthologues of genes associated with ring sideroblasts in human MDS, including *ABCB7* and *TMEM14C*, were not aberrantly spliced in SF3B1-K700E mice. These significant difference of mis-spliced events between mouse and human could be

contributed to the relatively poor interspecies conservation of intronic sequences able to function as aberrant splice sites [158].

Recently, Pellagatti et al. and Shiozawa et al. performed a systematical and comprehensive transcriptome analysis of aberrant RNA splicing associated with the presence of splicing factor mutations, including *SF3B1* mutations, on large cohorts of MDS patient [148, 149]. Shiozawa et al. found small but widespread reduction of intron-retaining isoforms is the most frequent splicing alteration in *SF3B1* mutated samples. It was shown that the decreased intron retention was more pronounced in the cytoplasm compared with the nucleus of *SF3B1* mutant cells, suggesting an impaired nuclear export of intron-retaining transcripts. *SF3B1* mutation was also associated with 3' splice site alterations, leading to the most pronounced reduction of canonical transcripts. Target genes include tumor suppressors and genes of mitochondrial iron metabolism or heme biosynthesis [148]. Pellagatti et al. assessed the aberrant 3' splice site events associated with *SF3B1* mutation and similarly found a more pronounced reduction of gene expression levels for NMD-sensitive events. A reduction of intron-retaining isoforms was consistently observed in *SF3B1* mutant samples. It is reported that there was a concomitant usage of an alternative 3' splice site for several of the most affected retained intron events, suggesting that reduced intron retention might be linked to the ability of SF3B1 mutants to select an upstream 3' splice site [149].

3.2 *SRSF2* mutation

3.2.1 Clinical significance

SRSF2 is a member of the SR-rich protein family that contributes to both constitutive and alternative splicing by binding to ESE sequences within pre-mRNA through its RNA recognition motif domain (RRM). After *SF3B1*, it is the second-most commonly mutated splicing factor, which occurs in ~15% of MDS and 40%-50% of CMML [41, 43, 70, 73]. The majority of mutations are heterozygous missense substitutions at P95 codon, which is immediately downstream of the RRM domain, implying a gain of function. A small minority of in-frame insertions or deletions affects the same region [122, 130, 159]. *SRSF2* mutations co-occur with several other mutations, many of which are also frequently found in CMML, including *TET2*, *ASXL1*, *CUX1*, *IDH2*, and *STAG2*. In contradistinction to patients with *SF3B1* mutations, patients with *SRSF2* mutations tend to have more dysplasia in the granulocytic lineage and less in the erythroid lineage, and patients consequently tend to have a less prominent transfusion requirement, more enrichment in the RAEB subtypes. In contrast to the favorable prognosis associated with *SF3B1* mutations, patients harboring *SRSF2* mutations have a significantly shorter overall survival a greater risk of progression to AML [70, 71, 121, 160].

3.2.2 Biological mechanisms

In a mouse model study, Kim et al. showed that the SRSF2-P95H mutation confers a gain/alteration in function that results in key phenotypes resembling those seen in human MDS, which include expansion of haematopoietic stem and progenitor cell in heterozygous SRSF2-P95H mutant mice with impaired differentiation, altered cell cycle kinetics, and increased apoptosis resulting in peripheral cytopaenias and morphologic dysplasia [126]. Transcriptional analysis revealed that SRSF2 mutations cause alteration of normal ESE recognition activity in both human and murine cells. Mutant SRSF2 preferentially recognizes a C-rich CCNG motif versus a G-rich GGNG motif, whereas wild-type SRSF2 binds to both motifs with similar affinity. These alterations in the RNA-binding activity of SRSF2 promote or repress recognition of exons containing C-rich or G-rich ESEs [126, 161, 162]. Altered ESE recognition causes widespread splicing changes in hundreds of genes. Several of these downstream mis-spliced genes are themselves recurrent mutational targets in myeloid malignancies, such as *EZH2*. SRSF2 mutations promoted inclusion of a ‘poison’ cassette exon of *EZH2*, which undergoes NMD and results in reduced expression of *EZH2* protein and histone H3 lysine 27 trimethylation (H3K27me3) level in SRSF2 mutant cells. Importantly, defective haematopoiesis in SRSF2 mutant cells could be partially rescued by overexpression of normally spliced *EZH2* mRNA [126]. Loss-of-function mutations in *EZH2* occur in MDS, and *SRSF2* and *EZH2* mutations are known to be mutually exclusive in MDS [41, 43, 163]. Therefore, reduced *EZH2* expression associated with

the mutant SRSF2 may provide a mechanistic basis for this mutual exclusivity [126, 139].

A recent knock-in mice model study also described the abnormal RNA splicing and compromised HSC function induced by SRSF2-P95H mutation, contributing to the deregulated haematopoiesis that recapitulates the MDS phenotypes. Although the SRSF2 mutant mice did not develop MDS under the steady-state condition, when their stem cells were transplanted into lethally irradiated mice, the recipients developed anaemia, leukopenia, and erythroid dysplasia. RNA sequencing of the *SRSF2* mutated haematopoietic stem and progenitor cells (HSPCs) revealed a number of abnormal splicing events and differentially expressed genes, including several potential targets implicated in the pathogenesis of haematopoietic malignancies, such as *CSF3R*, *FYN*, *GNAS*, *NSD1*, *HNRNPA2B1*, and *TRP53BP1*. Among the mutant SRSF2 associated splicing events, the most commonly observed were the enhanced inclusion and/or exclusion of cassette exons, which were caused by the altered consensus motifs for the recognition of exonic splicing enhancers [164]. The effect of splicing factor mutations remains largely unexplored in the different lineages affected in MDS. Pellagatti et al. observed aberrant splicing of some regulating genes in bone marrow erythroid and myeloid precursors of *SRSF2* mutant MDS patients. For example, aberrant splicing of *CSF3R*, a critical regulator of granulopoiesis was identified in *SRSF2* mutant granulocyte precursors. Similarly, *CSF1R*, a controller of macrophage production and

function, was also found to be aberrantly spliced in *SRSF2*-mutated monocyte/macrophage precursors [149].

3.3 *U2AF1* mutation

3.3.1 Clinical significance

U2AF1 encodes an auxiliary factor in the U2 spliceosome that is responsible for recognizing the AG splice acceptor dinucleotide at the 3' end of introns. The frequency of *U2AF1* mutations is about 7–11% in MDS patients [41, 43]. Similar to other commonly mutated splicing factor genes, the two most common mutations are both heterozygous missense substitutions, again implying a gain of function, and both appear to alter sequence specificity of pre-mRNA binding and splicing [165]. The two hotspot mutations of *U2AF1* are in separate zinc finger DNA binding domains, one at codon S34 toward the N-terminal of the protein, and one at codon Q157 toward the C terminal. S34F mutations account for 60–80% of *U2AF1* mutations in MDS [69, 122]. Some studies have shown that mutations of *U2AF1* in MDS were associated with shorter survival and an increased tendency to evolve into secondary AML [69, 70, 121, 166].

3.3.2 Biological mechanisms

Transcriptomic studies revealed that aberrant cassette exon skipping or inclusion is the predominant abnormal splicing event induced by *U2AF1* mutations in MDS and AML [130]. *U2AF1* normally recognizes the AG-dinucleotide at the 3' splice site during the

early steps of splicing catalysis in a sequence-specific manner. Mutations in U2AF1 have been found to alter the 3' splice site recognition activity associated with different consensus nucleotides around the AG-dinucleotide at the 3' splice site [165]. The U2AF1-S34F mutation recognizes the thymidine (uridine) at the -3 position of the 3' splice site less frequently than cytosine compared with wild-type U2AF1. The U2AF1-Q157 mutation is associated with preferential recognition of guanine instead of adenosine at the +1 position at the 3' splice site [165, 167]. The different recognition preference for 3' splice site consensus sequences may provide an explanation for the different effects of the S34 and Q157 mutations on pre-mRNA splicing [143, 165, 167]. Moreover, these studies also showed that U2AF1 mutations in AML result in differential splicing of genes and dysregulated pathways involved in cellular processes such as epigenetic regulation, DNA damage response, pre-mRNA splicing, RNA localization and transport, and cell cycle [143, 150, 167]. Shirai et al examined in vivo haematopoietic consequences of the U2AF1-S34F mutation using a doxycycline-inducible transgenic mouse model. Mice expressing mutant U2AF1-S34F display several phenotypes associated with MDS, including leukopenia, expansion of myeloid progenitor cells, abnormal mature cell lineage distribution and increased apoptosis in the bone marrow. However, these mice do not show dysplasia, nor do they develop MDS or AML. Integration with human RNA-seq datasets determined that common mutant U2AF1-induced splicing alterations are enriched in RNA processing genes,

ribosomal genes, and recurrently mutated MDS and acute myeloid leukaemia-associated genes [125].

However, none of these previous studies have shown that an alternatively spliced mRNA induced by U2AF1 mutations was functionally linked to the transformed phenotype. Yip et al. revealed the functional roles of U2AF1-S34F mutation in lineage-specific splicing alterations in MDS. U2AF1-S34F expression in human haematopoietic progenitors impairs erythroid differentiation and skews granulomonocytic differentiation toward granulocytes. RNA-seq analysis showed that U2AF1-S34F induced a higher number of cassette exon splicing events in granulomonocytic cells than in erythroid cells. Functional studies using human haematopoietic progenitors expressing aberrant isoforms of H2AFY and STRAP resulted in recapitulated phenotypes associated with U2AF1-S34F expression in erythroid and granulomonocytic cells, suggesting a causal link. Interestingly, modulated isoform of H2AFY and STRAP rescued the erythroid differentiation defect in U2AF1-S34F MDS cells, suggesting that splicing modulators could be used therapeutically [151]. A recent study by Park et al. proposed a model of oncogenic activity for U2AF1-S34F mutations. RNA-Seq analysis of a murine cell line stably expressing the U2AF1-S34F mutation showed that the most common type of aberrant RNA processing event associated with the mutant U2AF1 was increased usage of a distal cleavage and polyadenylation (CP) site. Increased distal CP site use of a specific

pre-mRNA results in a longer 3' UTR of *ATG7*, which encodes an essential autophagy factor. Previous studies have shown that mice with haematopoietic cells lacking *ATG7* develop an MDS-like syndrome [168]. Park et al. found that the longer *ATG7* mRNA was inefficiently translated, which led to autophagy defect that predisposes cells to secondary mutations and MDS transformation. Moreover, increased use of the *ATG7* distal CP site was similarly observed in bone marrow cells of MDS and AML patients harboring the U2AF1-S34F mutation [169].

4. Spliceosome mutations and hyperactivation of NF- κ B signalling

Lee et al. reported that *SF3B1* and *SRSF2* mutations result in hyperactive NF- κ B signalling via aberrant splicing of specific target genes. They found that *SF3B1* mutation promoted mis-spliced events of *MAP3K7*, for which an intron-proximal 3' splice site was used by mutant *SF3B1* in both human and mouse haematopoietic cells. This aberrant 3' splice site recognition is predicted to result in an out-of-frame transcript that undergoes nonsense-mediated decay. Consistent with this, a reduction of *MAP3K7* protein level was observed in isogenic cell lines as well as haematopoietic cells from MDS and CLL patients with *SF3B1* mutations compared to those without splicing-factor mutations. *MAP3K7* encodes a kinase that mediates tumor necrosis factor α (TNF α), interleukin-1 β (IL-1 β), and Toll-like receptor signalling through the NF- κ B, JNK, and MAPK pathways. The effects of *MAP3K7* loss have been extensively studied and can result in loss or promotion of inflammation depending on cellular context. It

has been suggested that SF3B1-K700E mutant human myeloid or lymphoid leukaemia cells stimulated with LPS had enhanced NF- κ B activation compared with SF3B1-WT controls. Functional studies confirmed NF- κ B signalling hyperactivation following shRNA-mediated downregulation of MAP3K7 in NALM-6 and K562 cells, and reduced NF- κ B signalling following re-expression of MAP3K7 in K562 SF3B1-K700E cells. Overall, these data suggest that the effects of SF3B1-K700E mutation on induction of NF- κ B signalling are mediated through aberrant splicing of *MAP3K7*. Furthermore, *MAP3K7* aberrant splicing was restricted to *SF3B1* mutant cells but was not found in *SRSF2* mutant cells. Interestingly, Lee et al, found aberrant splicing of *CASP8*, a gene encoding a key activator of NF- κ B, that was recurrently mis-spliced in AML and CMML patients with *SRSF2* mutations, but not those bearing *SF3B1* mutations. *SRSF2* mutations were associated with exclusion of a cassette exon resulting in an mRNA encoding a truncated CASP8 isoform that promotes NF- κ B signalling [170].

In a recent study, Smith et al. performed a global analysis of exon usage in AML samples revealed distinct molecular subsets containing alternative spliced isoforms of innate immune and NF- κ B signalling, correlating with worse AML prognosis. interleukin-1 receptor-associated kinase 4 (IRAK4), a serine/threonine kinase that mediates signalling downstream of toll-like receptors, was the dominant alternatively spliced isoform in U2AF1 mutant MDS cells. Aberrant *IRAK4* splicing is characterized

by a longer isoform that retains exon 4, encoding the full-length protein, IRAK4-Long (IRAK4-L) that interacts with MyD88 facilitating the assembly of the myddosome complex and results in maximal activation of NF- κ B as well as MAPK activation. Expression of IRAK4-L is mediated by mutant U2AF1 and is associated with oncogenic signalling in MDS/AML. Inhibition of IRAK4-L suppressed leukaemic growth, particularly in AML cells with higher expression of the IRAK4-L isoform. Inhibiting IRAK4 by small molecule treatments in U2AF1-mutant MDS PDX showed a decrease in MDS cell engraftment and negligible engraftment of MDS cells in secondary recipients. Collectively, *U2AF1* mutations in MDS/AML induce the expression of therapeutically targetable isoform of IRAK4-L and provide a genetic link to activation of chronic innate immune signalling in MDS and AML [171].

Interestingly, in another study, Pollyea et al. reported that mutations in *SF3B1*, *U2AF1*, and *SRSF2* enhance NF- κ B activity and LPS-induced inflammatory cytokine production in macrophages, patient-derived cell lines, and mouse and human myeloid cells. These observations suggest that MDS-associated spliceosome mutations might enhance innate immunity, which could contribute to disease pathogenesis [172]. Previous studies have demonstrated that *MyD88* gene encodes two isoforms, a long isoform (MyD88-L) plays a role in activating NF- κ B signalling, and a shorter isoform (MyD88-S) inhibits NF- κ B signalling [173]. Pollyea et al. found that *SF3B1* knockdown led to increased production of anti-inflammatory MyD88s, which reduced

inflammatory cytokine production. U2AF1 or SRSF2 knockdown in RAW264.7 macrophages also increased MyD88s mRNA levels without altering MyD88l. However, none of the three spliceosome mutations altered MyD88 splicing in K562 cells, suggesting that altered splicing of other TLR pathway genes drives the enhanced inflammatory cytokine production in these cells. While knockdown of spliceosome genes weakens innate immunity, expression of mutant spliceosome genes induced the opposite phenotype. This is consistent with studies indicating that these mutations are largely neomorphic rather than loss-of-function. By generating mouse models inducibly expressing U2AF1-S34F, it has been shown that IL6 expression was significantly increased in mature myeloid cells carrying *U2AF1* mutation but was unchanged or even slightly decreased in *U2AF1* mutant haematopoietic stem and progenitor cells. Moreover, this study confirms the conclusion of altered splicing of MAP3K7 or CASP8 that associates with enhanced innate immune signalling in spliceosome mutant cells by Lee et al [170, 172].

Increased inflammation in MDS may be a common feature that is not limited to patients with spliceosome mutations. Increased inflammatory signal is thought to impair normal haematopoiesis and contribute to the pathogenesis by altering HSPC survival, proliferation, differentiation, and apoptosis [174]. Specifically, NF- κ B signalling is known to play a key role in the determination of the quiescence versus the active state of HSCs, and that constitutive activation of NF- κ B in HSCs leads to pancytopenia and

bone marrow failure [175]. Overall, these studies provide evidence that splicing of many genes including *MAP3K7*, *CASP8* and *IRAK4* contributes to the enhanced innate immune signalling induced by SF3B1, SRSF2 and U2AF1 mutations.

5. R-loop formation is associated with splicing factor mutations

Chen et al. firstly reported that splicing factors mutations, including *SRSF2* and *U2AF1*, can cause cell growth defects by promoting the invasion of nascent RNA into template DNA, leading to the formation of R-loops (aberrant RNA:DNA hybrid structures). R-loops induce increased DNA damage, replication stress and activation of ATR-Chk1 pathway. It was further demonstrated that enhanced R-loops result from impaired transcription pause release because the mutant SRSF2 loses its ability to extract the RNA polymerase II (Pol II) C-terminal domain kinase, the positive transcription elongation factor complex (P-TEFb), from the 7SK complex. Elevated R-loops are linked to compromised proliferation of bone-marrow-derived blood progenitors, which can be partially rescued by RNase H overexpression, suggesting a direct contribution of augmented R-loops to the MDS phenotype [176]. A study by Nguyen et al. showed that expression of the U2AF1-S34F mutation in human cancer cell lines leads to accumulation of R-loops. ATR inhibitors promoted DNA damage and cell death in U2AF1-S34F expressing cells, with spliceosome inhibitors enhancing these effects, suggesting that ATR may be a novel therapeutic target in patients with MDS carrying the U2AF1 mutation and potentially other malignancies harboring spliceosome

mutations [177]. Similarly, a recent study by Singh et al. has shown that mutations induce the accumulation of R-loops and associated DNA damage, resulting in activation of the ATR pathway in MDS and leukaemia cells. The suppression of R-loops rescued cellular defects including DNA damage and ATR-Chk1 activation. It is also been suggested that Chk1 inhibitor, alone or in combination with splicing modulators, may represent a novel therapeutic strategy to target splicing factor mutant MDS and leukaemia [178]. Interestingly, aberrant splicing of several genes involved in the regulation of R-loop formation, including *SETX* and *ATR*, was shown in CD34+ cells of splicing factor mutant MDS patients [149]. R-loop induced DNA damage may give rise to deleterious mutations contributing to the clonal advantage of these cells.

6. Spliceosome Inhibitors

6.1 Categories of spliceosome inhibitors

The discovery of mutations in splicing factors in myeloid leukaemias has prompted interest to understand if malignant cells bearing splicing factor mutations might have preferential sensitivity to compounds that modulate splicing. The rationale for these studies is the aforementioned fact that splicing factor mutations occur in a mutually exclusive manner in cancer and consistently occur in a heterozygous state, suggesting that homozygous splicing factor mutations are likely to be lethal to the cell and that cells harboring splicing factor mutations cannot tolerate further perturbation to normal pre-mRNA splicing catalysis [130]. Recent work has revealed that the survival of

splicing factor mutant cells depends on presence of the wildtype allele [179-181]. Pharmacological inhibition of the spliceosome is expected to result in global splicing inhibition both in splicing factor mutant and wildtype cells [182]. Wildtype cells are able to tolerate a certain degree of inhibition of the spliceosome because they still have sufficient canonical splicing products, while splicing factor mutant cells have a higher proportion of aberrantly spliced mRNAs and no longer have enough canonical mRNAs to survive [141, 183] (Figure 1.7).

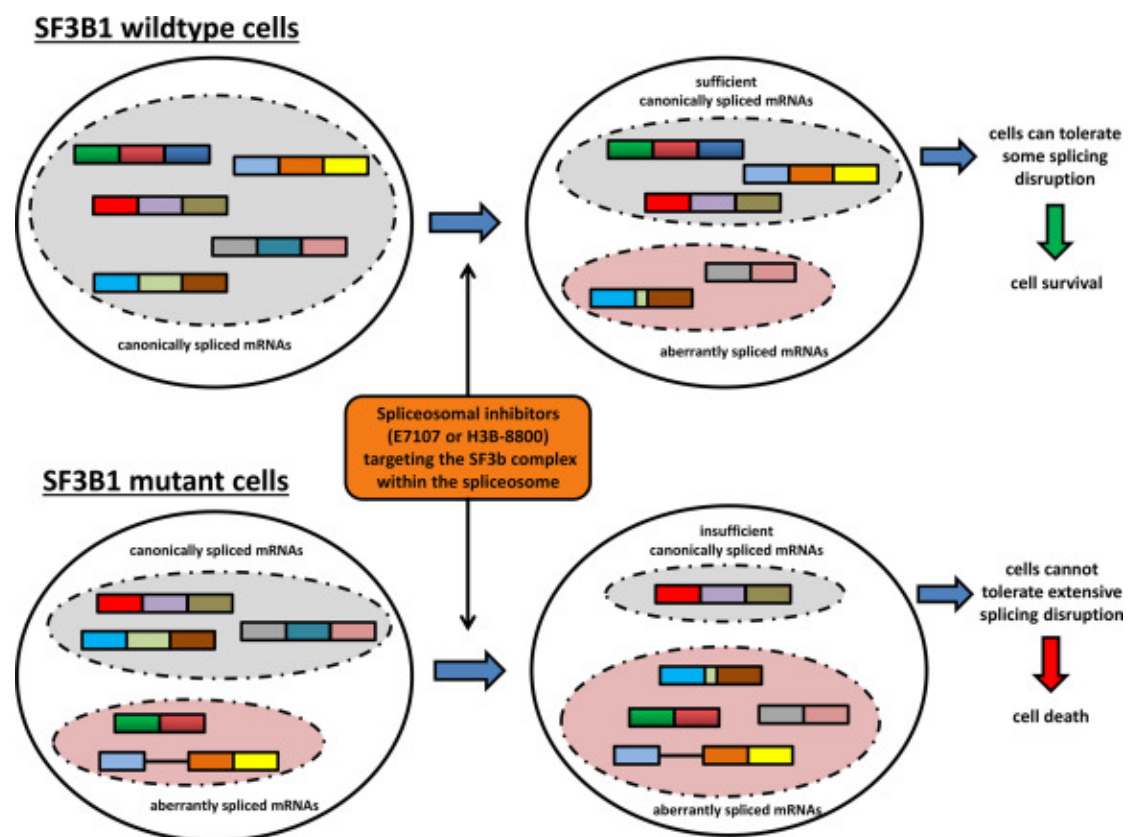


Figure 1.7. Rationale for the use of spliceosome inhibitors in splicing factor mutant cells (Adapted from Pellagatti et al [141]).

The assembly of the spliceosome is a complex process that involves over 100 proteins that potentially could be therapeutically targeted with small molecules. Several natural compounds derived from bacteria and their analogues have been identified to target the SF3B complex, interrupting early stages of spliceosome assembly [183]. These inhibitors can be grouped into three generalized families that are based on founding compounds that have structural distinctions: FR901464, pladienolide and herboxidiene. The FR901464 family includes spliceostatin A (SSA), Meayamycins, Sudemycins and Thailanstatins. The architecture of these compounds is based on two highly functionalized tetrahydropyran rings connected by a diene chain and differs the most in structure relative to the other two compounds. The analogous compounds of the pladienolide B family, including pladienolide B, pladienolide D, FD-895 and E7107, are polyketides with a large macrolide ring. The herboxidiene family of compounds, including GEX1A (herboxidiene), GEX1Q1-5 and 18-deoxyherboxidiene (RQN-18690A), has a similar structure, but with a smaller tetrahydropyran ring [184]. Despite these structural differences, the compounds exhibit similar effects on spliceosome assembly, presumably by targeting SF3b and modulating splicing function in cells [185].

However, the mechanism by which splicing inhibitors interact and target SF3B1 within the SF3b subcomplex is incompletely understood. A study by Corrionero et al. revealed a mechanism that SSA prevents the SF3b subcomplex from binding the pre-mRNA,

thus preventing the recruitment of U2 snRNP to the 5' of the branch point, and influencing splicing fidelity by causing inaccurate 3' splice site selection [186]. Similarly, E7107 also prevents spliceosome assembly by blocking the SF3b complex from binding the pre-mRNA branch point [187]. Teng et al. used site-directed mutagenesis and X-ray crystallography to show that the SF3b components, PHF5A and SF3B1, form a central node for binding to splicing modulators. Moreover, PHF5A-Y36, SF3B1-K1071, SF3B1-R1074 and SF3B1-V1078 were identified as key residues for the binding and efficacy of tested splicing inhibitors (GEX1A, E7107, SSA and Sudemycin D6). Mutations in these residues confer resistance to splicing modulators, however these mutation have not been reported in patients with myeloid malignancies [188]. There is mounting evidence that splicing inhibitors are selectively lethal to tumor cells but display less toxicity to normal cells in vitro. Splicing inhibitors are typically effective at inhibiting cancer cell growth at nanomolar concentrations. Furthermore, doses of splicing inhibitors that kill tumor cells in mouse xenografts have minimal toxicities on the host animal [183, 184].

6.2 Spliceosome inhibitors in pre-clinical models of MDS

Several studies have investigated the therapeutic potential of spliceosome inhibitors for treating myeloid malignancies harboring splicing factor mutations in pre-clinical models. Two of these compounds, E7107 and H3B-8800, have demonstrated potential therapeutic applicability, which has led to rapid clinical interest.

In vivo treatment of E7107 in isogenic leukaemic SRSF2P95H/+ mice (transformed by MLL/AF9 over-expression) showed a substantial reduction in leukaemic burden and longer survival compared to leukaemic SRSF2+/+ mice. Patient-derived xenografts (PDXs) were also generated using primary leukaemia cells from AML patients with or without splicing factor (*SRSF2* or *ZRSR2*) mutations. Similarly, when treated with E7107, a higher reduction in leukaemic burden was shown in splicing factor mutant PDX mice compared to wildtype PDX mice. Whereas E7107 treatment of mice resulted in widespread intron retention and cassette exon skipping in leukaemic cells regardless of *SRSF2* genotype, the magnitude of splicing inhibition following E7107 treatment was greater in SRSF2 mutant than in SRSF2 wild-type leukaemia, consistent with the differential effect of E7107 on survival [128, 181]. In another study, Obeng et al. demonstrated that HSPCs expressing SF3B1-K700E have an increased sensitivity to E7107 both in vitro and in vivo [157].

Sudemycins have also undergone pre-clinical evaluation as therapeutics of myeloid malignancies. Shirai et al. showed that haematopoietic cells expressing mutant U2AF1-S34F, including primary patient cells, have an increased sensitivity to in vitro Sudemycin treatment relative to controls. In vivo sudemycin treatment of U2AF1-S34F transgenic mice alters splicing and reverts haematopoietic progenitor cell expansion induced by mutant U2AF1 expression. In addition, RNA-seq analysis showed that the effects of Sudemycin D6 treatment on splicing and gene expression can be independent

of the effects of mutant U2AF1 expression, indicating that the increased sensitivity of mutant U2AF1 cells to Sudemycin D6 treatment may be due to cumulative aberrant splicing effects caused by drug treatment and mutant U2AF1. This study strongly suggested that Sudemycins predominantly modulate splicing changes rather than inhibit splicing [189].

More recently, Seiler et al. identified a novel, orally available modulator of the SF3b complex, H3B-8800, which potently and preferentially kills spliceosome-mutant haematologic tumor cells. The killing effects of H3B-8800 are due to its direct interaction with the SF3b complex, as evidenced by loss of H3B-8800 activity in drug-resistant cells bearing mutations in genes encoding SF3b components. Although H3B-8800 binds and modulates splicing catalysis of both wild type and mutant SF3B1 spliceosomes in in vitro assays of biochemical splicing, H3B-8800 showed preferential killing of SF3B1-mutant cells. Furthermore, in vivo studies suggested that H3B-8800 treatment inhibited tumor growth in NSG mice bearing xenografts of isogenic K562 cells with SF3B1-K700E, and reduced the leukaemic burden in mice xenografted with cells from a SF3B1-K700E mutant AML patient or with CD34⁺ cells from SRSF2-P95H mutant CMML patients. RNA-seq analysis of isogenic SF3B1-K700E K562 cells treated with H3B-8800 showed that retention of short, GC-rich introns are enriched for genes encoding spliceosome components, suggesting that the preferential killing of

splicing factor mutant cells is due to the dependency of these cells on normal pre-mRNA splicing [190, 191].

6.3 Clinical Trials of Splicing Inhibitors in MDS

The safety of E7107 treatment has been previously evaluated in two parallel phase I clinical trials involving patients with advanced solid tumors. Twenty-six patients with advanced solid tumors were enrolled in an open-label, single-arm and dose-escalation study. E7017 treatment was generally well tolerated with manageable systemic side effects, with the most common adverse events noted were nausea, vomiting, diarrhea, and fatigue. However, severe vision loss was noted in two patients that was possibly related to the E7107 treatment, leading to study termination. The exact underlying mechanism for the visual toxicity remains unclear; however, in the patient with permanent visual loss, it is believed that the optic nerve was the target of the toxicity. In PBMCs from patients in this study, an increase in unspliced introns was clearly shown and suggests that E7107 may globally inhibit spliceosomes [192]. In another study, forty patients with solid tumors refractory to standard therapies or with no standard treatment available were enrolled. Reasons for study drug discontinuation were primarily lack of efficacy (72%) or the occurrence of adverse events (15%). Frequently occurring side effects were mainly gastrointestinal. One patient experienced reversible grade 4 blurred vision that improved once E7107 treatment was stopped (and following high-dose corticosteroid treatment) [193]. It is not known whether vision

toxicity is a direct and general effect of SF3b inhibition or specifically associated with E7107. The development of infrequent, but severe ophthalmologic adverse events, at least some of which were reversible, currently hampers further clinical development of E7107.

A phase I clinical trial (NCT02841540) aiming to explore the safety, pharmacokinetics and pharmacodynamics of H3B-8800 in patients with MDS, AML and CMML harboring splicing factor mutations is under way. As of June 16th, 2019, 84 patients were enrolled at 24 centers in the US and Europe. Given the above-mentioned toxicity associated with E7017, patients with a known prior or ongoing retinal or optic nerve disease were excluded from the H3B-8800 trial. Preliminary results from this first-in-human trial suggested the favorable safety profile of H3B-8800. The most observed treatment related adverse events were Grade 1 or 2, including diarrhea, vomiting, QTc prolongation, nausea and fatigue. No ophthalmic adverse events were observed. Although no objective CR or PR were achieved, decreased red blood cell or platelet transfusion requirements were observed in 12 (14%) of enrolled patients [194].

C. Drug repurposing and connectivity mapping

1. Defining Drug repurposing

Traditional drug development strategy is costly, time consuming and burdened by a very low success rate. At present, the conventional de novo drug discovery process requires an average of about 13-15 years and \$2-3 billion to approve and launch a drug with a low success rate of approximately 2% only [195, 196]. Drug repurposing (also called drug repositioning, reprofiling or re-tasking) is a strategy for identifying new uses for approved or investigational drugs that are outside the scope of the original medical indication [197]. It mostly involves developing approved, discontinued, shelved and experimental drugs. Drug repositioning has helped to mitigate failures in drug discovery and has been associated with therapeutic breakthroughs. In recent years, approximately 30% of newly FDA approved drugs result from drug repurposing, and repurposed drugs currently generate around 25% of the annual revenue for the pharmaceutical industry [198]. This strategy decreases the overall cost of bringing a repurposed drug to market, which has been estimated to be \$300 million on average, compared with an estimated \$2–3 billion for a new chemical entity. Drug repurposing offers various advantages over developing an entirely new drug for a given indication, which dramatically reduces drug development time, cost and safety risk [199, 200]. The main advantage of repurposed candidates is that in many cases the safety and pharmacokinetic profiles of the repositioned candidates are already established in preclinical models and at early-stage trials in humans, thus being less likely to fail from

a safety point of view in subsequent efficacy trials. In the case of approved drugs, the data including safety, bioavailability, route and formulation of administration, pharmacokinetic and pharmacodynamic (PK-PD) profile, and associated adverse effects have been fully evaluated to pass clinical trials and regulatory process, and have already undergone post-marketing surveillance. Hence, the repurposing candidates have been 'de-risked' during the development stages of new therapeutic indications, which accelerates the evaluation process and thus saves money, although this will vary greatly depending on the development stage and process. Also, repurposed drugs may reveal new targets and pathways that can be further exploited [201].

Historically, most cases of successfully repositioned drugs were realized through serendipitous observations or unorganized discovery processes. Once a drug was found to have an off-target effect or a newly recognized on-target effect, it was taken forward for repositioning exploitation, for example, sildenafil, minoxidil, aspirin, valproic acid are some of best-known repurposed drugs. These successes have encouraged the development of more systematic approaches to identify repurposable compounds and associated mechanism of action. In drug repurposing, three major steps are involved, which consist of (1) identification of target candidate for new indication; (2) exploration of mechanism or signalling pathway involved in drug or disease; (3) proving the efficacy of drug in phase 2 and 3 clinical trials. Among all steps, identification of lead candidate is one of the most critical steps. This is the step where the most sophisticated

and systematic approaches are used for generation of new hypothesis in drug repurposing. It has been widely recognized that most small-molecule drugs interact with more than one target protein [202]. The main issue in drug repositioning is understanding the polypharmacology, and therefore, detecting novel drug-disease connections. To address this issue, various experimental and computational approaches have been proposed for the identification of repurposable drug candidates. Experimental approaches, such as chromatography or mass spectrometry based affinity binding assays, have been used to identify drug-target interactions [203]. High-throughput direct binding or catalytic assays, in which small-molecule–kinase binding is analysed in a kinome-wide fashion using a variety of in vitro and increasingly organism-based assays, are used to generate heat maps of biologically important interactions [204, 205]. Phenotypic screening, either cell-based or organism-based experimental method, have also been utilized in repositioning studies to identify compounds having disease-relevant effects in model systems without prior knowledge of the targets affected [206]. Computational approach is further subclassified as either drug-centric or disease-centric. In the drug-centric approach, new indications are found for existing drugs, while in the disease-centric approach, new drugs are identified for an existing disease [196]. Computational strategies involve systematic analysis of data of any type, which can then lead to the formulation of repurposing hypotheses. Recent advances in technology such as next-generation sequencing and high-throughput screening methods have enabled the process of generating and integrating large

quantities of multi-omics database and public chemical database, including genomics, transcriptomes, proteomics, metabolomics, chemical structure and many more, that could be useful for computational evaluation of drug repurposing. Various computational approaches can be used individually or in combination to systematically analyse different types of large-scale data to obtain meaningful interpretations for repurposing hypotheses. The most commonly used computational approaches are shown in Table 1.5.

Table 1.5 Computational methods used in drug repurposing (Adapted from Pushpakom et al. [201])

Methods	Description	Reference
Signature matching	Comparing the ‘signature’ of a drug, characteristics such as its transcriptomic, structural or adverse effect profiles, with that of another drug or disease phenotype	[207-212]
Computational molecular docking	Structure-based computational strategy to predict binding site complementarity between a ligand (small molecule or antibody) and a therapeutic target (protein)	[213-216]
Genome-wide association studies	Genes that are associated with a disease may prove to be potential drug targets	[217, 218]
Pathway or network mapping	Network analysis using genetic, protein or disease data can aid identification of repurposing targets	[219-222]

Retrospective clinical analysis	Systematic analysis of EHRs, clinical trial data and post-marketing surveillance data could inform drug repurposing	[222-224]
Novel data sources	Large-scale in vitro drug screens with paired genomic data, biobank and self-reported patient data are novel methods to exploit for drug repurposing	[225-230]

These systematic drug repositioning strategies based on in silico methods have been successfully applied for the identification of a number of promising candidate drugs, some of which are in advanced stages of clinical trials, with the potential for use in the treatment of both common and rare diseases. For example, Sanseau et al. used genome-wide association studies (GWAS) data to identify targets for potential repurposing of antibodies and small molecules drugs in multiple therapeutic areas, including cancers, cardiovascular diseases, neurological disorders, immunological and inflammatory conditions and respiratory disorders [217]. Specifically, signature matching based method is one of the most commonly used computational approaches. Signature matching is based on the comparison of the unique characteristics or 'signature' of a drug against that of another drug, disease or clinical phenotype [211, 231]. The signature of a drug could be derived from three general types of data: transcriptomic (RNA-seq or Microarray), proteomic, metabolomic data or chemical structures. Here,

we will focus on transcriptomic signature for drug repurposing, which is discussed in the following sections.

2. Application of transcriptome signature in drug repositioning

In the last few years, advances in bioinformatics suggest that any perturbation (e.g. drug treatment, genetic alteration) to the transcriptome of the cell can be summarised by a proper ‘signature’: a pattern of altered gene expression defined by a set of differentially expressed genes. Gene expression signatures have been assembled and compared based on transcriptome-wide similarity metrics, unveiling ‘connections’ among drugs, pathways and phenotype of disease by matching corresponding signatures of disease and drug responses [200, 232]. By using gene expression data, it has been possible to infer new potential therapeutic applications for drug repositioning. The strategy of matching transcriptomic signatures is based on the principle that shared gene expression patterns, between compounds and disease or between different compounds, could hint at potential efficacy or similarity of MoA [199, 233].

In drug–disease comparisons (estimating drug–disease similarity), one promising strategy is called ‘signature reversion’ which aims to identify drugs which induce a transcriptional response opposite to that of a given disease (that is, the drug signature will be closer to that obtained for the healthy state). Drugs identified using this approach may be capable of reverting the gene expression signature, and hence the phenotype of

a disease [200, 232]. The transcriptional profile of a particular drug is derived by comparing the gene expression profile of cells or tissue, before and after treatment with the drug. The drug-induced differential expressed genes (the transcriptional signature of the drug) can then be compared with a disease-associated expression profile that has been similarly obtained through differential expression analysis of disease versus healthy conditions. If they are negatively correlated (i.e. the genes upregulated in the disease are downregulated with the drug and vice versa), then it is reasonable to hypothesise that the drug may have a potential effect on the disease [234]. Despite this principle being quite simplistic, it has been successfully employed and experimentally validated to identify drugs that could be repositioned in a wide range of therapeutic areas from cancer to neurodegenerative disorder and metabolic disorders [235-237].

Alternatively, drug-drug similarity approaches aim to identify shared MoA of otherwise dissimilar drugs (drugs that belong to different classes or that are structurally dissimilar). A shared transcriptomic signature between two drugs could therefore imply that they also share a therapeutic application, regardless of the similarity or dissimilarity in their chemical structures [238]. This principle has proved effective even when comparing transcriptional signatures that are reflective of a secondary mode of action, which could be shared, for example, by a group of pharmacologically diverse mild correctors of a given disease phenotype [239].

3. Connectivity mapping methods and platforms

The Connectivity mapping (CMap) was introduced by Lamb et al. in 2006 [207]. CMap is a collection of comprehensive, large-scale drug perturbation databases containing gene expression profiles of dozens of cultured human cells treated with a large number of chemical compounds serving as reference databases. A query gene signature is a list of highly representative phenotype-specific genes generated from biological experiments investigating a particular biological condition. Differentially expressed genes between disease and normal conditions are often used to form a signature for the disease. The basic concept of CMap is to use a pattern-matching algorithm to compare a reference database containing drug-specific gene expression profiles with a disease-specific gene signature, which enables the discovery of connections, similarities or dissimilarities among drugs, genes and diseases. The CMap method is performed by simply submitting a list of genes thought to be relevant to a particular disease (i.e. disease-specific gene signature), then the reference dataset of drug-specific expression profiles can be queried with a disease gene signature to obtain a list of compounds that might revert or mimic those gene expression changes. The extent of positive or negative correlation between the gene expression signature of the drug and that of the disease can be quantitatively assessed and ranked by algorithms [207, 240]. Based on the ‘signature reversion’ principle, drugs that significantly induce gene expression in a direction opposite to that observed in the disease state, might represent new therapeutic agents [207, 240]. The principal workflow of CMap is shown in Figure 1.8.

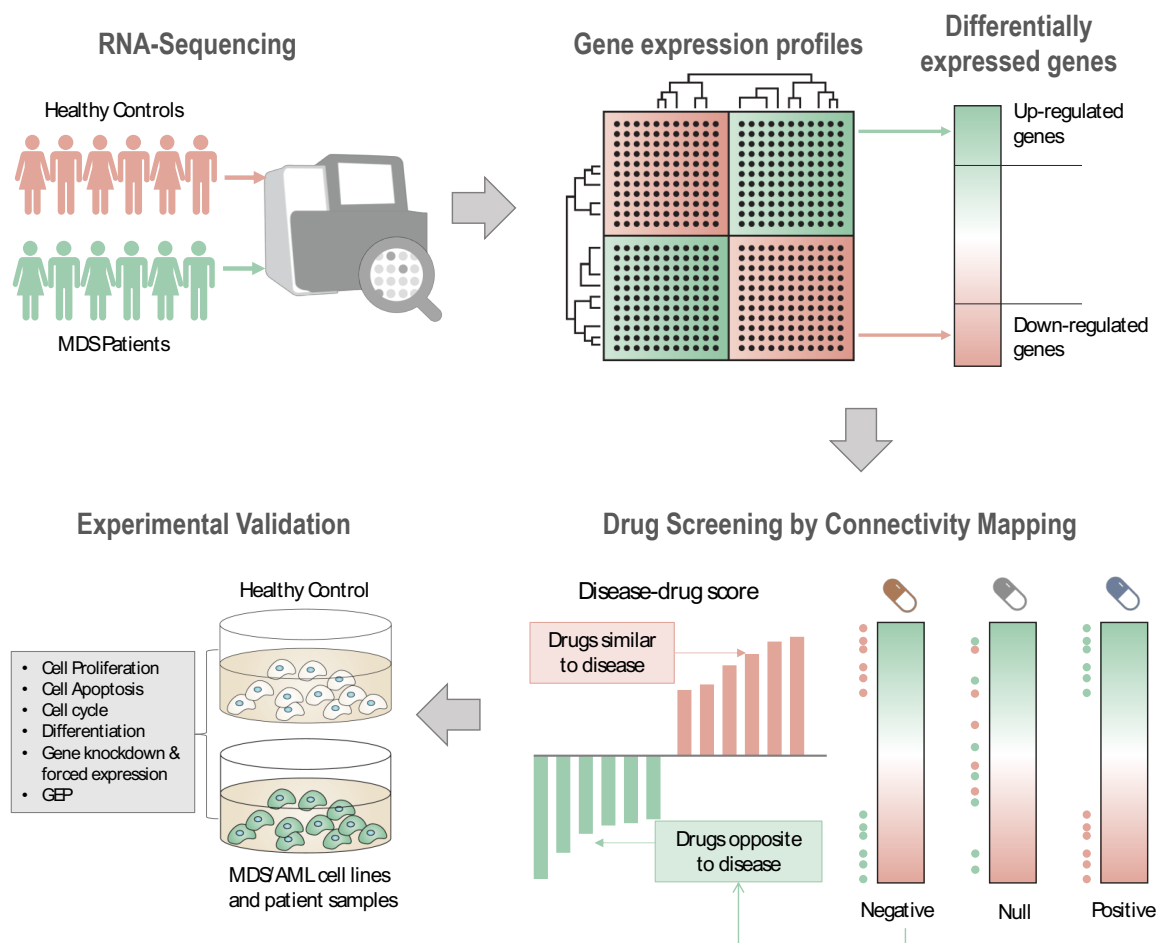


Figure 1.8 Workflow of GEP-based drug screening through connectivity mapping.

Connectivity Map (CMap) database is a collection of gene expression profiles derived from the treatment of cultured human cells with a large number of compounds. RNA-Seq or microarray based GEP datasets of patients are applied to determine the lists of DEG for CMap queries. Identified drugs that have high potential to reverse the deregulated gene signatures might be repositioned as potential new therapeutic agents.

The original version of CMap database (CMap Build 1.0) developed by Lamb et al. is a collection of paired gene expression profiles representing a series of structured microarray experiments. Microarray platform based gene expression profiles were collected for 165 small molecules applied to five human cell lines (MCF7, PC3, SKMEL5, HL60 and ssMCF7) at varying concentrations (42 different concentrations) against vehicle controls, and gene expression was measured at two time points for a

total of 455 instances. In Build 2.0 (broadinstitute.org/cmap), the initial database was expanded to cover in total 6100 instances with more drugs (1309, mostly FDA-approved) and concentration (156) but the same cell lines, for a parallel series of analysis. To compare the similarity between signatures, a so-called connectivity score is computed using the gene set enrichment analysis method [241], which is an adaptation of the Kolmogorov–Smirnov test for comparing a ranked list to an unordered reference set of matching elements. The resultant ‘connectivity score’ is normalized from -1 to $+1$ to reflect the connection between the expression profiles. A positive connectivity score, obtained for having most of the downregulated genes at the top of the reference profile and most of the upregulated genes at the bottom, represents the degree of similarity between a query signature and a reference profile (i.e. drug treatment can mimic the expression pattern of a disease state). In contrast, a negative connectivity score stands for a reversed mapping, obtained for having that most of the upregulated genes are at the bottom of the reference profile and most of the downregulated genes are at the top. A null connectivity score occurs when the up- and downregulated genes are randomly distributed over the reference profile. Overall, the results are obtained as a list of connectivity scores for all small molecules in the reference database, one connectivity score for each small molecule. Finally, the top-scoring drugs are selected by sorting all connectivity scores in descending order and identifying a relevance threshold [242]. Query gene signatures are created by end users as input to connectivity mapping analysis, so signature quality is critically important to

the success of the process. However, so far relatively little research has been done on methods for creating high standard gene signatures for connectivity mapping, in particular when they are to be generated from multiple datasets. The guidance provided from the original developers of CMap suggested that they found gene signatures with between 10 and 500 genes performed well. Users of connectivity mapping normally use their own experience and knowledge to create their own query gene signature [243]. Although there is no precise way of creating optimal gene signatures for CMap query, the conventional approach is to identify and use the DEG that are statistically significant displaying an association with a given phenotype. Ideally, the gene signature list is suggested to generate by using the microarray-based gene expression profiles to fit the current CMap build. Zhang et al. assessed the feasibility of incorporating RNA-Seq data into the current connectivity mapping framework by using adapted RNA-seq gene signature to query the microarray-based reference profiles. This allowed for the establishment of connections between the RNA-seq gene signature and those microarray reference profiles, alleviating the associated incurring cost of re-creating drug profiles with RNA-seq technology [244].

Connectivity mapping emerged as an alternative to the mainstream structural and molecular biology approaches for drug discovery. Instead of aiming to discover individual biomarker or target bound by small molecules, CMap approaches focus on global effects of a compound on the cells, as measured by gene expression changes,

providing a proxy to predict the phenotypic effects of the compound on cells and revealing uncharacterized MoA of small molecules. The CMap resource has been successfully applied to identify novel therapeutics for repositioning in various contexts, including leukaemia [245], solid tumours [246-248], obesity [249], and inflammatory bowel disease [250]

4. Connectivity Mapping Resource, variations and extensions

As discussed, there are three key components of a connectivity map resource 1) a large reference database containing gene expression profiles representing a change in cellular state in response to a drug, a disease, or a type of systematic perturbation; 2) a query signature; and 3) a computational method to relate the query signature with the reference database to discover connections [200]. Since the initial success of the CMap principle and methodology, there have been numerous applications of this approach by many research groups with a particular focus in drug discovery and development. The original build of CMap database provided a proof-of-concept tool, in which genes, drugs, and disease states are connected by virtue of common gene-expression signatures. Efforts to extend or scale up the CMap resources were also undertaken by many groups. For example, Zhang et al. developed a statistically significant connections map, termed sscMap (<http://purl.oclc.org/NET/sscMap>), providing a simple and robust method for connectivity mapping with increased specificity and sensitivity [208, 251]. The benefits of the method implemented in this application include a more principled statistical

procedure, effective safeguards against false connections, and an increased sensitivity. sscMap analysis led to the identification of entinostat as a candidate drug for the treatment of HOX-TALE-related leukaemia [252]. Mode of Action by NeTwoRk Analysis (MANTRA) is another cMAP extension that exploits similarity in gene expression profiles following drug treatment, across multiple cell lines and dosages, to predict similarities in drug effect and MoA (<http://mantra.tigem.it>). A “drug network” of 1,302 nodes (drugs) with 41,047 edges (indicating similarities between pair of drugs) was constructed based on network theory and non-parametric statistics on gene expression data. In order to study a novel drug, users have to give to MANTRA one or more genome-wide ranked list of genes sorted according to their differential expression in a treatment with the drug. On the basis of this input, MANTRA automatically integrates this novel drug in a huge network of compounds in which the topology reveals similarities and differences in MoA to predict their therapeutic effects. It is also possible to directly search the neighbourhood of a ‘seed’ compound with a desired MoA for identifying safer and repositionable drugs that have not been previously linked to that MoA [253, 254].

The Library of Integrated Network-Based Cellular Signatures (LINCS) is a National Institutes of Health (NIH) program, aiming to create a network-based understanding of biology by cataloguing changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents, and by using

computational tools to integrate this diverse information into a comprehensive view of normal and disease states that can be applied for the development of new biomarkers and therapeutics [209]. A central component of the NIH LINCS Consortium is a larger collection of gene expression profiling assay termed the L1000, introduced by the Broad Institute in 2017 [210]. L1000 provides more than a 1,000-fold scale-up of the CMap, made possible by a low-cost, high-throughput reduced representation expression profiling method. The L1000 assay measures the expression level of 978 mRNAs termed the landmark genes via a Luminex bead-based probe hybridization assay, which is highly reproducible and comparable to RNA sequencing. The expression level for the remaining non-measured genes is inferred by a computational model using additional data from the Gene Expression Omnibus [210, 255]. L1000 can be used to discover mechanism of action of small molecules, functionally annotate genetic variants of disease genes, and inform clinical trials [210]. The expanded L1000 CMap has profiled over 1.3 million samples to generate an initial set of signatures, which is available on the cloud-based CLUE software platform (<https://clue.io>). A total of 27,927 perturbagens have been profiled to produce 476,251 expression signatures. Approximately half of those signatures make up the Touchstone dataset generated from testing well-annotated genetic and small-molecular perturbagens in a core panel of cell lines. The remainder make up the Discover dataset, generated from profiling uncharacterized small molecules in a variable number of cell lines. While the L1000 data set is relatively new, several studies have already utilized this massive resource to

identify small molecules for drug discovery and repurposing. L1000 queries have been used to uncover potential novel treatments for a variety of diseases, such as non-alcoholic steatohepatitis [256], colorectal cancer [256], melanoma [257], and pancreatic ductal adenocarcinoma [258].

D. Objectives of the research

This research project involves the analysis of in-house generated large-scale microarray-based and RNA-seq based dataset on bone marrow CD34+ cells from MDS patients and healthy control donors. We performed transcriptomic profiling and connectivity mapping analysis in order to identify potential drugs for repurposing to MDS.

The aims of this project are:

- 1) To determine the lists of DEG associated with the presence of the most common gene mutations, such as splicing factor mutations, in MDS or MDS subtypes.
- 2) To analyse these lists of DEG by using connectivity mapping to identify candidate drugs that might be repositioned as potential new therapeutic agents for MDS.
- 3) To investigate the efficacy of the identified drugs in leukaemia cell lines and induced pluripotent stem cells (iPSCs) derived from MDS patients and cultured MDS patient bone marrow cells.
- 4) To investigate the potential MoA of the candidate drugs.

Chapter II Materials and Methods

1. Patient samples for RNA-sequencing

The RNA-seq dataset from MDS patient samples was part of our previous study published by Pellagatti et al [149]. The sample preparation and cDNA library preparation were carried out by Dr Richard N Armstrong and Dr Violetta Steeples (Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford).

1.1 Patient samples

CD34⁺ cells were isolated from the bone marrow samples of 73 patients with MDS and 8 healthy control individuals using MACS columns (Miltenyi Biotec, Germany). Of the 73 MDS cases, 28 were SF3B1 mutant, 6 U2AF1 mutant (U2AF1mut), and 39 had no mutations in splicing factor genes. The mutational profile of the patients with MDS was determined using a next-generation sequencing-based myeloid gene panel [43]. The research was approved by the relevant institutional review boards or ethics committees, and all participants gave written informed consent.

1.2 RNA-sequencing

Total RNA from CD34⁺ cells was extracted using TRIzol (Thermo Scientific), with a linear acrylamide carrier, treated with DNase I (Life Technologies) and purified using Agencourt RNAClean XP beads (Beckman Coulter). RNA quality was assessed using a 2100 Bioanalyzer (Agilent), and RNA integrity numbers were greater than 8 for all samples. cDNA libraries were produced using a SMARTer library preparation protocol (Clontech). Sequencing was performed on an Illumina HiSeq4000 with 100-bp paired-end reads. The reads were mapped to human genome GRCh37 using HISAT2 version

2.0.0-beta. Uniquely mapped read pairs were counted using featureCounts, included with subread v1.5.0.41. Quality control was performed on the mapped files using Picard CollectRnaSeqMetrics (<http://broadinstitute.github.io/picard>). The gene expression data discussed in this research has been deposited in the NCBI's Gene Expression Omnibus (GEO) repository (GEO accession number: GSE114922) and has been published by Pellagatti et al [149].

2. Identification of differentially expressed genes

Differential gene expression analyses were performed using R/Bioconductor package DESeq2 (<http://www.bioconductor.org/>). DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the FDR [259]. The significance of the gene expression difference was determined with an adjusted P-value (p_{adj}) < 0.05 found by DESeq2.

3. Identifying candidate drugs using connectivity map (CMap).

3.1 CMap analysis

CMap is a database with a collection of gene expression data sets from cultured human cell lines treated with various small molecules. CMap utilizes a gene-set enrichment metric (Kolmogorov-Smirnov) statistic to rank order individual treatment instances by their similarity to a given query signature. The output consists of small-molecule compounds with assigned enrichment scores (which lie between -1 and 1). This enrichment scores represent the relative correlations between query signature and reference gene expression profiles of individual drugs in the CMap database. The CMap analysis was performed by submitting lists of differentially expressed genes obtained

from the analysis of our transcriptomic data sets, then the reference dataset of drug-specific expression profiles was queried with the MDS-specific gene signatures to return a list of compounds that might revert or mimic those gene expression changes [207]. Based on the ‘signature reversion’ principle, drugs that significantly induce gene expression in a direction opposite to that observed in the disease state, might represent new therapeutic agents. The CMap analyses in Chapter III were carried out using microarray-based GEP platform (CMap Build 2.0) provided at <http://www.broad.mit.edu/cmap/>. CMap build 2.0 version is now maintained in an archived query-only mode and is not supported or updated by the Broad Institute. In Chapter V, L1000 assay was used for CMap analysis (<https://clue.io>).

3.2 MANTRA drug network

Mode of Action by NeTwork Analysis (MANTRA) is a cMAP extension that exploits similarity in gene expression profiles following drug treatment, across multiple cell lines and dosages, to predict similarities in drug effect and MoA (<http://mantra.tigem.it>). A “drug network” of 1,302 nodes (drugs) with 41,047 edges (indicating similarities between pair of drugs) was constructed based on network theory and non-parametric statistics on gene expression data. MANTRA was used to search the neighbourhood of a ‘seed’ compound with a desired MoA for identifying repositionable drugs never linked to that MoA before [254, 260].

4. Cell lines and cell culture

All leukaemia cell lines used in this study are shown in Table 4.1 in Chapter IV with detailed biological and clinical information. The culture conditions of these cell lines are shown in Table 2.1. Most leukaemia cell lines were maintained in 90% RPMI 1640

(Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich) and 1% Penicillin-Streptomycin-L-Glutamine (PSG; Sigma-Aldrich), except SKM-1 cells cultured in 90% RPMI 1640 supplemented with 20% FBS. SKM1, F36P TF-1 and MDS-L cell line were further supplemented with GM-CSF or IL-3 (Miltenyi Biotec) at specific concentration shown in Table 2.1.

Table 2.1 The cell culture conditions of leukaemia cell lines

Leukaemia lines	Type of leukaemia	cell culture
K-562	CML	90% RPMI 1640 + 10% FBS
SKM-1	MDS (RAEB-T) → AML-M5	80% RPMI 1640 + 20% FBS (+ 1 ng/ml GM-CSF)
HL60	AML-M3 → AML-M2	90% RPMI 1640 + 10% FBS
F36P	MDS (RAEB) → AML-M6	90% RPMI 1640 + 10% FBS + 1 ng/ml GM-CSF or 1 ng/ml IL-3
TF-1	AML-M6	90% RPMI 1640 + 10% FBS + 5 ng/ml GM-CSF
MDS-L	MDS (RARS/RAEB) (prior to leukaemic transformation)	90% RPMI 1640 + 10% FBS + 10 ng/ml IL-3
MEG-01	CML	90% RPMI 1640 + 10% FBS
JURKAT	T-ALL	90% RPMI 1640 + 10% FBS

The HUDEP2 cell line was grown in StemSpan™ serum-free haematopoietic cell expansion media (Stemcell technologies) supplemented with 1% PSG, 50 ng/ml SCF (Stemcell technologies), 3 IU/ml EPO (Stemcell technologies), 1 uM Dexamethasone (Stemcell technologies) and 1 ug/ml Doxycycline (Stemcell technologies). All cells were grown in an incubator at 37°C and 5% CO₂.

5. Genetically modified cell lines

HUDEP2 cell with heterozygous U2AF1^{Q157} mutation was a gift from Prof Richard Gibbons and Prof Doug Higgs (WIMM, University of Oxford). The genetically modified HUDEP2 cell line was generated by plasmid-based methodology to introduce short (<20 bp) DNA sequences into the genome of HUDEP-2 cells using the CRISPR/Cas9 system combined with single-stranded oligodeoxynucleotide (ssODN) donors. The protocol was described in detail by Moir-Meyer et al [261]. K562 cells with the SF3B1^{K700E} mutation and isogenic SF3B1^{K700K} K562 cells were purchased from Horizon Discovery. K562 cells overexpressing U2AF1^{S34F} were generated by Dr Shalini Singh (Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford) using retroviral pGCDNsam-IRES-EGFP plasmids containing U2AF1^{S34F} cDNA. Vector plasmids were co-transfected with Gag/Pol plasmids and vesicular stomatitis virus glycoprotein envelope using jetPRIME transfection reagent (Polyplus transfection) into HEK293T cells as described previously to obtain high-titer retrovirus stock. Retroviruses were harvested after 48h and 72h of transfection. Harvested retroviruses were concentrated by ultracentrifugation (Beckman Coulter, Brea, CA, USA; Ultracentrifuge Rotor SW28) at 28000 rpm for 3h at 4°C. Spinoculation of K562 cells was performed in the presence of 8 µg/ml polybrene (Sigma-Aldrich) at 800 ×g, 32°C for 2 hours. The transduced K562 cells were selected after overnight incubation with retroviruses using FACS sorting for GFP positive cells. All cell lines used in this study tested negative for mycoplasma.

6. Isolation of CD34⁺ cells from bone marrow mononuclear cells

Bone marrow samples from MDS patients were obtained with informed consent and collected with approval of research ethics committees (Oxford, REC 17/YH/0322; Oxford MDSBio Study). Healthy human bone marrow mononuclear cells were purchased from Lonza. Frozen bone marrow mononuclear cells from MDS patients and healthy samples were thawed and treated with DNase 1 (100U/ml) (ThermoFisher Scientific) in Dulbecco's PBS (pH 7.0-7.2) (Sigma-Aldrich), containing 0.5 mM MgCl₂, 1 mM CaCl₂ and 0.5% BSA (Sigma-Aldrich) for 45 mins at room temperature. CD34⁺ cells were isolated using the CD34 MicroBead Kit UltraPure (Miltenyi Biotec), according to the manufacturer's protocol. The purity of the isolated CD34⁺ cells was measured by flow cytometry (using CD34-PE, human, Clone: AC136, Miltenyi Biotec) and was >85% for all samples tested. CD34⁺ cells were cultured in StemSpan Serum-Free Expansion Medium (Stemcell Technologies) containing StemSpan CD34⁺ Expansion Supplement (Stemcell Technologies) for 48 hours before performing experiments.

7. Generation and characterization of induced pluripotent stem cells (iPSCs)

7.1 Generation of iPSCs

The iPSCs used in this study was generated by Dr Doaa Ahmed and Dr Dharamveer Tatwavedi (Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford). Cryopreserved bone marrow CD34⁺ cells of one healthy control (Lonza), two MDS patients (one sample harbouring an SF3B1-K700E and a DNMT3A-R882H mutation, and one CMML patient (harbouring SRSF2-P95L and TET2-V1371D mutation) were thawed and cultured in StemSpan SFEM media (Stemcell Technologies) supplemented with 100 ng/mL Stem Cell Factor, 10 ng/mL

IL-3, 100ng/ml Flt3, 40 µg/mL lipids (Stemcell Technologies), 2 mM L-glutamine, and 1% penicillin-streptomycin solution (Sigma-Aldrich). All cytokines were obtained from Miltenyi Biotec unless otherwise stated. The cells were cultured for 3-5 days to induce cell proliferation. CytoTune 2.0 Sendai reprogramming kit (Life Technologies) was used to generate iPSCs, according to the manufacturer's instructions. After 3-4 weeks, multiple colonies with human embryonic stem cell (hESC)-like morphology were picked and expanded. One iPSC clone harboring the SF3B1 and the DNMT3A mutations and one iPSC clone without these mutations from an MDS patient sample, one iPSC clone harboring SRSF2 and TET2 mutation from another MDS patient sample, and one iPSC clone from the healthy control sample, were selected for characterization and experimental work. The mutation status was confirmed in patient and control iPSCs using Sanger sequencing. iPSC lines were cultured feeder-free in Matrigel (Corning Inc) pre-coated dishes using mTeSR hESC media (Stem Cell Technologies) culture conditions, according to manufacturer's guidelines in a hypoxic incubator (5% O₂).

7.2 Confirmation of clearance of Sendai virus reprogramming factors

The absence of viral transgenes by CytoTune 2.0 Sendai virus in the iPSC genome was evaluated according to manufacturer's guidelines (Life Technologies). It was confirmed that the transduced genes were not expressed at the mRNA level in the established iPSC clones using real-time quantitative PCR.

7.3 Characterization of iPSCs

The iPSC clones retained typical characteristics of pluripotent stem cells such as expression of immature ES cell markers (SSEA4, TRA181, TRA160), as examined by

immunocytochemistry and flow cytometry, and the ability to differentiate into the three germ layers in vitro. Karyotypic analysis by G-banding of the iPSC clones derived from the MDS patient and the healthy control showed a normal karyotype.

7.4 Immunocytochemistry for expression of immature ES cell markers

iPSCs were grown in 24-well cell culture plates coated with Matrigel until they reached a confluency of 80%. Cells were fixed in 4% paraformaldehyde (ChemCruz biochemical) for 15 minutes at room temperature. Fixed cells were washed in PBS and blocking of nonspecific binding sites was performed in 5% BSA (Sigma-Aldrich) in PBS. Cells were incubated with primary antibodies [TRA-1-81 (Stemgent, Cambridge, Massachusetts, United States, Cat 09-0011, 1:100), TRA-1-60 (Merck Millipore, Burlington, Massachusetts, United States, Cat# MAB4360, 1:200), SSEA4 (Merck Millipore, Cat# MAB4304, 1:200)] at 4°C overnight. Cells were then incubated with anti-mouse antibody (Invitrogen, Carlsbad, California, United States, Cat# A-21202) for 1 h at room temperature in the dark. For nuclear visualization, cells were incubated with 2 µM DAPI (Invitrogen) for 5 minutes. Fluorescence microscopy was performed with an EVOS FL2 Auto fluorescence microscope (Invitrogen) and processed with ImageJ.

7.5 Flow Cytometry for expression of immature ES cell markers

Cells were washed with PBS and stained with fixable viability dye eFlour 780 (Invitrogen, Cat# 65-0865-14) according to manufacturer's instructions, and were subsequently fixed in 2% paraformaldehyde (ChemCruz biochemical). Cells were incubated with antibodies [(TRA-1-81 (Invitrogen Cat# 12-8883-82), TRA-1-60-R antibody (BioLegend, San Diego, California, United States, Cat# 330605), SSEA4

(Invitrogen, Cat # 53-8843-42)] for 30 minutes at 4°C in the dark and washed with PBS. Flow cytometry was performed on a BD LSRII instrument (Becton Dickinson). Data analysis was performed using FlowJo 10 Software version 10.6.1 (Tree Star).

7.6 Trilineage differentiation of iPSCs analysed by flow cytometry

In vitro trilineage differentiation was carried out using STEMdiff Trilineage Differentiation Kit (Stemcell Technologies), according to manufacturer's guidelines. Cells were washed with PBS and stained with different markers depending on the differentiation lineage. Mesodermal differentiation potential was assessed by the formation of endothelial cells (CD144⁺) (MACS Miltenyi Biotec, Cat# 130-100-713) and the presence of smooth muscle cells (CD140b⁺) (MACS Miltenyi Biotec, Cat# 130-105-322). Endodermal differentiation was assessed by the presence of definitive endoderm cells CD184⁺ (MACS Miltenyi Biotec, Cat# 130-109-886) and Sox17⁺ (MACS Miltenyi Biotec, Cat# 130-111-147). Ectodermal differentiation potential was assessed by the formation of neuroectoderm cells Pax6⁺ (MACS Miltenyi Biotec, Cat# 130-107-829) and Sox2⁺ (MACS Miltenyi Biotec, Cat# 130-104-993). Cells were stained as per manufacturer's instruction. Flow cytometry was performed on a BD LSRII device instrument (Becton Dickinson). Data analysis was performed using FlowJo 10 Software version 10.6.1 (Tree Star Inc.).

8. Compounds

The compounds identified through CMap analysis used for experimental validation and other compounds used in this study are shown in Table 2.2. The drugs were dissolved in dimethylsulphoxide (DMSO) and aliquoted and stored at -20°C. A fresh aliquot of

DMSO dissolved compounds was added into cell medium to make a series of dilutions for cell viability tests.

Table 2.2 The names and suppliers of the compounds

Drug name	Supplier
Quinacrine dihydrochloride	Sigma-Aldrich
Digitoxin	MP Biomedicals UK
Digoxin	Fisher Scientific
Niclosamide	Abcam plc
Anisomycin	MP Biomedicals UK
BX795	Enzo Life Sciences (UK) Ltd
Anisomycin	Cambridge Bioscience Ltd
Harmine	Insight Biotechnology Ltd
Simvastatin	Fluorochem Ltd
Famprofazone	Insight Biotechnology Ltd
Kaempferol	MP Biomedicals UK
Narciclasine	Insight Biotechnology Ltd
Kaempferol	Cambridge Bioscience Ltd
Celastrol	Stratech Scientific Ltd
Harmaline	Alfa Aesar
Pyrvinium pamoate hydrate	Sigma-Aldrich-Aldrich Co Ltd
Lenalidomide	Fluorochem Ltd
BX-912	Sigma-Aldrich-Aldrich Co Ltd
GPP 78	Cambridge Bioscience Ltd
Lycorine Hydrochloride	Enzo Life Sciences (UK) Ltd
Spiperone hydrochloride	Abcam plc
Pristimerin	Insight Biotechnology Ltd
Piperlongumine	Cambridge Bioscience Ltd
Thapsigargin	Insight Biotechnology Ltd
Parthenolide	Cambridge Bioscience Ltd

Sulfadoxine (Sulphadoxine)	Sigma-Aldrich
Parbendazole	Sigma-Aldrich
Bufexamac	Sigma-Aldrich
Syrosingopine	Sigma-Aldrich
Fenbendazole	Sigma-Aldrich
Rifabutin	Sigma-Aldrich
Hycanthone	Medchemtronica
Lasalocid sodium	Medchemexpress
Withaferin A	Fisher Scientific UK Ltd
Azacitidine	Sigma-Aldrich
Sudemycin C1	Gifted from a collaborator
Isogingketin	Gifted from a collaborator
Madrasin	Gifted from a collaborator

9. Cell viability assay

Cell lines or primary cells were seeded in a 96 well suspension culture plate (Corning Inc). Cells were treated with varying concentrations of different compounds, or with vehicle (DMSO) as control, for 48 or 72 hours. The cells were then transferred to an opaque white 96-well plate (Thermo Fisher Scientific), and the cell viability was analysed by using CellTiter-Glo Luminescent Cell Viability Assay Kit (Promega) and SpectraMax i3x plate reader (Molecular Devices) according to manufacturer's protocol. Percentages of viable cells were calculated by normalizing to the DMSO controls. The nonlinear regression curves were drawn using GraphPad Prism. Synergy scores were generated using SynergyFinder (<https://synergyfinder.fimm.fi>) using the HSA model [262].

10. Cell cycle analysis

Cells were seeded in 6 well plates and exposed to Celastrol (500 nM), Withaferin A (1000 nM), Parbendazole (200 nM) or Fenbendazole (700 nM) for 12 hours. Fixation

and permeabilization of cells were performed by dropwise addition of 70% ethanol while vortexing and incubation on ice for at least 1 hour. Permeabilized cells were washed with PBS and treated with 0.04mg/ml propidium iodide (Biolegend) and 0.1 mg/ml RNaseA (Sigma-Aldrich) for 30 minutes at room temperature in the dark. The DNA content was analysed via flow cytometer (BD LSR II, Becton Dickinson). The flow cytometry data were analysed using FlowJo software.

11. Cell Apoptosis analysis

Cells were seeded in 6 well plates and treated with Celastrol (500 nM), WA (1000 nM), PBD (200 nM) or FBD (700 nM) for 24 hours. The FITC Annexin V/PI dual staining apoptosis assay with FITC annexin V and PI for flow cytometry provides a rapid and convenient assay for apoptosis. The apoptotic cells were determined by staining with Annexin V-FITC and propidium iodide (Biolegend) according to manufacturer's protocol. A single-positive staining for Annexin-V is an early indicator of apoptosis (Annexin-V+/PI-), and the double-positive staining for Annexin and PI indicated late apoptosis or necrosis (Annexin-V+/PI+). The cells were analysed on a flow cytometer (BD LSR II, Becton Dickinson), and the acquired data were analysed using FlowJo software.

12. Cell differentiation assays

Cells were seeded in 6 well plates and treated with PBD or FBD at approximate IC₂₀ for each cell line for 4 days (HL60: PBD 50 nM, FBD 300 nM; SKM-1: PBD 100 nM, FBD 600 nM; TF1: PBD 30 nM, FBD 200 nM). Cell lineage marker expression was evaluated using flow cytometry. The expression of differentiation markers of myeloid cells was assessed by using antibodies against CD11b (PE, Biolegend) and CD14 (APC,

Biolegend). The expression of differentiation markers of erythroid cells was assessed by using antibodies against CD36 (PE, Biolegend), CD71 (PE-Cyanine7, eBiosciences) and CD235a (APC, eBiosciences). DAPI (Sigma-Aldrich) was used as viability dye. Cells were incubated with antibodies for 30 mins at 4°C in dark. Cells were subsequently washed with FACS buffer (PBS with 1% Bovine serum albumin) and resuspended in FACS buffer with DAPI for flow cytometry analysis. DAPI negative (viable) cells were used for subsequent analysis.

13. Colony-Forming Cell Assay

The CD34⁺ cells from a healthy control donor (Lonza) were cultured with PBD at 150 nM or FBD at 900 nM for 24h. 3000 cells were plated on methylcellulose containing 0.65µg/ml puromycin in triplicate (MethoCult H4434 Classic, Stemcell Technologies) for colony-forming cell assay according to the manufacturer's protocol. Plates were incubated in a humidified incubator at 37°C and 5% CO₂. Colonies were counted after 14 days of incubation.

14. Real-Time Quantitative Polymerase Chain Reaction

K562 cells or SKM1 cells were treated with selected drugs of interest for 4 h. Total RNA was isolated using the TRIzol reagent (Life Technologies) and treated by DNase I (Life Technologies) for 10 min. Total RNA was subjected to cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit (Life Technologies), according to the manufacturer's instructions. Quantitative reverse transcriptase (RT)-PCR was performed using LightCycler480 SYBR Green Master Mix (Roche) and the LightCycler 96 Instrument (Roche). Primer sequences used in this study were designed

as shown in Table 3. All reactions were run in triplicate, and data were analyzed using the $2^{-\Delta\Delta CT}$ method [263]. B2M was used as an internal standard.

Table 2.3 Primers used for RT-PCR

IL-6	Forward	5'-AGACAGCCACTCACCTCTTCAG-3'
	Reverse	5'-TTCTGCCAGTGCCTCTTTGCTG-3'
CCND1	Forward	5'-TCTACACCGACAACCTCCATCCG-3'
	Reverse	5'-TCTGGCATTCTTGGAGAGGAAGTG-3'
BCL2	Forward	5'-ATCGCCCTGTGGATGACTGAGT-3'
	Reverse	5'-GCCAGGAGAAATCAAACAGAGGC-3'
MYC	Forward	5'-CCTGGTGCTCCATGAGGAGAC-3'
	Reverse	5'-CAGACTCTGACCTTTTGCCAGG-3'
TNIK	Forward	5'-ACAGTGGCTGTCAGCGACATAC-3'
	Reverse	5'-ATACTGCCGCTGAAACTGTCCG-3'
AXIN2	Forward	5'-CAAACCTTCGCCAACCGTGGTTG-3'
	Reverse	5'-GGTGCAAAGACATAGCCAGAACC-3'
LEF1	Forward	5'-CTACCCATCCTCACTGTCAGTC-3'
	Reverse	5'-GGATGTTCTGTTTGACCTGAGG-3'
B2M	Forward	5'-CCACTGAAAAAGATGAGTATGCCT-3'
	Reverse	5'-CCAATCCAAATGCGGCATCTTCA-3'

15. Molecular docking

Molecular docking is a computational tool in structural molecular biology and computer-assisted drug design. The goal of molecular docking is to predict the predominant binding modes of a ligand with a protein of known three-dimensional structure [264]. Molecular Docking studies were performed in SwissDock (<http://swissdock.vital-it.ch/>) and DockingServer (<http://www.dockingserver.com>). The crystal structure of IRAK4 was obtained from the Protein data bank (PDB) database with the accession code 2NRU (<https://www.rcsb.org/>) for docking. Docking step was firstly performed using SwissDock web service based on the docking software EADock DSS. The ligand and receptor files in the appropriate formats were uploaded to the web-based server for docking study. A grid was designed in which many binding modes were generated for the most favorable bindings. Simultaneously, their CHARMM energies are estimated on the grid. The binding modes with the most

favorable energies were evaluated with Fast analytical continuum treatment of solvation (FACTS), and clustered. Binding modes were scored using their FullFitness and clustered. Clusters were then ranked according to the average FullFitness of their elements. Results of the SwissDock were visualized by UCSF Chimera package. Docking calculations and ligand-protein interaction modes were carried out using DockingServer. The MMFF94 force field was used for energy minimization of ligand molecule. Gasteiger partial charges were added to the ligand atoms. Non-polar hydrogen atoms were merged, and rotatable bonds were defined. Docking calculations were carried out on IRAK4 protein model. Essential hydrogen atoms, Kollman united atom type charges, and solvation parameters were added with the aid of AutoDock tools. Affinity (grid) maps of $26 \times 28 \times 28$ Å grid points and 0.375 Å spacing were generated using the Autogrid program. AutoDock parameter set- and distance-dependent dielectric functions were used in the calculation of the van der Waals and the electrostatic terms, respectively. Docking simulations were performed using the Lamarckian genetic algorithm (LGA) and the Solis & Wets local search method. Initial position, orientation, and torsions of the ligand molecules were set randomly. Each docking experiment was derived from 2 different runs that were set to terminate after a maximum of 250000 energy evaluations. The population size was set to 150. During the search, a translational step of 0.2 Å, and quaternion and torsion steps of 5 were applied.

16. Phospho-kinase array assay

To identify the drug-induced phosphorylation changes of kinase, we cultured SKM1 cells untreated or treated with PBD or FBD for 24 h. The level of phosphorylation of 46 kinase phosphorylation sites in the cell lysates was detected using the Human

Phospho-Kinase Array (R&D Systems) according to the manufacture's instruction. Capture and control antibodies have been spotted in duplicate on nitrocellulose membranes. The protein concentration of the cell lysates was quantified by total protein assay using Pierce BCA Protein Assay Kit (Thermo Scientific). Cell lysates are diluted and incubated overnight with the Human Phospho-Kinase Array. The array is washed to remove unbound proteins followed by incubation with a cocktail of biotinylated detection antibodies. Streptavidin-HRP and chemiluminescent detection reagents were applied, and a signal is produced at each capture spot corresponding to the amount of phosphorylated protein bound.

17. Kinase inhibition assay

The TRAF2 and NCK-interacting kinase (TNIK) kinase inhibition assay was performed using ADP-Glo™ Kinase Assay Kit and TNIK Kinase Enzyme System (Promega) according to the manufacture's instruction. The enzyme, substrate, ATP and inhibitors (PBD or FBD) was diluted in kinase buffer. The enzyme, inhibitors and substrate/ATP mix were added into the wells of 384-well low volume plate to make a reaction mixture containing 30 ng of TNIK, 25 uM ATP, and 0.1 ug/ul MBP, and incubated at room temperature for 60 min. Add 5 µl of ADP-Glo™ Reagent and Kinase Detection Reagent were subsequently added and incubated (40 min and 30 min, respectively), then the luminescence were recorded using SpectraMax i3x plate reader (Molecular Devices). Percentages of enzyme activity were calculated by normalizing to the DMSO controls. The dose response was created using GraphPad Prism to determine the potency of the inhibitor (IC50).

18. Statistics

Statistical analysis was performed using GraphPad Prism 6.01. A P-value <0.05 was considered significant. The type of statistical test used and the results including P values, means and standard error of the mean are shown in the figures and figure legends.

Chapter III Gene expression profiling of MDS patient CD34 cells and computational drug screening via connectivity mapping

A. Introduction and aims

In previous studies, the interconnections between mutations, gene expression profiles, clinical variables and patient outcome in MDS were investigated. It has been showed that the mutational status shapes the gene expression landscape, and deregulated genes associated with the most common gene mutations and disease subtypes in MDS were identified [87, 94-98, 116]. Connectivity mapping is a process used to recognise novel pharmacological properties in small molecules by comparing their gene expression signatures with others in a database [207, 240]. A number of resources have been developed that allow to perform a connectivity mapping analysis using gene expression profiling data. The CMap resource has been successfully applied to identify novel therapeutics for repositioning in various disease contexts [237, 244, 246, 248, 265, 266]. In this chapter, we used CMap tool (Build 2.0) to perform connectivity mapping analysis on RNA-seq and microarray-based GEP datasets on MDS CD34+ cells in order to identify candidate known drugs for repositioning as potential new therapeutic agents for MDS. We firstly analysed the in-house generated RNA-seq data of MDS patients with splicing factor gene mutations (*SF3B1* or *U2AF1* mutation) to obtain the query gene signature for computational drug screening via CMap analysis. A query gene signature is a set of highly representative disease-specific genes, selected from the DEG between disease and normal conditions. The CMap method is performed by simply submitting the gene signature list to query against the reference dataset of drug-specific

expression profiles, which then returns a list of compounds that might revert or mimic those gene expression changes. The drugs were ranked by quantifying the extent of positive or negative correlation between the gene signature of the drug and that of the disease. Based on the 'signature reversion' principle, drugs that significantly induce gene expression in a direction opposite to that observed in the disease state, might represent new therapeutic agents.

According to the 2008 WHO classification, the category Refractory Anaemia with Excess Blasts-2 (RAEB-2) describes cases in which there are 10% to 19% blasts in the marrow or 5% to 19% blasts in the blood, or both. The presence of Auer rods in blasts qualifies as RAEB-2 irrespective of the blast percentage. Patients with RAEB-2 have an unfavourable prognosis, which is considered as a high-risk subtype of MDS. The RAEB-2 type transforms to an acute leukaemia in approximately 33% of cases. There is an urgent need to discover therapeutic compound for the treatment of RAEB-2 [14]. Our previous studies of large-scale microarray-based GEP dataset on CD34+ cells from RAEB-2 patients has led to the identification of lists of significant deregulated genes and pathways associated with disease subtypes and karyotypic groups, patient prognosis and presence of the most common gene mutations. By analysing the microarray-based gene signatures using connectivity mapping method, we also identified candidate drugs that might be repositioned in RAEB-2 cases of MDS.

The aims of this chapter are 1) to determine the lists of DEG associated with the presence of the most common gene mutations (e.g. splicing factor mutations) or the specific disease subtype (e.g. RAEB-2) in MDS; 2) to analyse these lists of DEG by using connectivity mapping to recover a list of compounds that reverts or mimics those

gene expression signatures; 3) to select drug candidates for wet-lab experiment validation.

B. Results

1. RNA-seq based transcriptome analysis of MDS patients with splicing factor mutations

Splicing factor gene mutations are the most common mutations in patients with MDS. To determine the dysregulated genes in MDS cases with mutations of two splicing factor genes *SF3B1* and *U2AF1*, we performed an analysis of in-house generated RNA-Seq data on bone marrow CD34+ cells obtained from MDS patients (73 cases) and healthy control donors (9 cases). Among these MDS patients, 28 cases had *SF3B1* mutations, 6 cases had *U2AF1* mutations, whereas 39 cases had no mutations in splicing factor genes (*SF3B1*, *SRSF2*, *ZRSR2* or *U2AF1*). Other common gene mutations, such as epigenetic modulators *TET2* and *DNMT3A* were also detected by targeted next-generation sequencing (Figure 3.1).

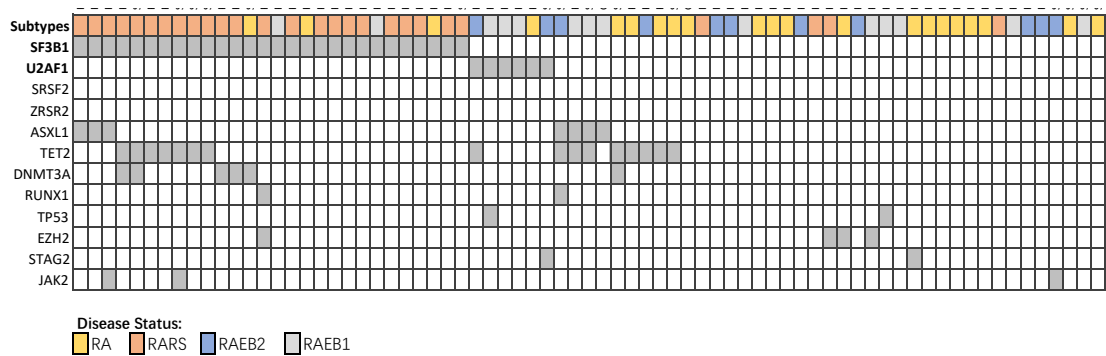


Figure 3.1 Disease subtypes and mutations status of MDS patients. MDS Patients were diagnosed as refractory anemia (RA), RA with ring sideroblasts (RARS), RA with excess blasts (RAEB1 and RAEB-2) according to 2008 World Health Organization (WHO) classification system. Splicing factor gene mutations and other common gene mutations were detected by targeted next-generation sequencing (A panel of 111 genes were sequenced as described by Papaemmanuil et al. [43]).

To generate lists of DEG for connectivity mapping queries, we used the DEseq2 package to analyse and normalise the RNA-Seq data of MDS patients carrying *SF3B1* or *U2AF1* gene mutations. In a given comparison, genes were considered to be differentially expressed if their fold change (Log2fold change) was >1 or <-1 , and the false discovery rate (FDR) $< 5\%$. As shown in Figure 3.2, we performed this analysis for the following comparisons: SF3B1-mutated MDS versus healthy controls (SF3B1^{Mut} vs HC), SF3B1-mutated MDS versus SF3B1 wild-type patients (SF3B1^{Mut} vs SF3B1^{WT}), U2AF1-mutated MDS versus healthy controls (U2AF1^{Mut} vs HC), and U2AF1-mutated MDS versus U2AF1 wild-type patients (U2AF1^{Mut} vs U2AF1^{WT}). There were 128 DEG identified in SF3B1^{Mut} vs HC SF3B1^{WT} comparison, 1573 DEG in SF3B1^{Mut} vs comparison, 1249 DEG in U2AF1^{Mut} vs HC comparison, and 64 DEG in U2AF1^{Mut} vs U2AF1^{WT} comparison. Not surprisingly, the lists of DEG contain many genes with established roles in MDS. For example, the iron transporter *ABCB7* is a well-recognized gene showing marked downregulation in MDS with ring sideroblasts, which is strongly associated with *SF3B1* mutations.

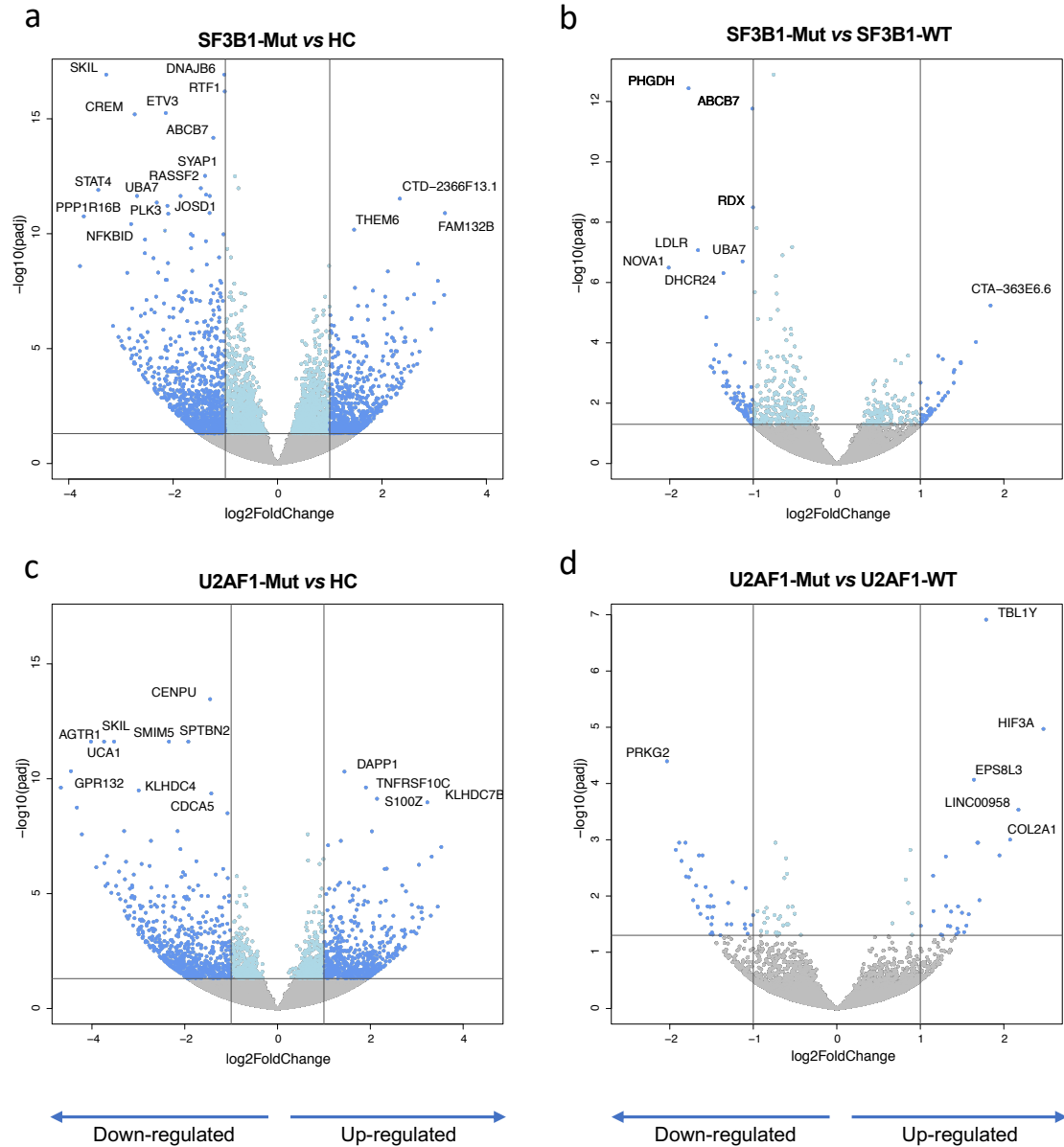


Figure 3.2 Differentially expressed genes in MDS with splicing factor gene mutations. RNA-Seq data of MDS patients was analysed and normalised by DEseq2 package in four different comparisons: **a)** SF3B1^{Mut} vs HC; **b)** SF3B1^{Mut} vs SF3B1^{WT}; **c)** U2AF1^{Mut} vs HC; **d)** U2AF1^{Mut} vs U2AF1^{WT}. Each plot represents a gene symbol, some gene symbols were labelled as examples. Vertical lines represent the cut-off of log2fold change equal to 1 and -1, and the horizontal indicates an FDR equal to 0.05.

2. Generation of RNA-Seq based DEG lists used as input to the connectivity mapping analysis

All DEG were extracted and converted to Affymetrix Probeset IDs that would be usable in the CMap analysis (CMap build 2.0). DEG lists that subsequently served as input to the connectivity mapping analysis were generated from each of the above comparisons, except the comparison of U2AF1^{Mu} vs U2AF1^{WT} that yielding an insufficient number of DEG. By referencing previous CMap studies [246, 248, 258, 267], we generated three query gene lists based on both expression level changes (log2fold change) and statistical significance (FDR): **1) Top 200:** Top 200 most statistically significant genes, regardless of up- or down-regulation; **2) Log2fold:** Genes with a Log2fold change lower than -1 or higher than 1, and an FDR<5% (or FDR<1% if the probe size is larger than 500); **3) Up-down 100:** Top 100 most up-regulated (over-expressed) and top 100 down-regulated (under-expressed) genes with an FDR<5%. It can be seen that there are some overlapping genes among the three lists, 12.9% overlap in SF3B1^{Mut} vs HC comparison, 19.4% in SF3B1^{Mut} vs SF3B1^{WT} and 31.8% in U2AF1^{Mut} vs HC (Figure 3.3).

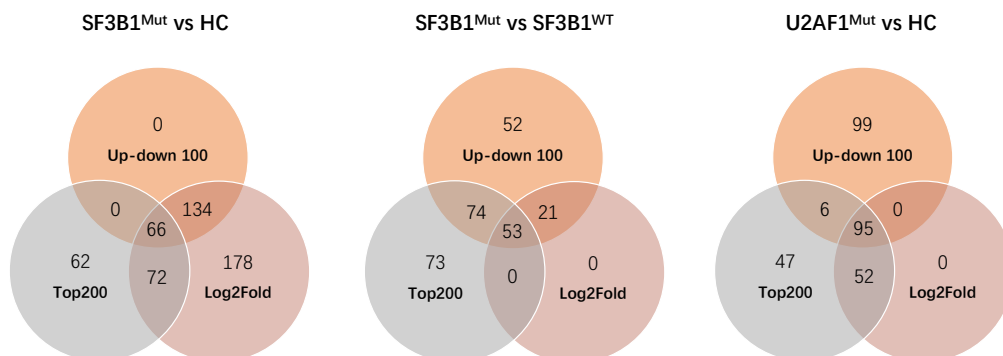


Figure 3.3 Venn diagrams show the overlapping genes of the three DEG lists for CMAP queries in three comparisons.

3. Candidate compounds identified by Connectivity map analysis of RNA-Seq based DEG lists

3.1 Drug screening by Connectivity map

We used the three lists of DEG identified above as input queries for CMap. The output drugs given by CMap were ranked by a ‘connectivity score’ normalized from -1 to $+1$ to reflect the connection between the expression profiles. A positive connectivity score suggesting the drug treatment can mimic the expression pattern of a disease state. In contrast, a negative connectivity score stands for a reversed transcriptomic mapping of disease state to drug treatment. Here, only those compounds with a negative connectivity score that were identified by the analysis of all three gene lists were considered of interests. These drugs were ranked by the absolute value of connectivity score in descending order, and only top-scoring drugs were used for the selection of candidates for wet-lab experiments. In light of existing data concerning chemotherapy drugs, epigenetic modulators, and other drugs used for MDS treatment, compounds without validated mechanisms that modulate dysregulated pathways in haematological malignancies were prioritised for experimental validation.

In the SF3B1^{Mut} vs SF3B1^{WT} and SF3B1^{Mut} vs HC comparisons, out of hundreds of overlapping compounds that were identified, Celastrol showed the highest absolute enrichment score, and was selected for experimental validation (Figure 3.4a, b). Similarly, CMap scored Withaferin A (WA) as one of the top-ranking candidates in U2AF1^{Mut} vs HC comparison (Figure 3.4c). The chemical structures of Celastrol and WA are shown in Figure 3.4d. Celastrol is a pentacyclic triterpene extracted from the *Tripterygium Wilfordii* plant; WA is a steroidal lactone isolated from the medicinal

plant *Withania somnifera*. The detailed drug lists with connectivity scores are shown in Supplementary table 3.1.

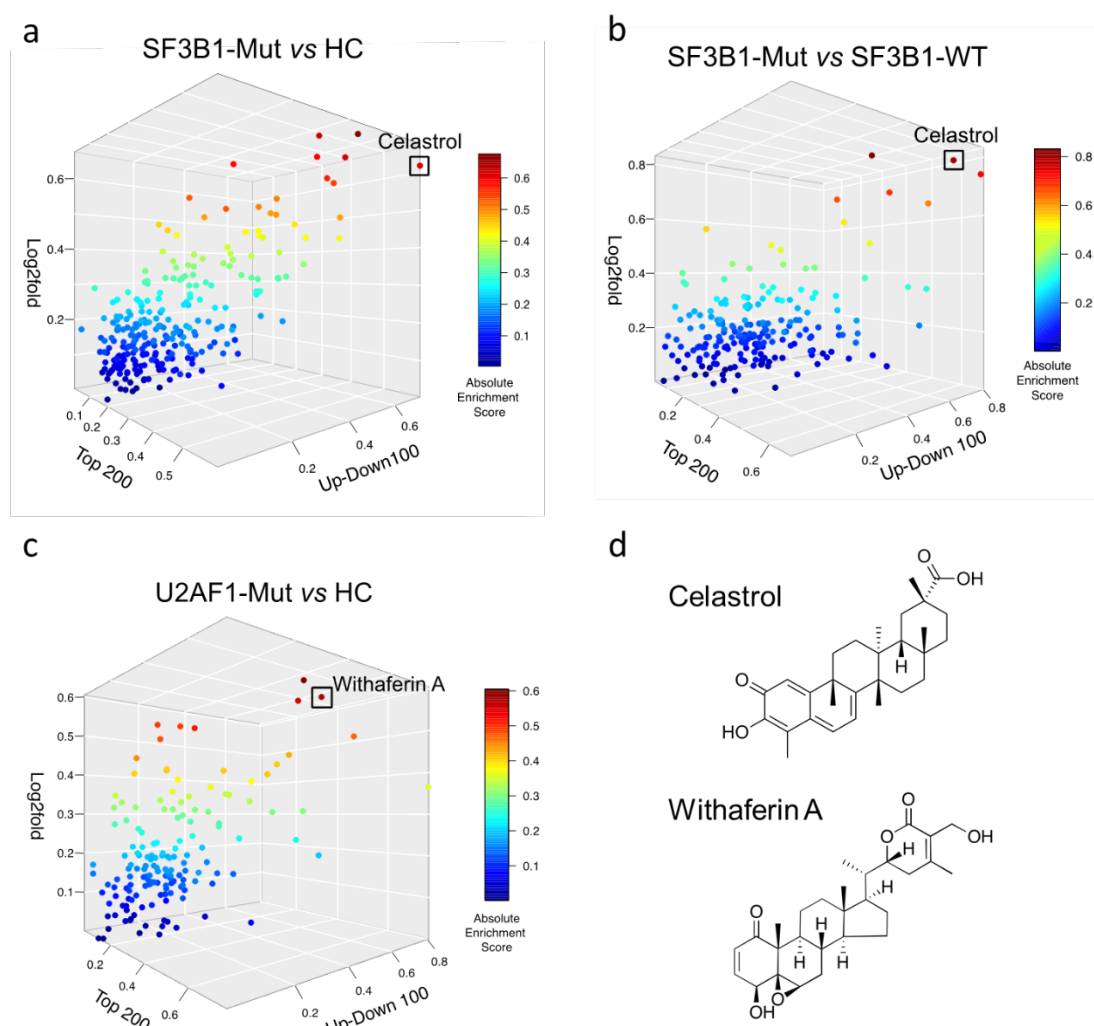


Figure 3.4. Celastrol and Withaferin A were identified as top-ranking candidates by CMAP. The 3D scatter plots show the overlapped compounds that discovered in three different CMAP queries. Each data point represents a compound. Blue to red color represents low-to-high absolute enrichment score. a, b) Celastrol was identified as one of the top-ranking candidates in SF3B1^{Mut} vs HC and in SF3B1^{Mut} vs SF3B1^{WT} comparison, respectively; c) WA was identified as one of the top-ranking candidates U2AF1^{Mut} vs HC comparison; d) Chemical structure of Celastrol and WA.

3.2 Gene expression profile based ‘Drug network’ of Celastrol and WA

In 2010, a publicly available online ‘drug network’ tool, MANTRA, was constructed by Iorio et al. based on the similarity in GEP generated by CMap drug treatment [260]. The drugs were connected and grouped into communities that were enriched for similar mode of action (MoA), or acting on the same pathway, and can be used to identify the compound-targeted biological pathways. New compounds can be integrated into the network to predict their therapeutic and off-target effects. Here, we used both Celastrol and WA as the input ‘core drugs’ to generate a GEP-based drug network (Figure 3.5).

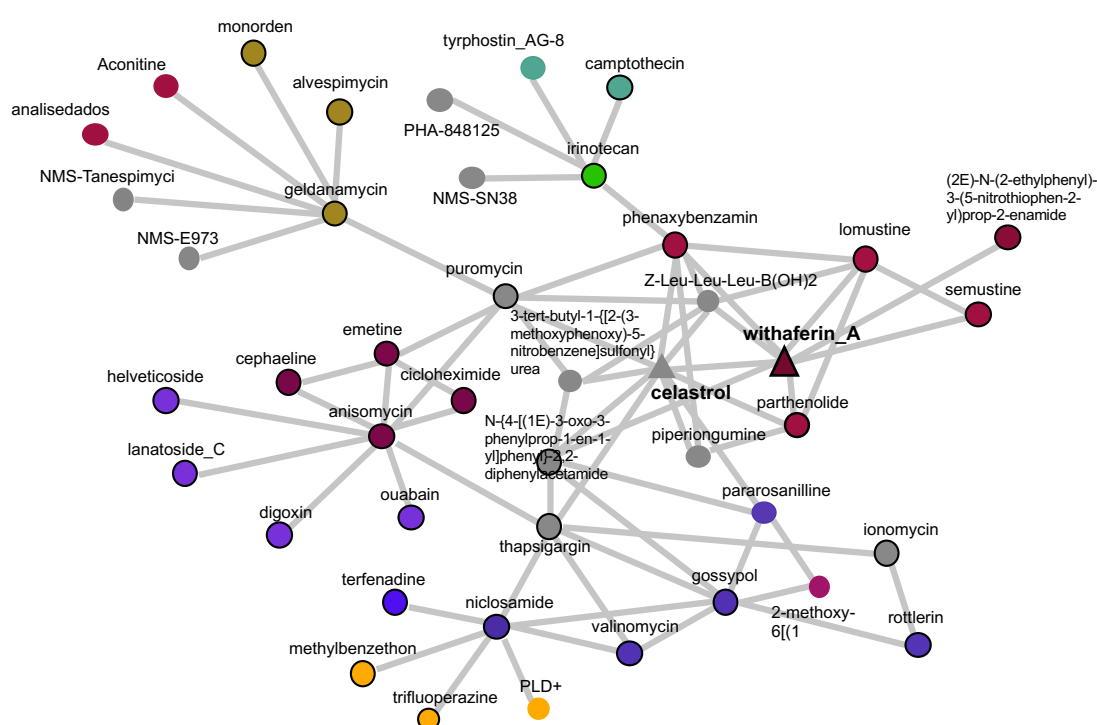


Figure 3.5. Drug networks of Celastrol and WA (MANTRA). The connected drugs were clustered into different communities based on gene expression similarity (Distance threshold: 0.5, Network Depth: 3).

The neighbouring drugs with the lowest distance score connecting to both Celastrol and WA suggest they may induce similar gene expression pattern, indicating similar MoA and therapeutic potential (Table 3.1). Interestingly, Celastrol and WA are closely connected, but each belongs to a specific ‘drug community’ in the MANTRA drug network, accordingly. Although structurally different, most of these connected drugs were clustered into either of these two ‘communities’. For example, Parthenolide is a sesquiterpene lactone that occurs naturally in the plant feverfew and has been shown to have anti-leukaemia properties, preferentially targeting AML progenitor and stem cell populations. The molecular mechanism of Parthenolide-mediated apoptosis is strongly associated with inhibition of NF- κ B, proapoptotic activation of p53, and increased reactive oxygen species (ROS) [268]. Piperlongumine was also found having selective cytotoxicity to cancers, including myeloid leukaemia, by several in vitro and xenograft mouse model experiments [269, 270].

Table 3.1 Distance table of Celastrol and WA generated by Mantra (Distance threshold: 0.5)

Selected Drug	Distance score	Connected Drug	Community
Celastrol	0.352	N-{4-[(1E)-3-oxo-3-phenylprop-1-en-1-yl]phenyl}-2,2-diphenylacetamide	40
	0.369	3-tert-butyl-1-{[2-(3-methoxyphenoxy)-5-nitrobenzene]sulfonyl}urea	40
	0.401	Z-Leu-Leu-Leu-B(OH) ₂	40
	0.406	thapsigargin	40
	0.431	puromycin	40
	0.443	withaferin_A	104
	0.474	parthenolide	104
	0.481	pararosaniline	62
	0.481	phenoxybenzamine	104
	0.482	piperlongumine	40
	0.49	disulfiram	40
Withaferin_A	0.326	lomustine	104
	0.413	parthenolide	104
	0.442	(2E)-N-(2-ethylphenyl)-3-(5-nitrothiophen-2-yl)prop-2-enamide	104
	0.443	celastrol	40
	0.455	phenoxybenzamine	104
	0.467	semustine	104
	0.47	Z-Leu-Leu-Leu-B(OH) ₂	40
	0.489	PLD+	100
	0.494	N-{4-[(1E)-3-oxo-3-phenylprop-1-en-1-yl]phenyl}-2,2-diphenylacetamide	40

As shown in Figure 3.6, the similarity of gene signatures was demonstrated by comparing the top 500 up-regulated and down-regulated genes induced by these four drugs (Celastrol, WA, Parthenolide, Piperlongumine) extracted from CMap database using MANTRA tool. There were 27.8% (139/500) overlapping up-regulated genes and 15.4% (77/500) down-regulated genes between Celastrol and WA, but only 9.2% (46/500) overlapping up-regulated genes and 2.2% (11/500) down-regulated genes among all four drugs (Supplementary table 3.2).

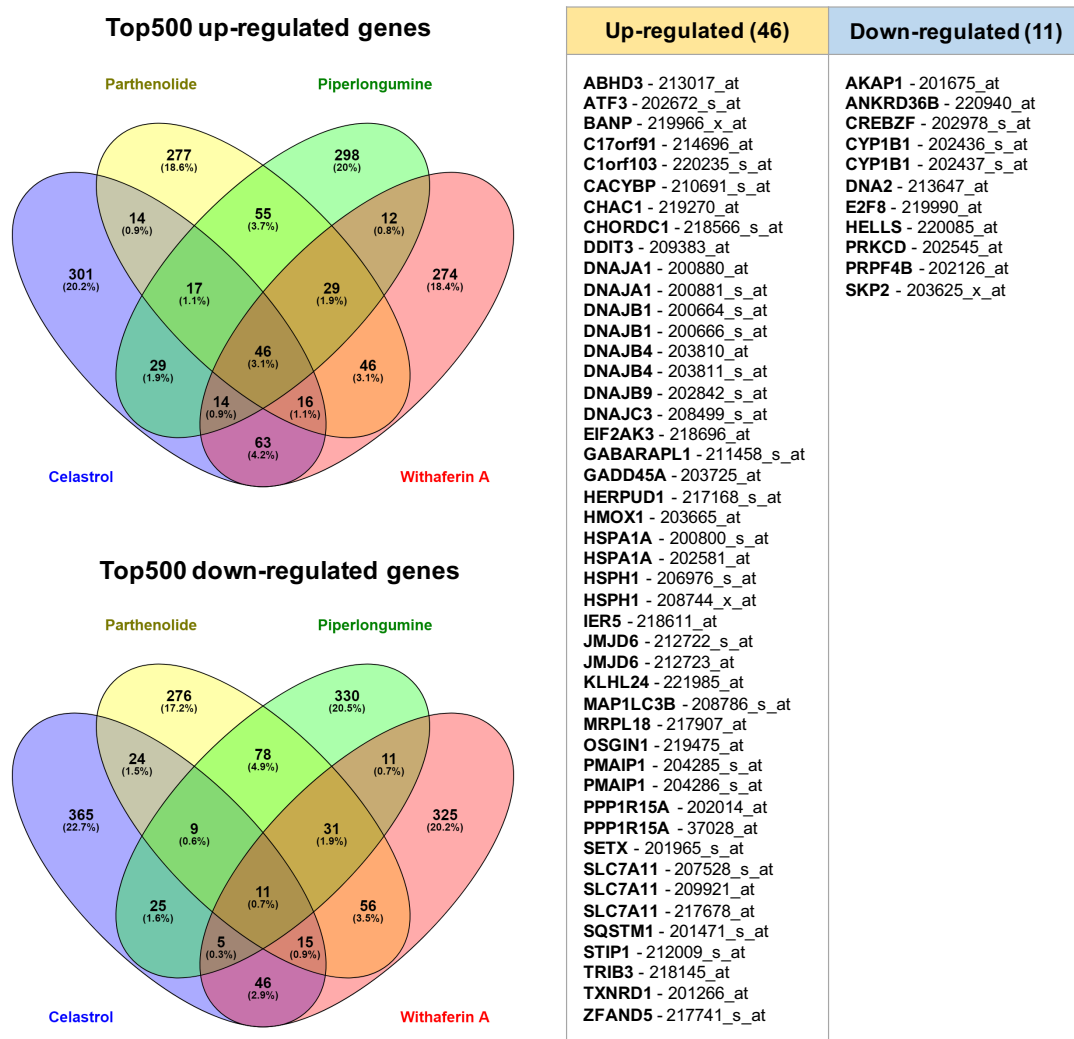


Figure 3.6 Top 500 up-regulated and down-regulated genes induced by four drugs from CMAP database. a) Venn diagrams show the common genes among the top 500 up-regulated and down-regulated genes in CMAP cell lines induced by Celastrol, WA, Piperlongumin, and Parthenolide. b) 46 overlapped up-regulated genes and 11 overlapped down-regulated genes.

4. Generation of microarray based DEG lists used as input to the connectivity mapping analysis (MDS-RAEB-2 patients)

In 2010, our group conducted a microarray-based GEP study on the bone marrow CD34+ cells of 183 MDS patients and 17 healthy controls. 43 patients with RAEB-2 were included in the study [116]. Using the same dataset, the transcriptome analysis has led to the identification of a DEG list that is associated with the RAEB-2 subtype. 1280 genes were up-regulated (normalised foldchange>1) and 978 genes were down-regulated (normalised foldchange<1) by comparing the gene expression level of RAEB-2 patient with healthy control (FDR<5%). By applying similar criteria as described above, we generated the three query gene lists based on both expression level changes (normalised fold change) and statistical significance (FDR) for CMap analysis:

1) Top 200: Top 200 most statistically significant genes, regardless of up- or down-regulation; **2) Foldchange:** Genes with a normalised fold change lower than 0.5 or higher than 2, and an FDR<5%; **3) Up-down 100:** Top 100 most up-regulated (over-expressed) and top 100 down-regulated (under-expressed) genes with an FDR<5%.

Figure 3.7 shows overlapping up- and down-regulated probes of the three DEG lists for CMap queries.

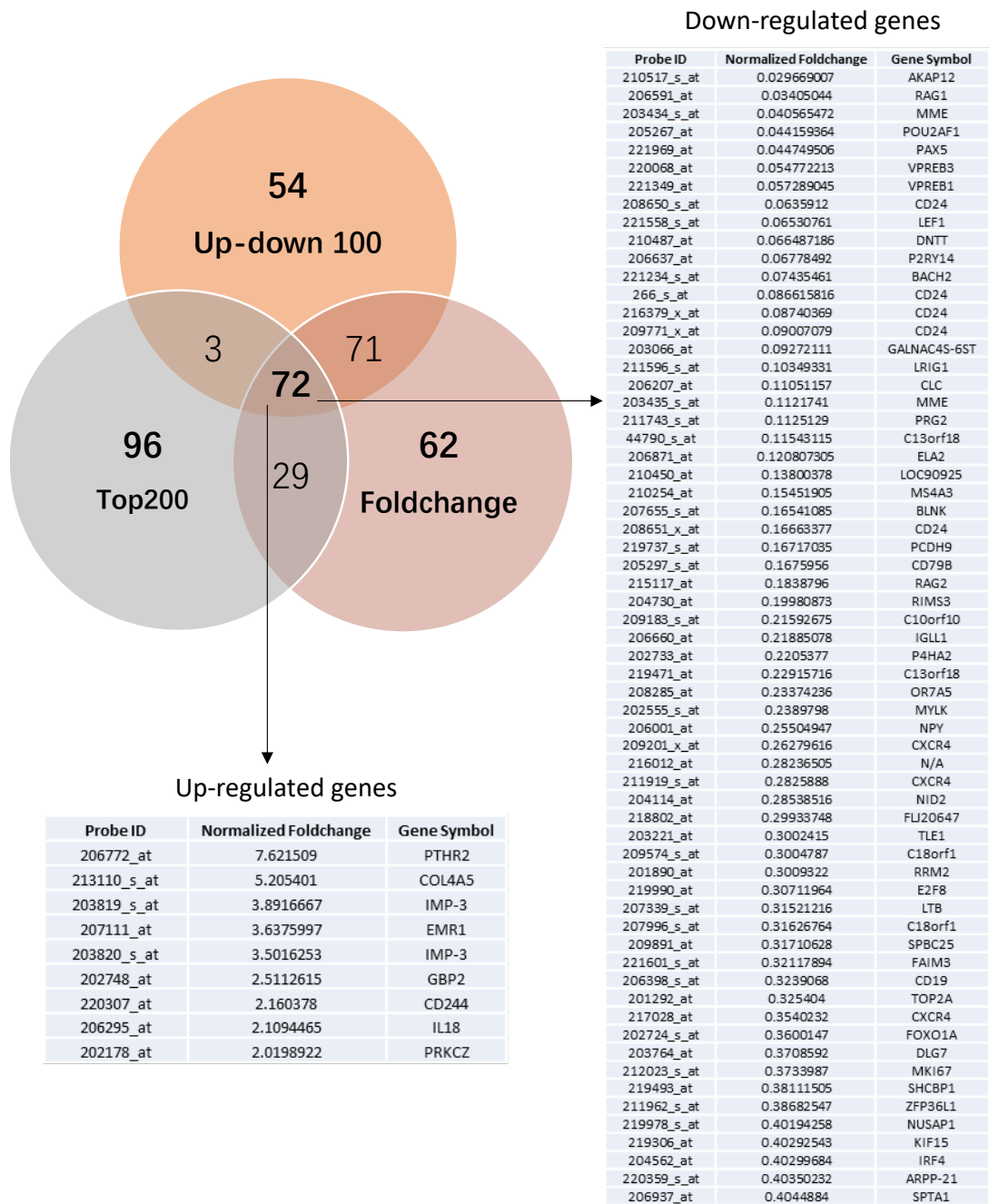


Figure 3.7. The overlapping genes of the three microarray-based DEG lists from MDS-RAEB-2 patients for CMAP queries. The tables show the overlapping up-regulated and down-regulated genes.

5. Candidate compounds identified by Connectivity map analysis of Microarray based DEG lists

5.1 Drug screening by Connectivity map

By applying similar connectivity mapping methods and computational drug screening criteria as described in RNA-seq based CMap analysis, we obtained a list of negatively connected drugs from CMap database ranked by absolute value of connectivity score.

Table 3.2 shows the top 30 drugs of the CMAP output list.

Table 3.2 Top 30 CMap drugs identified by using microarray based DEG in MDS-RAEB-2 patients

Drug name	up-down100	Foldchange	Top200	Sum
12,13-EODE	0.659	0.791	0.493	1.943
(-)-catechin	0.848	0.494	0.453	1.795
moxonidine	0.783	0.514	0.41	1.707
2-deoxy-D-glucose	0.538	0.665	0.42	1.623
5140203	0.571	0.62	0.403	1.594
dopamine	0.523	0.526	0.469	1.518
NS-398	0.597	0.296	0.565	1.458
Prestwick-1080	0.517	0.481	0.436	1.434
difenidol	0.649	0.541	0.133	1.323
vigabatrin	0.31	0.651	0.346	1.307
meropenem	0.69	0.448	0.118	1.256
tropine	0.331	0.549	0.364	1.244
kanamycin	0.527	0.429	0.257	1.213
diazoxide	0.365	0.422	0.385	1.172
methylethergometrine	0.497	0.38	0.293	1.17
16,16-dimethylprostaglandin E2	0.493	0.276	0.396	1.165
sulfadoxine	0.661	0.39	0.094	1.145
iopanoic acid	0.514	0.496	0.126	1.136
harmol	0.459	0.337	0.304	1.1
parbendazole	0.493	0.342	0.249	1.084
ethionamide	0.44	0.422	0.22	1.082
lycorine	0.281	0.438	0.352	1.071
pancuronium bromide	0.597	0.412	0.06	1.069

verteporfin	0.716	0.072	0.278	1.066
nialamide	0.573	0.174	0.277	1.024
sulfaphenazole	0.449	0.304	0.271	1.024
letrozole	0.486	0.387	0.145	1.018
aminogluthethimide	0.461	0.078	0.479	1.018
piretanide	0.473	0.368	0.173	1.014
molindone	0.656	0.15	0.176	0.982

Top-scoring drug candidates were selected for experimental validation, by prioritising the drugs that have not been already evaluated in AML/MDS studies, and/or that have not been shown to have high toxicity and narrow therapeutic window in previous studies (e.g. high in vitro IC₅₀ >20 μ M). Based on these criteria, Parbendazole (PBD), an antihelminthic benzimidazole family drug, was selected for experimental validation, which will be discussed in next chapter (Figure 3.8a).

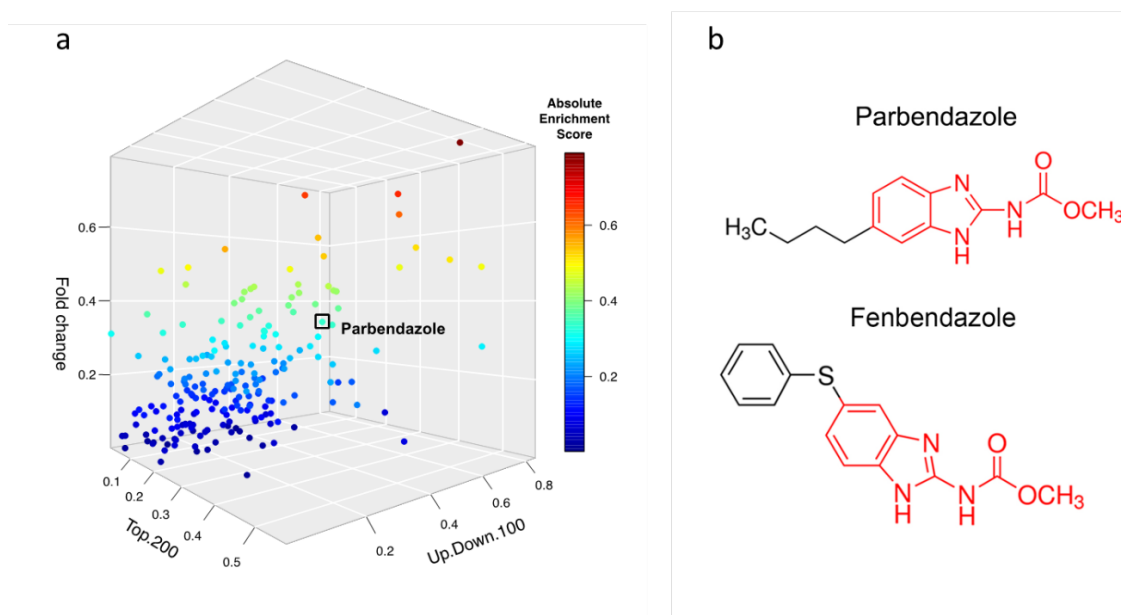


Figure 3.8. Drugs identified from the CMap analysis in RAEB-2. a) Parbendazole was one of top-scoring drug candidates and was selected for experimental validation. b) Fenbendazole is structurally similar as Parbendazole connected by MANTRA drug network, the chemical structures of Parbendazole (PBD) and Fenbendazole (FBD) were shown.

5.2 Gene expression profile based ‘Drug network’ of Parbendazole

By querying PBD to MANTRA, we found that Fenbendazole (FBD) is directly connected to parbendazole by setting a threshold distance score of 0.7 (Table 3.3). FBD is structurally similar to PBD, containing a common benzimidazole backbone (Figure 3.8b). Interestingly, several benzimidazole family drugs (e.g. Mebendazole, and Nocodazole) and our previous identified drugs by CMap analysis of RNA-seq based DEG (i.e. Celastrol and WA) were further connected to FBD and N-[2-(1,3-benzoxazol-2-yl)phenyl]-4-methylbenzenesulfonamide (Figure 3.9). These ‘-dazole’ drugs have also been identified in our RNA-seq based drug candidate lists, which suggested their possible association with splicing factor modulation (Supplementary table 3.1). FBD was the most closely connected drug to PBD and was selected for wet-lab experiments. The Top 500 up-regulated and down-regulated genes induced by PBD and FBD treatment were also extracted from CMap database using MANTRA tool, showing 10.8% (54/500) overlapping up-regulated genes and 14.8% (74/500) down-regulated genes (Supplementary table 3.3).

Table 3.3 Distance table of Parbendazole generated by Mantra (Distance threshold: 0.8)

Selected Drug	Distance score	Connected Drug	Community
parbendazole	0,654	fenbendazole	69
	0,746	N-[2-(1,3-benzoxazol-2-yl)phenyl]-4-methylbenzenesulfonamide	48
	0,751	vinblastine	70
	0,751	oxaprozin	7
	0,756	santonin	22
	0,758	paclitaxel	48
	0,76	rotenone	62
	0,762	nialamide	22
	0,764	Vitamin K-2	22
	0,766	(+)-chelidonine	40
	0,766	vinburnine	40
	0,774	2-{[4-(6-chloro-2,3-dihydroindol-1-yl)-6-methoxyquinazolin-7-yl]oxy} ethanol	11
	0,779	quinethazone	22
	0,78	chlorambucil	22
	0,781	medrysone	90
	0,781	sulconazole	90
	0,782	promethazine	90
	0,784	sanguinarine	89
	0,788	3-(2-{5-[(pyridin-3-yl)amino]-2H-pyrazol-3-yl} ethyl)-N-[4-(pyrrolidin-1-yl)-3-(trifluoromethyl)phenyl]benzamide	11
	0,789	mitoxantrone	14
	0,789	DF1270Y_cel_ultimate	90
	0,79	phthalylsulfathiazole	90
	0,791	rimexolone	32
	0,791	N-[1-(2,2-dimethylpropyl)-5-oxopyrrolidin-3-yl]-3-(furan-2-yl)propanamide	6
	0,792	doxorubicin	14
	0,792	colchicine	69
	0,793	beta-escin	90
	0,794	ciclacillin	22
	0,795	doxazosin	52
	0,796	gliclazide	61

comparison, respectively. Parbendazole was selected from top-ranking CMap candidates by querying the gene signature of MDS RAEB-2 patients.

The identification of bioactive compounds derived from medicinal plants has attracted considerable interest in recent years, including Celastrol and WA. Many studies have reported their anti-inflammatory, anti-cancer, and neuroprotective properties by modulating a variety of molecular targets for potential therapeutic applications [271-278]. Celastrol, also called tripterine, is a pentacyclic triterpenoid that belongs to a small category of natural products of triterpene quinone methides. It has been suggested that Celastrol could affect protein function through the formation of covalent Michael adducts (electrophilic sites in the structure of quinone methide rings that susceptible to the nucleophilic attack of thiol groups of cysteine residues to form covalent protein adducts) [279]. WA is a steroidal lactone, the first member of the withanolide class of ergostane type product to be discovered. Structurally, withanolides consist of a steroid backbone bound to a lactone or one of its derivatives [278, 280]. WA has been widely studied for its anti-inflammatory properties, more recent studies characterizing the biological activities of WA have expanded to include the determination of its anti-cancer properties. Regulation of the activity of NF- κ B signalling is one of the key molecular mechanisms shared by Celastrol and WA [279-283]. Piperlogumine and Parthenolide identified by MANTRA drug network connecting to Celastrol and WA, are also natural bioactive small molecules isolated from plants, which were proved to be a promising anti-cancer agent in single or combination therapy [269, 284, 285].

Parthenolide is a sesquiterpene lactone of the germacranolide class, which occurs naturally in the plant feverfew (*Tanacetum parthenium*) [269]. Piperlongumine is an alkaloid extracted from *Piper longum* Linn [285]. PBD and FBD (identified by MANTRA) are synthetic benzimidazoles with potent inhibition of microtubule assembly. They exhibit a broad-spectrum anthelmintic activity, in the same class as FDA-approved anthelmintics mebendazole, flubendazole, and others [286, 287]. The anti-cancer properties of benzimidazoles have been demonstrated in various pre-clinical studies across a number of different cancer types, including melanoma, leukaemia, medulloblastoma, breast cancer, prostate cancer and colorectal cancers [288-293]. However, the repurposing potential of benzimidazole-based anthelmintics have not been studied in MDS or MDS derived AML thus far.

Concerning the methods employed in this chapter, several aspects of empirical findings are noteworthy. Firstly, our approach did not involve drug screening based on the targeting of single molecules. The connectivity mapping approach that we used has broad application for the identification of candidate new drugs for leukaemia or solid cancers for which large-scale gene expression datasets exist. By looking at the similarity of GEP generated by connected drugs, it is possible to find previously unrecognized MoA, and genes and pathways that are targets of drugs that could be repositioned in a different disorder. That is the reason why the compounds identified in our studies and other CMap-based repositioning studies are typically ‘multitargeting’ agents, exhibiting action on multiple cellular targets and thus may be repurposed in a

wide spectrum of diseases (discussed further in Chapter VI). As shown in Figure 3.6, the similarity of gene signatures was demonstrated by comparing the top 500 up-regulated and down-regulated genes induced by the four drugs from CMap database. Although these gene signatures do not directly indicate the MoA of these compounds in MDS due to differences between cell lines and primary MDS cells, dosage effects and differences in experimental settings, some of the overlapping genes were identified as dysregulated in MDS. For example, *DNAJB* chaperone subfamily genes involved in protein folding were significantly down-regulated in our MDS gene signatures, which might be conversely up-regulated by these drugs. Also, pro-apoptotic transcription factor gene *DDIT3* and antioxidant gene *HMOX1* have been shown to be down-regulated and aberrantly methylated in MDS [294-296].

Importantly, as a key input to the connectivity mapping process, the CMap query genes are crucial in returning biologically meaningful results. The determination of query gene signatures is a recognized variable for drug connectivity mapping. In practice, query gene signatures are created by end users as input to connectivity mapping analysis based on their own experience. Limited guidance from the original developers of CMap and a lack of systematic research on methods for creating high standard gene signatures for connectivity mapping represent obstacles to the application of CMap [243]. Some examples of the variability in CMap analysis are the use of sets of genes between 10 and 500 as query input (as recommended by the developers), the difference in size between up- and down-regulated gene sets, and the use of either up- or down-

regulated gene sets only in a query list. A commonly accepted approach is to generate query gene signatures that are statistically significant and are associated with a given disease or phenotype. However, a limitation of this approach is the selection of a subset of DEG, which is arbitrarily determined by a cutoff of FDR rate and expression fold change from a large number of genes. Moreover, the CMap database employs five different human cancer cell lines (MCF7, PC3, SKMEL5, HL60 and ssMCF7) treated with a wide range of compounds at varying concentrations and times to generate the reference database. The difference of the gene expression pattern between a disease of interest and drug-treated CMap cell lines may lead to deviation or mismatching of ‘drug-phenotype’ connection.

Chapter IV Investigation of the effects of drug candidates on cell viability, cell apoptosis, cell cycle, and differentiation

A. Introduction and aims

In this chapter, we explored the therapeutic potential of these drug candidates by testing them on cell lines and patient primary samples. Several leukaemia cell lines were used for experimental validation, including MDS cell line (MDS-L), AML/MDS cell lines (i.e., derived from an AML patient with a preceding MDS; SKM1, F36P), acute T cell leukaemia cell line (Jurkat), chronic myeloid leukaemia cell line (MEG-01), acute promyelocytic leukaemia cell line (HL60), erythroleukaemia cell line (TF-1). Table 4.1 shows the clinical and biological information of the leukaemia cell lines. Human umbilical cord blood-derived erythroid progenitor cell line 2 (HUDEP2), derived from healthy CD34⁺ haematopoietic stem/progenitor cells in umbilical cord blood, was used as a non-leukaemia control cell line (Table 4.1). Besides the above-mentioned leukaemia cell lines, MDS patient bone marrow mononuclear cells and purified progenitor CD34⁺ cells were also used for validation experiments.

The aims of this chapter are to assess the effects of candidate drugs of interest on cell viability, apoptosis, cell cycle, and differentiation in different leukaemia cell lines and in primary MDS samples. We first assessed the Celastrol, WA, Piperlogumine, and Parthenolide identified by analysing RNA-seq data. And then tested PBD and FBD discovered from microarray-based screening.

Table 4.1: Clinical and biological information of leukaemia cell lines

Leukaemia lines	Type of leukaemia	Cell type	Clinical data	Immunoprofile	Key genetic alterations	Differentiation
K-562	CML	erythroid	Patient: 53-year-old female Treatment status: at blast crisis Specimen site: pleural effusion Established: 1970	T-/NK cell: CD1-, CD2-, CD3-, CD4-, CD5-, CD7-, CD8-, CD28-, CD56-, CD57-, TCR $\alpha\beta$ -, TCR δ - B-cell: CD10-, CD19-, CD20-, CD21-, CD22-, CD40-, CD80+, CD86- Myelomonocytic: CD13+, CD14-, CD15+, CD33+, CD65+ Ery-megakaryocytic: CD9+, CD36-, CD41(+), CD42b-, CD61+/-, CD235a+, PPO-, vWF- Non-lineage/stem cell: CD34-, CD38-, CD45+, CD70-, CD71+, CD133-, HLA-DR-, TdT- Adhesion: CD11b+/-, CD11c-, CD18-, CD44+, CD54+, CD58+	Specifics Ph+ t(9;22)(q34;q11) → BCR-ABL1 b3-a2 fusion gene Gene alteration ABL amplification, P15INK4B deletion, P16INK4A deletion, P53 mutation	hemin → ery differentiation; TPA → meg differentiation
SKM-1	MDS (RAEBT) → AML M5	/	Patient: 76-year-old Japanese male Treatment status: refractory Specimen site: peripheral blood Established: 1989	T-/NK cell: CD2-, CD3-, CD4+, CD5-, CD7-, CD8- B cell: CD19- Myelomonocytic: CD13+, CD14(+)/-, CD15+, CD33+ Non-lineage/stem cell: CD34+, HLA-DR+ Adhesion: CD11b+	P53 mutation	/
HL-60	initially AML M3, later changed to AML M2	myelocytic cell	Patient: 35-year-old female Disease diagnosis: Treatment status at diagnosis Specimen site: peripheral blood Established 1976	T-/NK cell: CD1-, CD2-, CD3-, CD4+, CD5-, CD6-, CD7-, CD8-, CD28-, CD56+, CD57-, TCR $\alpha\beta$ -, TCR $\gamma\delta$ - B-cell: CD10-, CD19-, CD20-, CD21-, CD22-, CD23-, CD24+, CD40-, CD80-, CD86+/-, sIg- Myelomonocytic: CD13+, CD14-, CD15+, CD16-, CD33+, CD65+, CD203c- Ery-megakaryocytic: CD9+, CD41a-, CD42b-, CD61-, CD235a-, vWF- Non-lineage/stem cell: CD30+, CD34-, CD38+/-, CD45+, CD71+, CD133-, HLA-DR-, TdT- Adhesion: CD11b-, CD11c+, CD18+, CD44+ CD54-, CD62L-	Amplification of MYC, NEU, NRAS; P53 alteration	inducible to neutro, mono, macro, eosino, baso differentiation with various reagents
F-36P	MDS (RAEB) → AML M6	/	Patient 68-year-old male Disease diagnosis MDS (RAEB) → AML M6 Treatment status at diagnosis Specimen site pleural effusion Established 1989	T-/NK cell: CD3-, CD4-, CD5-, CD8- B-cell: CD10-, CD19- Myelomonocytic: CD13+, CD14-, CD15-, CD33+ Ery-megakaryocytic: CD41a+/-, CD42b-, CD71+, CD235a+, PPO(+) Non-lineage/stem cell: CD34+, CD45+	P16INK4A deletion, P53 mutation	EPO → ery differentiation

TF-1	AML M6	erythroid	Patient 35-year-old male Disease diagnosis AML M6 Treatment status at diagnosis Specimen site bone marrow Established 1987	T-/NK cell: CD3-, CD4-, CD5-, CD7-, CD8-, CD57- B-cell: CD10-, CD19- Myelomonocytic: CD13+, CD14-, CD15-, CD33+, CD68(+) Ery-megakaryocytic: CD41+, CD42a(+), CD42b+, CD61+, CD62-, CD235a(+/-), PPO+ Non-lineage/stem cell: CD34+, CD38+, CD71+, HLA-DR+ Adhesion: CD11a-, CD11b+, CD11c+, CD18+, CD44+, CD54+, CD56-, CD62L-, CD106-	P53 mutation	hemin, δ -ALA $\rightarrow$ ery differentiation; TPA $\rightarrow$ macro or meg differentiation
MDS-L	MDS (RARS/R AEB) (prior to leukaemic transformation)	/	Patient: 52-year-old male Treatment status: at diagnosis Specimen site: bone marrow Established: 1992	T-/NK cell: CD2-, CD3-, CD7- B-cell: CD19-, CD20(+) Myelomonocytic: CD13+, CD14(+), CD33+ Ery-megakaryocytic: CD41(+), CD61-, CD235a- Non-lineage/stem cell: CD34+, HLA-DR+ Adhesion: CD11b+	del(5)(q13q35), NRAS mutation	G-CSF, GM-CSF, IL-3, SCF $\rightarrow$ neutro differentiation
MEG-01	CML	megakaryocytic cell	Patient: 55-year-old male Treatment status: at megakaryocytic blast crisis Specimen site: bone marrow Established: 1983	T-/NK cell: CD1-, CD2-, CD3-, CD4+/-, CD5-, CD7-, CD8-, CD28-, CD56-, CD57(+) B-cell: CD10-, CD19-, CD20-, CD21+, CD22-, CD23-, CD203c+ Myelomonocytic: CD13+, CD14+/-, CD15+, CD16-, CD33+, CD68(+/-) Ery-megakaryocytic: CD9+, CD36+, CD41a+, CD42a-, CD42b(+), CD61+/-, CD62P-, CD235a-, PPO+, vWF+ Non-lineage/stem cell: CD34(+/-), CD38-, CD71-, CD133-, HLA-DR+/-, TdT- Adhesion: CD11b+, CD44+	Ph+ t(9;22)(q34;q11) $\rightarrow$ BCR-ABL1 b2-a2 fusion gene P15INK4B deletion, P16INK4A deletion, P53 mutation	TPA $\rightarrow$ meg differentiation
Jurkat	T-ALL	T-cell	Patient: 14-year-old male Treatment status: at 1st relapse Specimen site: peripheral blood Established: 1976	T-/NK cell: CD1+, CD2+, CD3+, cyCD3+, CD4+, CD5+, CD6+, CD7+, CD8-, CD27+, CD28+, CD57(+), TCR $\alpha\beta$ +, cyTCR α +, cyTCR β +, TCR $\gamma\delta$ -, cyTCR γ - B-cell: CD9+, CD10-, CD19-, CD20-, CD21-, CD24-, CD40- Myelomonocytic: CD13-, CD14-, CD15-, CD16-, CD33- Ery-megakaryocytic: CD41a-, CD61- Non-lineage/stem cell: CD30+, CD34(+), CD38+, CD45RA+, CD45RO-, CD71+, CD133-, HLA-DR-, TdT+ Adhesion CD11b-, CD29+	IFNA deletion, P15INK4B deletion, P16INK4A deletion, P53 mutation	/

B. Results

1. Celastrol and Withaferin A selectively inhibited the proliferation of leukaemia cell lines

Cell viability assay is a common method to determine the number of viable cells in culture based on quantitation of the ATP present, which signals the presence of metabolically active cells. In time-and dose-course drug treatment assays, when normal cells undergo apoptosis and necrosis, the measurement of ATP will be characteristically decreased, indicating the drug-induced cytotoxicity and inhibited cell proliferation. As shown in *Chapter III, section B-3.1*, Celastrol and WA were identified as top-ranking compounds from the CMap analysis of RNA-seq based DEG in MDS patients with splicing factor mutations. First, we assessed the effects of Celastrol and WA on cell viability in different leukaemia cell lines (MDS-L, SKM1, F36P, Jurkat, MEG-01), in the HUDEP2 cell line and in purified bone marrow CD34⁺ cells from a healthy donor. Cell viability was measured by ATP-based assay after 48 h and 72h, and the extent of drug-induced viability inhibition was normalised to DMSO-treated control. Figure 4.1 shows the drug response curves in different cells. Both Celastrol and WA were shown to be considerably cytotoxic against leukaemic cell lines. Treatment with these drugs resulted in a potent reduction of cell viability in a dose-dependent manner, particularly in the cell lines derived from myeloid malignancies, including MDS-L, F36P, and MEG-01. However, Jurkat cells were less sensitive to both Celastrol and WA treatment. HUDEP2 cell line, an immortalized human CD34⁺ cell line derived from umbilical

cord blood harbouring no leukaemic mutations, exhibited almost no suppression in viability. Also, the healthy bone marrow CD34⁺ cells showed a comparably modest sensitivity to these drugs. Importantly, the inhibition doses of both Celastrol and WA in leukaemia cells were at micromolar range, demonstrating an IC₅₀ < 1200 nM after 48 h (Table 4.2). These results indicate that Celastrol and WA exhibit preferential toxicity in leukaemic cells, and have a wide potential therapeutic window.

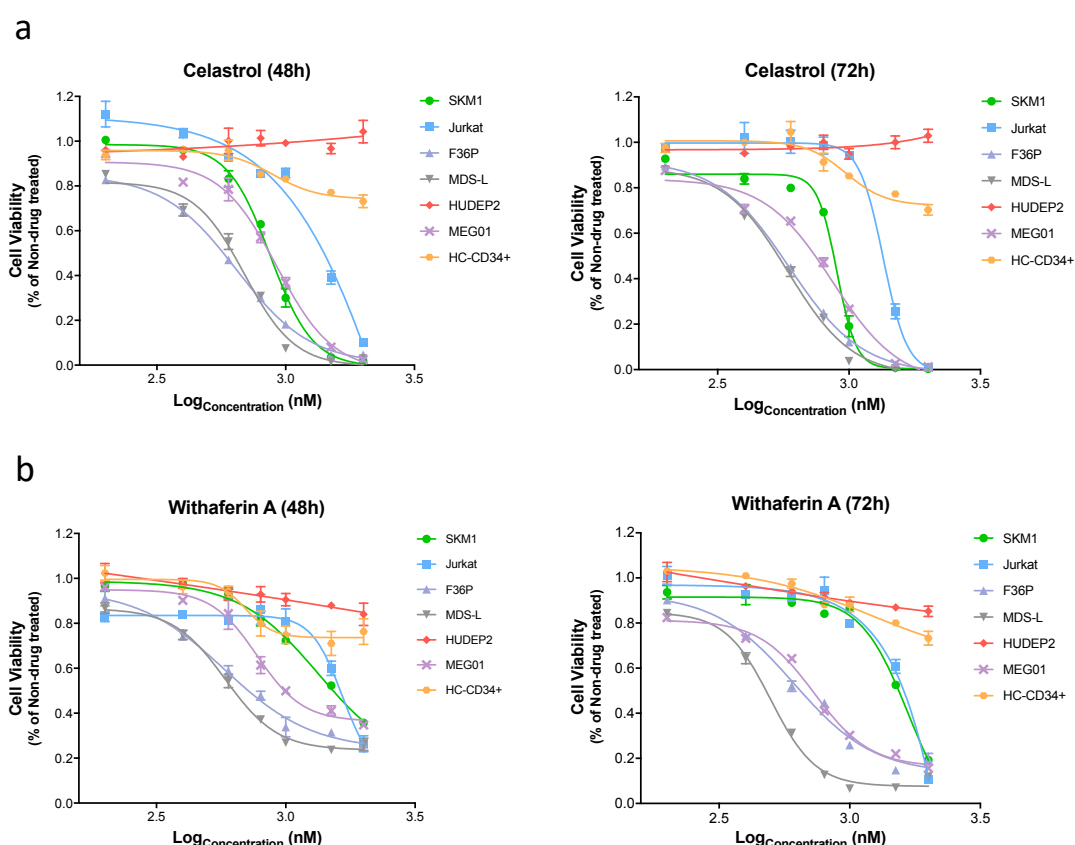


Figure 4.1 Drug dose-response curves in cells by Celastrol or WA treatment. Cell viability of cell lines measured by ATP-based assay normalised to vehicle-treated control. Drug doses are represented as logarithm base 10 of nanomolarity. Each point represents the average from triplicate samples and error bars represent SD. a) Cell viability of cell lines after 48 h and 72h Celastrol treatment. b) Cell viability of cell lines 48 h and 72h after WA treatment.

Table 4.2 The IC₅₀ of 72h drug treatment in cell lines (nM)

	Celastrol		Withaferin A	
	48h	72h	48h	72h
SKM-1	835.4-908.3	871.6-927.3	1104-1627	1313-2146
Jurkat	Very Wide	1273-1429	1273-2146	Very wide
F36P	621.3-666.4	574.8-609.3	519.8-647.6	545.6-687.0
MDS-L	644.7-743.5	550.6-623.0	553.7-615.4	474.2-522.8
MEG01	855.9-998.7	791.1-945.5	726.1-826.1	692.0-776.3
HUDEP2	Very wide	Very wide	Very wide	Very Wide
CD34+	725.2-1106	805.1-1178	612.8-794.9	420.6-3452

2. Celastrol and Withaferin A induced cell cycle arrest at G2/M phase and apoptosis in leukaemia cell lines

The assessment of cell cycle and apoptosis provides important information on the mechanism of action as well as the therapeutic efficacy of anti-cancer drugs. During cell cycle progression, proliferating cells sequentially undergo a transition of G1→S→G2→M phases for synthesis of DNA, preparation of cell division and subsequent mitosis process. To investigate whether selected drugs of interest suppress cell growth by hindering cell cycle progression, we conducted cell cycle assays by staining the DNA content with Propidium iodide (PI) in leukaemia cell lines and HUDEP2 cells after drug treatments.

Figure 4.2 shows the proportion of cells at different cell cycle stages with or without drug exposure at a specific dose. The cells were cultured for 12 h because more prolonged exposure leads to cell apoptosis and death. Interestingly, both Celastrol and

WA induced cell cycle arrest at the G2/M phase in SKM-1, F36P and MDS-L cells. The cell population at the G2/M phase was significantly increased (by approximately more than 30%) in cells treated with WA at 1000 nM. Although less than WA, cells treated with 500nM Celastrol also showed significant increase in cell population at G2/M phase (by approximate 10%-20%). In agreement with the cell viability data, cell cycle analysis of HUDEP2 cells did not show significant differences associated with the treatment with Celastrol or WA.

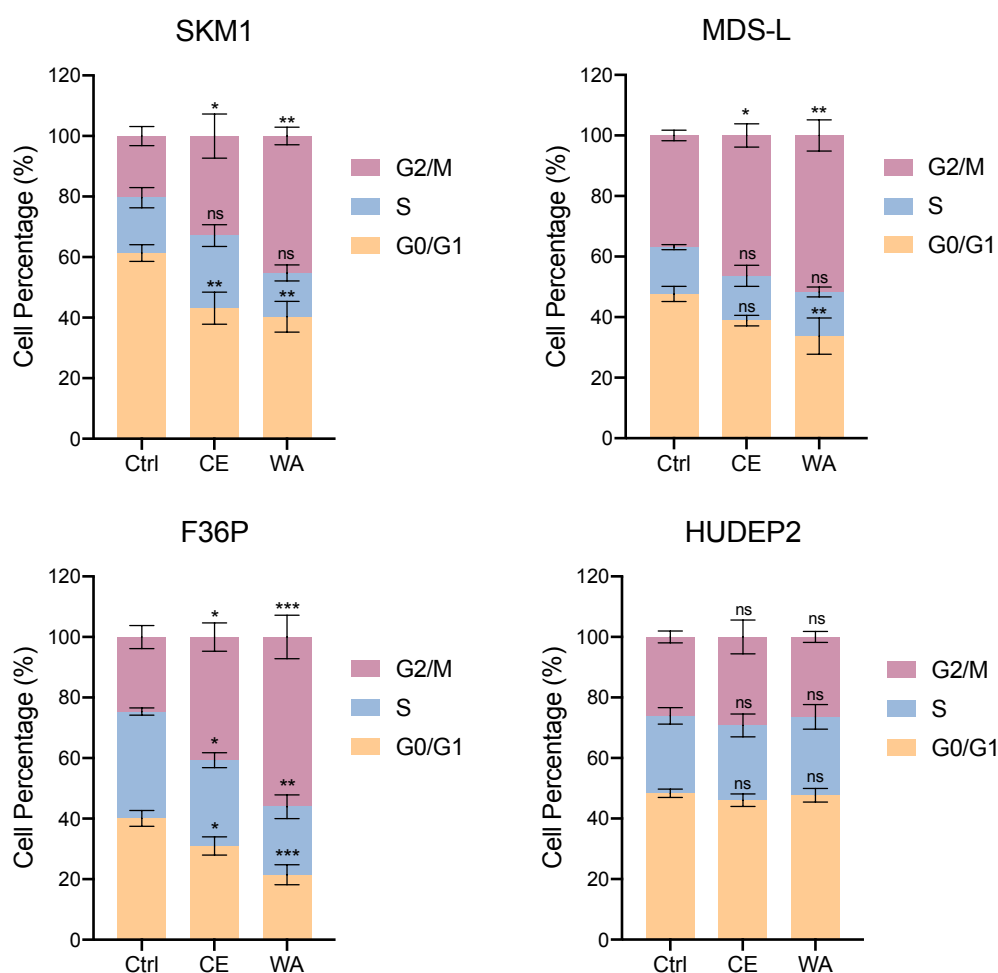


Figure 4.2 Cell cycle analyses by PI staining. Cell lines were treated with Celastrol (500 nM) and WA (1000 nM) for 12 h, and the cells were stained by PI and analyzed by flow cytometry. All values are plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

Apoptosis is a carefully regulated process of cell death distinguished from necrosis, or accidental cell death, by characteristic morphological and biochemical changes, including compaction and fragmentation of the nuclear chromatin, shrinkage of the cytoplasm, and loss of membrane asymmetry. We also performed apoptosis assays to evaluate whether Celastrol and WA convey their anti-proliferative effects in leukaemia cell lines through the induction of apoptosis. The cells were treated at same concentration as cell cycle analysis (Celastrol, 500 nM; WA, 1000 nM) for 24 h to determine their sensitivity to each drug. Annexin V-FITC/PI dual staining showed an increased apoptotic cell population after drug treatment in F36P, MDS-L and SKM1 cell lines (Figure 4.3). In line with the cell viability data, MDS-L cells were most susceptible to Celastrol and WA, which demonstrates a strong induction of apoptosis and necrosis after treatment. In contrast, Celastrol and WA induced apoptosis in HUDEP2 cells to a much lesser extent. (Figure 4.3).

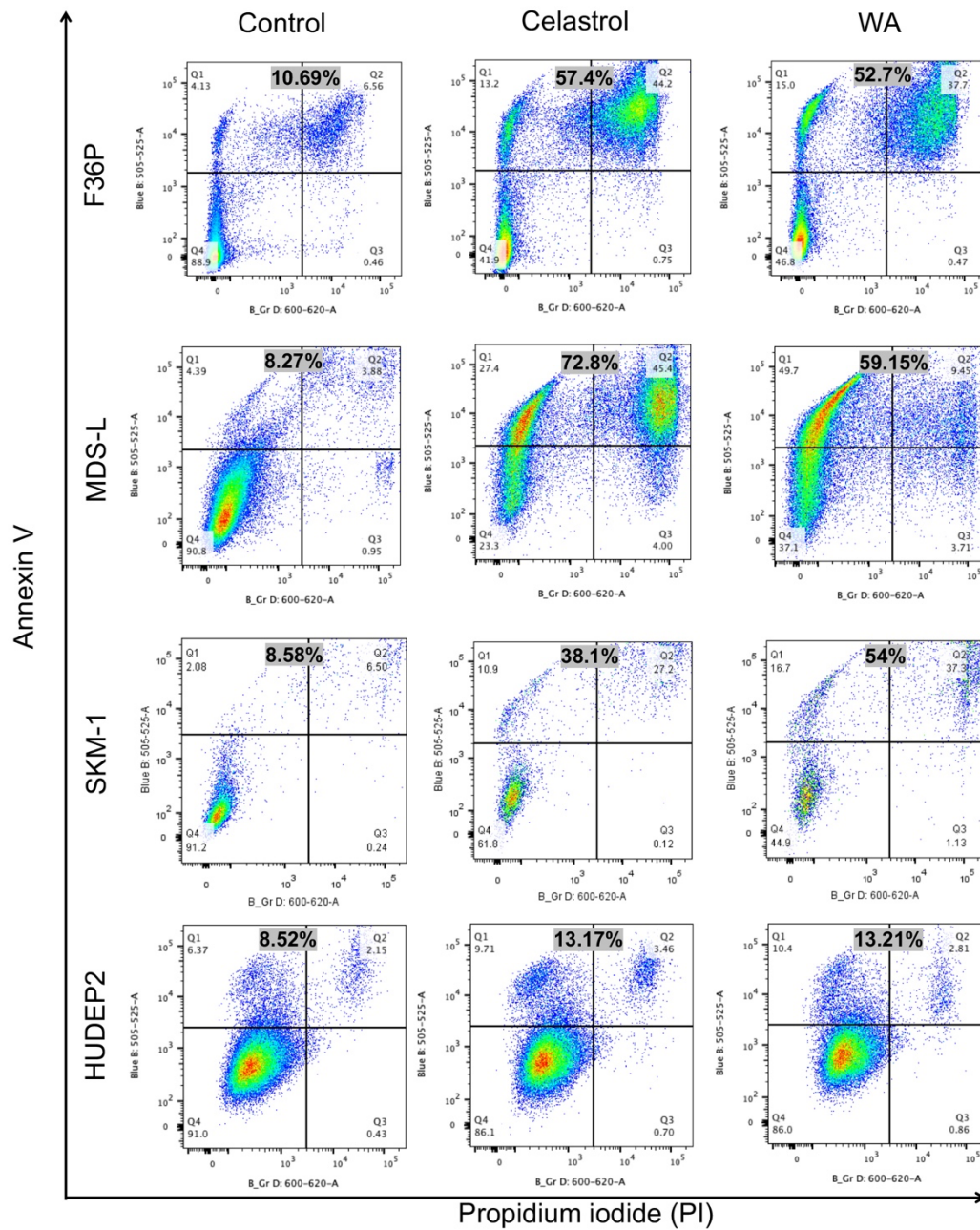
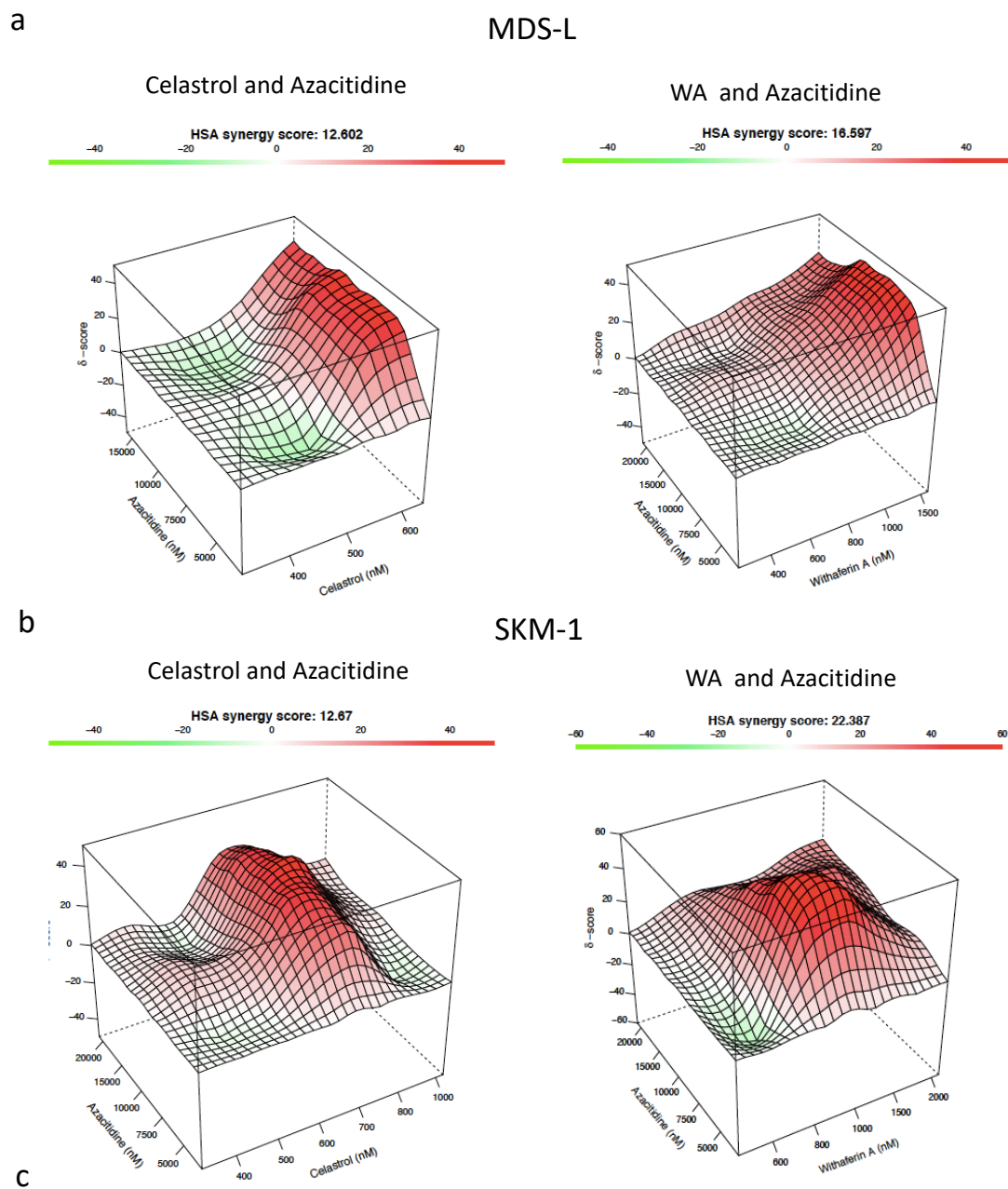


Figure 4.3. Cell apoptosis induced by Celastrol and WA. F36P, SMK1, MDS-L and HUDEP2 cells were treated with 500 nM Celastrol or 1000 nM WA for 24 h. Cell apoptosis was measured by using Annexin-V/PI dual staining. A single-positive staining for Annexin-V is an early indicator of apoptosis (Annexin-V+/PI-, upper-left area), and the double-positive staining for Annexin and PI implies late apoptosis or necrosis (Annexin-V+/PI+, upper-right area).

3. Drug combinations showed synergistic effects in inhibiting leukaemia cell proliferation

We conducted drug combination assays to determine whether there are any synergistic or additive effects of Celastrol and WA, and of either drug in combination with drugs that are used as therapy for MDS. The cytotoxic effects of Celastrol or WA on leukaemia cells were enhanced by the hypomethylating agent Azacitidine (Figure 4.4). We have also determined the synergy scores of Azacitidine and Celastrol, and Azacitidine and WA on SKM1 and MDS-L cells. The synergistic effects were measured by the highest single agent (HSA) model, where the synergy score quantifies the excess over the highest single drug response. A positive synergy score suggests synergic effects. Various dose combinations showed a positive synergy score (δ -score) in SKM1 and MDS-L cells, indicating synergy of Azacitidine with Celastrol and WA. The δ -score of Celastrol and Azacitidine combination is 12.68 in SKM1 cells and 12.602 in MDS-L cells, whereas WA and Azacitidine combination treatment exhibited a higher δ -score with 22.456 in SKM1 cells and 16.692 in MDS-L cells (Figure 4.4). Interestingly, combination treatment using Celastrol and WA showed a strong synergistic effect in SKM-1 cells (δ -score=9.742) and MDS-L cells (δ -score=4.438), and also in HUDEP2 cells to a lesser extent (δ -score=1.668) (Figure 4.5). The synergistic effects of Celastrol and WA indicate that these two drugs may have different targets or mechanisms of action.



Cell line	Drug combination	Synergy score	Most synergistic area score
MDS-L	Celastrol- Azacitidine	12.6	22.62
MDS-L	Withaferin A - Azacitidine	16.6	29.09
SKM1	Celastrol- Azacitidine	12.67	25.14
SKM1	Withaferin A - Azacitidine	22.46	40.75

Figure 4.4 Visualization of 3D synergy maps showing overall synergy scores (δ -score) for combination treatment of Celastrol or WA and Azacitidine on (a) SKM1 cells and (b) MDS-L cells. A positive synergy score (δ -score) (Red) denotes synergy. (c) Table showing the synergy score and most synergistic area score of drug combinations in SKM-1 and MDS-L cells.

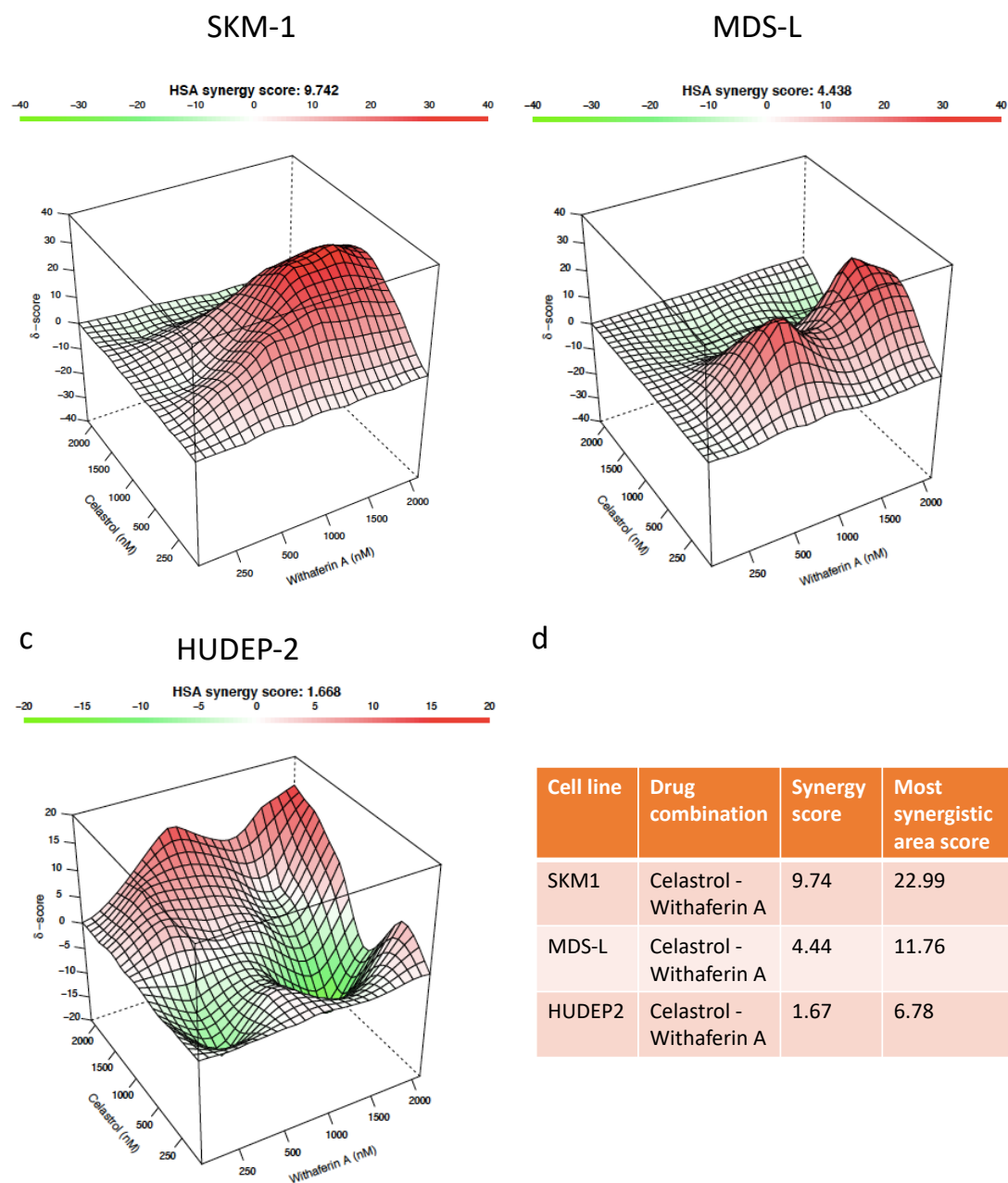


Figure 4.5 Visualization of 3D synergy maps showing overall synergy scores (δ -score) for combination treatment of Celestrol and WA on (a) SKM1 cells and (b) MDS-L cells (c) HUDEP-2 cells. A positive synergy score (δ -score) (Red) denotes synergy. (d) Table showing the synergy score and most synergistic area score of drug combinations in SKM-1, MDS-L, and HUDEP2 cells.

4. Celastrol and Withaferin A preferentially inhibit the proliferation of MDS patient CD34+ cells

We tested the effects of Celastrol and WA on the viability of CD34+ haematopoietic stem and progenitor cells isolated from the bone marrow of MDS patients (n=3) and healthy individuals (n=3). Bone marrow mononuclear cells were cultured in CD34+ expansion medium for two to three days to promote the expansion of the CD34+ population. Then, the CD34+ cells were isolated from bone marrow samples by using CD34 magnetic beads, and the purity of CD34+ cells was verified by flow cytometry. After 72h drug treatment, both Celastrol and WA showed inhibition of cell viability in a dose-dependent manner. Notably, these compounds showed preferential inhibition of CD34+ cells from MDS patients compared to CD34+ cells from healthy control. The IC₅₀ of Celastrol and WA in CD34+ patient cells was also at nanomolar range (Figure 4.6).

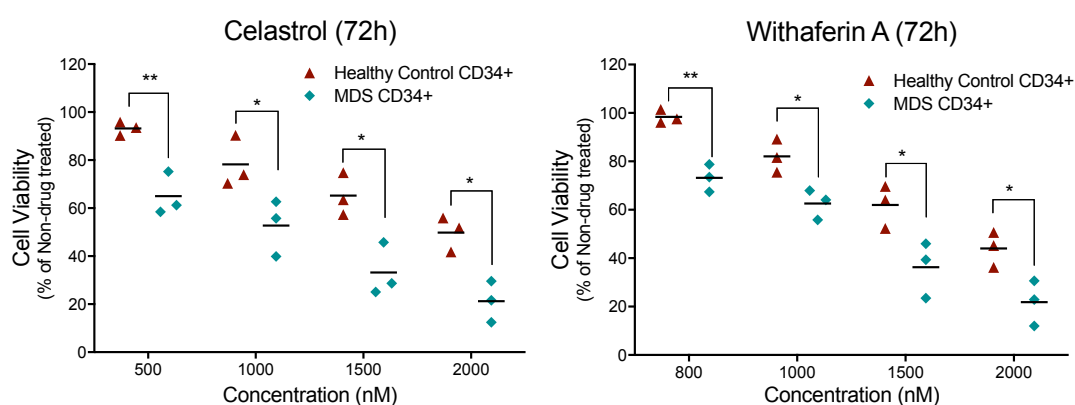


Figure 4.6 Cell viability of CD34+ cells from MDS patients and healthy donors after 72h Celastrol or WA treatment with indicated concentrations. All values are plotted as mean $\pm$ SEM. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

5. Piperlogumine and Parthenolide inhibited the proliferation of leukaemia cell lines

As shown in *Chapter III, section B-3.2*, several drugs, such as Piperlogumine and Parthenolide, showed a strong connection with Celastrol and WA in a drug network. We tested the sensitivity of leukaemia cells to treatment with Piperlogumine and Parthenolide, and found that these drugs elicited a dose-dependent inhibition of cell viability of similar magnitude to that observed with Celastrol and WA. The proliferation of leukaemic cell lines was markedly reduced by Piperlogumine and Parthenolide, whereas HUDEP2 cells exhibited a modest reduction of viability (Figure 4.7). However, the inhibition doses of both Piperlogumine and Parthenolide in leukaemia cells are higher than Celastrol and WA, with an IC₅₀ span at micromolar ranges not higher than 15 μ M (Table 4.3).

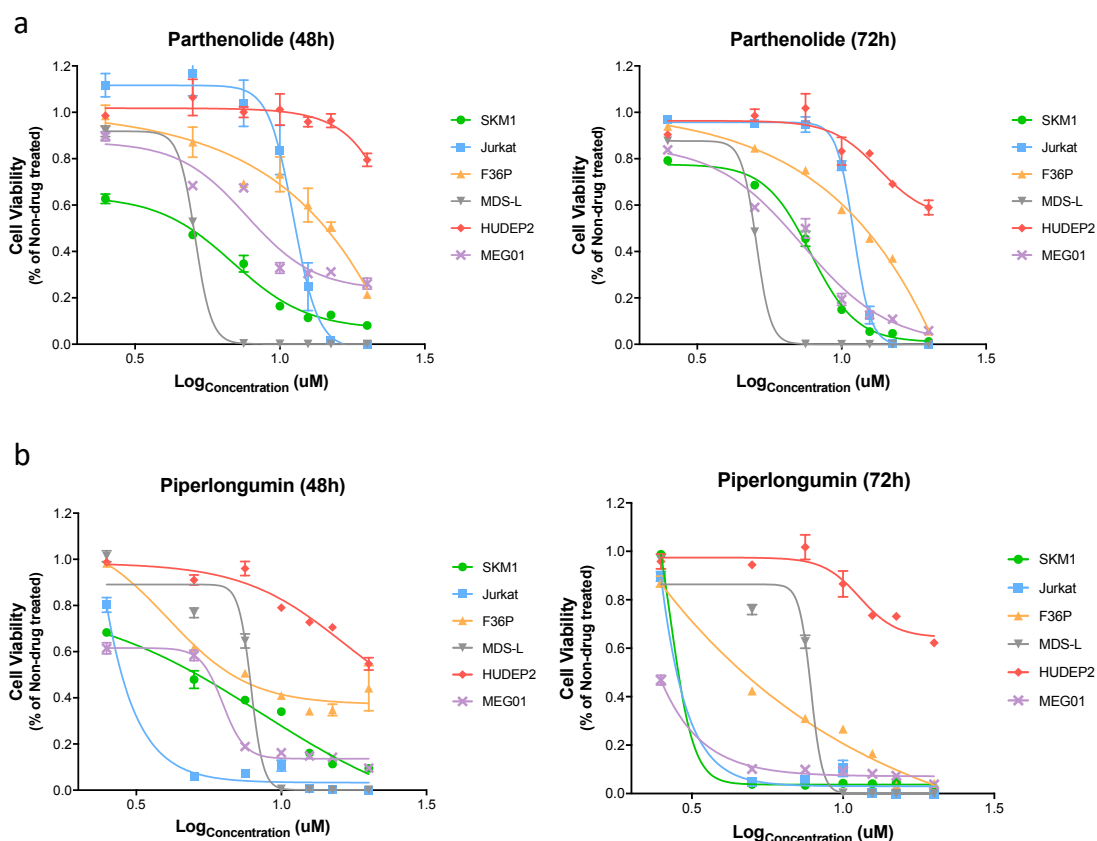


Figure 4.7 Drug dose-response curves in cells treated by Piperlogumine or Parthenolide. Cell viability of cell lines measured by ATP-based assay normalised to vehicle-treated control. Drug doses are represented as logarithm base 10 of micromolarity. Each point represents the average from triplicate samples and error bars represent SD. a) Cell viability of cell lines after 48 h and 72h Parthenolide treatment. b) Cell viability of cell lines 48 h and 72h after Piperlongumin treatment.

Table 4.3 The IC50 of Parthenolide and Piperlongumin treatment in cell lines (µM)

	Parthenolide		Piperlongumin	
	48h	72h	48h	72h
SKM-1	6.130-7.723	7.501-8.132	4.975-14.71	Very wide
MDS-L	4.965-5.235	4.931-5.5197	6.811-9.047	6.521-9.399
Jurkat	10.40-11.90	10.81-11.22	Very wide	Very wide
MEG01	6.431-9.185	6.284-8.466	5.782-6.812	Very wide
F36P	Very wide	Very wide	2.073-7.957	Very wide
HUDEP2	Very wide	10.10-18.10	6.235-42.5	9.766-13.19

6. PBD and FBD inhibit the proliferation of leukaemia cell lines

As shown in Chapter III, section 4, PBD and FBD were identified from the CMap and MANTRA analysis of microarray-based DEG in MDS patients with RAEB-2. We assessed the effects of PDB and FBD on cell viability in different myeloid leukaemia cell lines (MDS-L, SKM1, F36P, TF-1, HL60, K562) and HUDEP2 cells. Cell viability was measured by ATP-based assay at 48 h and 72h after drug treatment, and the extent of drug-induced viability inhibition was normalised to DMSO-treated control. Figure 4.8 shows the drug dose-response curves of PBD or FBD treatment in different cell lines. Both PBD and FBD showed dose-dependent cytotoxicity against all leukaemic cell lines, especially in MDS-L and F36P cells. However, K562 and HUDEP2 cell lines exhibited relatively lower sensitivity to PBD and FBD (Figure 4.8). These results indicate that benzimidazole-based anthelmintics have a potent anti-leukaemia toxicity, with a wide in vitro therapeutic window (Table 4.4).

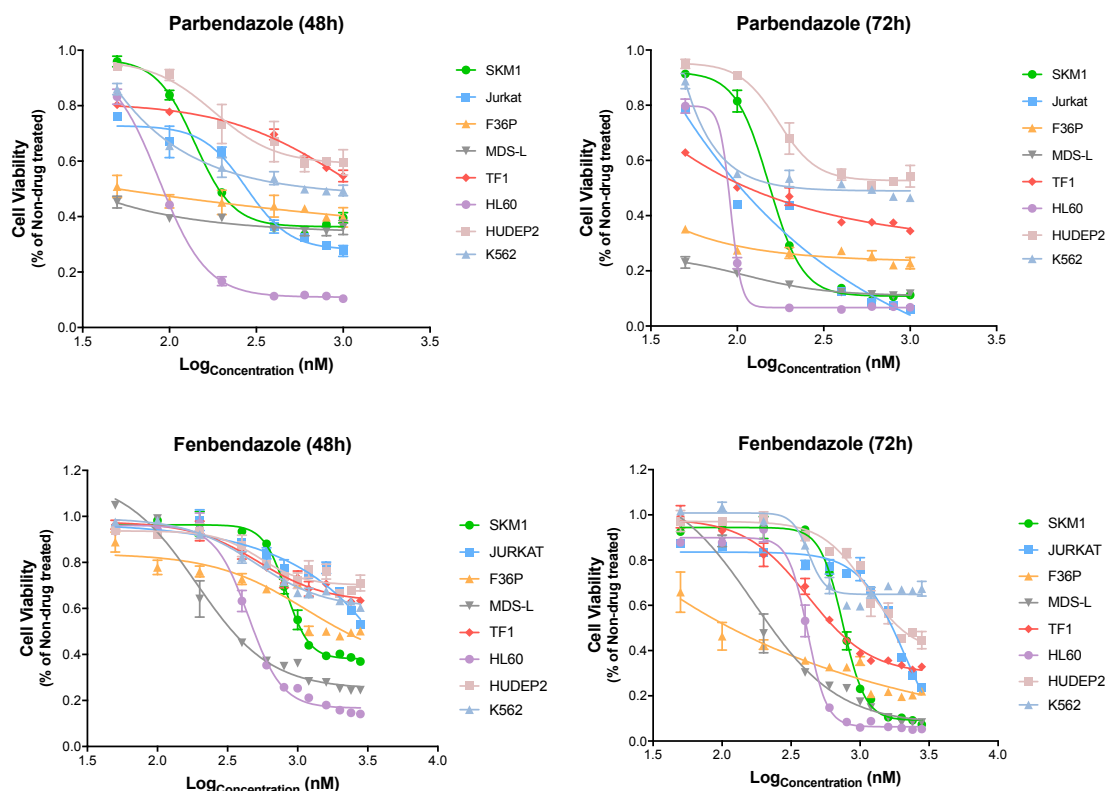


Figure 4.8 Drug dose-response curves in cells treated with PBD or FBD. Cell viability of cell lines measured by ATP-based assay normalised to vehicle-treated control. Drug doses are represented as logarithm base 10 of nanomolarity. Each point represent the average from triplicate samples and error bars represent SD. a) Cell viability of cell lines 48 h and 72h after PBD treatment. b) Cell viability of cell lines 48 h and 72h after FBD treatment.

Table 4.4 The IC₅₀ of PBD and FBD treatment in cell lines (nM)

	PBD		FBD	
	48h	72h	48h	72h
SKM-1	130.0-150.3	145.7-159.4	790.0-871.9	726.9-774.3
Jurkat	238.2-321.5	Very wide	Very wide	Very wide
F36P	Very wide	Very wide	Very wide	Very wide
MDS-L	Very wide	64.29-188.4	148.6-243.4	118.1-243.6
TF1	Very wide	Very wide	401.2-689.5	363.0-483.2
HL60	80.72-93.82	Very wide	418.9-476.2	400.2-434.8
K562	Very wide	Very wide	Very wide	Very wide
HUDEP2	Very wide	Very wide	Very wide	958.3-1306

7. PBD and FBD induce cell cycle arrest at G2/M phase and apoptosis in leukaemia cell lines

To investigate whether the benzimidazole drugs suppress cell growth by hindering cell cycle progression, we conducted cell cycle analysis by Propidium iodide (PI) staining in cell lines after 12 hours of treatment with PBD and FBD. Both PBD and FBD induced a marked cell cycle arrest at the G2/M phase in SKM1, HL60 and MDS-L cells. The proportion of cells at the G2/M phase was significantly increased (by approximately more than 20%) in cells treated with PBD at 200 nM or FBD 700 nM for 12h. However, the cell cycle progression of HUDEP2 cells was not significantly affected by benzimidazole drug treatment (Figure 4.9).

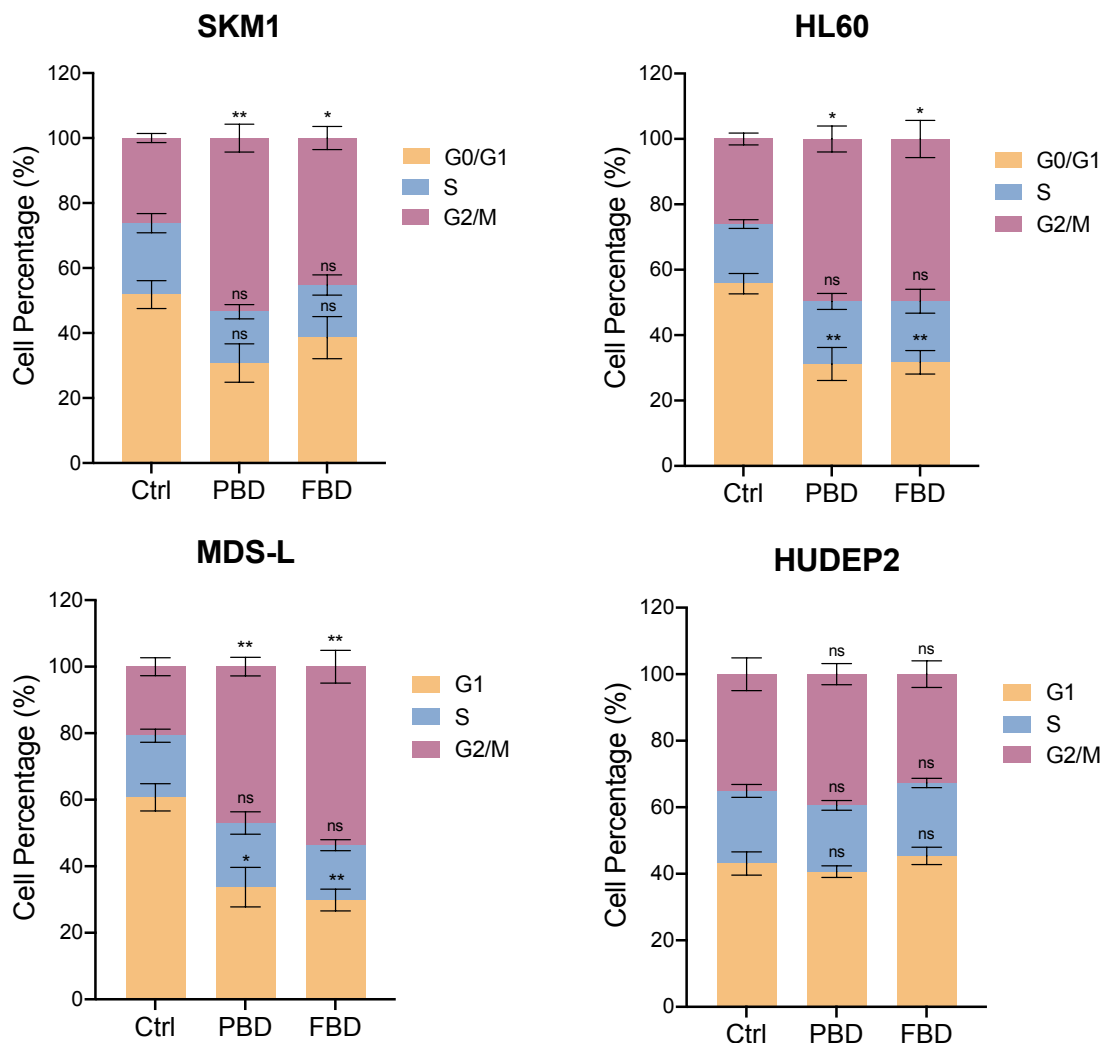


Figure 4.9 Cell cycle analyses by PI staining. Cell lines were treated with PBD (200 nM) and FBD (700 nM) for 12 h, and the cells were stained by PI and analyzed by flow cytometry. All values are plotted as mean $\pm$ SD. P values were obtained using multiple t-test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

We used HL60, SKM1 and HUDEP2 cells to explore whether benzimidazole anthelmintics convey their anti-proliferative effects in leukaemia cell lines through the induction of apoptosis. These cell lines were treated for 24 h with PBD (200 nM) or FBD (700 nM). Annexin V/PI dual staining showed an increase of apoptotic cells after drug treatment in HL60 and SKM1 cell lines. In contrast, both drugs only slightly

induced apoptosis in HUDEP2 cells. In agreement with the cell viability data, HL60 cells are the most sensitive to benzimidazoles, with a > 40% increase in apoptotic and necrotic cells after either PBD or FBD treatment (Figure 4.10).

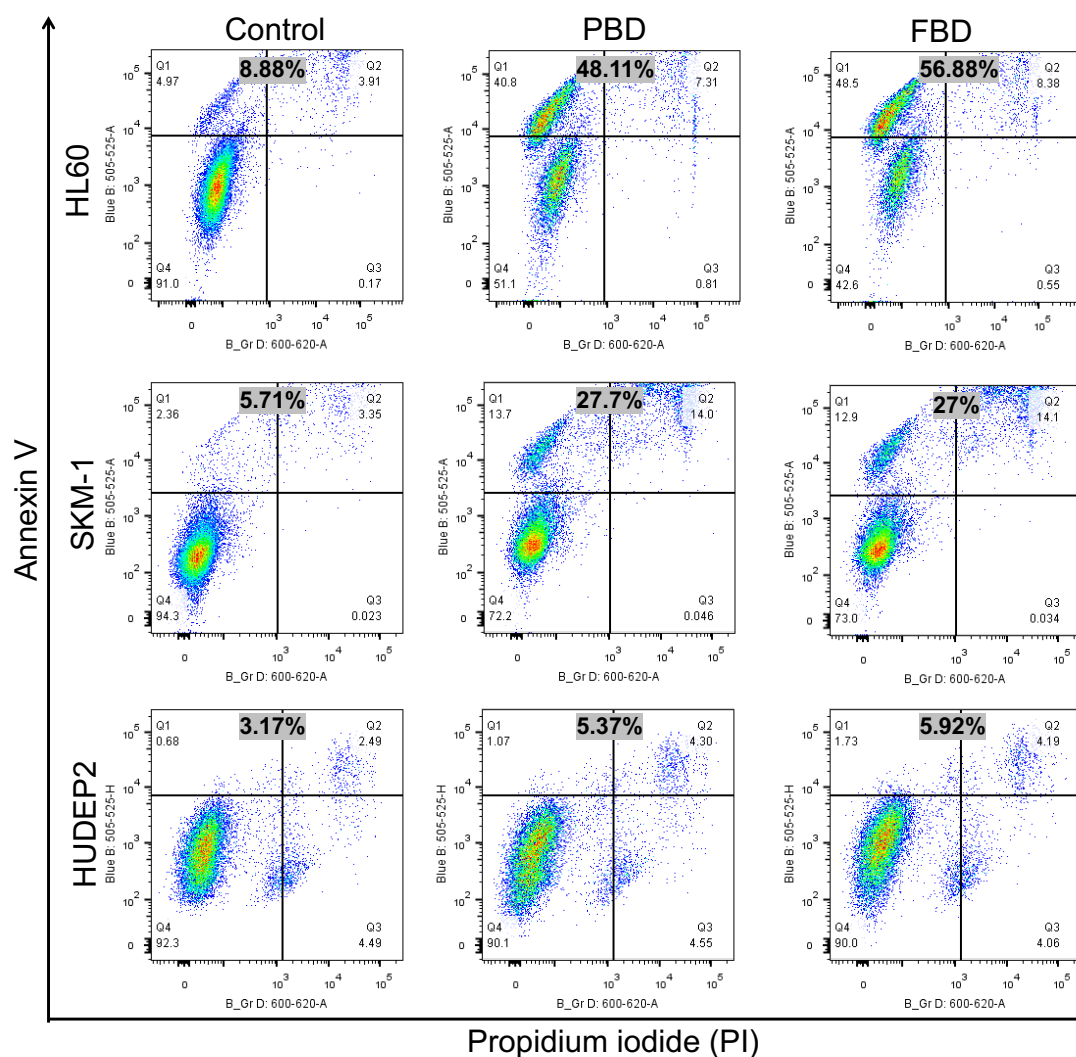


Figure 4.10. Cell apoptosis induced by PBD and FBD. HL60, SMK1 and HUDEP2 cells were treated with 200 nM PBD or 700 nM FBD for 24 h. Cell apoptosis was measured by using Annexin-V/PI dual staining. A single-positive staining for Annexin-V is an early indicator of apoptosis (Annexin-V+/PI-, upper-left area), and the double-positive staining for Annexin and PI indicates late apoptosis or necrosis (Annexin-V+/PI+, upper-right area).

8. PBD and FBD inhibit the proliferation of the mononuclear cells and CD34+ cells from MDS patients

We compared the sensitivity of bone marrow mononuclear cells from an MDS patient and from a healthy individual to PBD and FBD. The sample was confirmed with mutations in IDH2, SRSF2, ASXL1, RUNX1 by targeted sequencing. Mononuclear cells from the MDS patient were more sensitive to these drugs than mononuclear cells from a healthy donor (Figure 4.11a). The IC₅₀ of PBD treatment was 127.2 nM for MDS mononuclear cells and 179.1 nM for healthy control cells; the IC₅₀ of FBD was 828.6 nM for MDS mononuclear cells and 1060 nM for healthy control cells. Moreover, purified bone marrow CD34⁺ cells from MDS patients (n=3) and healthy controls (n=3) were also used to test their sensitivity to benzimidazoles (Figure 4.11b). Both PBD and FBD showed preferential inhibition on MDS CD34⁺ cells compared to CD34⁺ cells from healthy controls in a dose-dependent manner. Notably, PBD showed a nanomolar range IC₅₀ lower than 200 nM in inhibiting MDS primary cells.

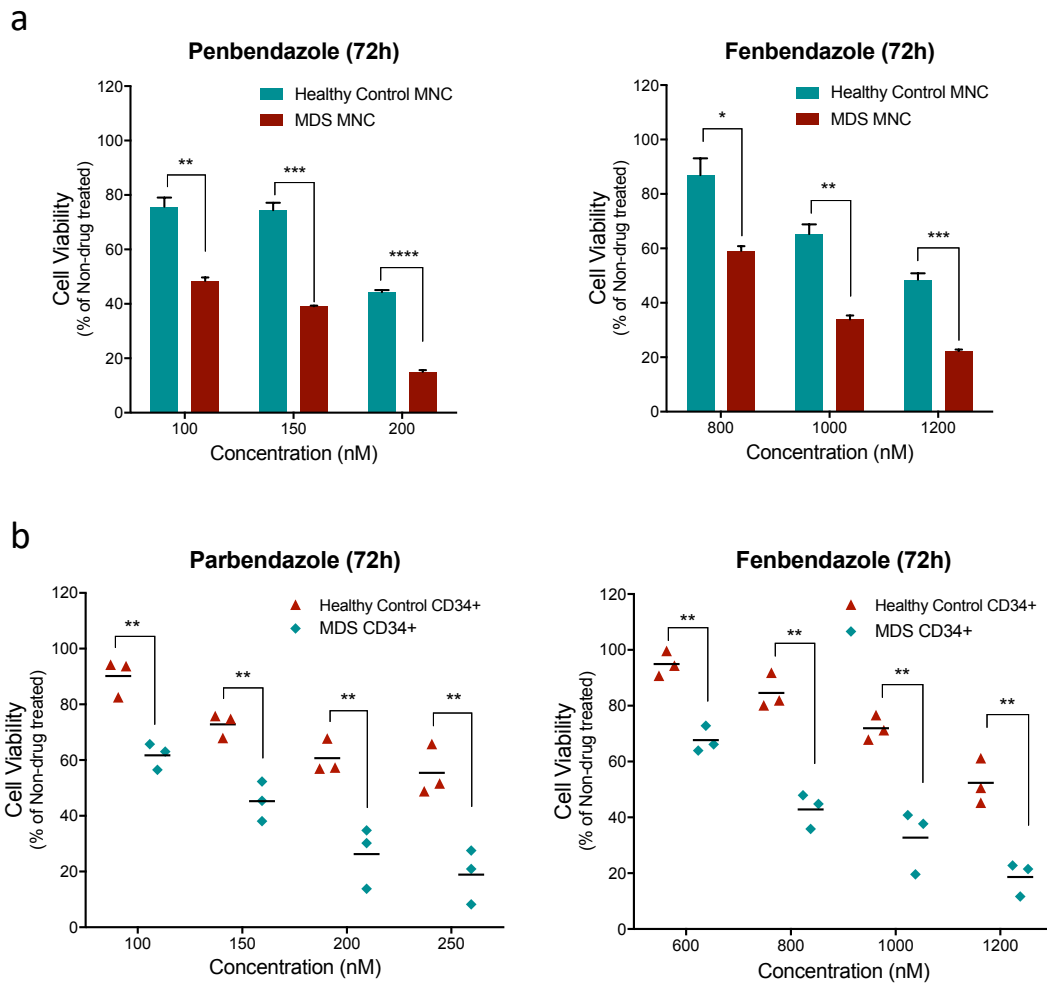


Figure 4.11. Cell viability of MDS patient cells after treatment with PBD and FBD. (a) Bone marrow mononuclear cells from an MDS patient and healthy donors treated with PBD or FBD for 72h. The sample was confirmed with mutations in IDH2, SRSF2, ASXL1, RUNX1 by targeted sequencing. (b) CD34⁺ cells from MDS patients (n=3) and healthy donors (n=3) treated with PBD or FBD for 72h. All values are plotted as mean $\pm$ SEM. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

9. PBD and FBD induce myeloid differentiation in MDS/AML cell lines

To investigate whether PBD and FBD induce myeloid differentiation in leukaemia cells, we measured the expression of myeloid antigen markers after treatment with these drugs. CD11b and CD14 are recognised as myeloid lineage markers, and are presented on the surface of monocytes, granulocytes and macrophages. Considering that higher drug doses would lead to significant cell apoptosis, we used a drug concentration at

approximate IC20 for each cell line. Flow cytometry results showed that PBD and FBD significantly increased the proportion of CD11b and CD14 positive cells in HL60, SKM-1 and TF-1 cells (Figure 4.12).

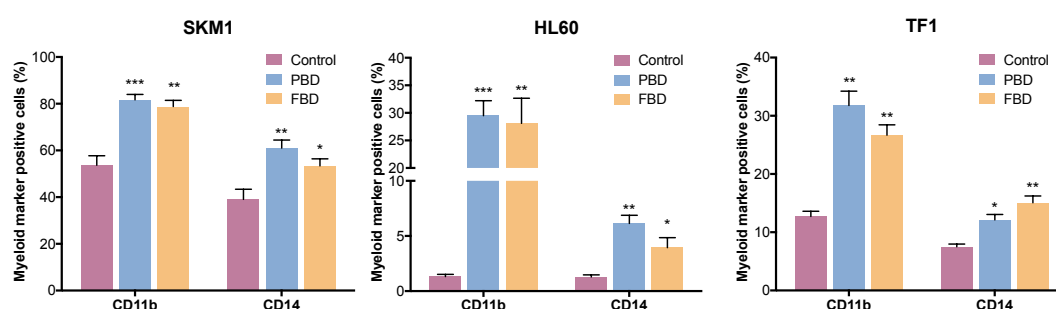


Figure 4.12 FBD and PBD induced myeloid differentiation in leukaemia cell lines. CD14 and CD11b antigen expression were measured after PBD or FBD treatment for 4 days at approximate IC20 doses for each cell line (HL60: PBD 50 nM, FBD 300 nM; SKM-1: PBD 100 nM, FBD 600 nM; TF1: PBD 30 nM, FBD 200 nM). P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

We also performed a colony-forming assay for PBD and FBD treated healthy CD34 cells. The CD34⁺ cells were purified from a healthy bone marrow mononuclear sample and cultured with PBD and FBD at approximate IC20 for 24 h. The enriched progenitor cells were then seeded on methylcellulose-based medium with no drugs added and cultured for ten days. The number of colonies formed on the methylcellulose-based medium was reduced after drug treatment. The proportion of CFU-M, CFU-G and CFU-GM was increased whereas the proportion of BFU-E, CFU-E and CFU-GEMM was decreased, suggesting the higher proportion of megakaryocyte, granulocyte and monocyte differentiation (Figure 4.13).

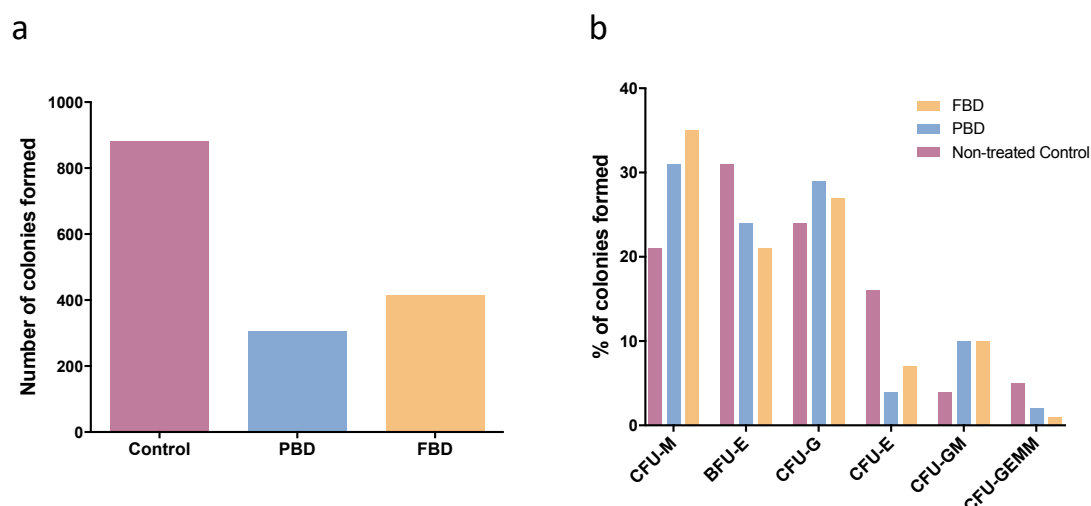


Figure 4.13 Colony-forming assay of CD34⁺ cells treated with PBD and FBD. CD34⁺ cells from a healthy donor were cultured with PBD at 150 nM or FBD at 900 nM for 24h. a) Number of colonies formed after 10 days culture on methylcellulose-based medium. b) Proportion of different types of colon-forming unit (CFU).

10. Other drugs selected for experimental validation

Besides Celastrol, Withaferin A, PBD and FBD, several other top-ranking compounds identified by connectivity mapping were selected for experimental validations as well. We selected more than 30 compounds identified in our in-silico screen for cell viability in leukaemia cells and HUDEP2 cells (Table 4.5). We chose these compounds because most of them have not been comprehensively investigated in blood cancers (preclinical and clinical stages); and/or their anti-cancer mechanism has not been well elucidated. We tested the changes of cell viability after drug treatment as a primary screening method. However, these drugs inhibited the growth and proliferation of both leukaemia cells and HUDEP2 cells, and/or showed an IC₅₀ higher than 20 μ M in leukaemia cells. So, these candidates were not selected for further cellular and molecular mechanism studies.

Table 4.5 Drug selected for experimental validation

Drug name	Provider
Quinacrine dihydrochloride	Sigma-Aldrich
Digitoxin	MP Biomedicals UK
Digoxin	Fisher Scientific
Niclosamide	Abcam plc
Anisomycin	MP Biomedicals UK
BX795	Enzo Life Sciences (UK) Ltd
Anisomycin	Cambridge Bioscience Ltd
Harmine	Insight Biotechnology Ltd
Simvastatin	Fluorochem Ltd
Famprofazone	Insight Biotechnology Ltd
Kaempferol	MP Biomedicals UK
Narciclasine	Insight Biotechnology Ltd
Kaempferol	Cambridge Bioscience Ltd
Celastrol	Stratech Scientific Ltd
Harmaline	Alfa Aesar
Pyrrvinium pamoate hydrate	Sigma-Aldrich-Aldrich Co Ltd
BX-912	Sigma-Aldrich-Aldrich Co Ltd
GPP 78	Cambridge Bioscience Ltd
Lycorine Hydrochloride	Enzo Life Sciences (UK) Ltd
Spiperone hydrochloride	Abcam plc
Piperlongumine	Cambridge Bioscience Ltd
Thapsigargin,	Insight Biotechnology Ltd
Parthenolide	Cambridge Bioscience Ltd
Sulfadoxine (Sulphadoxine)	Sigma-Aldrich
Parbendazole	Sigma-Aldrich
Bufexamac	Sigma-Aldrich
Syrosingopine	Sigma-Aldrich
Fenbendazole	Sigma-Aldrich
Rifabutin	Sigma-Aldrich
Hycanthone	Medchemtronica
Lasalocid sodium	Medchemexpress
Withaferin A	Fisher Scientific UK Ltd

C. Discussion

In this chapter, we investigated the effects of candidate drugs on cell viability, apoptosis, cell cycle regulation, and differentiation using different cell lines and primary MDS patient samples. We found that Celastrol, WA, PBD and FBD exhibited anti-tumour activity by suppressing cell proliferation, inducing cell cycle arrest in G2/M phase and increasing both the early and late stage of apoptosis in leukaemia cell lines. It would have been interesting to evaluate the expression and phosphorylation of cell cycle and apoptosis relevant protein and signalling (such as p21, p53, p38, p27, cyclins, CDKs, and cleaved-PARP, caspase-3, caspase-8, Bax, Bcl-2 and etc.), and changes of intracellular reactive oxygen species (ROS) levels after drug treatment; however, these experiments could not be performed due to the Covid-19 disruption. The differential expressions of these proteins and ROS in leukaemia cells have been suggested to play critical roles in the G2/M cell cycle arrest that blocks cell proliferation and induces apoptosis [293, 297-299]. For example, increased intracellular ROS accumulation in ovarian cancer cells contributed to the Celastrol-induced apoptosis [297]. Pre-treatment with ROS scavenger N-acetyl-cysteine completely blocked the apoptosis induced by Celastrol [297]. Withaferin A has been shown to trigger apoptosis of glioblastoma cells by up-regulating expression of Bim and Bad, and arrested cell cycle at the G2/M phase through dephosphorylating Thr161 of CDK1 by activating p53-independent p21 up-regulation [300]. Knockdown of p21 restored cell cycle progression and cell viability by down-regulating the expression of Bad [300].

Next, we investigated whether Celastrol, WA, PBD and FBD could inhibit cell proliferation in primary MDS patient samples. The proliferation of bone marrow mononuclear cells from an MDS patient harbouring mutations in IDH2, SRSF2, ASXL1, RUNX1 was preferentially inhibited by benzimidazoles, compared to the bone marrow mononuclear cells from a healthy donor. Bone marrow CD34⁺ cells from three MDS patients showed higher sensitivity to Celastrol, WA, PBD and FBD treatment compared to CD34⁺ cells from healthy controls. The mutational profile of these three MDS patients was not available. Hence, further experiments using CD34⁺ cells from MDS cases with known mutation status are required to understand the role of these drugs in inhibiting mutation specific HPSCs.

Our data shown that benzimidazoles induced the increased expression of myeloid differentiation marker in leukaemia cell lines. Colony-forming assays showed a higher proportion of megakaryocyte, granulocyte and monocyte differentiation in PBD and FBD treated healthy CD34 cells. However, the roles of Celastrol and WA on cell differentiation have not been explored. Previous studies have suggested that Celastrol and WA inhibited the expression of Myb target genes and induce myeloid differentiation (CD11b expression) in HL60 cells [301, 302]. Interestingly, when WA was combined with ATRA, a known differentiation inducer, a more than additive effect on CD11b expression was observed in HL60 cells, indicating that WA induces differentiation cooperatively with ATRA [302]. It would be interesting to assess

whether Celastrol, WA, PBD and FBD can induce myeloid differentiation of MDS patient HSCs using colony-forming assay and flow cytometry analysis.

Chapter V Investigation of the molecular mechanisms of drug candidates

A. Introduction and aims

Our in vitro data have demonstrated the therapeutic potential of some compounds (Celastrol, WA, PBD, FBD) that might be repositioned in MDS and AML. Although some studies have explored the mechanism of action of these candidate drugs in different cellular content and disease models, their molecular effects in myeloid leukaemia or haematopoietic progenitor cells have not been fully elucidated [271, 278, 293, 303-305]. In this chapter, we aimed to investigate the signalling pathways or proteins that might be regulated or targeted in leukaemia cells by these candidate compounds (Celastrol, WA, PBD, FBD). Primarily, the drug sensitivity was explored by performing cell viability assays in model cell lines: splicing factor mutant cell lines and induced pluripotent stem cells (iPSC). The splicing factor mutant cell lines used here are HUDEP2-U2AF1^{Q157P} cells, K562-SF3B1^{K700E} cells and K562 cells overexpressing U2AF1^{S34F} cells (K562-U2AF1^{S34F}).

We also aimed to find the molecular targets that might be inhibited or modulated by these candidate drugs. Previous studies have revealed the roles of Celastrol and WA in immunosuppressive and anti-inflammatory modulation, especially their inhibitory effects on NF- κ B signalling pathways in autoimmune disease, cancer, and diabetes. Celastrol and WA have been shown to carry out their potent anti-arthritis and anti-

cancer activity via suppressing NF- κ B signalling activation and the production of pro-inflammatory cytokines [271, 278, 281, 303]. Importantly, recent studies suggested that splicing factor mutations in MDS/AML result in hyperactive NF- κ B signalling and increased inflammatory cytokine production via aberrant splicing of specific target genes, including MAP3K7, CASP8 and IRAK4 [170-172]. These observations indicate that MDS-associated spliceosome mutations enhance the activity of innate immune signalling, which inhibits apoptosis and induces pro-inflammatory cytokines, further contributing to disease pathogenesis. Therefore, inhibiting the aberrantly spliced genes involved in NF- κ B signalling may have therapeutic potential in spliceosome mutant malignancies. For example, inhibiting IRAK4 by small-molecule treatments has showed therapeutic effects in U2AF1-mutant MDS PDX model [170]. Where, we performed molecular docking simulation to model the interaction of IRAK4 with Celastrol or WA, as well as assessed the inhibitory effects of Celastrol and WA on NF- κ B signalling target gene expression in splicing factor mutant cells.

Canonical Wnt signalling has been implicated in the regulation of normal haematopoiesis as well as the aberrant self-renewal of HSCs, suggesting an involvement in leukaemogenesis [306]. Several studies of haematopoietic stem and progenitor cells confirmed a role of aberrant Wnt signalling in the pathogenesis of myeloid leukaemia, including MDS and AML [116, 306-311]. Moreover, dysregulated Wnt signalling activation in HSCs has been directly implicated in self-renewal of leukaemia stem cells and is associated with a poorer outcome in mouse models and in

AML patients [310]. A recent study has shown that mebendazole, a benzimidazole anthelmintic agent, has inhibitory activity on TNIK mediated Wnt signalling pathway [312]. Here, we also studied PBD and FBD, compounds of benzimidazole anthelmintic family, for their interaction with TNIK and modulation of Wnt signalling in leukaemia cells.

B. Results

1. Celastrol and WA preferentially inhibit the proliferation of splicing factor mutant cells.

In chapter III, Celastrol was identified as a top-ranking drug candidate from the analysis of gene expression data of SF3B1 mutant MDS patients compared to wildtype cases or healthy controls, and WA was the top-ranking drug in the CMap analysis of U2AF1 mutant MDS versus healthy control comparison. To investigate whether Celastrol and WA preferentially inhibit cell proliferation in association with the presence of splicing factor gene mutations, we tested the cytotoxicity of these drugs in HUDEP2 cells with U2AF1^{Q157P} mutation introduced via CRISPR/Cas9, K562 with SF3B1^{K700E} mutation, and K562 cells overexpressing U2AF1^{S34F}.

First, we evaluated the sensitivity of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} cells to splicing modulators that have not been previously tested on HUDEP2 cells. Sudemycin C1, Isogingketin, and Madrasin showed preferential inhibition on the U2AF1 mutant cells (Figure 5.1).

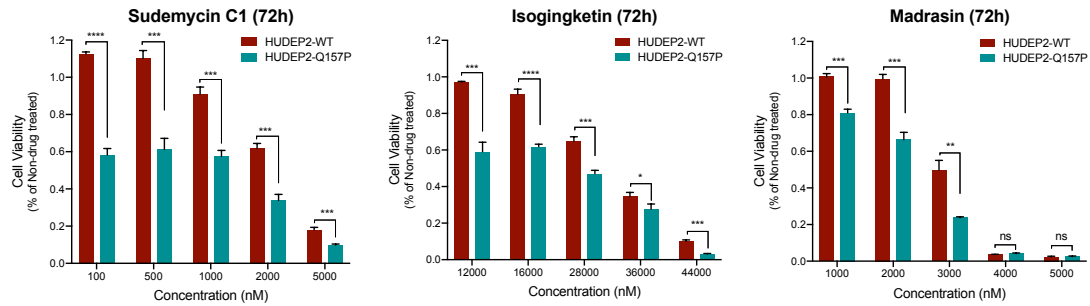


Figure 5.1 Responses of HUDEP2 cells to drug treatments. The Viability of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} after 72h to Sudemycin C1, Isogingketin, or Madrasin treatment was measured by ATP-based assay, normalised to DMSO-treated control. All values are plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

Figure 5.2 shows the cell responses to Celastrol and WA treatment. Since HUDEP2 cells were less sensitive than MDS/AML cell lines to Celastrol and WA treatment, we used a higher dose range to test the cell viability after 72h (Celastrol: 1.5-4 μ M; WA: 2-8 μ M). HUDEP2-U2AF1^{Q157P} cells were significantly more sensitive to Celastrol and WA than the U2AF1 wildtype HUDEP2 cells (Figure 5.2a). Similarly, when exposed to a micromolar range of drug concentrations, K562 cells also showed inhibition to drug treatment (Celastrol: 0.8-3 μ M; WA: 1-6 μ M). K562-SF3B1^{K700E} cells showed preferential sensitivity to these drugs compared with isogenic K562- SF3B1^{K700K} cells (Figure 5.2b). K562 cells overexpressing U2AF1^{S34F} similarly demonstrated higher sensitivity to Celastrol and WA compared to empty vector transfected K562 cells. (Figure 5.2c).

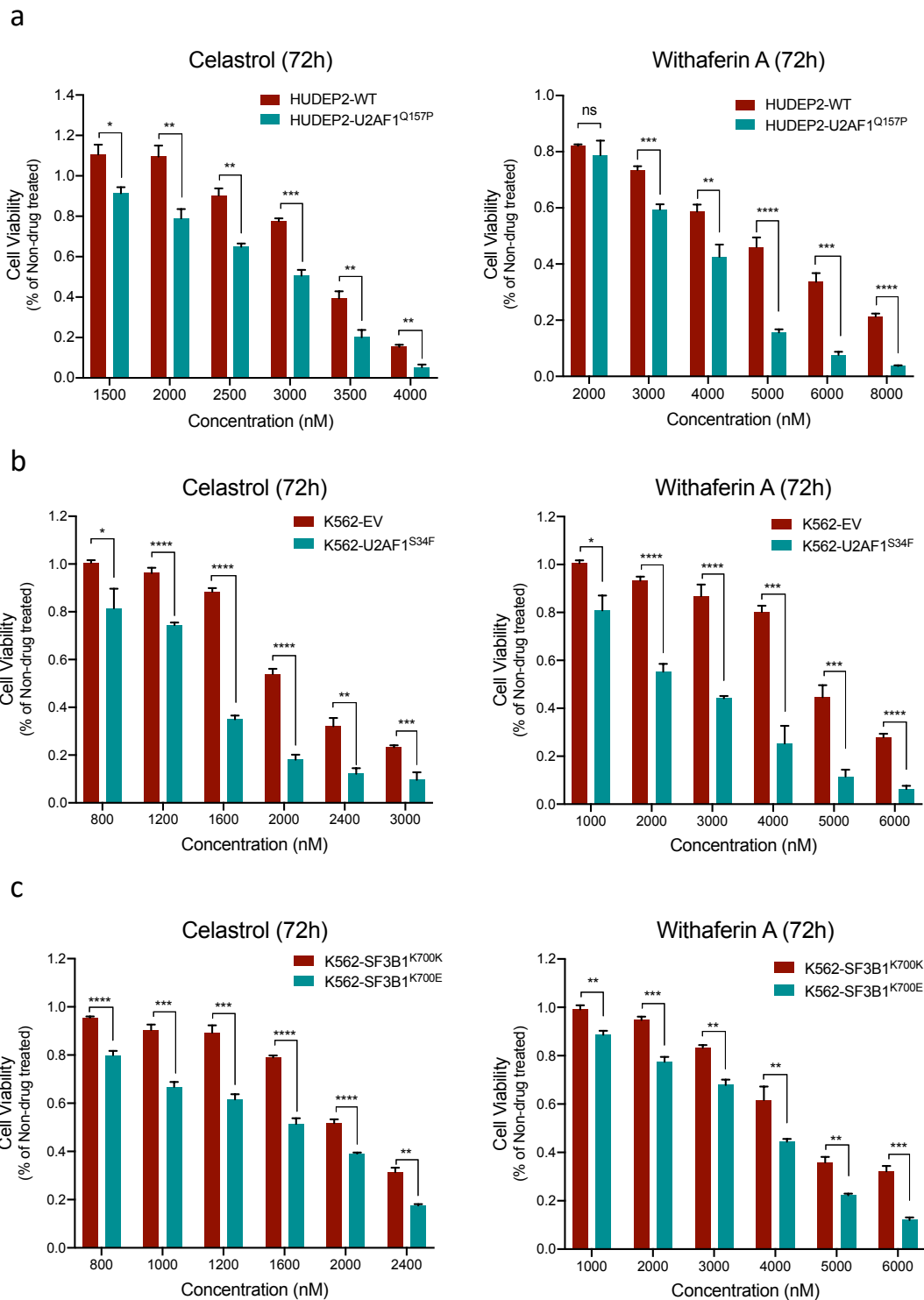


Figure 5.2 Responses of HUDEP2 cell and K562 cell to drug treatments. Cell viability was measured by ATP-based assay normalised to DMSO-treated control. (a) Viability of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} after 72h Celastrol or WA treatment. (b) Viability of K562-U2AF1^{EV} and K562-U2AF1^{S34F} after 72h Celastrol or WA treatment. (c) Viability of K562-SF3B1^{K700K} and K562-SF3B1^{K700E} after 72h Celastrol or WA treatment. All values are plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

The sensitivity of HUDEP2-U2AF1^{Q157} cells towards Piperlongumine and Parthenolide was also examined (Figure 5.3). The U2AF1^{Q157} mutant cells were more sensitive to Piperlongumine and Parthenolide compared to wildtype HUDEP2 cells when treated with a higher range of doses than the doses used on leukaemia cells (Piperlongumine: 7.5-50 μ M; Parthenolide: 2.5-50 μ M).

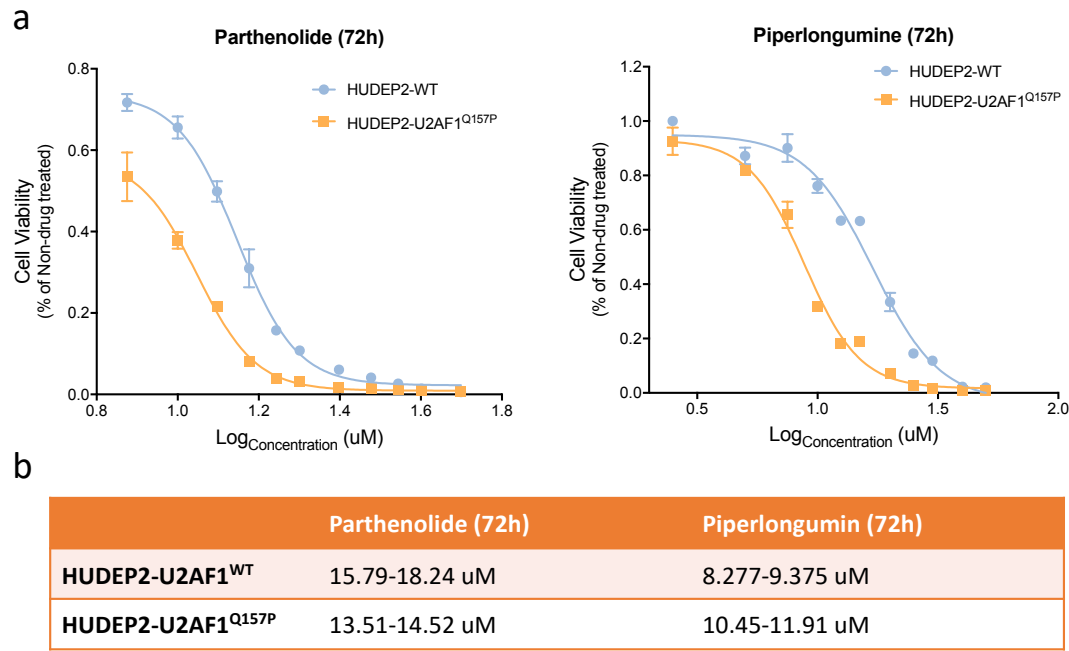


Figure 5.3. Responses of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} cells to drugs a) Viability of HUDEP2 cells after 72h. After 72h Parthenolide or Piperlongumin treatment, the viability of cell lines with indicated concentrations was measured by ATP-based assay and normalised to DMSO-treated control. Drug doses are represented as logarithm base 10 of micromolarity. b) The IC₅₀ of Parthenolide and Piperlongumin treatment in HUDEP2 cells (μ M).

2. Effects of Celastrol and WA on NF- κ B signalling and inflammatory cytokine production

Previous studies have demonstrated that spliceosome gene mutations enhance innate immune signalling, contributing to the pathogenesis of MDS and AML [171]. Mutations in SF3B1, U2AF1, and SRSF2 can enhance NF- κ B activity and

inflammatory cytokine production (such as IL6) in myeloid malignancy-relevant cells [172]. Interestingly, Celastrol and WA have been widely validated as an inhibitor of NF- κ B signalling pathway [273, 279, 280, 282, 313]. IL-6 has been evidenced as a downstream target gene of NF- κ B, and one of the best-studied pro-tumorigenic cytokines, the expression of which is elevated in many different cancers [314-318]. We evaluated the mRNA expression level of IL-6 in splicing factor mutant cells. Our qPCR results show that the untreated U2AF1^{Q157P} mutant HUDEP2 cells, SF3B1^{K700E} mutant K562 cells and U2AF1^{S34F} overexpressing K562 cells had significantly higher basal levels of IL-6 mRNA expression, a downstream target gene of NF- κ B signalling, compared to the spliceosome wildtype cells. Celastrol or WA treatment showed a significant suppression of IL-6 expression levels in both spliceosome mutant and wild-type K562 cells (Figure 5.4).

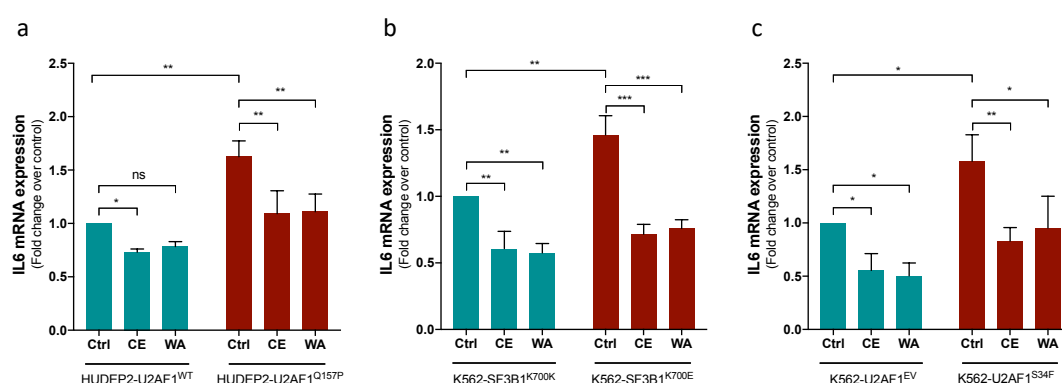


Figure 5.4 IL-6 mRNA expression in spliceosome mutant cells after drug treatment. The IL-6 mRNA expression levels were measured by syber green-based qPCR and normalised to DMSO-treated control. a) HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} cells; b) K562-EV and K562-U2AF1^{S34F} cells; c) K562-SF3B1^{K700K} and K562-SF3B1^{K700E} cells after 4h Celastrol (1000 nM) or WA (2000 nM) treatment. All values are plotted as mean $\pm$ SD. P values were obtained using two-way ANOVA with the Tukey's multiple comparisons test. *P < 0.05; **P < 0.01; ***P < 0.001.

A molecular docking simulation using Swissdock provided a possible mechanism of Celastrol and WA interacting with IRAK4 (Interleukin-1 receptor-associated kinase 4), a critical kinase in regulating and activating NF- κ B signalling in myeloid leukaemia [171]. The docking model showed the ATP-binding cleft of IRAK4 kinase domain was occupied by Celastrol or WA, sandwiched between a bilobal structure (Figure 5.5a). Figure 5b shows the ligand-protein interaction modes and docking calculations carried out by using DockingServer. Celastrol binding to IRAK4 might be mediated by numerous forms of interaction, including a hydrogen bond with MET265, polar interaction with LYS313, ASN316, ALA211, and hydrophobic interaction with LEU318, MET192, VAL200 and ALA315. WA may bind to IRAK-4 through interacting with VAL200, SER269, MET192, PHE197, LEU271, ALA315 and ASP272. These amino acids compose the core structure of the ATP-binding site for the IRAK4 kinase domain [319]. The simulation model also predicted the ligand-protein interaction energy and inhibition constant as shown in Figure 5.5b.

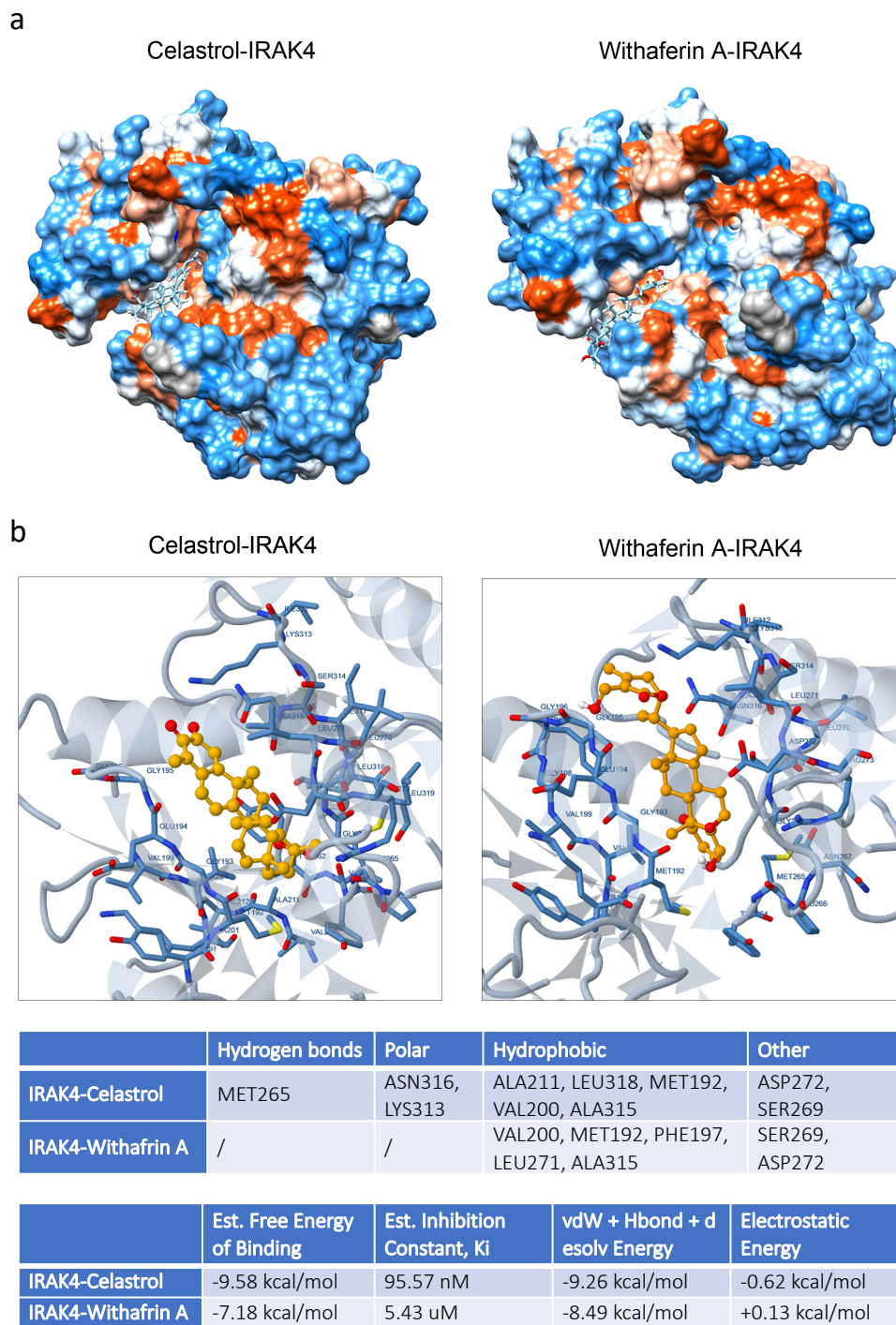


Figure 5.5 Molecular docking of Celastrol or WA with IRAK4. a) Molecular docking of Celastrol or WA with IRAK4 simulated by Swissdock. The colors indicate the hydrophobicity of the binding surface, from dodger blue for the most hydrophilic, to white, to orange red for the most hydrophobic. b) Binding modes and docking calculations of Celastrol or WA interacting with IRAK4 carried out by using DockingServer. The drugs were shown as a stick-and-ball model (orange balls are C atoms and red ball is O atom). The interactions between Celastrol or WA and amino acids in ATP-binding site were highlighted.

3. PBD and FBD preferentially inhibit the proliferation of splicing factor mutant cells.

PBD and FBD were selected from the CMap and MANTRA analysis of microarray-based data on CD34⁺ cells from advanced MDS cases. Hence, the drug-induced gene expression pattern was not directly associated with the presence of splicing factor mutations. However, there is some evidence suggesting that splicing factor mutant leukaemia cells may be sensitive to benzimidazoles. In a transcriptome analysis of K562-SF3B1^{WT} and K562-SF3B1^{K700E} mutant cells (GEO: GSE70959) [320], we used the same method described in Chapter III to generate the DEG between SF3B1^{K700E} mutant K562-cells and wildtype K562 cells for downstream CMap analysis. Here, we used a new version of CMap database, CMap L1000 assay (<https://clue.io/cmap>), to identify the compounds that may reverse the expression pattern of the gene signatures of interest. Figure 5.6a shows the volcano plot of the differentially expressed genes. Query lists of gene signatures were built based on both log2fold change and statistical significance: **1) Top 200:** Top 200 most statistically significant genes; **2) Log2fold:** Genes with a Log2fold change lower than -1 or higher than 1, and an FDR<5% or FDR<1% (if the probe size large than 500); **3) Up-down 100:** Top 100 most up-regulated and top 100 down-regulated genes with an FDR<5%. However, the Log2fold gene list contained 862 up-regulated genes and 793 down-regulated genes (FDR<1%), which is beyond the technical limitation of 150 gene input of the new CMap database. Hence, we only applied the Top200 list and Up-down 100 gene list for CMap analysis (Figure 5.6b).

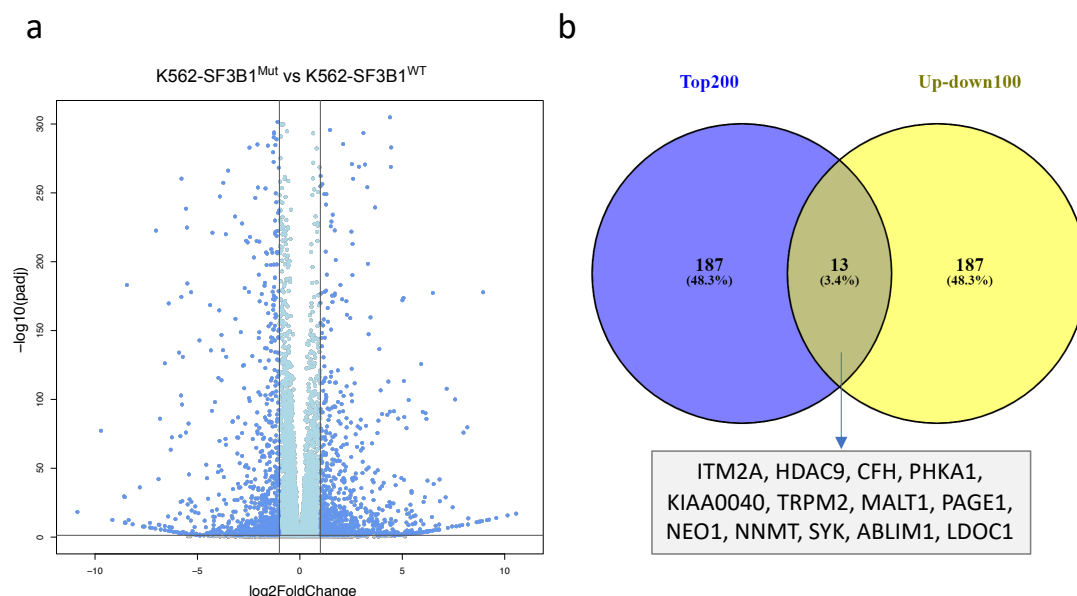


Figure 5.6. Differentially expressed genes in K562 cells with SF3B1-K700E mutation. a) Volcano plot showing DEG in K562-SF3B1^{K700E} mutant cells. RNA-Seq data of K562 cells was analysed and normalised by DEseq2 package. Vertical lines represent the cut-off of log2fold change equals to 1 and -1, and the horizontal indicates an FDR equals to 0.05. b) Venn diagrams show the overlapping genes of the two gene lists for CMAP queries.

The output compound lists ranked by the absolute value of enrichment score are shown in Table 5.1 and Table 5.2 for the Top200 gene list and Up-down 100 gene list, respectively. Several benzimidazole anthelmintic compounds were ranked among top candidates in Table 5.1, including albendazole, flubendazole, and fenbendazole. In Table 5.2, Mebendazole was also identified by CMap analysis of the Up-down 100 gene queries, although the overall absolute score is lower than the Top200 gene query.

Table 5.1 Top 30 compounds generated from the CMap analysis of Top200 gene list

Rank	Absolute Score	Name	Description
1	94.26	MK-1775	Wee1 kinase inhibitor
2	92.66	vinorelbine	Tubulin inhibitor
3	92.32	albendazole	Anthelmintic
4	91.37	flubendazole	Tubulin inhibitor
5	90.26	ABT-751	Tubulin inhibitor
6	89.31	JX-401	p38 MAPK inhibitor
7	88.7	ON-01910	PLK inhibitor
8	86.44	celestrol	Anti-inflammatory
9	86.39	TUL-XXI039	Serine/threonine kinase inhibitor
10	85.21	promazine	Dopamine receptor antagonist
11	84.05	diethylstilbestrol	Estrogen receptor agonist
12	83.48	SA-792574	Microtubule inhibitor
13	83.37	LY-2183240	FAAH inhibitor
14	82.18	acitretin	Retinoid receptor agonist
15	80.98	VX-222	HCV inhibitor
16	80.43	indirubin	CDK inhibitor
17	77.81	podophyllotoxin	Microtubule inhibitor
18	77.61	tamoxifen	Estrogen receptor antagonist
19	77.56	suloctidil	Adrenergic receptor antagonist
20	75.94	SA-63133	Tubulin inhibitor
21	75.86	phensuximide	Succinimide antiepileptic
22	75.77	niclosamide	DNA replication inhibitor
23	75.45	vincristine	Tubulin inhibitor
24	73.44	CO-101244	Ionotropic glutamate receptor antagonist
25	72.73	metrizamide	Radiopaque medium
26	71.36	BRD-K41143549	Glutamate receptor antagonist
27	71.26	lapatinib	EGFR inhibitor
28	71.24	fenbendazole	Tubulin inhibitor
29	70.83	lasalocid	Bacterial permeability inducer
30	70.17	serdemetan	MDM inhibitor

Table 5.2 Top 30 compounds generated from the CMap analysis of Up-down 100 gene list

Rank	Score	Name	Description
1	74.35	VU-0415374-1	Glutamate receptor modulator
2	67.85	panobinostat	HDAC inhibitor
3	67.55	LY-303511	Casein kinase inhibitor
4	67.18	apicidin	HDAC inhibitor
5	63.97	mebendazole	Tubulin inhibitor
6	62.48	baeomycetic-acid	Lipoxygenase inhibitor
7	58.01	nimetazepam	GABA receptor agonist
8	55.89	SKF-96365	Calcium channel blocker
9	45.03	SA-63133	Tubulin inhibitor
10	44.57	deltaline	Acetylcholine receptor antagonist
11	36.53	vinorelbine	Tubulin inhibitor
12	36.07	vinblastine	Microtubule inhibitor
13	34.29	1-monopalmitin	P-glycoprotein inhibitor
14	33.45	LY-2183240	FAAH inhibitor
15	33.12	BRD-K63784565	Topoisomerase inhibitor
16	32.6	mifepristone	Glucocorticoid receptor antagonist
17	32.07	rifampicin	RNA polymerase inhibitor
18	30.31	prunetin	Breast cancer resistance protein inhibitor
19	30.16	TG-101348	FLT3 inhibitor
20	27.2	LDN-193189	Serine/threonine kinase inhibitor
21	26.48	brefeldin-a	Protein synthesis inhibitor
22	26.38	hinokitiol	Tyrosinase inhibitor
23	25.58	phenylbutyrate	HDAC inhibitor
24	24.58	SKF-89976A	GABA uptake inhibitor
25	24.55	methyl-angolensate	Apoptosis inhibitor
26	23.59	lasalocid	Bacterial permeability inducer
27	22.32	WYE-354	MTOR inhibitor
28	19.06	scoulerine	Adrenergic receptor antagonist
29	18.08	ceforanide	Penicillin binding protein inhibitor
30	17.91	nisoldipine	Calcium channel blocker

Given the result of CMap analysis, the therapeutic potential of benzimidazoles in cells presenting splicing factor gene mutations was investigated. We tested the cytotoxicity of PDB and FBD in HUDEP2 cells with heterozygous U2AF1^{Q157P} mutation, K562 with heterozygous SF3B1^{K700E} mutation, and K562 cells overexpressing U2AF1^{S34F}. HUDEP2-U2AF1^{Q157P} cells were more sensitive to benzimidazoles than wild-type HUDEP2 cells (Figure 5.7a). U2AF1^{S34F} overexpressing K562 cells similarly demonstrated higher sensitivity to these drugs than empty vector or U2AF1^{WT} plasmid transfected K562 cells (Figure 5.7b). SF3B1^{K700E} K562 cells showed preferential sensitivity to the drugs compared with isogenic SF3B1^{K700K} K562 cells (Figure 5.7c).

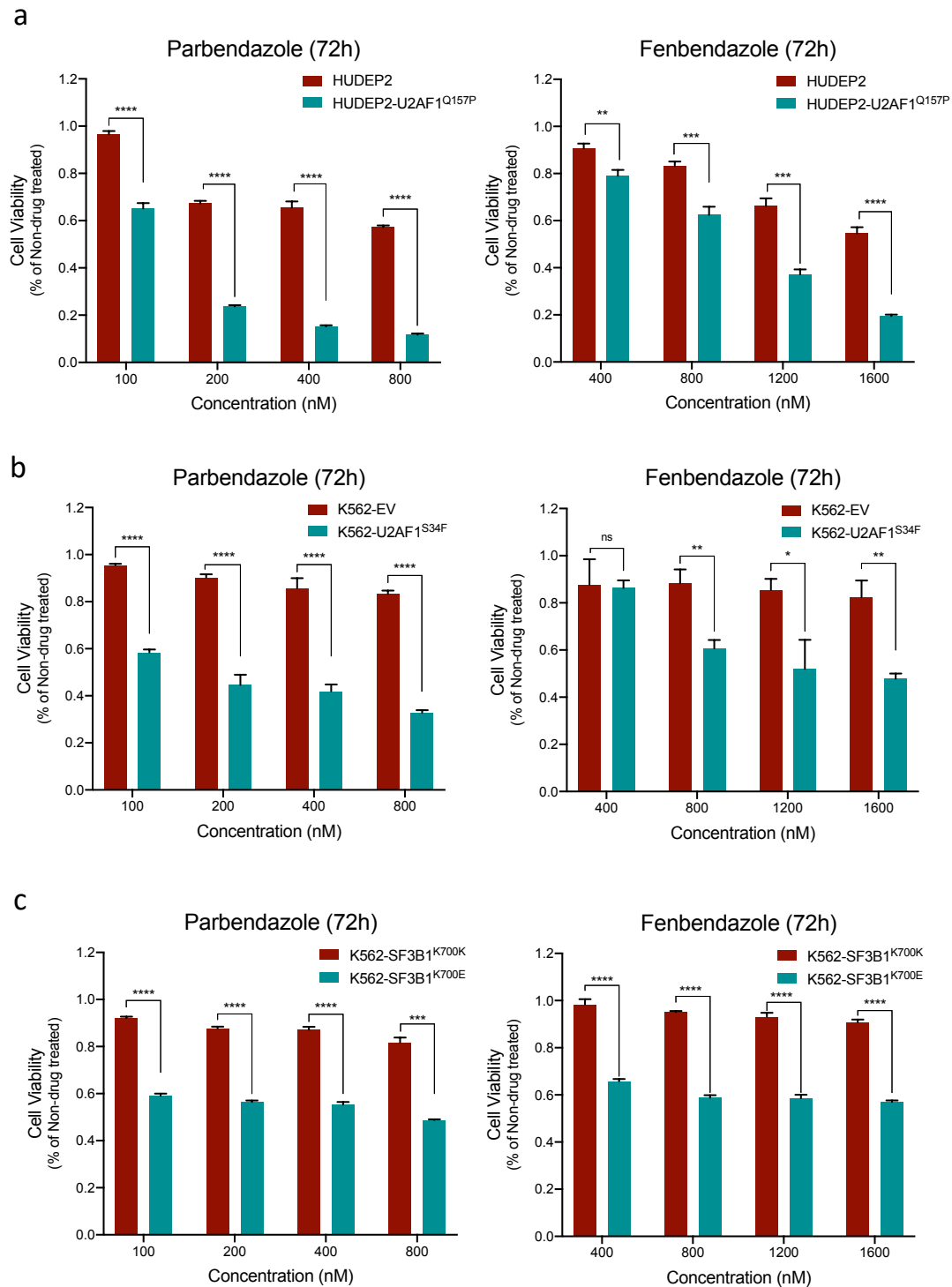


Figure 5.7 Responses of HUDEP2 cell and K562 cell to drug treatments. The viability was measured by ATP-based assay and normalised to DMSO-treated control. (a) Viability of HUDEP2-U2AF1^{WT} and HUDEP2-U2AF1^{Q157P} after 72h PBD or FBD treatment with indicated concentrations. (b) Viability of K562-U2AF1^{EV} and K562-U2AF1^{S34F} after 72h PBD or FBD treatment. (c) Viability of K562-SF3B1^{K700K} and K562-SF3B1^{K700E} after 72h PBD or FBD treatment. All values were plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

4. PBD and FBD preferentially inhibit iPSC derived from MDS patient cells with splicing factor mutations

Next, we analysed iPSCs that were generated from bone marrow CD34⁺ cells of one SF3B1 mutant MDS patient and of one healthy control. Two clones were generated from the MDS sample, one had the SF3B1^{K700E} and DNMT3A^{R882H} mutations (MDS-SF3B1^{Mut}), and the other one was wild-type for these mutations (MDS-SF3B1^{WT}). The MDS-SF3B1^{WT} and healthy control iPSCs showed similar responses to benzimidazole treatment, but cells from the MDS-SF3B1^{Mut} clone were much more sensitive to benzimidazoles than the other two groups, showing significantly decreased cell viability after drug treatment (Figure 5.8).

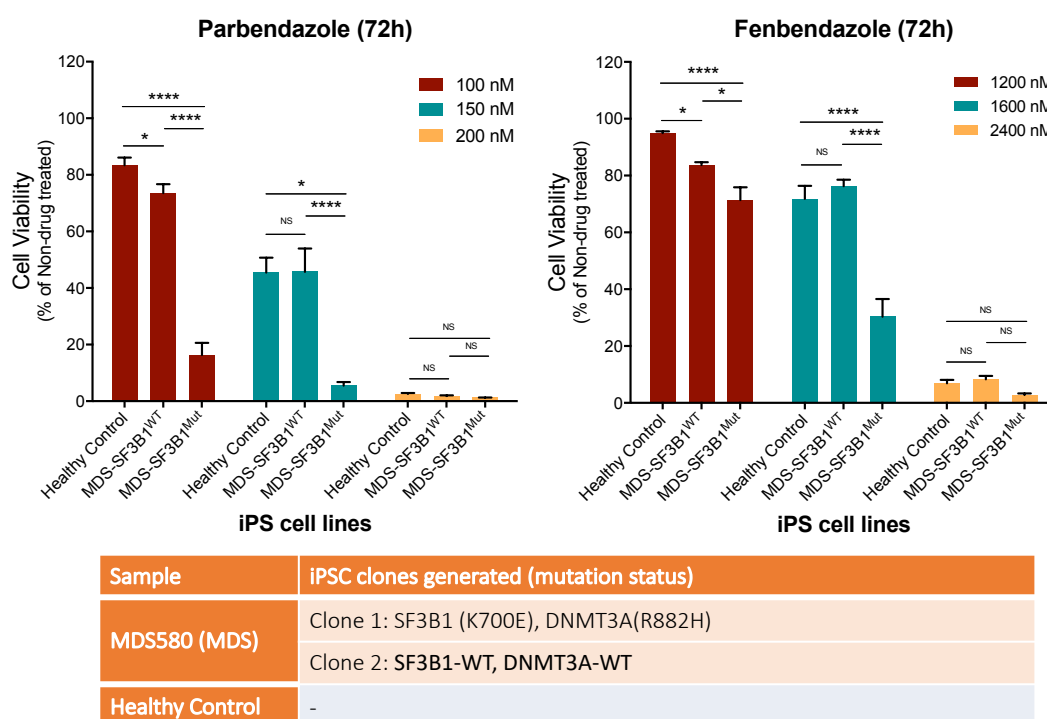


Figure 5.8 Responses of iPSCs to drug treatments. The viability of iPSCs after 72h PBD or FBD treatment was measured by ATP-based assay and normalised to DMSO-treated control. The table below shows the mutation status of two clones of MDS-derived iPSC. All values were plotted as mean $\pm$ SD. P values were obtained using two-way ANOVA with the Tukey's multiple comparisons test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

We also used another iPSC line generated from a CMML patient sample harboring a SRSF2^{P95L} mutation to test the sensitivity to PBD and FBD. A statistically significant difference in drug sensitivity between the MDS-derived iPSCs and healthy control iPSCs was observed (Figure 5.9). These data indicate that PBD and FBD preferentially inhibit iPSC derived from MDS patient cells with splicing factor mutations.

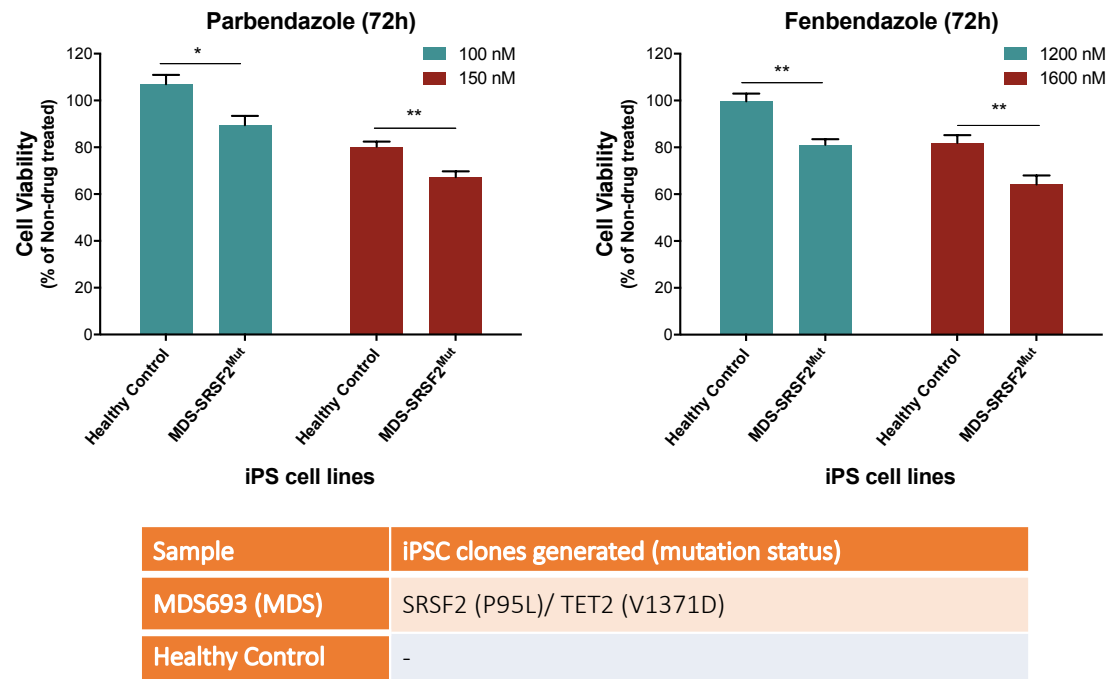


Figure 5.9 Responses of iPSCs to drug treatments. The viability of iPSCs after 72h PBD or FBD treatment was measured by ATP-based assay and normalised to DMSO-treated control. The table below shows the mutation status of the MDS-derived iPSC. All values were plotted as mean $\pm$ SD. P values were obtained using multiple t-test. **P < 0.01. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

5. Investigation of the MoA of PBD and FBD in leukaemia cells

To better understand the MoA of benzimidazole drugs in MDS, phospho-kinase array assays were performed to detect the relative level of site-specific phosphorylation of kinases after drug treatment. The phosphorylation level of STAT5a-Y694 and STAT5b-Y699 were significantly reduced after 24 hours treatment with PBD or FBD in SKM1 cells (Figure 5.10).

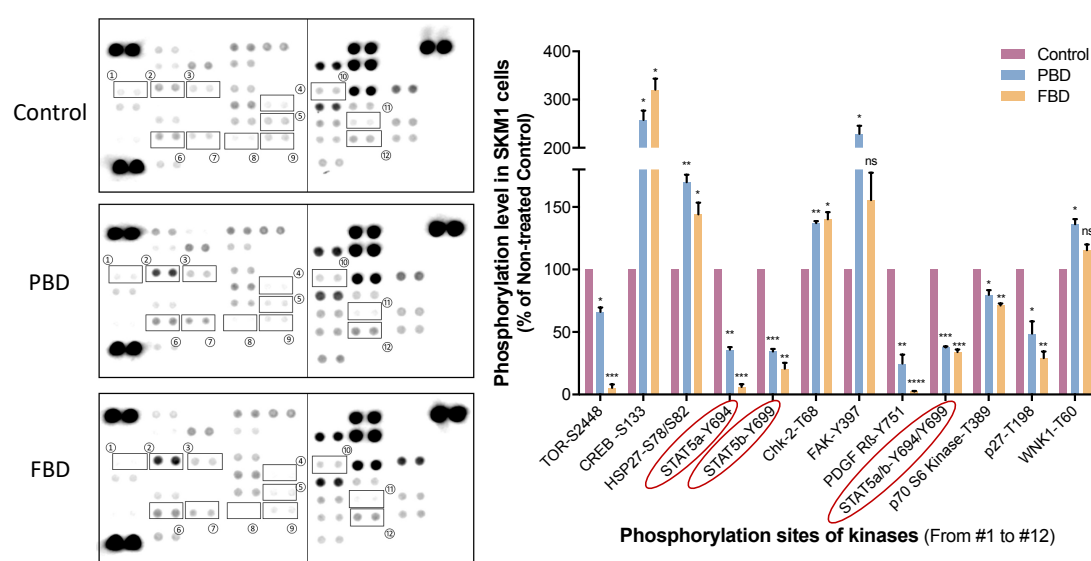


Figure 5.10 Phospho-kinase array assays of SKM1 cells. The site-specific phosphorylation of kinases in SKM1 cells after 24h PBD or FBD treatment was measured by membrane-based sandwich immunoassay. The phosphorylation level was calculated by ImageJ and normalised to DMSO-treated control (n=2). All values were plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

We also measured the transcription changes of selected STAT5a/b target genes in order to determine whether the expression levels of these genes were affected by benzimidazole treatment. The expression levels of *CCND1*, *BCL-2* and *MYC* in SKM1 cells were inhibited after 6 h treatment with PBD or FBD, as quantified by qPCR (Figure 5.11).

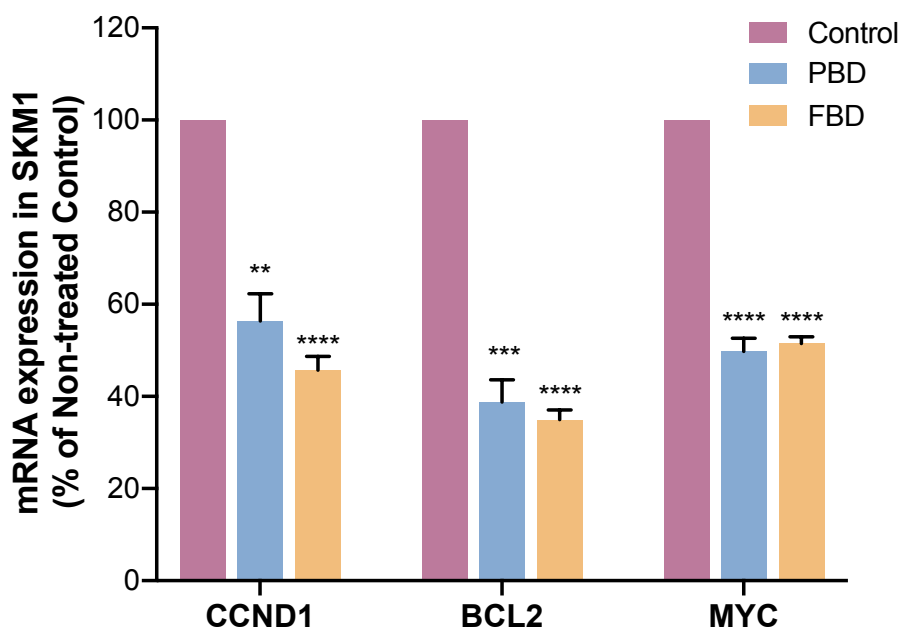


Figure 5.11 Expression levels of STAT5a/b target genes in SKM1 cells after treatment with PBD or FBD. The mRNA level was measured by syber green-based qPCR method and normalised to DMSO-treated control. All values were plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

It has been reported that Mebendazole could selectively inhibit 91.8% of the TNIK signal at 10 μ M, with a dissociation constant (Kd) of $\sim$ 1 μ M [312]. Mebendazole binds to the ATP-binding pocket of TNIK in a similar fashion as dasatinib (CYS108) [312]. Since PBD and FBD are structurally similar to Mebendazole, containing a benzimidazole backbone, we assessed whether PBD and FBD can inhibit the TNIK kinase activity. TNIK kinase is characterised by an N-terminal kinase domain and a C-terminal germinal center kinase homology region that serve a regulatory function. TNIK is autophosphorylated in a manner dependent upon Lys54 in the ATP-binding pocket of its kinase domain and plays a main role in cytoskeleton regulation [321]. The docking model carried out by SwissDock showed that the ATP-binding pocket of the TNIK kinase domain could be occupied by PBD or FBD (Figure 5.12a). Figure 5.12b

shows the ligand-protein interaction modes and docking calculations carried out by using DockingServer. It can be seen that PBD or FBD interacted with the ATP-binding pocket in the N-terminal kinase domain of TNIK. The binding of the PBD to TNIK might be mediated by numerous forms of interaction with TRY36 (polar); VAL39, MET105, VAL31, LEU73, ALA52, LEU103 (hydrophobic); LEU160, GLU69, ASP115 (other interaction). FBD may bind to TNIK through interaction with TRY36, VAL31, ASP115 (hydrogen bonds); SER112 (polar); CYS108, VAL39, LEU160, ALA52 (hydrophobic); GLN157, THR114 (other interaction). The estimation of the ligand-protein interaction energy and inhibition constant is also shown in Figure 5.12b.

To validate and measure the TNIK kinase activity inhibition by PBD and FBD, We performed a luminescent ADP detection assay to quantify the amount of ADP produced during a kinase-substrate reaction. Both PBD and FBD inhibited TNIK kinase activity at micromolar range (PBD: IC₅₀=8.35 μ M; FBD: IC₅₀=17.81 μ M) (Figure 5.13).

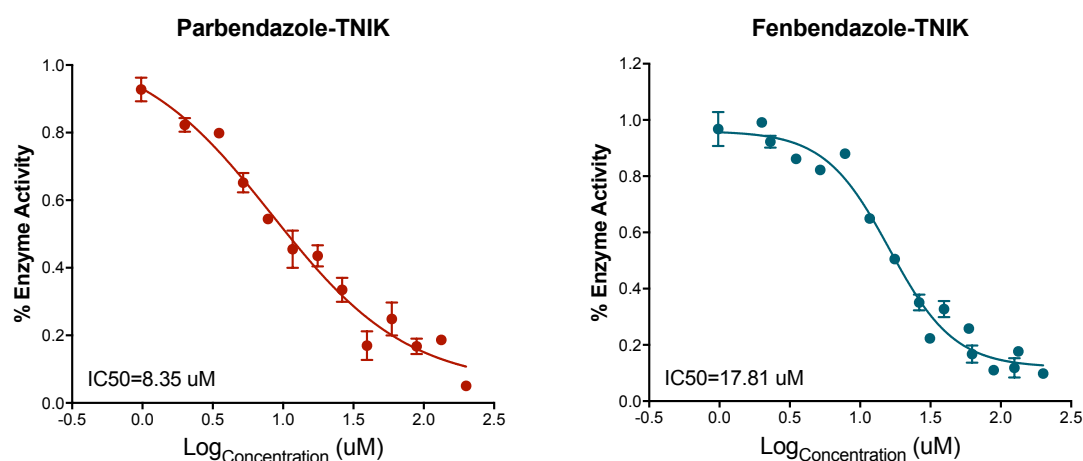


Figure 5.13 TNIK kinase activity inhibited by PBD and FBD. The kinase activity was measured by quantifying the luminescent of ADP produced during a kinase-substrate reaction and was normalised to DMSO-treated control. All values were plotted as mean $\pm$ SD.

We then assessed the gene expression level of Wnt signalling target genes in SKM1 cells after treatment with PBD and FBD. The expression of the Wnt target genes *AXIN2*, *LEF1*, as well as that of *TNIK*, was suppressed after 6 h benzimidazoles treatment (Figure 5.14). Moreover, *MYC* and *CCND1* were also proved as Wnt target genes, the expression of which were down-regulated after benzimidazoles treatment (Figure 5.11). These results suggest that PBD and FBD might regulate the Wnt signalling pathway through inhibition of TNIK.

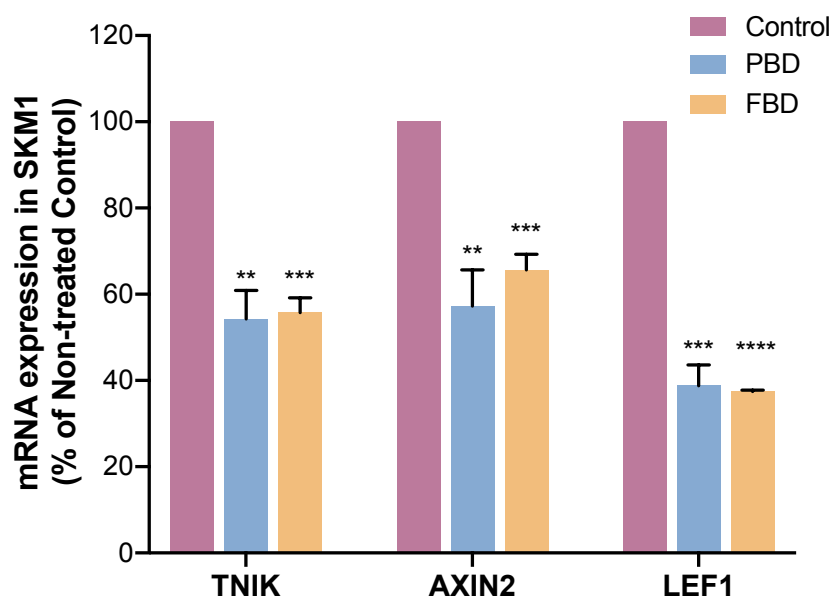
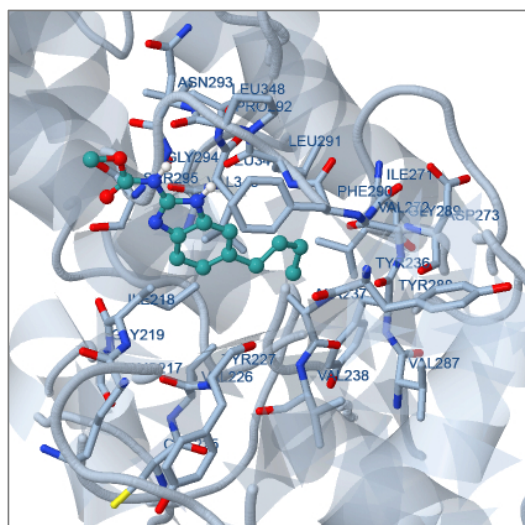


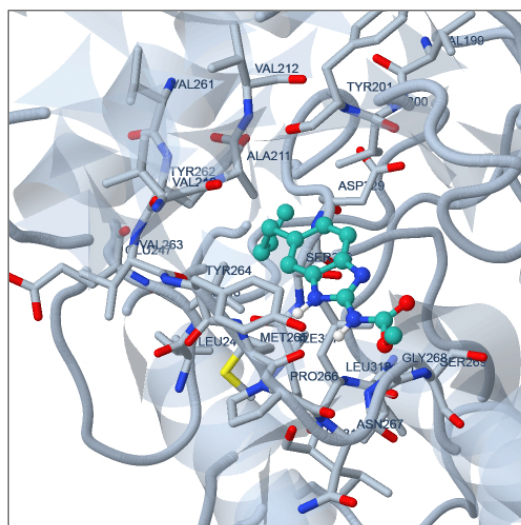
Figure 5.14 Expression levels of Wnt target genes in SKM1 cells after treatment with PBD or FBD. The mRNA level was measured by syber green-based qPCR method and normalised to DMSO-treated control. All values were plotted as mean $\pm$ SD. P values were obtained using multiple t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

The IRAK1/4 inhibitor I, N-[1-[2-(4-morpholinyl)ethyl]-1H-benzimidazol-2-yl]-3-nitro-benzamide, is a well-studied benzimidazole derivate that has been found to selectively disrupt the activity of IRAK1 and IRAK4 with IC₅₀ values of 0.3 and 0.2 μ M, respectively (PubChem CID: 11983295). To investigate whether benzimidazoles interact with IRAK1/4, we also performed the simulation of molecular docking mode of PBD and FBD interacting with IRAK1/4. Figure 5.15 and Figure 5.16 show the docking mode and calculations of PBD or FBD interacting with IRAK1/4 carried out by DockingServer, respectively. The inhibitory effects of PBD and FBD on IRAK1/4 were predicted at a micromolar range (<40 μ M).

Parbendazole-IRAK1



Parbendazole-IRAK4



	Hydrogen bonds	Polar	Hydrophobic	Other
IRAK1-PBD	LEU291, PRO292	/	ILE218, LEU347, TRY288, ALA237	PHE290
IRAK4-PBD	MET265	/	TRY262, VAL200, LEU318, ALA211, VAL246	SER328, TRY264

	Est. Free Energy of Binding	Est. Inhibition Constant, Ki	vdW + Hbond + desolv Energy	Electrostatic Energy
IRAK1-PBD	-6.02 kcal/mol	38.92 uM	-7.19 kcal/mol	+0.02 kcal/mol
IRAK4-PBD	-6.27 kcal/mol	25.39 uM	-7.37 kcal/mol	-0.04 kcal/mol

Figure 5.15 Molecular docking of PBD with IRAK1/4. Binding modes and docking calculations of PBD interacting with IRAK1/4 carried out by DockingServer. PBD is shown as a stick-and-ball model (green balls are C atoms and blue balls are N atoms). The interactions between PBD and amino acids are highlighted and listed in the table below. The prediction of the ligand-protein binding energy and inhibition constant are also shown.

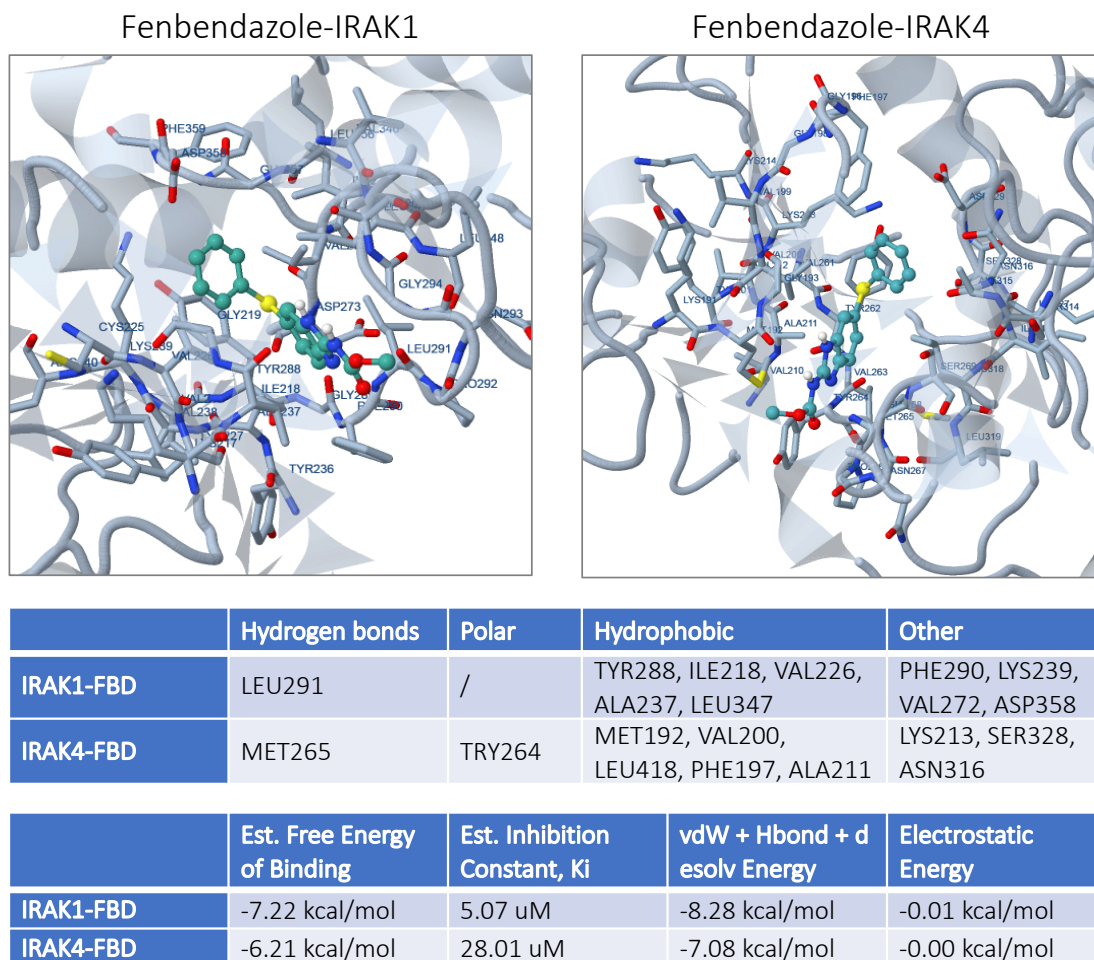


Figure 5.16 Molecular docking of FBD with IRAK1/4. Binding modes and docking calculations of FBD interacting with IRAK1/4 carried out by DockingServer. FBD is shown as a stick and ball model (green balls are C atoms, blue balls are N atoms, and yellow ball is O atom). The interactions between FBD and amino acids are highlighted and listed in the table below. The prediction of the ligand-protein binding energy and inhibition constant are also shown.

C. Discussion

1. MoA of Celastrol and WA

Our study provided evidence of the involvement of splicing modulation in the anti-proliferative activity of Celastrol and WA in leukaemia cells. Higher cytotoxicity induced by Celastrol or WA treatment was observed in splicing factor mutant cells; however, the underlying mechanism needs to be further investigated. Emerging evidence has revealed that hyperactive NF- κ B signalling is caused by spliceosome mutations and aberrant splicing events of specific target genes (e.g. *MAP3K7* and *CASP8*) in splicing factor mutant MDS [170]. Pollyea et al. reported that mutations in *SF3B1*, *U2AF1*, and *SRSF2* enhance NF κ B activity and LPS-induced inflammatory cytokine production in myeloid lineage cells [172]. More recently, a pivotal paper by Smith et al. reported that IRAK4, a serine/threonine kinase regulating the NF- κ B signalling pathway, is alternatively spliced in U2AF1 mutant MDS/AML patients [171]. Mis-spliced IRAK4 is characterized by the retention of exon 4 (IRAK4-long), interacting with MyD88 and facilitating the assembly of the myddosome complex, resulting in maximal activation of NF- κ B as well as MAPK activation. IRAK4-long is associated with oncogenic signalling in MDS/AML and essential for leukaemic cell function. Inhibition of IRAK4 was shown to abrogate leukaemic cell growth in vitro and in vivo [171]. These observations suggest that MDS-associated spliceosome mutations might enhance innate immunity, which inhibits apoptosis and induces pro-inflammatory cytokines, and consequently contributes to disease pathogenesis [172]. Increased inflammatory signal is thought to impair normal haematopoiesis and

contribute to the pathogenesis by altering HSPC survival, proliferation, differentiation, and apoptosis [174]. Specifically, NF- κ B signalling is known to play a key role in the determination of the quiescence versus the active state of HSCs, and that constitutive activation of NF- κ B in HSCs leads to pancytopenia and bone marrow failure [175].

We performed the molecular docking simulation to model the interaction of IRAK4 with Celastrol or WA. We showed that the ATP-binding pocket of IRAK4 kinase domain was occupied by Celastrol and WA, indicating their potential role as IRAK4 inhibitors. The inhibitory effects of Celastrol and WA on NF- κ B signalling target gene expression (IL-6) in splicing factor mutant cells (HUDEP2-U2AF1^{Q157P} cells, K562-U2AF1^{S34F} cells, K562-SF3B1^{K700E}) were assessed. Our qPCR data showed that the splicing factor mutant cells had significantly higher basal levels of IL-6 mRNA expression than spliceosome wildtype cells. Celastrol or WA treatment suppressed the IL-6 expression levels in splicing factor mutant cells. These data indicate that Celastrol and WA may interact with IRAK4 and attenuate the expression of NF- κ B target genes, illuminating the molecular mechanism of their preferential inhibitory effects on spliceosome mutant cells. However, we did not get direct evidence showing that NF- κ B signalling is activated in the splicing factor mutant cell lines, or that IRAK4 is aberrantly spliced in splicing factor mutant cell lines. We were planning to perform NF- κ B luciferase reporter assays or western blotting for I κ B α degradation together with phospho-p65 accumulation to detect NF- κ B activation, and qPCR to measure the expressing level of IRAK4-long/IRAK4-short in splicing factor mutant cell lines.

However, we were not able to complete these experiments due to the disruption caused by the Covid-19 situation.

Another possible mechanism is that Celastrol or WA affects kinase function by Michael addition thioalkylation reactions. The electrophilic sites within the unsaturated rings of WA (epoxide or lactone ring) or Celastrol (quinone methide rings) may act as Michael acceptor and react with nucleophilic thiol groups of cysteine residues (e.g. targeting Cys-179 in kinase domain activation loop of I κ B kinase β (IKK β), a central coordinator of NF- κ B signalling and inflammatory responses) [279, 322-326]. Several other kinases or proteins have also been suggested to be susceptible to thioalkylation reaction in the catalytic domain, indicating that WA and Celastrol may target cysteine residues of various kinases or phosphatases that affected the phosphorylation status of multiple signalling regulators, such as p38, MEK/ERK, JNK, Akt, Vimentin and NF- κ B [324, 327]. Some studies also suggested that the kinase-inhibitory effects of WA or Celastrol via Michael addition thioalkylation reactions could be neutralized by pre-loading thiol-containing agents [313, 328].

2. MoA of Benzimidazoles

This is the first study showing that leukaemia cells harboring splicing factor mutations are preferentially sensitive to benzimidazoles. However, the underlying mechanism has not been elucidated yet. Benzimidazole derivatives are known to act against a range of kinases and non-kinase target and thus gained clinical applications in a wide spectrum of diseases [329]. Here, we modeled the interacting mode of benzimidazoles with TNIK

and IRAK1/4, as well as assessed the TNIK activity via kinase inhibition test. Mebendazole, an approved anthelmintic benzimidazole-containing drug, was recently identified as a selective inhibitor of TNIK and is under clinical evaluation (dissociation constant $K_d \sim 1 \mu\text{M}$) [312]. TNIK regulates Wnt signalling, and its pharmacological inhibition seems to be a promising therapeutic approach in many cancers [312]. TNIK kinase mediated Wnt signalling pathway is inhibited by PBD/FBD but showing much higher IC₅₀ doses (micromolar range) than their cytotoxicity doses (nanomolar range) in leukaemia cells. A possible explanation is that PBD/FBD may act against multi-targets that leading to synthetic lethality in leukaemia cells at nanomolar doses.

Moreover, we found that the phosphorylation level of STAT5a-Y694 and STAT5b-Y699 and the expression level of target genes of STAT5 signalling were significantly reduced by benzimidazole treatment in leukaemia cells. *CCND1*, *BCL-2* and *MYC* have been considered as common target genes of STAT5 signalling [330]. The *CCND1* gene encodes cyclin D1 protein that functions in the regulation of CDK kinases in the cell cycle [331]. The Bcl-2 protein is the founding member of the Bcl-2 family of apoptosis regulators and was the first apoptosis modulator to be associated with haematological malignancies [332]. *BCL-2* gene encodes an integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes [333]. *MYC* is a proto-oncogene coding for a transcription factor that activate expression of many pro-proliferative genes. Constitutive upregulation of the *MYC* gene is frequently observed in numerous human cancers [334]. Overexpression and deregulation of these

genes have been recurrently found in many types of human lymphoma and leukaemia and are thus viewed as attractive therapeutic targets [335-338].

Consistent with the recent discovery that benzimidazole derivatives are highly potent and selective IRAK1/4 inhibitors [319, 339, 340], our molecular docking models showed that PBD and FBD resided in the ATP-binding pocket of IRAK1/4 kinase with a micromolar range inhibitory dynamic. The first example of the crystal structure of IRAK4 has been co-crystallized with a benzimidazole inhibitor (N-acyl 2-aminobenzimidazole) by the Amgen group, which has a structure of 3-nitro phenyl moiety, connected by an amide linker to a benzimidazole scaffold (Figure 5.17a, 5.17b) [319]. N-Acyl-2-aminobenzimidazole series of inhibitors were discovered via high throughput screening (HTS) and subsequent structure-activity relationship (SAR) studies. It has been proposed that the preference for the amide carbonyl linker indicates its possible involvement in hydrogen bonding to the hinge backbone of IRAK-4, while the requirement for the amide N-H may indicate a possible hydrogen bond to the kinase, or potentially, the requirement that the acyl-benzimidazole be able to populate a tautomeric form via abstraction of this hydrogen (Figure 5.17c) [339].

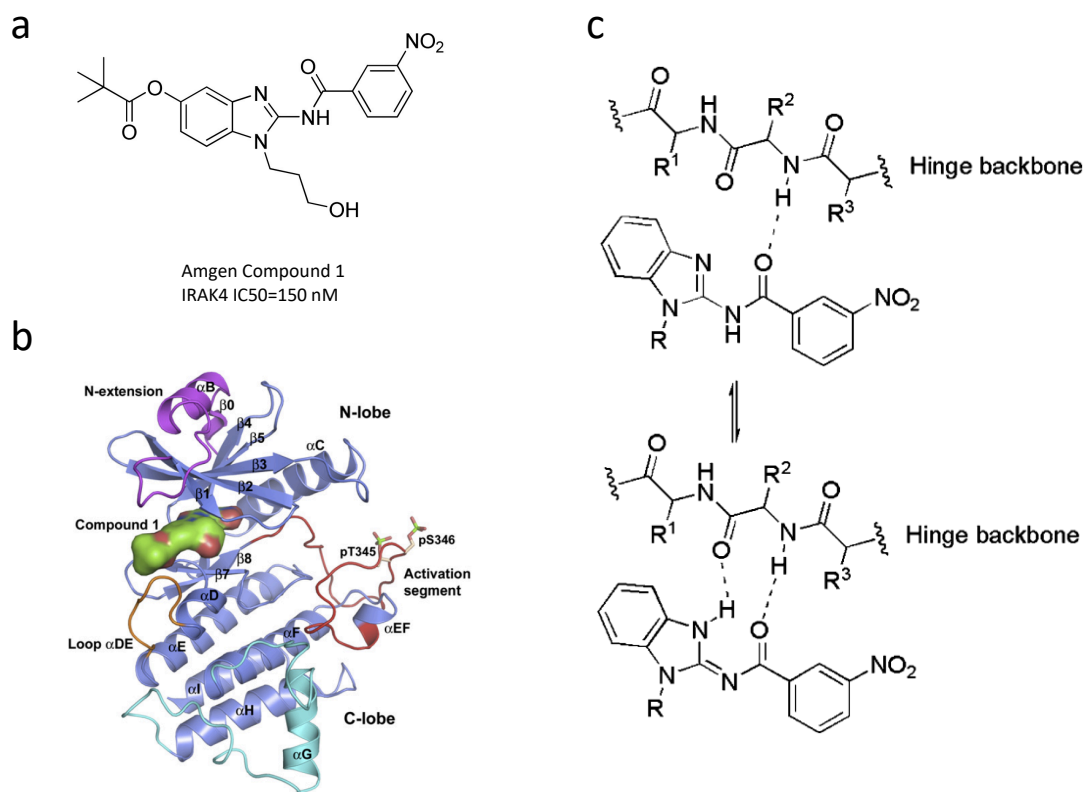


Figure 5.17 N-Acyl-2-aminobenzimidazole inhibitors of IRAK4 discovered via high-throughput screening by Amgen. a) Compound 1 of N-Acyl-2-aminobenzimidazole inhibitor series of IRAK4 discovered via HTS and SAR studies; b) The crystal structure of the human IRAK-4 kinase domain (residues 154–460 of the full-length protein of 460 amino acids) in complex with Compound 1; c) Proposed binding of acyl-benzimidazole tautomeric forms to IRAK-4 [319, 339].

In summary, Celastrol, WA and benzimidazoles (PBD and FBD) discovered by CMap analysis hold high promise as novel anti-leukaemia therapeutics. Our study primarily suggested that the splicing pathway modulation might be involved in the anti-proliferative activity of these compounds, but the underlying MoA largely remains unclear. Further studies are required to understand the mechanisms of these candidate compounds in suppressing haematological malignant cell growth and regulating splicing process.

Chapter VI Discussion and Future directions

A. Discussion

Traditional drug development through high-throughput screening strategy is costly, time consuming and burdened by a very low success rate. Drug repositioning (repurposing) is a potential alternative to novel drug discovery by redirecting existing drugs for treating another disease. It is an attractive strategy that can dramatically reduce drug development time, cost and safety risk [199, 200]. In the last few years, gene expression signatures have been assembled and compared based on genome-wide similarity metrics, unveiling ‘connections’ among drugs, pathways and disease phenotypes by matching corresponding signatures of disease and drug responses [200, 232]. Some bioinformatic databases and tools (such as CMap) that contain transcriptomics data of cell lines treated with small molecules were developed, greatly facilitating the discovery of new therapeutic applications of existing compounds for drug repositioning. The CMap resource has been successfully used to identify novel therapeutics for repositioning in various cancer contexts, including myeloid malignancies [245, 341] and solid tumours [246-248].

In this study, we analysed the RNA-seq or microarray-based transcriptome data from a cohort of MDS patients with splicing factor mutations and from a cohort of MDS patients with advanced disease (RAEB-2 subtype), respectively. Based on previous CMap studies and our experience, we generated several lists of gene signatures and used them as input for computational drug screening, i.e., connectivity mapping. By

applying the strategy of ‘signature reversion’, compounds inducing a transcriptional response opposite to that of MDS cases were identified as potential new drugs for the treatment of MDS [200, 232, 239]. The compounds were ranked by the absolute value of connectivity scores and were primarily screened by comprehensive literature review based on their MoA and toxicity profiles in the context of leukaemia, cancer and other disease. Several top-ranking candidate drugs for repositioning to MDS were identified and validated by in vitro experiments. Celastrol and WA were identified from the analysis of splicing factor mutant MDS patient RNA-seq data, and Parbendazole was discovered from the analysis of MDS-RAEB-2 microarray-based gene signatures.

1. Drugs identified by CMap analysis of RNA-seq GEP

First, in RNA-seq data based CMap analysis, Celastrol and WA were among the top-ranked compounds identified and were selected for experimental validation. Both Celastrol and WA were found to exhibit anti-tumor activity by suppressing cell proliferation, inducing cell cycle arrest in G2/M phase and increasing both the early and late stage of apoptosis in leukaemia cell lines. Differential signalling activation and accumulated intracellular ROS induced by Celastrol and WA have been suggested to play critical roles in arresting cell cycle arrest and inducing apoptosis in cancers (Table 6.1). Moreover, we observed that Celastrol or WA synergised with the hypomethylating agent Azacitidine inhibiting leukaemia cell growth. The combination of Celastrol and WA also showed synergistic effects on leukaemia cells. Notably, Celastrol and WA showed preferential toxicity against CD34⁺ cells from primary MDS patient samples,

suggesting their potential therapeutic applications for MDS. The therapeutic properties of Celastrol and WA in many diseases and their toxicological profile have been broadly investigated and reviewed [271, 278, 342, 343]. Celastrol is a bioactive constituent extracted from *Tripterygium wilfordii*, belongs to a class of natural compounds known as triterpene quinone methides and is made up of five cyclic rings, further classifying it as a pentacyclic triterpenoid. Celastrol possesses a broad range of biological activities and curative properties in the treatment of cancer, obesity, inflammatory, auto-immune and neurodegenerative diseases through the modulation of a variety of molecular targets [343, 344]. WA is a steroidal lactone derived from *Withania somnifera*, and has been shown to display multiple activities including anti-inflammatory, antiproliferative, proapoptotic, and antitumorigenic effects [278, 303]. However, the molecular mechanism underlying these effects has not been elucidated. Clinical trials using *Tripterygium wilfordii* or *Withania somnifera* plant extracts (containing Celastrol or WA) as natural immunomodulators and anti-inflammatory agents have already been conducted in some diseases, including schizoaffective disorder, rheumatoid arthritis, inflammatory bowel diseases, Crohn's disease, and lupus nephritis [271]. However, by searching ClinicalTrial.gov, no registered clinical trials directly using these two compounds have been reported.

Many studies have described the roles of WA and Celastrol in immunosuppressive, anti-inflammatory and anti-cancer scenarios via suppression of NF- κ B signalling activation [283, 313, 345, 346]. The NF- κ B signalling pathway is one of the main

players involved in normal and pathophysiological immune responses. NF- κ B target genes have been associated with the regulation of cell survival as well as apoptosis [347-349]. Deregulation of NF- κ B signalling activity contributes to the pathogenic processes of various inflammatory diseases [347, 348]. Interestingly, activation of NF- κ B may function as a tumor promoter in inflammation-associated cancer, suggesting that NF- κ B is a key player linking chronic inflammation and cancer [350] (Figure 6.1). Therefore, various lines of research focus on the development of therapeutics that target the NF- κ B pathway. WA has been shown to exhibit anti-inflammatory activity via inhibition of NF- κ B activation by preventing the TNF-induced activation of I κ B kinase β (IKK β) [313]. This occurs via a thioalkylation-sensitive redox mechanism, coinciding with MEK1/ERK-dependent Ser-181 hyperphosphorylation [313]. WA prevents IKK β phosphorylation and degradation, which subsequently blocks nuclear translocation of NF- κ B, NF- κ B/DNA binding, and downstream gene transcription [313]. Further biochemical analysis showed that WA potently blocks I κ B phosphorylation and inhibits I κ B kinase by targeting cysteine 179 in IKK β [280]. A cytokine array study investigated the regulation of inflammatory cytokines released by monocyte-derived THP-1 cells following treatment with WA, suggesting that WA concurrently targets NF- κ B and the inflammasome complex, leading to inhibition of IL-1 β and IL-18, respectively, in addition to differential expression of multiple cytokines/chemokines [351]. Molecular analysis revealed that Celastrol also inhibits NF- κ B by targeting I κ B kinase. Similarly to WA, Celastrol suppressed the NF- κ B activation by inhibiting IKK β activity, possibly

through directly targeting cysteine 179 in the activation loop of IKK β . Mutation of cysteine 179 could abolish the IKK β sensitivity towards to Celastrol [279].

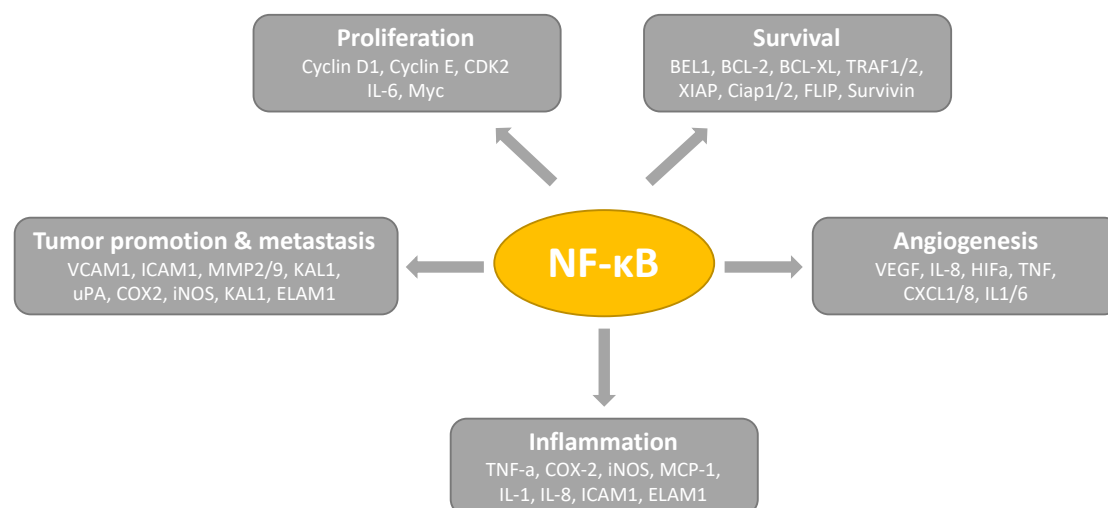


Figure 6.1 NF- κ B signalling in cancer and inflammatory disease. NF- κ B signalling regulates the transcription of genes involved in cell proliferation, survival, angiogenesis, inflammation and tumor promotion and metastasis in cancer and inflammatory disease.

NF- κ B signalling regulates the transcription of various inflammatory cytokines, such as IL-1, IL-2, IL-6, IL-8, IL-12 and TNF- α , as well as genes encoding COX-2, iNOS (pro-inflammatory enzymes promote cytokine production), immunoreceptors, cell adhesion molecules, haematopoietic growth factors, and growth factor receptors [154]. The anti-inflammatory and anti-tumor properties of Celastrol and WA have been partly attributed to the decreased production of proinflammatory cytokines and chemokines induced by drug treatment. For example, Celastrol has been shown to block IL-6 gene expression via downregulation of NF- κ B in many disease contexts, including prostate carcinoma cells [273]. Knockdown of IL-6 attenuated the anti-proliferative effect of Celastrol on PC-3 cells [273]. WA can also attenuate NF- κ B activity, as well as the consequent secretion of TNF α , IL-1 β and IL-18, and the expression COX-2, and iNOS

induced by LPS [351, 352]. Aberrant or constitutive activation of NF- κ B signalling has been shown to be associated with a variety of cancer and inflammatory diseases [315, 347, 348, 350]. NF- κ B signalling regulates molecular pathogenesis of cancer and inflammation not only directly by increasing the production of inflammatory cytokines, chemokines and adhesion molecules, but also regulating the transcription of genes involved in cell proliferation, apoptosis, angiogenesis, and tumor promotion and metastasis. Evidence showed that Celastrol treatment suppressed TNF-induced gene expression that is involved in antiapoptosis (IAP1, IAP2, Bcl-2, Bcl-XL, c-FLIP, and survivin), proliferation (cyclin D1 and COX-2), invasion (MMP-9), and angiogenesis (VEGF) by modulating the NF- κ B pathway. Celastrol suppressed both inducible and constitutive NF- κ B activation [353]. WA has also been shown to inhibit the potential targets involved in proliferative and anti-apoptotic signalling during cancer development [303]. All these studies suggest that Celastrol and WA are valuable compounds for modulation of NF- κ B-dependent pathological conditions.

Splicing factor mutations and aberrant mRNA splicing are the most common molecular alterations in MDS, differentiating its molecular pathogenesis from other haematological malignancies. In this study, computational drug screening hints the connection of Celastrol and WA to splicing process modulation. Celastrol and WA were top hits ranked by CMap analysis, and the input gene signature was generated by comparing the transcriptomes of splicing factor mutant MDS CD34⁺ cells to the transcriptomes of splicing factor wild-type MDS or healthy control CD34⁺ cells. Our

study provided evidence of the involvement of splicing modulation in the anti-proliferative activity of Celastrol and WA in leukaemia cells. Notably, spliceosome mutant cells, including HUDEP2-U2AF1^{Q157} cells, K562-SF3B1^{K700E} cell and K562-U2AF1^{S34F} cells, showed a significantly higher sensitivity to Celastrol and WA treatment than wild-type cells. In a previous chapter (*Chapter V, Section C-1*), we proposed a possible mechanism that Celastrol and WA could inhibit IRAK4-mediated NF-κB signalling activation in splicing factor mutant cells. Many studies have revealed that mutations in splicing factor genes and aberrant splicing events lead to hyperactive NF-κB signalling and inflammatory cytokine production in MDS [170, 172]. In line with these observations, higher basal levels of IL-6 (NF-κB signalling target gene) mRNA expression were observed in splicing factor mutant cells compared to wildtype cells were observed in our study. We also found that Celastrol or WA treatment inhibited the IL-6 mRNA expression levels in splicing factor mutant cells. IRAK4 is a critical kinase regulator of NF-κB signalling pathway, which has been shown to be alternatively spliced in *U2AF1* mutant MDS/AML patients [171]. Mis-spliced IRAK4 is characterized by the retention of exon 4 (IRAK4-long), interacting with MyD88 and facilitating the assembly of the myddosome complex, resulting in maximal activation of NF-κB as well as MAPK activation. Inhibition of IRAK4 was shown to abrogate leukaemic cell growth in vitro and in vivo [171]. Hereby, we performed molecular docking demonstrating the interaction mode of IRAK4 with Celastrol or WA, and we found that the ATP-binding pocket of IRAK4 kinase domain was occupied by Celastrol and WA, indicating their potential role as IRAK4 inhibitors. These observations

primarily support our assumption that Celastrol or WA attenuates IRAK4-mediated NF- κ B signalling activation in splicing factor mutant cells, illuminating the molecular mechanism of their preferential inhibitory effects on splicing factor mutant cells.

Besides NF- κ B signalling, several pathways have been revealed that contributed to the anti-cancer properties of Celastrol and WA in tumor cell lines and animal cancer models (Table 6.1). These pathways have been associated with the effects of cell death activation, cell cycle arrest, angiogenesis inhibition and treatment and radiotherapy sensitizing action. It has been found that Celastrol or WA can induce cell apoptosis via generation of reactive oxygen species (ROS), disruption of mitochondrial membrane potential and inactivation of Akt in several types of cancers, including osteosarcoma, gastric cancer, head and neck squamous cell carcinoma [272, 275, 276, 354-357]. Also, Celastrol and WA have been shown to inhibit HSP90, block proteasome system, and attenuate endoplasmic reticulum (ER) stress, leading to growth inhibition of cancer cells [249, 282, 295, 358-361]. A recent study reported that Celastrol suppressed the proliferative potential of AML cells by inhibiting the Myb/p300 interaction [301].

Moreover, by applying the CMap-based ‘drug connection’ network, we suggested that Piperlogumine and Parthenolide could exert similar anti-proliferative effects on leukaemic cell lines. Multiple mechanisms have been proposed to explain the anti-tumor effects of Piperlogumine and Parthenolide, which mainly involve the induction of ROS accumulation and inhibition of the NF- κ B pathway [362-365]. Besides,

Piperlogumine and Parthenolide could eradicate AML cells by targeting aberrant glutathione metabolism, which includes inhibiting key glutathione metabolic enzyme (GCLC and GPX1), as well as direct depletion of glutathione (Table 6.1) [269, 270].

Table 6.1 Molecular mechanisms of candidate compounds in malignancies

Drugs	Molecular Mechanisms
Celastrrol	● Suppresses the proliferation of AML CD34+ cells by inhibiting the Myb-C/EBPβ-p300 transcriptional module [301]. ● Downregulates AML1-ETO fusion protein and c-KIT kinases, and inhibiting AKT, STAT3 and Erk1/2 [295]. ● Inhibits Hsp90 (via disrupting interaction with cdc37 and co-chaperone p23), blocks proteasome, IKK, NF-κB signalling and suppresses the production of inflammatory cytokines [273, 279, 282, 353]. ● Induces JNK activation and ROS generation [275]. ● Inhibits PI3K/AKT/mTOR [366]. ● ER stress-mediated apoptosis [358, 367, 368]. ● Inhibits topoisomerase II [369].
Withaferin A	● Induces HMOX1 and autophagy-related protein LC3A/B expression [296]. ● Induces ROS production and disrupts mitochondrial membrane potential [276, 355, 356]. ● Induces vimentin disassembly [327, 370]. ● Inhibits JAK/STAT3 signalling [371]. ● Endoplasmic (ER) stress-mediated induction of apoptosis [359, 361]. ● Inhibits PI3K/Akt/mTOR signalling [357, 372]. ● Inhibits Hsp90, IKK, proteasome and NF-κB signalling [280, 283].
Piperlongumine	● Inhibits glutathione and thioredoxin systems and increases ROS production [285]. ● Activation of ROS-p38/JNK pathways [270]. ● Inhibit NF-κB signalling [373]. ● Induces ER stress response in HCC cells by targeting thioredoxin reductase 1 (TrxR1) [374]. ● Target aberrant glutathione metabolism in AML, induce almost complete glutathione depletion and severe cell death in CD34+ AML cells [269]. ● Inhibits STAT3 activation [375]. ● Inhibition of PI3K/Akt/mTOR signalling [284, 376].
Parthenolide	● Targets several components of the NF-κB signalling (directly binds to NF-κB subunits, inhibits IKK complex) [362, 364]. ● Induces ROS production [363, 365]. ● Interferes with microtubule formation [377]. ● Target aberrant glutathione metabolism in AML (inhibit key glutathione metabolic enzyme GCLC and GPX1) [269]. ● Inhibits STAT3 activation [378]

Taken together, our data suggest that Celastrol and WA hold high promise as novel therapeutic drugs for the treatment of splicing factor mutant MDS and other myeloid malignancies. Further studies are needed to unravel the exact mechanisms of these candidate compounds in suppressing the growth of haematological malignant cells.

2. Drugs identified by CMap analysis of microarray-based GEP

Benzimidazole drugs, including PBD, FBD, Mebendazole and Nocodazole, were identified by microarray-based CMap and CMap drug network analysis as potential new drugs for patients with advanced MDS (RAEB-2). PBD and FBD were selected for *in vitro* experiment validation by testing in cell lines, primary patient samples (both bone marrow mononuclear cells and CD34⁺ cells), and MDS patient derived iPSCs. We showed that PBD and FBD are potent inhibitors that can suppress the growth of MDS and leukaemia cells, with half maximal inhibitory concentration (IC₅₀) values in the nanomolar range. These drugs markedly inhibited proliferation, arrested cell cycle, and induced apoptosis of leukaemia cells. Moreover, benzimidazoles demonstrated significant inhibition of the phosphorylation of STAT5a/STAT5b and decreased the expression of downstream target genes involved in differentiation (*MYC*), apoptosis (*BCL-2*) and cell cycle (*CCND1*).

Emerging data has suggested that strategies to inhibit or modulate spliceosomal genes may have therapeutic value in the treatment of spliceosome-mutant disorders [128, 129].

Although FBD and PBD were not directly identified from the analysis of gene expression changes associated with the presence of splicing factor mutations, several benzimidazole anthelmintic compounds (including albendazole, flubendazole, Mebendazole and fenbendazole) were ranked among top candidates by querying the gene signature of K562-SF3B1^{K700E} to the new version of CMap database. Our data show for the first time that leukaemia cells harboring splicing factor mutations are preferentially sensitive to benzimidazoles. Splicing factor mutant cells, including HUDEP2-U2AF1^{Q157} cells, K562-SF3B1^{K700E} cell and K562-U2AF1^{S34F} cells, showed a significantly higher sensitivity to PBD and FBD treatment than wild-type cells.

Repurposing of well characterised non-leukaemia drugs in MDS therapy may have several advantages, including good safety profiles and recognized pharmacokinetic properties in term of absorption, metabolism and toxicity, which may accelerate the clinical translation of preclinical results obtained with these drugs. Some benzimidazole-based anthelmintic family drugs, such as albendazole, mebendazole and flubendazole, are being considered for repurposing in cancer therapy [288-291, 379]. These drugs are widely used in animals and humans as safe and effective treatments for intestinal parasites, but there is also evidence of their anti-tumour effects in melanoma, leukaemia, medulloblastoma, breast and colorectal cancers, through mechanisms involving alteration of microtubule organization and formation of irregular mitotic spindles [293]. Although some studies have reported the potential of these candidate drugs as anti-leukaemia therapeutics, their MoA in tumour suppression remains unclear.

It has been shown that mebendazole could induce c-MYB degradation via the proteasome, by interfering with the HSP70 chaperone system to inhibit AML cell growth in vitro [380]. Mebendazole reduced AML progression in vivo and prolonged survival in an AML xenograft mouse model [380]. Furthermore, mebendazole treatment inhibited the activation of Akt and Erk signalling in AML cells [290]. Flubendazole inhibited microtubule function through a mechanism distinct from that differed from Vinca alkaloids and showed anti-tumour property in leukaemia and myeloma. In addition, Flubendazole synergized with the Vinca alkaloids vinblastine and vincristine in vitro and in vivo, and could be useful in overcoming some forms of resistance to Vinca alkaloids [381]. The effects of other benzimidazoles, such as PBD and FBD, have not been studied yet in MDS and other myeloid malignancies with specific genetic backgrounds. In conclusion, our data provide evidence that PBD or FBD are potential candidate drugs for repurposing to splicing factor mutant myeloid malignancies.

B. Future Directions

1. Drug efficacy and MoA studies

Future work would aim at testing candidate drugs (Celastrol, WA, PBD, FBD) in bone marrow HSCs from a larger cohort of MDS patients (and also from healthy individuals as controls) to evaluate their effects on cell growth and proliferation, apoptosis, and cell cycle. Colony forming cell assays could also be performed to determine whether the drug treatment improves the erythroid and/or myeloid differentiation of MDS patient cells. It would be also interesting to test the drugs identified from the CMap analysis in combination with existing MDS therapeutic agents on MDS primary samples to determine whether there are any synergistic or additive effects. To further reveal the MoA of candidate drugs, especially their roles in gene splicing regulation, it would be interesting to explore the pre-mRNA splicing and expression changes in splicing factor mutated cells treated with drugs of interest. RNA-seq analysis of drug treated MDS primary cells would help to determine whether the disease signatures identified in MDS patients could be reversed by the drugs. A variety of cell lines expressing different splicing factor gene mutations have been generated in our laboratory, including myeloid leukaemia cells edited using CRISPR/Cas9 and MDS patient-derived iPSCs. Our group has optimised a protocol for the differentiation of iPSCs to CD34⁺ cells. It would be interesting to perform RNA-seq on the above-mentioned splicing factor mutant myeloid leukaemia cells and iPSC-derived CD34⁺ cells after treatment with the candidate drugs identified from the CMap analysis. The study of the transcriptome of

these cells, before and after drug treatment, would allow to determine the underlying mechanisms of the candidate compounds in modulating the splicing process. To investigate the kinase selectivity of these candidate drugs, a comprehensive kinase array assay could be conducted using a panel of hundreds of recombinant human protein kinases.

Future studies would involve testing the identified new candidate drugs in mouse xenografts with MDS/AML cell lines or patient cells to assess their *in vivo* efficacy, and ultimately in the context of MDS clinical trials. Importantly, the existing toxicology and safety data available for these drugs may accelerate their future introduction into the clinic.

2. Single-cell analysis of haematopoietic stem cell in SF mutant MDS

In our previous studies, gene expression microarrays and RNA-seq were used to investigate the transcriptome in MDS patients, and these studies identified some critical dysregulated genes and pathways that drive the molecular pathogenesis [95, 103, 105, 116, 149, 151, 382]. However, the analysis of bulk bone marrow CD34⁺ cells can only provide a view of gene expression levels of the whole cell population studied, representing an average of gene expression patterns across thousands to millions of cells. Recent advances in single-cell RNA sequencing (scRNA-seq) technologies enable the study of the transcriptomic landscape at unprecedented resolution. scRNA-seq analysis can provide detailed information about the heterogeneity of gene

expression between cells within a tumour or complex tissues, with the potential to reveal subpopulations with differing sensitivities to therapy or novel therapeutic targets [383-387]. Recent studies have illustrated the power of single-cell technology in the identification of potential therapeutic pathways and cancer-cell-specific targets [383-387]. For example, by performing scRNA-seq analysis of 135,929 CD34+ HSPCs from myelofibrosis patients, Psaila et al. identified potential targets for the inhibition of pathological megakaryocyte differentiation and megakaryocyte-induced fibrosis and validate G6B as a cell surface marker that may enable specific ablation of myelofibrosis cells using immunotherapy [386].

Rodriguez-Meira et al. developed TARGET-seq, a novel method for the high-sensitivity detection of multiple mutations within single cells from both genomic and coding DNA, in parallel with unbiased whole-transcriptome analysis. In this study, the analysis of over 4,000 single cells by TARGET-seq demonstrated how this technique could uniquely resolve the transcriptional and genetic tumor heterogeneity in stem and progenitor cells from patients with myeloproliferative neoplasms. Our group is currently using TARGET-seq to determine the genetic and transcriptomic heterogeneity of single HSPCs from MDS patients with SF mutations [386, 388, 389]. It would be interesting to use the CMap tools on dysregulated genes identified from the analysis of MDS HSPC single-cell data, as this may lead to the identification of other potential drugs for repurposing to MDS.

Reference

1. Garcia-Manero, G., *Myelodysplastic syndromes: 2014 update on diagnosis, risk-stratification, and management*. Am J Hematol, 2014. 89(1): p. 97-108.
2. Fenaux, P., et al., *Myelodysplastic syndromes: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Ann Oncol, 2014. 25 Suppl 3: p. iii57-69.
3. Fey, M.F., M. Dreyling, and E.G.W. Grp, *Acute myeloblastic leukaemias and myelodysplastic syndromes in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology, 2010. 21: p. v158-v161.
4. Kuendgen, A., et al., *Myelodysplastic syndromes in patients younger than age 50*. Journal of Clinical Oncology, 2006. 24(34): p. 5358-5365.
5. Morel, P., et al., *Cytogenetic analysis has strong independent prognostic value in de novo myelodysplastic syndromes and can be incorporated in a new scoring system: a report on 408 cases*. Leukemia, 1993. 7(9): p. 1315-23.
6. Radlund, A., et al., *Incidence of myelodysplastic syndromes in a Swedish population*. Eur J Haematol, 1995. 54(3): p. 153-6.
7. Greenberg, P.L., *The multifaceted nature of myelodysplastic syndromes: clinical, molecular, and biological prognostic features*. J Natl Compr Canc Netw, 2013. 11(7): p. 877-84; quiz 885.
8. Ma, X., et al., *Myelodysplastic syndromes - Incidence and survival in the United States*. Cancer, 2007. 109(8): p. 1536-1542.
9. Ravipati, H.P., S. Annangi, and V. Kota, *Incidence and Survival Outcomes Of Myelodysplastic Syndrome In The United States: A SEER Analysis 2001-2010*. Blood, 2013. 122(21).
10. Matsuda, A., et al., *Difference in clinical features between Japanese and German patients with refractory anemia in myelodysplastic syndromes*. Blood, 2005. 106(8): p. 2633-2640.
11. Qu, S.Q., et al., *Impacts of cytogenetic categories in the Revised International Prognostic Scoring System on the prognosis of primary myelodysplastic syndromes: results of a single-center study*. Leukemia & Lymphoma, 2012. 53(5): p. 940-946.
12. Mufti, G.J., et al., *Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts*. Haematologica, 2008. 93(11): p. 1712-7.
13. Bennett, J.M., et al., *Proposals for the classification of the myelodysplastic syndromes*. Br J Haematol, 1982. 51(2): p. 189-99.
14. Vardiman, J.W., et al., *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes*. Blood, 2009. 114(5): p. 937-51.
15. Delacretaz, F., et al., *Histopathology of myelodysplastic syndromes. The FAB classification (proposals) applied to bone marrow biopsy*. Am J Clin Pathol, 1987. 87(2): p. 180-6.
16. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO)*

- classification of the myeloid neoplasms*. Blood, 2002. 100(7): p. 2292-2302.
17. Howe, R.B., et al., *The WHO classification of myelodysplastic syndromes (MDS) does make a difference*. Blood, 2002. 100(11): p. 790a-791a.
 18. van Spronsen, M.F., et al., *Validation of and proposals for refinements of the WHO 2016 classification for myelodysplastic syndromes*. Am J Hematol, 2017. 92(11): p. E631-E634.
 19. Maassen, A., et al., *Validation and proposals for a refinement of the WHO 2008 classification of myelodysplastic syndromes without excess of blasts*. Leuk Res, 2013. 37(1): p. 64-70.
 20. Vardiman, J., *The classification of MDS: From FAB to WHO and beyond*. Leukemia Research, 2012. 36(12): p. 1453-1458.
 21. Barbui, T., et al., *The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion*. Blood Cancer Journal, 2018. 8.
 22. Swerdlow, S.H., et al., *The 2016 revision of the World Health Organization classification of lymphoid neoplasms*. Blood, 2016. 127(20): p. 2375-2390.
 23. Leonard, J.P., P. Martin, and G.J. Roboz, *Practical Implications of the 2016 Revision of the World Health Organization Classification of Lymphoid and Myeloid Neoplasms and Acute Leukemia*. Journal of Clinical Oncology, 2017. 35(23): p. 2708-+.
 24. Greenberg, P.L., et al., *Revised international prognostic scoring system for myelodysplastic syndromes*. Blood, 2012. 120(12): p. 2454-65.
 25. Malcovati, L., et al., *Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes*. J Clin Oncol, 2007. 25(23): p. 3503-10.
 26. Malcovati, L., et al., *Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS)*. Haematologica, 2011. 96(10): p. 1433-40.
 27. Malcovati, L., et al., *Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: a basis for clinical decision making*. J Clin Oncol, 2005. 23(30): p. 7594-603.
 28. Greenberg, P., et al., *International scoring system for evaluating prognosis in myelodysplastic syndromes*. Blood, 1997. 89(6): p. 2079-88.
 29. Sperr, W.R., et al., *Proposed score for survival of patients with myelodysplastic syndromes*. European Journal of Clinical Investigation, 2013. 43(11): p. 1120-1128.
 30. Steensma, D.P., *Myelodysplastic syndromes current treatment algorithm 2018*. Blood Cancer J, 2018. 8(5): p. 47.
 31. Fenau, P., et al., *Myelodysplastic syndromes: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology, 2014. 25: p. 57-69.
 32. Malcovati, L., et al., *Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet*. Blood, 2013. 122(17): p. 2943-64.
 33. Corey, S.J., et al., *Myelodysplastic syndromes: the complexity of stem-cell diseases*. Nat Rev Cancer, 2007. 7(2): p. 118-29.
 34. Platzbecker, U., *Who benefits from allogeneic transplantation for myelodysplastic syndromes?: new insights*. Hematology Am Soc Hematol Educ Program, 2013. 2013: p. 522-8.
 35. Vaughn, J.E., B.L. Scott, and H.J. Deeg, *Transplantation for myelodysplastic syndromes 2013*. Curr Opin Hematol, 2013. 20(6): p. 494-500.

36. Fenaux, P., et al., *Luspatercept in Patients with Lower-Risk Myelodysplastic Syndromes*. N Engl J Med, 2020. 382(2): p. 140-151.
37. Platzbecker, U., et al., *Luspatercept for the treatment of anaemia in patients with lower-risk myelodysplastic syndromes (PACE-MDS): a multicentre, open-label phase 2 dose-finding study with long-term extension study*. Lancet Oncol, 2017. 18(10): p. 1338-1347.
38. Haase, D., et al., *New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients*. Blood, 2007. 110(13): p. 4385-95.
39. Ogawa, S., *Genetics of MDS*. Blood, 2019. 133(10): p. 1049-1059.
40. Nakajima, H., *Molecular Pathogenesis and Treatment of Myelodysplastic Syndromes*. Intern Med, 2021. 60(1): p. 15-23.
41. Haferlach, T., et al., *Landscape of genetic lesions in 944 patients with myelodysplastic syndromes*. Leukemia, 2014. 28(2): p. 241-7.
42. Pellagatti, A. and J. Boultonwood, *The molecular pathogenesis of the myelodysplastic syndromes*. Eur J Haematol, 2015. 95(1): p. 3-15.
43. Papaemmanuil, E., et al., *Clinical and biological implications of driver mutations in myelodysplastic syndromes*. Blood, 2013. 122(22): p. 3616-27; quiz 3699.
44. Sperling, A.S., C.J. Gibson, and B.L. Ebert, *The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia*. Nature Reviews Cancer, 2017. 17(1): p. 5-19.
45. Lindsley, R.C., et al., *Acute myeloid leukemia ontogeny is defined by distinct somatic mutations*. Blood, 2015. 125(9): p. 1367-1376.
46. Bejar, R., et al., *Clinical effect of point mutations in myelodysplastic syndromes*. N Engl J Med, 2011. 364(26): p. 2496-506.
47. Kwok, B., et al., *MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance*. Blood, 2015. 126(21): p. 2355-2361.
48. Bejar, R., et al., *TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients*. Blood, 2014. 124(17): p. 2705-12.
49. Itzykson, R., et al., *Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias*. Leukemia, 2011. 25(7): p. 1147-52.
50. Bejar, R., et al., *Somatic mutations predict poor outcome in patients with myelodysplastic syndrome after hematopoietic stem-cell transplantation*. J Clin Oncol, 2014. 32(25): p. 2691-8.
51. Della Porta, M.G., et al., *Clinical Effects of Driver Somatic Mutations on the Outcomes of Patients With Myelodysplastic Syndromes Treated With Allogeneic Hematopoietic Stem-Cell Transplantation*. J Clin Oncol, 2016. 34(30): p. 3627-3637.
52. Bejar, R., et al., *Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes*. J Clin Oncol, 2012. 30(27): p. 3376-82.
53. Malcovati, L., et al., *Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms*. Blood, 2011. 118(24): p. 6239-46.
54. Kitagawa, M., et al., *p53 expression in myeloid cells of myelodysplastic syndromes. Association with evolution of overt leukemia*. Am J Pathol, 1994. 145(2): p. 338-44.
55. Lindsley, R.C., et al., *Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation*. N Engl J Med, 2017. 376(6): p. 536-547.

56. Yoshizato, T., et al., *Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation*. Blood, 2017. 129(17): p. 2347-2358.
57. Jadersten, M., et al., *TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression*. J Clin Oncol, 2011. 29(15): p. 1971-9.
58. Ok, C.Y., et al., *TP53 mutation characteristics in therapy-related myelodysplastic syndromes and acute myeloid leukemia is similar to de novo diseases*. Journal of Hematology & Oncology, 2015. 8.
59. Wong, T.N., et al., *Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia*. Nature, 2015. 518(7540): p. 552-555.
60. Kulasekararaj, A.G., et al., *TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis*. Br J Haematol, 2013. 160(5): p. 660-72.
61. Takahashi, K., et al., *Clinical implications of TP53 mutations in myelodysplastic syndromes treated with hypomethylating agents*. Oncotarget, 2016. 7(12): p. 14172-87.
62. Mossner, M., et al., *Prevalence, clonal dynamics and clinical impact of TP53 mutations in patients with myelodysplastic syndrome with isolated deletion (5q) treated with lenalidomide: results from a prospective multicenter study of the german MDS study group (GMDS)*. Leukemia, 2016. 30(9): p. 1956-9.
63. Sallman, D.A., et al., *Impact of TP53 mutation variant allele frequency on phenotype and outcomes in myelodysplastic syndromes*. Leukemia, 2016. 30(3): p. 666-73.
64. Bernard, E., et al., *Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes*. Nat Med, 2020. 26(10): p. 1549-1556.
65. Papaemmanuil, E., et al., *Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts*. N Engl J Med, 2011. 365(15): p. 1384-95.
66. Malcovati, L., et al., *SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts*. Blood, 2015. 126(2): p. 233-41.
67. Cazzola, M., M.G. Della Porta, and L. Malcovati, *The genetic basis of myelodysplasia and its clinical relevance*. Blood, 2013. 122(25): p. 4021-34.
68. Malcovati, L., et al., *SF3B1-mutant myelodysplastic syndrome as a distinct disease subtype - A Proposal of the International Working Group for the Prognosis of Myelodysplastic Syndromes (IWG-PM)*. Blood, 2020.
69. Graubert, T.A., et al., *Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes*. Nat Genet, 2011. 44(1): p. 53-7.
70. Thol, F., et al., *Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes*. Blood, 2012. 119(15): p. 3578-84.
71. Damm, F., et al., *Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes*. Blood, 2012. 119(14): p. 3211-8.
72. Patel, B.J., et al., *Genomic determinants of chronic myelomonocytic leukemia*. Leukemia, 2017. 31(12): p. 2815-2823.
73. Meggendorfer, M., et al., *SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML)*. Blood, 2012. 120(15): p. 3080-8.
74. Grossmann, V., et al., *Molecular profiling of chronic myelomonocytic leukemia reveals diverse*

- mutations in > 80% of patients with TET2 and EZH2 being of high prognostic relevance. *Leukemia*, 2011. 25(5): p. 877-879.
75. Bejar, R., et al., *Somatic Mutations in MDS Patients Are Associated with Clinical Features and Predict Prognosis Independent of the IPSS-R: Analysis of Combined Datasets from the International Working Group for Prognosis in MDS-Molecular Committee*. *Blood*, 2015. 126(23).
 76. Bejar, R., *Implications of molecular genetic diversity in myelodysplastic syndromes*. *Curr Opin Hematol*, 2017. 24(2): p. 73-78.
 77. Malcovati, L., et al., *Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia*. *Blood*, 2014. 124(9): p. 1513-1521.
 78. Karimi, M., et al., *High-throughput mutational screening adds clinically important information in myelodysplastic syndromes and secondary or therapy-related acute myeloid leukemia*. *Haematologica*, 2015. 100(6): p. e223-5.
 79. Steensma, D.P., et al., *Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes*. *Blood*, 2015. 126(1): p. 9-16.
 80. Walter, M.J., et al., *Clonal architecture of secondary acute myeloid leukemia*. *N Engl J Med*, 2012. 366(12): p. 1090-8.
 81. Takahashi, K., et al., *Dynamic acquisition of FLT3 or RAS alterations drive a subset of patients with lower risk MDS to secondary AML*. *Leukemia*, 2013. 27(10): p. 2081-3.
 82. Makishima, H., et al., *Dynamics of clonal evolution in myelodysplastic syndromes*. *Nature Genetics*, 2017. 49(2): p. 204-212.
 83. Hou, H.A., et al., *Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome*. *Blood Cancer J*, 2018. 8(4): p. 39.
 84. Nazha, A., et al., *Incorporation of molecular data into the Revised International Prognostic Scoring System in treated patients with myelodysplastic syndromes*. *Leukemia*, 2016. 30(11): p. 2214-2220.
 85. Nazha, A. and R. Bejar, *Molecular Data and the IPSS-R: How Mutational Burden Can Affect Prognostication in MDS*. *Curr Hematol Malig Rep*, 2017. 12(5): p. 461-467.
 86. Hurd, P.J. and C.J. Nelson, *Advantages of next-generation sequencing versus the microarray in epigenetic research*. *Brief Funct Genomic Proteomic*, 2009. 8(3): p. 174-83.
 87. Theilgaard-Monch, K., et al., *Gene expression profiling in MDS and AML: potential and future avenues*. *Leukemia*, 2011. 25(6): p. 909-20.
 88. Kwa, M., A. Makris, and F.J. Esteva, *Clinical utility of gene-expression signatures in early stage breast cancer*. *Nat Rev Clin Oncol*, 2017. 14(10): p. 595-610.
 89. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer*. *Nature*, 2002. 415(6871): p. 530-6.
 90. Lapointe, J., et al., *Gene expression profiling identifies clinically relevant subtypes of prostate cancer*. *Proc Natl Acad Sci U S A*, 2004. 101(3): p. 811-6.
 91. Metsis, A., et al., *Whole-genome expression profiling through fragment display and combinatorial gene identification*. *Nucleic Acids Res*, 2004. 32(16): p. e127.
 92. Fielden, M.R. and T.R. Zacharewski, *Challenges and limitations of gene expression profiling in*

- mechanistic and predictive toxicology*. Toxicol Sci, 2001. 60(1): p. 6-10.
93. Chibon, F., *Cancer gene expression signatures - the rise and fall?* Eur J Cancer, 2013. 49(8): p. 2000-9.
 94. Mills, K.I., et al., *Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome*. Blood, 2009. 114(5): p. 1063-72.
 95. Pellagatti, A., et al., *Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes*. J Clin Oncol, 2013. 31(28): p. 3557-64.
 96. Gerstung, M., et al., *Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes*. Nat Commun, 2015. 6: p. 5901.
 97. Kim, K., et al., *Gene expression signatures associated with sensitivity to azacitidine in myelodysplastic syndromes*. Sci Rep, 2020. 10(1): p. 19555.
 98. Shiozawa, Y., et al., *Gene expression and risk of leukemic transformation in myelodysplasia*. Blood, 2017. 130(24): p. 2642-2653.
 99. De Clara, E., et al., *Longitudinal Transcriptome Analysis of Myelodysplastic Syndrome Progression to Acute Myeloid Leukemia: Implications for Prognosis*. Blood, 2017. 130.
 100. Arindrarto, W., et al., *Comprehensive diagnostics of acute myeloid leukemia by whole transcriptome RNA sequencing*. Leukemia, 2021. 35(1): p. 47-61.
 101. Mano, H., *DNA micro-array analysis of myelodysplastic syndrome*. Leuk Lymphoma, 2006. 47(1): p. 9-14.
 102. Qian, J., et al., *Gene expression profiling of the bone marrow mononuclear cells from patients with myelodysplastic syndrome*. Oncol Rep, 2005. 14(5): p. 1189-97.
 103. Pellagatti, A., et al., *Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype*. Blood, 2006. 108(1): p. 337-45.
 104. Boultonwood, J., *Gene expression profiling in patients with myelodysplastic syndromes, including patients with the del(5q)*. Leukemia Research, 2009. 33: p. S3-S4.
 105. Pellagatti, A., et al., *Gene expression profiling in the myelodysplastic syndromes*. Hematology, 2005. 10(4): p. 281-287.
 106. Haeflrich, T., et al., *Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Subclassification of Leukemia: Report From the International Microarray Innovations in Leukemia Study Group*. Journal of Clinical Oncology, 2010. 28(15): p. 2529-2537.
 107. Boultonwood, J., et al., *The Role of the Iron Transporter ABCB7 in Refractory Anemia with Ring Sideroblasts*. Plos One, 2008. 3(4).
 108. Pondarre, C., et al., *Abcb7, the gene responsible for X-linked sideroblastic anemia with ataxia, is essential for hematopoiesis*. Blood, 2007. 109(8): p. 3567-3569.
 109. Cavadini, P., et al., *RNA silencing of the mitochondrial ABCB7 transporter in HeLa cells causes an iron-deficient phenotype with mitochondrial iron overload*. Blood, 2007. 109(8): p. 3552-3559.
 110. Nikpour, M., et al., *Gene expression profiling of erythroblasts from refractory anaemia with*

- ring sideroblasts (RARS) and effects of G-CSF*. Br J Haematol, 2010. 149(6): p. 844-54.
111. Dolatshad, H., et al., *Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes*. Leukemia, 2016. 30(12): p. 2322-2331.
 112. Fenaux, P. and C. Kelaidi, *Treatment of the 5q- syndrome*. Hematology Am Soc Hematol Educ Program, 2006: p. 192-8.
 113. Nilsson, L., et al., *The molecular signature of MDS stem cells supports a stem-cell origin of 5q - myelodysplastic syndromes*. Blood, 2007. 110(8): p. 3005-3014.
 114. Boultonwood, J., et al., *Gene expression profiling of CD34+ cells in patients with the 5q- syndrome*. Br J Haematol, 2007. 139(4): p. 578-89.
 115. Pellagatti, A., et al., *Haploinsufficiency of RPS14 in 5q-syndrome is associated with deregulation of ribosomal- and translation-related genes*. British Journal of Haematology, 2008. 142(1): p. 57-64.
 116. Pellagatti, A., et al., *Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells*. Leukemia, 2010. 24(4): p. 756-764.
 117. Sridhar, K., et al., *Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression*. Blood, 2009. 114(23): p. 4847-58.
 118. Pellagatti, A., et al., *Identification of Gene Expression-Based Prognostic Markers in the Hematopoietic Stem Cells of Patients With Myelodysplastic Syndromes*. Journal of Clinical Oncology, 2013. 31(28): p. 3557-3564.
 119. Boultonwood, J., et al., *The role of splicing factor mutations in the pathogenesis of the myelodysplastic syndromes*. Adv Biol Regul, 2014. 54: p. 153-61.
 120. Saez, B., M.J. Walter, and T.A. Graubert, *Splicing factor gene mutations in hematologic malignancies*. Blood, 2017. 129(10): p. 1260-1269.
 121. Makishima, H., et al., *Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis*. Blood, 2012. 119(14): p. 3203-10.
 122. Yoshida, K., et al., *Frequent pathway mutations of splicing machinery in myelodysplasia*. Nature, 2011. 478(7367): p. 64-9.
 123. Pellagatti, A. and J. Boultonwood, *Splicing factor gene mutations in the myelodysplastic syndromes: impact on disease phenotype and therapeutic applications*. Adv Biol Regul, 2017. 63: p. 59-70.
 124. Desterro, J., P. Bak-Gordon, and M. Carmo-Fonseca, *Targeting mRNA processing as an anticancer strategy*. Nat Rev Drug Discov, 2020. 19(2): p. 112-129.
 125. Shirai, C.L., et al., *Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo*. Cancer Cell, 2015. 27(5): p. 631-43.
 126. Kim, E., et al., *SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition*. Cancer Cell, 2015. 27(5): p. 617-30.
 127. Dolatshad, H., et al., *Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells*. Leukemia, 2015. 29(8): p. 1798.
 128. Lee, S.C. and O. Abdel-Wahab, *Therapeutic targeting of splicing in cancer*. Nat Med, 2016. 22(9): p. 976-86.

129. Salton, M. and T. Misteli, *Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy*. Trends Mol Med, 2016. 22(1): p. 28-37.
130. Taylor, J. and S.C. Lee, *Mutations in spliceosome genes and therapeutic opportunities in myeloid malignancies*. Genes Chromosomes Cancer, 2019. 58(12): p. 889-902.
131. Wahl, M.C., C.L. Will, and R. Luhmann, *The spliceosome: design principles of a dynamic RNP machine*. Cell, 2009. 136(4): p. 701-18.
132. Matera, A.G. and Z.F. Wang, *A day in the life of the spliceosome*. Nature Reviews Molecular Cell Biology, 2014. 15(2): p. 108-121.
133. Wahl, M.C., C.L. Will, and R. Luhmann, *The Spliceosome: Design Principles of a Dynamic RNP Machine*. Cell, 2009. 136(4): p. 701-718.
134. Yan, C.Y., R.X. Wan, and Y.G. Shi, *Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome*. Cold Spring Harbor Perspectives in Biology, 2019. 11(1).
135. Lee, Y. and D.C. Rio, *Mechanisms and Regulation of Alternative Pre-mRNA Splicing*. Annual Review of Biochemistry, Vol 84, 2015. 84: p. 291-323.
136. Will, C.L. and R. Luhmann, *Spliceosome structure and function*. Cold Spring Harb Perspect Biol, 2011. 3(7).
137. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes*. Nature, 2008. 456(7221): p. 470-476.
138. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*. Nature Genetics, 2008. 40(12): p. 1413-1415.
139. Dvinge, H., et al., *RNA splicing factors as oncoproteins and tumour suppressors*. Nature Reviews Cancer, 2016. 16(7): p. 413-430.
140. Obeng, E.A., C. Stewart, and O. Abdel-Wahab, *Altered RNA Processing in Cancer Pathogenesis and Therapy*. Cancer Discovery, 2019. 9(11): p. 1493-1510.
141. Pellagatti, A. and J. Boulton, *Splicing factor mutant myelodysplastic syndromes: Recent advances*. Adv Biol Regul, 2020. 75: p. 100655.
142. Jacob, A.G. and C.W.J. Smith, *Intron retention as a component of regulated gene expression programs*. Hum Genet, 2017. 136(9): p. 1043-1057.
143. Brooks, A.N., et al., *A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events (vol 9, e87361, 2014)*. Plos One, 2014. 9(4).
144. Chabot, B. and L. Shkreta, *Defective control of pre-messenger RNA splicing in human disease*. J Cell Biol, 2016. 212(1): p. 13-27.
145. Yip, B.H., et al., *Impact of Splicing Factor Mutations on Pre-mRNA Splicing in the Myelodysplastic Syndromes*. Curr Pharm Des, 2016. 22(16): p. 2333-44.
146. Malcovati, L. and M. Cazzola, *Recent advances in the understanding of myelodysplastic syndromes with ring sideroblasts*. Br J Haematol, 2016. 174(6): p. 847-58.
147. Mian, S.A., et al., *Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome*. Haematologica, 2013. 98(7): p. 1058-66.
148. Shiozawa, Y., et al., *Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia*. Nature Communications, 2018. 9.

149. Pellagatti, A., et al., *Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations*. Blood, 2018. 132(12): p. 1225-1240.
150. Przychodzen, B., et al., *Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms*. Blood, 2013. 122(6): p. 999-1006.
151. Yip, B.H., et al., *The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes*. J Clin Invest, 2017. 127(9): p. 3557.
152. Nikpour, M., et al., *The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts*. Leukemia, 2013. 27(4): p. 889-896.
153. Mian, S.A., et al., *SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment*. Nature Communications, 2015. 6.
154. Alsafadi, S., et al., *Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage*. Nat Commun, 2016. 7: p. 10615.
155. Darman, R.B., et al., *Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point*. Cell Rep, 2015. 13(5): p. 1033-45.
156. DeBoever, C., et al., *Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers*. PLoS Comput Biol, 2015. 11(3): p. e1004105.
157. Obeng, E.A., et al., *Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation*. Cancer Cell, 2016. 30(3): p. 404-417.
158. Mupo, A., et al., *Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts*. Leukemia, 2017. 31(3): p. 720-727.
159. Visconte, V., et al., *Emerging roles of the spliceosomal machinery in myelodysplastic syndromes and other hematological disorders*. Leukemia, 2012. 26(12): p. 2447-54.
160. Wu, S.J., et al., *The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution*. Blood, 2012. 120(15): p. 3106-11.
161. Zhang, J., et al., *Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities*. Proc Natl Acad Sci U S A, 2015. 112(34): p. E4726-34.
162. Daubner, G.M., et al., *A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well*. EMBO J, 2012. 31(1): p. 162-74.
163. Nikoloski, G., et al., *Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes*. Nature Genetics, 2010. 42(8): p. 665-667.
164. Kon, A., et al., *Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice*. Blood, 2018. 131(6): p. 621-635.
165. Okeyo-Owuor, T., et al., *U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing*. Leukemia, 2015. 29(4): p. 909-917.
166. Wu, S.J., et al., *Clinical implications of U2AF1 mutation in patients with myelodysplastic syndrome and its stability during disease progression*. American Journal of Hematology, 2013. 88(11): p. E277-E282.
167. Ilagan, J.O., et al., *U2AF1 mutations alter splice site recognition in hematological malignancies*. Genome Research, 2015. 25(1): p. 14-26.
168. Mortensen, M., et al., *The autophagy protein Atg7 is essential for hematopoietic stem cell*

- maintenance*. Journal of Experimental Medicine, 2011. 208(3): p. 455-467.
169. Park, S.M., et al., *U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3' End Formation*. Mol Cell, 2016. 62(4): p. 479-90.
 170. Lee, S.C., et al., *Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations*. Cancer Cell, 2018. 34(2): p. 225-241 e8.
 171. Smith, M.A., et al., *U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies*. Nat Cell Biol, 2019. 21(5): p. 640-650.
 172. Pollyea, D.A., et al., *Myelodysplastic syndrome-associated spliceosome gene mutations enhance innate immune signaling*. Haematologica, 2019. 104(9): p. e388-e392.
 173. De Arras, L. and S. Alper, *Limiting of the innate immune response by SF3A-dependent control of MyD88 alternative mRNA splicing*. PLoS Genet, 2013. 9(10): p. e1003855.
 174. Baldridge, M.T., K.Y. King, and M.A. Goodell, *Inflammatory signals regulate hematopoietic stem cells*. Trends in Immunology, 2011. 32(2): p. 57-65.
 175. Nakagawa, M.M. and C.V. Rathinam, *Constitutive Activation of the Canonical NF-kappaB Pathway Leads to Bone Marrow Failure and Induction of Erythroid Signature in Hematopoietic Stem Cells*. Cell Rep, 2018. 25(8): p. 2094-2109 e4.
 176. Chen, L., et al., *The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations*. Molecular Cell, 2018. 69(3): p. 412-+.
 177. Nguyen, H.D., et al., *Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes*. Cancer Res, 2018. 78(18): p. 5363-5374.
 178. Singh, S., et al., *SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications*. Leukemia, 2020.
 179. Fei, D.L., et al., *Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival*. Plos Genetics, 2016. 12(10).
 180. Zhou, Q.H., et al., *A Chemical Genetics Approach for the Functional Assessment of Novel Cancer Genes*. Cancer Research, 2015. 75(10): p. 1949-1958.
 181. Lee, S.C., et al., *Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins*. Nat Med, 2016. 22(6): p. 672-8.
 182. Yoshimi, A. and O. Abdel-Wahab, *Molecular Pathways: Understanding and Targeting Mutant Spliceosomal Proteins*. Clinical Cancer Research, 2017. 23(2): p. 336-341.
 183. Armstrong, R.N., et al., *Splicing factor mutations in the myelodysplastic syndromes: target genes and therapeutic approaches*. Adv Biol Regul, 2018. 67: p. 13-29.
 184. Effenberger, K.A., V.K. Urabe, and M.S. Jurica, *Modulating splicing with small molecular inhibitors of the spliceosome*. Wiley Interdisciplinary Reviews-Rna, 2017. 8(2).
 185. Effenberger, K.A., et al., *Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages*. Rna, 2016. 22(3): p. 350-359.
 186. Corrionero, A., B. Minana, and J. Valcarcel, *Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A*. Genes & Development, 2011. 25(5): p. 445-459.
 187. Folco, E.G., K.E. Coil, and R. Reed, *The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region*. Genes &

- Development, 2011. 25(5): p. 440-444.
188. Teng, T., et al., *Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex*. Nature Communications, 2017. 8.
 189. Shirai, C.L., et al., *Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome*. Nat Commun, 2017. 8: p. 14060.
 190. Seiler, M., et al., *H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers*. Nature Medicine, 2018. 24(4): p. 497-+.
 191. Buonamici, S., et al., *H3B-8800, an Orally Bioavailable Modulator of the SF3b Complex, Shows Efficacy in Spliceosome-Mutant Myeloid Malignancies*. Blood, 2016. 128(22).
 192. Hong, D.S., et al., *A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) spliceosome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors*. Invest New Drugs, 2014. 32(3): p. 436-44.
 193. Eskens, F.A., et al., *Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors*. Clin Cancer Res, 2013. 19(22): p. 6296-304.
 194. Steensma, D.P., et al., *Results of a Clinical Trial of H3B-8800, a Splicing Modulator, in Patients with Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML) or Chronic Myelomonocytic Leukemia (CMML)*. Blood, 2019. 134.
 195. Scannell, J.W., et al., *Diagnosing the decline in pharmaceutical R&D efficiency*. Nat Rev Drug Discov, 2012. 11(3): p. 191-200.
 196. Nishimura, Y. and H. Hara, *Editorial: Drug Repositioning: Current Advances and Future Perspectives*. Front Pharmacol, 2018. 9: p. 1068.
 197. Ashburn, T.T. and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs*. Nat Rev Drug Discov, 2004. 3(8): p. 673-83.
 198. Talevi, A. and C.L. Bellera, *Challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics*. Expert Opinion on Drug Discovery, 2020. 15(4): p. 397-401.
 199. Iorio, F., et al., *Transcriptional data: a new gateway to drug repositioning?* Drug Discov Today, 2013. 18(7-8): p. 350-7.
 200. Musa, A., et al., *A review of connectivity map and computational approaches in pharmacogenomics*. Brief Bioinform, 2017. 18(5): p. 903.
 201. Pushpakom, S., et al., *Drug repurposing: progress, challenges and recommendations*. Nat Rev Drug Discov, 2019. 18(1): p. 41-58.
 202. Paolini, G.V., et al., *Global mapping of pharmacological space*. Nat Biotechnol, 2006. 24(7): p. 805-15.
 203. Brehmer, D., et al., *Cellular targets of gefitinib*. Cancer Research, 2005. 65(2): p. 379-382.
 204. Karaman, M.W., et al., *A quantitative analysis of kinase inhibitor selectivity*. Nat Biotechnol, 2008. 26(1): p. 127-32.
 205. Davis, M.I., et al., *Comprehensive analysis of kinase inhibitor selectivity*. Nat Biotechnol, 2011. 29(11): p. 1046-51.
 206. Moffat, J.G., et al., *Opportunities and challenges in phenotypic drug discovery: an industry*

- perspective*. Nat Rev Drug Discov, 2017. 16(8): p. 531-543.
207. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease*. Science, 2006. 313(5795): p. 1929-35.
 208. Zhang, S.D. and T.W. Gant, *sscMap: An extensible Java application for connecting small-molecule drugs using gene-expression signatures*. BMC Bioinformatics, 2009. 10.
 209. Keenan, A.B., et al., *The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations*. Cell Systems, 2018. 6(1): p. 13-24.
 210. Subramanian, A., et al., *A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles*. Cell, 2017. 171(6): p. 1437-1452 e17.
 211. Keiser, M.J., et al., *Predicting new molecular targets for known drugs*. Nature, 2009. 462(7270): p. 175-81.
 212. Campillos, M., et al., *Drug target identification using side-effect similarity*. Science, 2008. 321(5886): p. 263-266.
 213. Huang, H.B., et al., *Reverse Screening Methods to Search for the Protein Targets of Chemopreventive Compounds*. Frontiers in Chemistry, 2018. 6.
 214. Pagadala, N.S., K. Syed, and J. Tuszynski, *Software for molecular docking: a review*. Biophys Rev, 2017. 9(2): p. 91-102.
 215. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: Methods and applications*. Nature Reviews Drug Discovery, 2004. 3(11): p. 935-949.
 216. Kharkar, P.S., S. Warriar, and R.S. Gaud, *Reverse docking: a powerful tool for drug repositioning and drug rescue*. Future Medicinal Chemistry, 2014. 6(3): p. 333-342.
 217. Sanseau, P., et al., *Use of genome-wide association studies for drug repositioning*. Nature Biotechnology, 2012. 30(4): p. 317-320.
 218. Zhang, J.H., et al., *Use of Genome-Wide Association Studies for Cancer Research and Drug Repositioning*. Plos One, 2015. 10(3).
 219. Smith, S.B., et al., *Identification of Common Biological Pathways and Drug Targets Across Multiple Respiratory Viruses Based on Human Host Gene Expression Analysis*. Plos One, 2012. 7(3).
 220. Greene, C.S. and B.F. Voight, *Pathway and network-based strategies to translate genetic discoveries into effective therapies*. Human Molecular Genetics, 2016. 25(R2): p. R94-R98.
 221. Iorio, F., J. Saez-Rodriguez, and D. di Bernardo, *Network based elucidation of drug response: from modulators to targets*. BMC Systems Biology, 2013. 7.
 222. Greene, C.S., et al., *Understanding multicellular function and disease with human tissue-specific networks*. Nature Genetics, 2015. 47(6): p. 569-576.
 223. Jensen, P.B., L.J. Jensen, and S. Brunak, *Mining electronic health records: towards better research applications and clinical care*. Nat Rev Genet, 2012. 13(6): p. 395-405.
 224. Paik, H., et al., *Repurpose terbutaline sulfate for amyotrophic lateral sclerosis using electronic medical records*. Scientific Reports, 2015. 5.
 225. Wicks, P., et al., *Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm*. Nature Biotechnology, 2011. 29(5): p. 411-U230.
 226. Iorio, F., et al., *A Landscape of Pharmacogenomic Interactions in Cancer*. Cell, 2016. 166(3):

- p. 740-754.
227. Seashore-Ludlow, B., et al., *Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset*. Cancer Discovery, 2015. 5(11): p. 1210-1223.
 228. Basu, A., et al., *An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules*. Cell, 2013. 154(5): p. 1151-1161.
 229. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2012. 483(7391): p. 603-607.
 230. Huang, Y.H. and C.R. Vakoc, *A Biomarker Harvest from One Thousand Cancer Cell Lines*. Cell, 2016. 166(3): p. 536-537.
 231. Hieronymus, H., et al., *Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators*. Cancer Cell, 2006. 10(4): p. 321-30.
 232. Qu, X.A. and D.K. Rajpal, *Applications of Connectivity Map in drug discovery and development*. Drug Discov Today, 2012. 17(23-24): p. 1289-98.
 233. Dudley, J.T., T. Deshpande, and A.J. Butte, *Exploiting drug-disease relationships for computational drug repositioning*. Brief Bioinform, 2011. 12(4): p. 303-11.
 234. McArt, D.G. and S.D. Zhang, *Identification of candidate small-molecule therapeutics to cancer by gene-signature perturbation in connectivity mapping*. PLoS One, 2011. 6(1): p. e16382.
 235. Wagner, A., et al., *Drugs that reverse disease transcriptomic signatures are more effective in a mouse model of dyslipidemia*. Molecular Systems Biology, 2015. 11(3).
 236. Hu, G. and P. Agarwal, *Human disease-drug network based on genomic expression profiles*. PLoS One, 2009. 4(8): p. e6536.
 237. Hsieh, Y.Y., et al., *Repositioning of a cyclin-dependent kinase inhibitor GW8510 as a ribonucleotide reductase M2 inhibitor to treat human colorectal cancer*. Cell Death Discovery, 2016. 2.
 238. Iorio, F., et al., *Identification of small molecules enhancing autophagic function from drug network analysis*. Autophagy, 2010. 6(8): p. 1204-+.
 239. Iorio, F., et al., *A Semi-Supervised Approach for Refining Transcriptional Signatures of Drug Response and Repositioning Predictions*. PLoS One, 2015. 10(10): p. e0139446.
 240. Lamb, J., *The Connectivity Map: a new tool for biomedical research*. Nat Rev Cancer, 2007. 7(1): p. 54-60.
 241. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proc Natl Acad Sci U S A, 2005. 102(43): p. 15545-50.
 242. Keenan, A.B., et al., *Connectivity Mapping: Methods and Applications*. Annual Review of Biomedical Data Science, Vol 2, 2019, 2019. 2: p. 69-92.
 243. Wen, Q., et al., *Connectivity mapping using a combined gene signature from multiple colorectal cancer datasets identified candidate drugs including existing chemotherapies*. BMC Systems Biology, 2015. 9.
 244. McArt, D.G., et al., *Connectivity Mapping for Candidate Therapeutics Identification Using Next Generation Sequencing RNA-Seq Data*. Plos One, 2013. 8(6).
 245. Sanda, T., et al., *Interconnecting molecular pathways in the pathogenesis and drug sensitivity of T-cell acute lymphoblastic leukemia*. Blood, 2010. 115(9): p. 1735-45.

246. Beck, A., et al., *Connectivity map identifies HDAC inhibition as a treatment option of high-risk hepatoblastoma*. *Cancer Biol Ther*, 2016. 17(11): p. 1168-1176.
247. Bhat-Nakshatri, P., et al., *Identification of FDA-approved Drugs Targeting Breast Cancer Stem Cells Along With Biomarkers of Sensitivity*. *Scientific Reports*, 2013. 3.
248. Cheng, H.W., et al., *Identification of thioridazine, an antipsychotic drug, as an antiglioblastoma and anticancer stem cell agent using public gene expression data*. *Cell Death & Disease*, 2015. 6.
249. Liu, J., et al., *Treatment of obesity with celastrol*. *Cell*, 2015. 161(5): p. 999-1011.
250. Dudley, J.T., et al., *Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease*. *Sci Transl Med*, 2011. 3(96): p. 96ra76.
251. Zhang, S.D. and T.W. Gant, *A simple and robust method for connecting small-molecule drugs using gene-expression signatures*. *Bmc Bioinformatics*, 2008. 9.
252. Kettyl, L.M.J., F.G. Liberante, and A. Thompson, *Rational Drug Repurposing Using sscMap Analysis in a HOX-TALE Model of Leukemia*. *Hox Genes: Methods and Protocols*, 2014. 1196: p. 349-370.
253. Iorio, F., et al., *Discovery of drug mode of action and drug repositioning from transcriptional responses*. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. 107(33): p. 14621-14626.
254. Carrella, D., et al., *Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis*. *Bioinformatics*, 2014. 30(12): p. 1787-1788.
255. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. *Nucleic Acids Res*, 2002. 30(1): p. 207-10.
256. Leung, S.W., et al., *An Integrated Bioinformatics Analysis Repurposes an Antihelminthic Drug Niclosamide for Treating HMGA2-Overexpressing Human Colorectal Cancer*. *Cancers (Basel)*, 2019. 11(10).
257. Fagone, P., et al., *Identification of novel chemotherapeutic strategies for metastatic uveal melanoma*. *Sci Rep*, 2017. 7: p. 44564.
258. Er, J.L., et al., *Identification of inhibitors synergizing gemcitabine sensitivity in the squamous subtype of pancreatic ductal adenocarcinoma (PDAC)*. *Apoptosis*, 2018. 23(5-6): p. 343-355.
259. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biol*, 2014. 15(12): p. 550.
260. Iorio, F., et al., *Discovery of drug mode of action and drug repositioning from transcriptional responses*. *Proc Natl Acad Sci U S A*, 2010. 107(33): p. 14621-6.
261. Moir-Meyer, G., et al., *Robust CRISPR/Cas9 Genome Editing of the HUDEP-2 Erythroid Precursor Line Using Plasmids and Single-Stranded Oligonucleotide Donors*. *Methods Protoc*, 2018. 1(3).
262. Ianevski, A., et al., *SynergyFinder: a web application for analyzing drug combination dose-response matrix data*. *Bioinformatics*, 2020. 36(8): p. 2645.
263. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method*. *Methods*, 2001. 25(4): p. 402-8.
264. Morris, G.M. and M. Lim-Wilby, *Molecular docking*. *Methods Mol Biol*, 2008. 443: p. 365-82.
265. Hassane, D.C., et al., *Discovery of agents that eradicate leukemia stem cells using an in silico*

- screen of public gene expression data. *Blood*, 2008. 111(12): p. 5654-5662.
266. McArt, D.G. and S.D. Zhang, *Identification of Candidate Small-Molecule Therapeutics to Cancer by Gene-Signature Perturbation in Connectivity Mapping*. *Plos One*, 2011. 6(1).
 267. Liberante, F.G., et al., *Identification and validation of the dopamine agonist bromocriptine as a novel therapy for high-risk myelodysplastic syndromes and secondary acute myeloid leukemia*. *Oncotarget*, 2016. 7(6): p. 6609-19.
 268. Guzman, M.L., et al., *The sesquiterpene lactone parthenolide induces apoptosis of human acute myelogenous leukemia stem and progenitor cells*. *Blood*, 2005. 105(11): p. 4163-9.
 269. Pei, S., et al., *Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells*. *J Biol Chem*, 2013. 288(47): p. 33542-58.
 270. Xiong, X.X., et al., *Piperlongumine induces apoptotic and autophagic death of the primary myeloid leukemia cells from patients via activation of ROS-p38/JNK pathways*. *Acta Pharmacol Sin*, 2015. 36(3): p. 362-74.
 271. Cascao, R., J.E. Fonseca, and L.F. Moita, *Celastral: A Spectrum of Treatment Opportunities in Chronic Diseases*. *Front Med (Lausanne)*, 2017. 4: p. 69.
 272. Chen, X., et al., *Celastral induces ROS-mediated apoptosis via directly targeting peroxiredoxin-2 in gastric cancer cells*. *Theranostics*, 2020. 10(22): p. 10290-10308.
 273. Chiang, K.C., et al., *Celastral blocks interleukin-6 gene expression via downregulation of NF-kappaB in prostate carcinoma cells*. *PLoS One*, 2014. 9(3): p. e93151.
 274. Lee, J., et al., *Withaferin A is a leptin sensitizer with strong antidiabetic properties in mice*. *Nat Med*, 2016. 22(9): p. 1023-32.
 275. Li, H.Y., et al., *Celastral induces apoptosis and autophagy via the ROS/JNK signaling pathway in human osteosarcoma cells: an in vitro and in vivo study*. *Cell Death Dis*, 2015. 6: p. e1604.
 276. Chang, H.W., et al., *Withaferin A Induces Oxidative Stress-Mediated Apoptosis and DNA Damage in Oral Cancer Cells*. *Front Physiol*, 2017. 8: p. 634.
 277. Dubey, S., et al., *Withaferin A Associated Differential Regulation of Inflammatory Cytokines*. *Front Immunol*, 2018. 9: p. 195.
 278. Hassannia, B., et al., *Withaferin A: From ayurvedic folk medicine to preclinical anti-cancer drug*. *Biochem Pharmacol*, 2020. 173: p. 113602.
 279. Lee, J.H., et al., *Inhibition of NF-kappa B activation through targeting I kappa B kinase by celastral, a quinone methide triterpenoid*. *Biochem Pharmacol*, 2006. 72(10): p. 1311-21.
 280. Heyninck, K., et al., *Withaferin A inhibits NF-kappaB activation by targeting cysteine 179 in IKKbeta*. *Biochem Pharmacol*, 2014. 91(4): p. 501-9.
 281. Ni, H., et al., *NF-kappa B modulation is involved in celastral induced human multiple myeloma cell apoptosis*. *PLoS One*, 2014. 9(4): p. e95846.
 282. Wang, Z., Z. Zhai, and X. Du, *Celastral inhibits migration and invasion through blocking the NF-kappaB pathway in ovarian cancer cells*. *Exp Ther Med*, 2017. 14(1): p. 819-824.
 283. Grover, A., et al., *Inhibition of the NEMO/IKK beta association complex formation, a novel mechanism associated with the NF-kappa B activation suppression by Withania somnifera's key metabolite withaferin A*. *Bmc Genomics*, 2010. 11.
 284. Shrivastava, S., et al., *Piperlongumine, an alkaloid causes inhibition of PI3 K/Akt/mTOR signaling axis to induce caspase-dependent apoptosis in human triple-negative breast cancer*

- cells. *Apoptosis*, 2014. 19(7): p. 1148-64.
285. Wang, H., et al., *Piperlongumine increases sensitivity of colorectal cancer cells to radiation: Involvement of ROS production via dual inhibition of glutathione and thioredoxin systems*. *Cancer Lett*, 2019. 450: p. 42-52.
 286. Horton, R.J., *Albendazole in treatment of human cystic echinococcosis: 12 years of experience*. *Acta Trop*, 1997. 64(1-2): p. 79-93.
 287. Feldmeier, H., et al., *Flubendazole versus mebendazole in intestinal helminthic infections*. *Acta Trop*, 1982. 39(2): p. 185-9.
 288. Canova, K., et al., *Flubendazole induces mitotic catastrophe and apoptosis in melanoma cells*. *Toxicol In Vitro*, 2018. 46: p. 313-322.
 289. Kralova, V., et al., *Flubendazole induces mitotic catastrophe and senescence in colon cancer cells in vitro*. *J Pharm Pharmacol*, 2016. 68(2): p. 208-18.
 290. He, L., et al., *Mebendazole exhibits potent anti-leukemia activity on acute myeloid leukemia*. *Exp Cell Res*, 2018. 369(1): p. 61-68.
 291. Hou, Z.J., et al., *Flubendazole, FDA-approved anthelmintic, targets breast cancer stem-like cells*. *Oncotarget*, 2015. 6(8): p. 6326-40.
 292. Larsen, A.R., et al., *Repurposing the antihelmintic mebendazole as a hedgehog inhibitor*. *Mol Cancer Ther*, 2015. 14(1): p. 3-13.
 293. Florio, R., et al., *The Benzimidazole-Based Anthelmintic Parbendazole: A Repurposed Drug Candidate That Synergizes with Gemcitabine in Pancreatic Cancer*. *Cancers (Basel)*, 2019. 11(12).
 294. Lin, J., et al., *Aberrant methylation of DNA-damage-inducible transcript 3 promoter is a common event in patients with myelodysplastic syndrome*. *Leuk Res*, 2010. 34(8): p. 991-4.
 295. Yu, X., et al., *Celastrol Induces Cell Apoptosis and Inhibits the Expression of the AML1-ETO/C-KIT Oncoprotein in t(8;21) Leukemia*. *Molecules*, 2016. 21(5).
 296. Okamoto, S., et al., *Withaferin A suppresses the growth of myelodysplasia and leukemia cell lines by inhibiting cell cycle progression*. *Cancer Sci*, 2016. 107(9): p. 1302-14.
 297. Xu, L.N., et al., *Celastrol Inhibits the Growth of Ovarian Cancer Cells in vitro and in vivo*. *Front Oncol*, 2019. 9: p. 2.
 298. Li, X., et al., *Celastrol strongly inhibits proliferation, migration and cancer stem cell properties through suppression of Pin1 in ovarian cancer cells*. *Eur J Pharmacol*, 2019. 842: p. 146-156.
 299. Zhou, X., et al., *Flubendazole inhibits glioma proliferation by G2/M cell cycle arrest and pro-apoptosis*. *Cell Death Discov*, 2018. 4: p. 18.
 300. Tang, Q., et al., *Withaferin A triggers G2/M arrest and intrinsic apoptosis in glioblastoma cells via ATF4-ATF3-CHOP axis*. *Cell Prolif*, 2020. 53(1): p. e12706.
 301. Uttarkar, S., et al., *Targeting acute myeloid leukemia with a small molecule inhibitor of the Myb/p300 interaction*. *Blood*, 2016. 127(9): p. 1173-82.
 302. Falkenberg, K.D., et al., *Withaferin A, a natural compound with anti-tumor activity, is a potent inhibitor of transcription factor C/EBPbeta*. *Biochim Biophys Acta Mol Cell Res*, 2017. 1864(7): p. 1349-1358.
 303. Lee, I.C. and B.Y. Choi, *Withaferin-A--A Natural Anticancer Agent with Pleiotropic Mechanisms of Action*. *Int J Mol Sci*, 2016. 17(3): p. 290.

304. Duan, Q., Y. Liu, and S. Rockwell, *Fenbendazole as a potential anticancer drug*. Anticancer Res, 2013. 33(2): p. 355-62.
305. Gao, P., C.V. Dang, and J. Watson, *Unexpected antitumorigenic effect of fenbendazole when combined with supplementary vitamins*. J Am Assoc Lab Anim Sci, 2008. 47(6): p. 37-40.
306. Luis, T.C., et al., *Wnt signaling strength regulates normal hematopoiesis and its deregulation is involved in leukemia development*. Leukemia, 2012. 26(3): p. 414-21.
307. Garg, S., et al., *Differential antigen expression and aberrant signaling via PI3/AKT, MAP/ERK, JAK/STAT, and Wnt/beta catenin pathways in Lin-/CD38-/CD34+ cells in acute myeloid leukemia*. Eur J Haematol, 2016. 96(3): p. 309-17.
308. Lane, S.W., et al., *Differential niche and Wnt requirements during acute myeloid leukemia progression*. Blood, 2011. 118(10): p. 2849-56.
309. Staal, F.J., et al., *Aberrant Wnt Signaling in Leukemia*. Cancers (Basel), 2016. 8(9).
310. Stoddart, A., et al., *Inhibition of WNT signaling in the bone marrow niche prevents the development of MDS in the Apc(del/+) MDS mouse model*. Blood, 2017. 129(22): p. 2959-2970.
311. Wang, Y., et al., *The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML*. Science, 2010. 327(5973): p. 1650-3.
312. Tan, Z., L. Chen, and S. Zhang, *Comprehensive Modeling and Discovery of Mebendazole as a Novel TRAF2- and NCK-interacting Kinase Inhibitor*. Sci Rep, 2016. 6: p. 33534.
313. Kaileh, M., et al., *Withaferin A strongly elicits I kappa B kinase beta hyperphosphorylation concomitant with potent inhibition of its kinase activity*. Journal of Biological Chemistry, 2007. 282(7): p. 4253-4264.
314. Grivennikov, S.I. and M. Karin, *Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage*. Ann Rheum Dis, 2011. 70 Suppl 1: p. i104-8.
315. Karin, M., *NF-kappaB as a critical link between inflammation and cancer*. Cold Spring Harb Perspect Biol, 2009. 1(5): p. a000141.
316. Libermann, T.A. and D. Baltimore, *Activation of interleukin-6 gene expression through the NF-kappa B transcription factor*. Mol Cell Biol, 1990. 10(5): p. 2327-34.
317. McFarland, B.C., et al., *NF-kappaB-induced IL-6 ensures STAT3 activation and tumor aggressiveness in glioblastoma*. PLoS One, 2013. 8(11): p. e78728.
318. Taniguchi, K. and M. Karin, *IL-6 and related cytokines as the critical lynchpins between inflammation and cancer*. Semin Immunol, 2014. 26(1): p. 54-74.
319. Wang, Z., et al., *Crystal structures of IRAK-4 kinase in complex with inhibitors: a serine/threonine kinase with tyrosine as a gatekeeper*. Structure, 2006. 14(12): p. 1835-44.
320. Kesarwani, A.K., et al., *Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures*. Oncogene, 2017. 36(8): p. 1123-1133.
321. Fu, C.A., et al., *TNIK, a novel member of the germinal center kinase family that activates the c-Jun N-terminal kinase pathway and regulates the cytoskeleton*. J Biol Chem, 1999. 274(43): p. 30729-37.
322. Yokota, Y., et al., *Development of withaferin A analogs as probes of angiogenesis*. Bioorg Med Chem Lett, 2006. 16(10): p. 2603-7.
323. Fuska, J., et al., *Novel cytotoxic and antitumor agents. IV. Withaferin A: relation of its structure to the in vitro cytotoxic effects on P388 cells*. Neoplasma, 1984. 31(1): p. 31-6.

324. Oh, J.H., et al., *Withaferin A inhibits iNOS expression and nitric oxide production by Akt inactivation and down-regulating LPS-induced activity of NF-kappaB in RAW 264.7 cells*. Eur J Pharmacol, 2008. 599(1-3): p. 11-7.
325. Sreeramulu, S., et al., *Molecular mechanism of inhibition of the human protein complex Hsp90-Cdc37, a kinome chaperone-cochaperone, by triterpene celastrol*. Angew Chem Int Ed Engl, 2009. 48(32): p. 5853-5.
326. Trott, A., et al., *Activation of heat shock and antioxidant responses by the natural product celastrol: transcriptional signatures of a thiol-targeted molecule*. Mol Biol Cell, 2008. 19(3): p. 1104-12.
327. Bargagna-Mohan, P., et al., *The tumor inhibitor and antiangiogenic agent withaferin A targets the intermediate filament protein vimentin*. Chem Biol, 2007. 14(6): p. 623-34.
328. Peng, B., et al., *HSP90 inhibitor, celastrol, arrests human monocytic leukemia cell U937 at G0/G1 in thiol-containing agents reversible way*. Mol Cancer, 2010. 9: p. 79.
329. Bistrovic, A., et al., *Design, synthesis and biological evaluation of novel benzimidazole amidines as potent multi-target inhibitors for the treatment of non-small cell lung cancer*. Eur J Med Chem, 2018. 143: p. 1616-1634.
330. Wingelhofer, B., et al., *Implications of STAT3 and STAT5 signaling on gene regulation and chromatin remodeling in hematopoietic cancer*. Leukemia, 2018. 32(8): p. 1713-1726.
331. Qie, S. and J.A. Diehl, *Cyclin D1, cancer progression, and opportunities in cancer treatment*. J Mol Med (Berl), 2016. 94(12): p. 1313-1326.
332. Perini, G.F., et al., *BCL-2 as therapeutic target for hematological malignancies*. J Hematol Oncol, 2018. 11(1): p. 65.
333. Cory, S. and J.M. Adams, *The Bcl2 family: regulators of the cellular life-or-death switch*. Nat Rev Cancer, 2002. 2(9): p. 647-56.
334. Kalkat, M., et al., *MYC Deregulation in Primary Human Cancers*. Genes (Basel), 2017. 8(6).
335. Delgado, M.D. and J. Leon, *Myc roles in hematopoiesis and leukemia*. Genes Cancer, 2010. 1(6): p. 605-16.
336. Luo, H., et al., *c-Myc rapidly induces acute myeloid leukemia in mice without evidence of lymphoma-associated antiapoptotic mutations*. Blood, 2005. 106(7): p. 2452-61.
337. Zhou, J.D., et al., *BCL2 overexpression: clinical implication and biological insights in acute myeloid leukemia*. Diagn Pathol, 2019. 14(1): p. 68.
338. Teixeira Mendes, L.S., et al., *Cyclin D1 overexpression in proliferation centres of small lymphocytic lymphoma/chronic lymphocytic leukaemia*. J Clin Pathol, 2017. 70(10): p. 899-902.
339. Powers, J.P., et al., *Discovery and initial SAR of inhibitors of interleukin-1 receptor-associated kinase-4*. Bioorg Med Chem Lett, 2006. 16(11): p. 2842-5.
340. Wang, Z., et al., *Discovery of potent, selective, and orally bioavailable inhibitors of interleukin-1 receptor-associate kinase-4*. Bioorg Med Chem Lett, 2015. 25(23): p. 5546-50.
341. Ramsey, J.M., et al., *Entinostat prevents leukemia maintenance in a collaborating oncogene-dependent model of cytogenetically normal acute myeloid leukemia*. Stem Cells, 2013. 31(7): p. 1434-45.
342. Vanden Berghe, W., et al., *Molecular insight in the multifunctional activities of Withaferin A*. Biochem Pharmacol, 2012. 84(10): p. 1282-91.

343. Ng, S.W., et al., *Molecular modulators of celastrol as the keystones for its diverse pharmacological activities*. Biomed Pharmacother, 2019. 109: p. 1785-1792.
344. Kannaiyan, R., M.K. Shanmugam, and G. Sethi, *Molecular targets of celastrol derived from Thunder of God Vine: potential role in the treatment of inflammatory disorders and cancer*. Cancer Lett, 2011. 303(1): p. 9-20.
345. Oh, J.H. and T.K. Kwon, *Withaferin A inhibits tumor necrosis factor-alpha-induced expression of cell adhesion molecules by inactivation of Akt and NF-kappa B in human pulmonary epithelial cells*. International Immunopharmacology, 2009. 9(5): p. 614-619.
346. Ichikawa, H., et al., *Withanolides potentiate apoptosis, inhibit invasion, and abolish osteoclastogenesis through suppression of nuclear factor-kappa B (NF-kappa B) activation and NF-kappa B-regulated gene expression*. Molecular Cancer Therapeutics, 2006. 5(6): p. 1434-1445.
347. DiDonato, J.A., F. Mercurio, and M. Karin, *NF-kappaB and the link between inflammation and cancer*. Immunol Rev, 2012. 246(1): p. 379-400.
348. Liu, T., et al., *NF-kappaB signaling in inflammation*. Signal Transduct Target Ther, 2017. 2.
349. Taniguchi, K. and M. Karin, *NF-kappaB, inflammation, immunity and cancer: coming of age*. Nat Rev Immunol, 2018. 18(5): p. 309-324.
350. Pikarsky, E., et al., *NF-kappaB functions as a tumour promoter in inflammation-associated cancer*. Nature, 2004. 431(7007): p. 461-6.
351. Dubey, S., et al., *Withaferin a associated Differential regulation of inflammatory cytokines*. Frontiers in Immunology, 2018. 9.
352. Martorana, F., et al., *Withaferin A Inhibits Nuclear Factor-kappaB-Dependent Pro-Inflammatory and Stress Response Pathways in the Astrocytes*. Neural Plast, 2015. 2015: p. 381964.
353. Sethi, G., et al., *Celastrol, a novel triterpene, potentiates TNF-induced apoptosis and suppresses invasion of tumor cells by inhibiting NF-kappaB-regulated gene products and TAK1-mediated NF-kappaB activation*. Blood, 2007. 109(7): p. 2727-35.
354. Samadi, A.K., et al., *Withaferin A, a cytotoxic steroid from Vassobia breviflora, induces apoptosis in human head and neck squamous cell carcinoma*. J Nat Prod, 2010. 73(9): p. 1476-81.
355. Zhang, H.L. and H. Zhang, *Withaferin-A Induces Apoptosis in Osteosarcoma U2OS Cell Line via Generation of ROS and Disruption of Mitochondrial Membrane Potential*. Pharmacogn Mag, 2017. 13(51): p. 523-527.
356. Xia, S., Y. Miao, and S. Liu, *Withaferin A induces apoptosis by ROS-dependent mitochondrial dysfunction in human colorectal cancer cells*. Biochem Biophys Res Commun, 2018. 503(4): p. 2363-2369.
357. Li, X., et al., *Synergistic antitumor activity of withaferin A combined with oxaliplatin triggers reactive oxygen species-mediated inactivation of the PI3K/AKT pathway in human pancreatic cancer cells*. Cancer Lett, 2015. 357(1): p. 219-230.
358. Ren, B., et al., *Celastrol induces apoptosis in hepatocellular carcinoma cells via targeting ER-stress/UPR*. Oncotarget, 2017. 8(54): p. 93039-93050.
359. Nishikawa, Y., et al., *Withaferin A Induces Cell Death Selectively in Androgen-Independent*

- Prostate Cancer Cells but Not in Normal Fibroblast Cells*. PLoS One, 2015. 10(7): p. e0134137.
360. Yu, Y., et al., *Withaferin A targets heat shock protein 90 in pancreatic cancer cells*. Biochem Pharmacol, 2010. 79(4): p. 542-51.
 361. Choi, M.J., et al., *Endoplasmic reticulum stress mediates withaferin A-induced apoptosis in human renal carcinoma cells*. Toxicol In Vitro, 2011. 25(3): p. 692-8.
 362. Dai, Y., et al., *The NF (Nuclear factor)-kappaB inhibitor parthenolide interacts with histone deacetylase inhibitors to induce MKK7/JNK1-dependent apoptosis in human acute myeloid leukaemia cells*. Br J Haematol, 2010. 151(1): p. 70-83.
 363. Herrera, F., et al., *Intracellular redox state regulation by parthenolide*. Biochem Biophys Res Commun, 2005. 332(2): p. 321-5.
 364. Kwok, B.H., et al., *The anti-inflammatory natural product parthenolide from the medicinal herb Feverfew directly binds to and inhibits IkappaB kinase*. Chem Biol, 2001. 8(8): p. 759-66.
 365. Steele, A.J., et al., *The sesquiterpene lactone parthenolide induces selective apoptosis of B-chronic lymphocytic leukemia cells in vitro*. Leukemia, 2006. 20(6): p. 1073-9.
 366. Zhu, Y., et al., *Celastral Suppresses Glioma Vasculogenic Mimicry Formation and Angiogenesis by Blocking the PI3K/Akt/mTOR Signaling Pathway*. Front Pharmacol, 2020. 11: p. 25.
 367. Feng, L., et al., *ER stress-mediated apoptosis induced by celastral in cancer cells and important role of glycogen synthase kinase-3beta in the signal network*. Cell Death Dis, 2013. 4: p. e715.
 368. Fribley, A.M., et al., *Celastral induces unfolded protein response-dependent cell death in head and neck cancer*. Exp Cell Res, 2015. 330(2): p. 412-422.
 369. Nagase, M., et al., *Apoptosis induction in HL-60 cells and inhibition of topoisomerase II by triterpene celastral*. Biosci Biotechnol Biochem, 2003. 67(9): p. 1883-7.
 370. Thaiparambil, J.T., et al., *Withaferin A inhibits breast cancer invasion and metastasis at sub-cytotoxic doses by inducing vimentin disassembly and serine 56 phosphorylation*. Int J Cancer, 2011. 129(11): p. 2744-55.
 371. Um, H.J., et al., *Withaferin A inhibits JAK/STAT3 signaling and induces apoptosis of human renal carcinoma Caki cells*. Biochem Biophys Res Commun, 2012. 427(1): p. 24-9.
 372. Suman, S., et al., *Withaferin-A suppress AKT induced tumor growth in colorectal cancer cells*. Oncotarget, 2016. 7(12): p. 13854-64.
 373. Wang, Y., et al., *Piperlongumine Suppresses Growth and Sensitizes Pancreatic Tumors to Gemcitabine in a Xenograft Mouse Model by Modulating the NF-kappa B Pathway*. Cancer Prev Res (Phila), 2016. 9(3): p. 234-44.
 374. Zhang, Q., et al., *Piperlongumine, a Novel TrxR1 Inhibitor, Induces Apoptosis in Hepatocellular Carcinoma Cells by ROS-Mediated ER Stress*. Front Pharmacol, 2019. 10: p. 1180.
 375. Bharadwaj, U., et al., *Drug-repositioning screening identified piperlongumine as a direct STAT3 inhibitor with potent activity against breast cancer*. Oncogene, 2015. 34(11): p. 1341-53.
 376. Makhov, P., et al., *Piperlongumine promotes autophagy via inhibition of Akt/mTOR signalling and mediates cancer cell death*. Br J Cancer, 2014. 110(4): p. 899-907.
 377. Fonrose, X., et al., *Parthenolide inhibits tubulin carboxypeptidase activity*. Cancer Res, 2007. 67(7): p. 3371-8.
 378. Carlisi, D., et al., *Parthenolide sensitizes hepatocellular carcinoma cells to TRAIL by inducing the expression of death receptors through inhibition of STAT3 activation*. J Cell Physiol, 2011.

- 226(6): p. 1632-41.
379. Castro, L.S., et al., *Albendazole as a promising molecule for tumor control*. Redox Biol, 2016. 10: p. 90-99.
 380. Walf-Vorderwulbecke, V., et al., *Targeting acute myeloid leukemia by drug-induced c-MYB degradation*. Leukemia, 2018. 32(4): p. 882-889.
 381. Spagnuolo, P.A., et al., *The antihelmintic flubendazole inhibits microtubule function through a mechanism distinct from Vinca alkaloids and displays preclinical activity in leukemia and myeloma*. Blood, 2010. 115(23): p. 4824-33.
 382. Pellagatti, A., et al., *Association Between Gene Expression Profiles and Commonly Mutated Genes In The Hematopoietic Stem Cells Of Patients With Myelodysplastic Syndromes*. Blood, 2013. 122(21).
 383. Couturier, C.P., et al., *Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy*. Nat Commun, 2020. 11(1): p. 3406.
 384. Katzenelenbogen, Y., et al., *Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer*. Cell, 2020. 182(4): p. 872-885 e19.
 385. Lang, C., et al., *Single-Cell Sequencing of iPSC-Dopamine Neurons Reconstructs Disease Progression and Identifies HDAC4 as a Regulator of Parkinson Cell Phenotypes*. Cell Stem Cell, 2019. 24(1): p. 93-106 e6.
 386. Psaila, B., et al., *Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets*. Mol Cell, 2020. 78(3): p. 477-492 e8.
 387. Zhang, L., et al., *Single-Cell Transcriptomics in Medulloblastoma Reveals Tumor-Initiating Progenitors and Oncogenic Cascades during Tumorigenesis and Relapse*. Cancer Cell, 2019. 36(3): p. 302-318 e7.
 388. Chaligne, R., A.S. Nam, and D.A. Landau, *TARGET-seq Takes Aim at Cancer Evolution through Multi-omics Single-Cell Genotyping and Transcriptomics*. Mol Cell, 2019. 73(6): p. 1092-1094.
 389. Rodriguez-Meira, A., et al., *TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing*. STAR Protoc, 2020. 1(3): p. 100125.