Global analysis of alternative splicing in human cells of myeloid origin

Lorenzini Paolo Alberto

2018

Lorenzini Paolo Alberto. (2018). Global analysis of alternative splicing in human cells of myeloid origin. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/74234>

<https://doi.org/10.32657/10356/74234>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**GLOBAL ANALYSIS OF ALTERNATIVE SPLICING
IN HUMAN CELLS OF MYELOID ORIGIN**

LORENZINI PAOLO ALBERTO
Interdisciplinary Graduate School
HealthTech

2018

GLOBAL ANALYSIS OF ALTERNATIVE SPLICING IN HUMAN CELLS OF MYELOID ORIGIN

LORENZINI PAOLO ALBERTO

**Interdisciplinary Graduate School
HealthTech**

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2018

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my main supervisor, Asst. Professor Xavier Roca, for his guidance, patience, enthusiasm and immense knowledge which gave me an invaluable help in the understanding of not only this project but also in general notions in molecular biology. Besides my main supervisor, I am particularly thankful to my co-supervisor Asst. Professor Zheng Jie for his precious suggestions regarding the bioinformatics aspects of this project. Indeed, I very grateful to him and his team for performing the maintenance of the school's server allowing us to run all the RNA-seq analysis very smoothly.

I am also profoundly grateful to Dr. Hongfei Liu for her fundamental role in setting up the bioinformatics pipeline and contributions in many of the experiments and for her role in RNA-seq data analysis. I would like to thank also Dr. Zhang Fan, which when we started this project was a Ph.D. student in my co-supervisor's lab and who helped us in setting up the RNA-seq data analysis.

I would like to express my gratitude also to my mentor, Dr. Ong Sin Tiong, for his precious comments during the TAC meetings and to Dr. Stefan Mockenhaupt for kindly providing to our lab some of the vectors and for his suggestions in designing the CRISPR experiments.

Indeed, I am extremely thankful to Dr. Daniela Moses and her very skilled team for the sequencing service provided to us. I am very thankful also to Dr. Fursham Hamid who gave me very insightful advice for the bioinformatics analysis during the second part of this Ph.D. and to Ms. Resilind Chew who was a very proficient FYP student.

I would like to thank also all my other lab mates, Ms. Luo Shufang, Ms. Ng Cher Belle, Ms. Justine Gabrielle Aczon Dacanay, Ms. Cheryl Tan, Dr. Tan Jiazi, Ms. Ly Phuong Thao, Ms. Ho Jia Xin Jessie, Dr. Bhadra Malini, Ms. Tang Sze Jing, Dr. Liu Jun and Dr. Shaohi Xu.

Lastly but not least, I am indeed extremely grateful to my family which have been very supportive in all these years that I have been in Singapore.

ABBREVIATIONS

3'ss	3' splice site
5'ss	5' splice site
A3SS	Alternative 3'ss
A5SS	Alternative 5'ss
ABCB7	ATP-binding cassette sub-family B member 7
AFMID	Arylformamidase
AKAP17A	A-Kinase Anchoring Protein 17A
AKT1	AKT serine/threonine kinase 1
ALS	Amyotrophic Lateral Sclerosis
AML	Acute myeloid leukemia
ANGPTL4	Angiopoietin-like protein 4
APC	Allophycocyanin
APCs	Antigen presenting cells
ApE	A plasmid editor
APP	Amyloid precursor protein
AS	Alternative splicing
ATCC	American type culture collection
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine triphosphate
AUF1	AU-rich element RNA-binding protein 1
Bcl-2	B cell leukemia/lymphoma 2
Bcl-xL	B-cell lymphoma-extra large
BCLAF1	Bcl-2-associated transcription factor 1
BD	Becton Dickinson
BFmedian	Bayes factor median
BFmin	Bayes factor minimum
BMDMs	Bone marrow-derived macrophages
BNIP3	BCL2 Interacting Protein 3
BNIP3L	BCL2 Interacting Protein 3 Like

BPS	Branch point sequence
BrdU	Bromodeoxyuridine
Bru-seq	Bromouridine sequencing
BSA	Bovine serum albumin
CA9	Carbonic anhydrase 9
CA12	Carbonic anhydrase 12
CCL1	Chemokine (C-C motif) ligand 1
CCL2	Chemokine (C-C motif) ligand 2
CCL7	Chemokine (C-C motif) ligand 7
CCL8	Chemokine (C-C motif) ligand 8
CCL22	Chemokine (C-C motif) ligand 22
CCL24	Chemokine (C-C motif) ligand 24
CCL25	Chemokine (C-C motif) ligand 25
CCL26	Chemokine (C-C motif) ligand 26
CCR	C-C chemokine receptor type 2
CD3	Cluster of differentiation 3
CD8	Cluster of differentiation 8
CD11b	Cluster of differentiation 11b
CD14	Cluster of differentiation 14
CD16	Cluster of differentiation 16
CD28	Cluster of differentiation 28
CD33	Cluster of differentiation 33
CD36	Cluster of differentiation 36
CD44	Cluster of differentiation 44
CD45	Cluster of differentiation 45
CD54	Cluster of differentiation 54
CD68	Cluster of differentiation 68
CD71	Cluster of differentiation 71
CD80	Cluster of differentiation 80
CD86	Cluster of differentiation 86
CD163	Cluster of differentiation 163

CD206	Cluster of differentiation 206
CD209	Cluster of differentiation 209
CD235a	Cluster of differentiation 235 alpha
c-Maf	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog
cDNA	Complementary DNA
CDKI	Cyclin-dependent kinase inhibitor
CELF2	CUGBP elav-like family member 2
CLL	Chronic lymphoid leukemia
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukemia
CMP	Common myeloid progenitor
CO ₂	Carbon dioxide
COL1A2	Collagen type I alpha 2 chain
co-IP	Co-Immunoprecipitation
CPT	Camptothecin
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CTD	Carboxy terminal domain
CTL-4	Cytotoxic T-lymphocyte associated protein 4
CTNND1	Catenin delta 1
CX3CL1	Chemokine (C-X3-C motif) ligand 1
CX3CR1	Chemokine (C-X3-C motif) receptor 1
CXCL2	Chemokine (C-X-C motif) ligand 2
CXCL9	Chemokine (C-X-C motif) ligand 9
CXCL10	Chemokine (C-X-C motif) ligand 10
CXCL11	Chemokine (C-X-C motif) ligand 11
CXCL13	Chemokine (C-X-C motif) ligand 13
CXCL16	Chemokine (C-X-C motif) ligand 16
CXCR3	C-X-C chemokine receptor type 3
CXCR4	C-X-C chemokine receptor type 4
DAG	Diacylglycerol

DCs	Dendritic cells
DDR	DNA damage response
DEGs	Differentially expressed genes
del(5q)	Deletion 5q
DMEM	Dulbecco's modified eagle's medium
DMSO	Dymethyl sulphoxide
DNA	Deoxyribonucleic acid
DNM1	Dynamin 1
DNM2	Dynamin 2
DNMT3B	DNA Methyltransferase 3 Beta
DREME	Discriminative Regular Expression Motif Elicitation
DRB	5,6-dichloro-1- β -Dribofuranosylbenzimidazole
DSBs	Double strand breaks
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGR1	Early growth response 1
EHMT2	Hystone-lysine N-methyltransferase 2
EMT	Epithelial-to-mesenchymal transition
ENO2	Enolase 2
EPB41	Erythrocyte Membrane Protein Band 4.1
EPO	Erythropoietin
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence activated cell sorting
FAM126	Family With Sequence Similarity 126
FBS	Fetal bovine serum
Fc	Fragment crystallizable
Fc γ	Fragment crystallizable gamma

FDR	False discovery rate
FDA	Food and Drug Administration
FITC	Fluorescein isothiocyanate
FMNL3	Formin Like 3
FN1	Fibronectin-1
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FSC	Forward scatter
FYP	Final year project
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GATA3	GATA binding protein 3
GFP	Green fluorescent protein
GM-CSF	Granulocyte macrophage colony stimulating factor
GMP	Granulocyte macrophage progenitor
GO	Gene ontology
GSEA	Gene Set Enrichment Analysis
GTFs	General transcription factors
H3K36me3	Histone H3 tri methyl K36
Hb	Hemoglobin
HBB	Hemoglobin beta
HeLa	Henrietta Lacks
HIF1A	Hypoxia-inducible factor 1-alpha
HIF2A	Hypoxia-inducible factor 2-alpha
HIF3A	Hypoxia-inducible factor 3-alpha
HLA-DR	Human Leukocyte Antigen – antigen D Related
hnRNP	Heterogeneous nuclear ribonucleoprotein
hnRNPL	Heterogeneous nuclear ribonucleoprotein L
HOX	Homeobox
HREs	Hypoxia-response elements
HRP	Horse radish peroxidase
HSCs	Hematopoietic stem cells
ICAM1	Intercellular adhesion molecule 1

ICSBP	Interferon consensus sequence binding protein
IFN- γ	Interferon gamma
IgG	Immunoglobulin G
IgG1 κ	Immunoglobulin 1 kappa
IgG2a	Immunoglobulin 2a
IgG2b	Immunoglobulin 2b
IgM	Immunoglobulin M
IGV	Integrative genomics viewer
IL-1 β	Interleukin 1 beta
IL-4	Interleukin 4
IL-7R α	Interleukin 7 receptor alpha
IL-10	Interleukin 10
IL-13	Interleukin 13
IL-32 γ	Interleukin 32 gamma
INDELs	Insertions and deletions
IP6K2	Inositol Hexakisphosphate Kinase 2
IPSDMs	Induced pluripotent stem cells
IRF-8	Interferon regulatory factor 8
IR	Intron retention
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
KLF1	Kruppel-like factor 1
KLF4	Kruppel-like factor 4
KO	Knock-out
LARP7	La Ribonucleoprotein Domain Family Member 7
LB	Lysogenic broth
LC-MS	Liquid chromatography–mass spectrometry
LCSs	Leukemic stem cells
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LEF1	Lymphoid enhancer binding factor 1

lncRNA	Long non coding RNA
LPS	Lipopolysaccharide
LSK	Lymphocyte specific protein tyrosine kinase
LT-HSC	Long term hematopoietic stem cell
LUC7L2	Luc7-like 2
M-CSF	Macrophage colony stimulating factor
M-MuLV	Moloney Murine Leukemia Virus
MafB	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog B
MAPK	Mitogen-activated protein kinase
MAP2K7	Mitogen-activated protein kinase kinase 7
MAP3K7	Mitogen-activated protein kinase kinase kinase 7
MAP4K7	Mitogen-activated protein kinase kinase kinase kinase 7
MATS	Multivariate analysis of transcript splicing
MAXENT	Maximum Entropy Modeling
MBNL1	Muscleblind Like Splicing Regulator 1
MDS	Myelodysplastic syndrome
MEME	Multiple EM for Motif Elicitation
MEP	Megakaryocyte erythroid progenitor
MFI	Minimum fluorescence intensity
MgCl ₂	Magnesium chloride
MHC	Major histocompatibility complex
miR210	Micro-RNA 210
MISO	Mixture of isoforms
MMP9	Matrix metalloproteinase 9
MMP	Multipotent progenitor
MMPs	Metalloproteinases
MRPL55	Mitochondrial Ribosomal Protein L55
Myb	v-myb avian myeloblastosis viral oncogene homolog
MyD88	Myeloid differentiation primary response 88
MSCs	Mesenchymal stem cells
MXE	Mutually exclusive exons

mTOR	Mammalian target of rapamycin
NaCl	Sodium chloride
NEB	New England Biolabs
NF- κ B	Nuclear factor kappaB
NGS	Next generation sequencing
NHEJ	Nonhomologous end-joining
NMD	Non sense mediated decay
NPCs	Neuronal precursor cells
NUS	National University of Singapore
O ₂	Oxygen
Opti-MEM	Optimum Minimal Essential Medium
ORF	Open reading frame
p38-MAPK	p38 mitogen-activated protein kinase
PAM	Protospacer adjacent motif
PABPs	Poly(A)-binding proteins
PARP	Poly ADP ribose polymerase
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCD	Programmed cell death
PCR	Polymerase chain reaction
PE	Phycoerythrin
PECAM1	Platelet endothelial cell adhesion molecule 1
Pen/Strep	Penicillin/Streptomycin mixture
PerCP-Cy5.5	Peridinin Chlorophyll-Cy5.5
PGK1	Phosphoglycerate Kinase 1
Ph.D.	Doctor of philosophy
PHD2	Prolyl hydroxylase domain-containing 2
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PKC	Protein kinase C
PMA	Phorbol myristate acetate

Pol II	Polymerase II (RNA)
PPI	Protein-protein interactions
PPT	Polypyrimidine tract
PRPF31	Pre-mRNA processing factor 31
PRPF40A	Pre-mRNA processing factor 40 homolog A
PRPF40B	Pre-mRNA processing factor 40 homolog B
PRPF8	Pre-mRNA processing factor 8
Pre-mRNA	Precursor messenger RNA
PSI	Percentage Splice In
PTBP1	Polypyrimidine tract binding protein 1
PTBP2	Polypyrimidine tract binding protein 2
PTC	Premature termination codon
PTMs	Post-translational modifications
QC	Quality control
QKI	Quaking
qRT-PCR	Quantitative real time polymerase chain reaction
RAI14	Retinoic Acid Induced 14
RANBP9	Ran-binding protein 9
RARS	Refractory anemia with ring sideroblasts
RAVD	Retinoic acid-vitamin D3
RBP	RNA binding proteins
RCOR3	REST Corepressor 3
RC	Reverse complement
RIN	RNA integrity number
RIPK2	Receptor Interacting Serine/Threonine Kinase 2
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROS	Reactive Oxygen Species
RPKM	Reads per kilobase per million mapped reads
RPMI	Roswell Park Memorial Institute
RPS14	Ribosomal Protein S14

RRM	RNA recognition motif
RS	Ring sideroblasts
RT-PCR	Reverse transcription PCR
RUNX1	Runt-related transcription factor 1
SAM68	Src-associated substrate in mitosis of 68kDa
SAP25	Sin3A Associated Protein 25
SBS	School of biological sciences
SCE	School of computer engineering
SDS	Sodium dodecyl sulfate
SCELSE	Singapore Centre on Environmental Life Sciences Engineering
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SE	Skipped exon
SF1	Splicing factor 1
SF3A1	Splicing factor 3a, subunit 1
SF3B1	Splicing factor 3b, subunit 1
sgRNA	Single guide RNA
shRNA	Small hairpin RNA
siRNA	Small interfering RNA
SILAC	Stable isotope labeling with amino acids in cell culture
SMA	Spinal muscular atrophy
SMN1	Survival of motor neuron 1
SMN2	Survival of motor neuron 2
SNPs	Single nucleotide polymorphisms
snRNAs	Small nuclear RNAs
snRNPs	Small nuclear ribonucleoproteins
snRNP70	Small nuclear ribonucleoprotein 70 kDa
SNVs	Single-nucleotide variations
SPTAN1	Spectrin Alpha, Non-Erythrocytic 1
SRSF2	Ser/Arg-rich splicing factor 2
SRSF3	Ser/Arg-rich splicing factor 3
sRANKL	Soluble receptor activator of nuclear factor kappaB ligand

SSC	Side scatter
STAT2	Signal Transducer And Activator Of Transcription 2
ST-HSC	Short term hematopoietic stem cell
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SUGP2	SURP And G-Patch Domain Containing 2
SYK	Spleen tyrosine kinase
Th1	T helper type 1
Th2	T helper type 2
T-ALL	T cell acute lymphoblastic leukemia
T4 PNK	T4 Polynucleotide Kinase
TBST	Tris-buffered saline with Tween 20
TCR	T cell receptor
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TGF- β	Transforming growth factor beta
TGF- β 3	Transforming growth factor beta 3
TLR4	Toll-like receptor 4
TNF- α	Tumor necrosis factor alpha
TRIS	Tris(hydroxymethyl)aminomethane
U2AF	U2 auxiliary factor
U2AF35	U2 auxiliary factor subunit 35
U2AF65	U2 auxiliary factor subunit 65
UCSC	University of California, Santa Cruz
UMV	Universal Minigene Vector
USA	United States of America
WB	Western blot
WHO	World Health Organization
WT	Wild-type
VD3	Vitamin D3
VEGF	Vascular endothelial growth factor
VHL	Von Hippel Lindau

Zfp326	Zinc finger protein 326
ZNF76	Zinc finger protein 76
ZNF678	Zinc finger protein 678
ZRSR2	Zinc finger RNA binding motif and Ser/Arg rich 2

LIST OF CONTENTS

1 INTRODUCTION	1
1.1 Constitutive and alternative pre-mRNA splicing	1
1.2 The major and the minor spliceosome	5
1.3 Intron and exon definition models	9
1.4 Modes of alternative splicing	10
1.5 Human hematopoiesis	14
1.6 Alternative splicing during normal hematopoiesis and in normal development.....	16
1.7 Alternative splicing in the immune system: B cells, T cells and DCs	18
1.8 Human Monocytes	19
1.9 Splicing in human diseases	22
1.10 MDS pathogenesis.....	24
1.11 MDS and the dysregulated spliceosome	28
1.12 PRPF40B and other proteins containing the WW and FF domains	31
1.13 Tumor Hypoxia	33
1.14 Hypoxia in normal hematopoiesis and in leukemia/MDS	37
1.15 Context of this project	38
1.16 Aims	39
2 MATERIALS AND METHODS	40
2.1 Cell culture, differentiation and polarization	40
2.2 Isolation of primary human macrophages and relative differentiation/polarization methods	40
2.3 Flow cytometry for assessment of macrophage differentiation (cell line and primary cells)	41

2.4 Live cell imaging.....	43
2.5 Total RNA extraction and sample preparation for RNA sequencing (RNA-seq).....	43
2.6 Strand-specific RNA-seq library preparation (performed by SCELSE) and RNA-seq	44
2.7 RNA-seq data analysis	44
2.8 cDNA synthesis and RT-PCR.....	45
2.9 Semi-quantitative (Semi-Q) RT-PCR	46
2.10 Protocol for CRISPR Knockout (KO) of human <i>PRPF40B</i> in K562 cell line.....	47
2.10.1 Cell Culture	47
2.10.2 CRISPR single guide RNA vector design	47
2.10.3 Nucleofection	50
2.10.4 Fluorescence-activated cell sorting (FACS)	50
2.10.5 Single cell clone selection	52
2.10.6 Genomic DNA extraction and genotyping CRISPR-edited <i>PRPF40B</i>	52
2.10.7 CRISPR qRT-PCR	54
.....	55
2.11 PCR mutagenesis of <i>PRPF40B</i>	55
2.12 Transfection with overexpression plasmids (Rescue experiments)	59
2.13 Total RNA extraction, sample preparation for RNA sequencing and RNA-seq data analysis.....	60
2.14 Additional Bioinformatics analysis:.....	61
2.15 Primers for the validation of the <i>PRPF40B</i> regulated AS events	62
2.16 qRT-PCR <i>PRPF40B</i> regulated DEGs validation	63
2.17 Splicing minigenes	64
2.18 Bacterial transformation and plasmid extraction.....	66
2.19 Inhibition of Pol II mRNA transcription	67

2.20 Western Blotting	67
2.21 Sample preparation for LC-MS/MS (Shotgun) and MS analysis.....	68
2.23 Measurement of Apoptosis, DNA damage and proliferation.....	68
3 RESULTS.....	69
3.1 Cell surface markers for macrophage differentiation and polarization of THP-1.....	69
3.1.1 Cell surface markers for THP-1 macrophage-like cell differentiation with PMA and VD3	69
3.1.2 Cell surface markers for polarization of M1-like and M2-like macrophages (cell line) 71	
3.2 Cell surface markers for macrophage differentiation and polarization of primary human blood monocytes	72
3.2.1 Cell surface markers for macrophage differentiation of primary human blood monocytes	74
3.2.2 Cell surface markers for macrophage polarization of primary human blood monocytes	76
3.3 Live cell images for macrophage differentiation and polarization	76
3.3.1 Live cell images for the differentiation and polarization of THP-1 cells	77
3.3.2 Live cell images for the differentiation and polarization of primary cells	79
3.4 Total RNA quality control before RNA-seq	80
3.5 RNA-seq data analysis results.....	80
3.6 Analysis of the DEG in cell line and primary cells.....	82
3.7 DEGs analysis of chemokines and cytokines in cell line and primary cells	83
3.8 Principal Component Analysis (PCA) of the DEGs in the cell line an primary cells.....	85
3.9 GO analysis of the DEGs in the cell line model an in primary cells.....	86
3.10 Identification of novel AS events that change upon differentiation and polarization of the cell line and primary cells	89
3.11 Validation of the novel AS events identified by RNA-seq	91

3.12 Knock out of PRPF40B by CRISPR	95
3.13 WB and qRT-PCR results of PRPF40B KO clones	98
3.14 Rescue experiments	98
3.15 RNA-seq analysis of the PRPF40KO and of the overexpression samples	102
3.16 Analysis of the DEG upon KO and overexpression	104
3.18 Validation of the DEG regulated by PRPF40B	109
3.19 Analysis of the expression of PRPF40B and its homologue PRPF40A in various human cancers	110
3.20 Assessment of PRPF40A protein level in the KO cells (WB)	114
3.21 Analysis of the AS events upon KO and overexpression	115
3.22 Validation of the AS events regulated by PRPF40B	118
3.23 Features of the AS events regulated by PRPF40B	125
3.24 Motifs analysis of the PRPF40B regulated cassette exon events	129
3.25 Splicing minigenes	130
3.25.1 UMV minigenes	130
3.25.2 pCDNA3.1+ minigenes	131
3.26 RNA Pol II inhibition	132
3.27 Assessment of Apoptosis, DNA damage and proliferation in PRPF40B KO cells	135
4 DISCUSSION AND FUTURE DIRECTIONS	136
4.1 Comparison between differentiation protocols of the cell line model and primary cells..	136
4.2 Analysis of the RNA-seq data of the differentiation of monocytes	141
4.3 PRPF40B CRISPR KO	144
4.4 Rescue experiments in PRPF40B KO cells	146
4.5 RNA-seq data analysis of the PRPF40B KO and its MDS mutations	147

4.6 Role of PRPF40B in regulating hypoxia and cellular differentiation	147
4.7 PRPF40A and PRPF40B roles in hypoxia and connection with AML drug therapy	149
4.8 Regulation of AS by PRPF40B	150
4.9 Role of PRPF40B in co-transcriptional splicing	153
5 CONCLUSION	156
REFERENCES	157

LIST OF FIGURES

Figure 1: Mechanism of pre-mRNA splicing.	2
Figure 2: AS and proteome complexity.	5
Figure 3: Schematic representation of minor and major spliceosome assembly.	7
Figure 4: Schematic representation of <i>cis</i> -acting elements and <i>trans</i> -acting splicing factors.	9
Figure 5: Exon and Intron definition models.	10
Figure 6: Patterns of AS.	11
Figure 7: Co-transcriptional splicing.	13
Figure 8: Human hematopoiesis.	16
Figure 9: Representation of macrophage heterogeneity.	21
Figure 10: Landscape of MDS mutations.	25
Figure 11: Mutated splicing factors in MDS and other hematological diseases.	31
Figure 12: Proteins involved in transcription and splicing.	32
Figure 13: PRPF40B protein structure.	33
Figure 14: Hypoxia pathways.	36
Figure 15: Bioinformatics pipeline.	45
Figure 16: CRISPR PRPF40B KO design.	48
Figure 17: CRISPR Workflow.	51

Figure 18: Genotyping PCRs.....	52
Figure 19: PRPF40B UCSC genome browser isoforms.....	56
Figure 20: PRPF40B vectors sequencing primers design.....	58
Figure 21: Minigenes design. Schematic describing the cloning of the <i>RIPK2</i> and <i>SPTAN1</i> alternative middle exons into the UMV vector (A) or into the pcDNA3.1+ vector (B).....	66
Figure 22: Flow cytometry results for differentiation/polarization of THP-1 cells.....	72
Figure 23: Procedure for isolation of human blood monocytes.....	73
Figure 24: Results for the isolation of primary cells.....	74
Figure 25: Flow cytometry results for differentiation/polarization of human monocytes.....	75
Figure 26: THP-1 live cell imaging.	78
Figure 27: Primary cells live cell imaging.....	79
Figure 28: Alignment results of the RNA-seq of cell line and primary cells.	81
Figure 29: Results of DEGs in cell line and primary cells.	82
Figure 30: Results of DEGs in common between cell line model and primary cells.	83
Figure 31: Cytokines/Chemokines DEG in cell line and primary cells.....	84
Figure 32: PCA in cell line and primary cells.....	85
Figure 33: GO Network of the DEGs during differentiation of THP-1 cells.	86
Figure 34: GO Reactome results.....	88
Figure 35: Changes in AS in primary cells.....	89
Figure 36: Changes in AS in THP-1.....	90
Figure 37: Common changes in AS between THP-1 and primary cells.	90
Figure 38: Validation of AS in cell line.....	92
Figure 39: GO Network of common changes in AS between THP-1 and primary cells.....	93
Figure 40: Results of the GO analysis by DAVID.	94
Figure 41: Results of the genotyping PCRs.....	96

Figure 42: Sequencing results of the PRPF40B KO clones.....	97
Figure 43: Results of WB and qRT-PCR of the PRPF40B KO.....	98
Figure 44: Sequencing results of PCR mutagenesis.	99
Figure 45: Rescue experiments results.	101
Figure 46: Alignment results of the RNA-seq of the PRPF40B KO and MDS mutations.....	103
Figure 47: Results of correlation between replicates.....	104
Figure 48: DEGs results of the PRPF40 KO and rescue experiments.....	105
Figure 49: Results of GO analysis of the DEGs regulated by PRPF40B	106
Figure 50: Heat map of DEGs in hypoxia.	107
Figure 51: DEGs with role in development and erythroid differentiation.....	108
Figure 52: qRT-PCR validation results.....	110
Figure 53: PRPF40A and PRPF40B mRNA expression in TCGA database.....	111
Figure 54: <i>PRPF40B</i> mRNA expression levels in various cancers.....	112
Figure 55: Inverse correlation of PRPF40A and PRPF40B mRNA expression levels in AML patients.....	113
Figure 56: Inverse correlation of HIF1A and PRPF40B mRNA expression levels in AML patients.....	114
Figure 57: PRPF40A WB results in the KO cells.....	115
Figure 58: RNA-seq data analysis of AS.....	115
Figure 59: PRPF40B regulated AS events.....	116
Figure 60: Cassette exon events regulated by PRPF40B.....	117
Figure 61: Analysis of the AS events regulated by mutations of PRPF40B.	118
Figure 62: Validation of the AS events regulated by PRPF40B.....	119
Figure 63: Visualization of SPTAN1 AS.....	120
Figure 64: PPI network of the AS events regulated by PRPF40B.....	121
Figure 65: Gene set enrichment analysis of the DEGs and AS events regulated by PRPF40B.	123

Figure 66: Gene set enrichment analysis of the DEGs and AS events regulated by PRPF40B by KEGG database.....	124
Figure 67: Features of the cassette events regulated by PRPF40B.....	126
Figure 68: Correlation of the PSI and 5'ss strengths of the cassette events regulated by PRPF40B.	127
Figure 69: Transcription elongation rate of the cassette exons events regulated by PRPF40B.	128
Figure 70: Motifs analysis of the cassette exons events regulated by PRPF40B.	129
Figure 71: Results of UMV minigenes	131
Figure 72: Results of pCDNA minigenes.	132
Figure 73: Pol II inhibition assay.....	133
Figure 74: Visualization of hnRNP DL exon 8.	134
Figure 75: Flow cytometry results for the assessment of apoptosis, DNA damage and proliferation in KO cells.	135
Figure 76: CRISPR KO and Knock-In strategies.	144
Figure 77: Overexpression experiments.	147
Figure 78: Roles of PRPF40A and PRPF40B in AML.	150
Figure 79: Model of the role of PRPF40B in co-transcriptional splicing.....	154

LIST OF TABLES

Table 1: MDS classification.....	28
Table 2: List of antibodies for flow cytometry.	42
Table 3: List of primers for the Semi-Q RT-PCR validation.....	47
Table 4: CRISPR sgRNAs designed online.....	49
Table 5: CRISPR sgRNAs ordered for cloning.	50
Table 6: List of primers for the two genotyping PCRs.....	53
Table 7: PCR conditions for the genotyping PCRs.	54

Table 8: List of primers for the qRT-PCR to assess PRPF40B mRNA expression level.....	55
Table 9: Real Time PCR conditions.	55
Table 10: List of primers used to clone and sequence PRPF40B isoform C.....	56
Table 11: List of primers to insert one NAG into the PRPF40B isoform C.....	57
Table 12: List of primers to introduce two MDS mutation into PRPF40B WT.....	57
Table 13: List of sequencing primers for the two MDS mutations and PRPF40B isoform C.....	58
Table 14: Amounts of PRPF40B WT overexpression plasmid and mock used for titration.....	60
Table 15: List of various types of RNA-seq data analysis and corresponding R packages.....	61
Table 16: List of primers for the validation of the AS events regulated by PRPF40B.....	62
Table 17: List of primers for the validation of the DEGs regulated by PRPF40B.	64
Table 18: List of primers for the cloning of the UMV minigenes.	65
Table 19: List of primers for the cloning of pcDNA3.1+ minigenes.	65
Table 20: List of primers for the assessment of the inhibition of Pol II.	67

ABSTRACT

Pre-mRNA splicing is one of the processing steps by which the genetic information flows in a biological system for functional gene expression. Alternative splicing plays an important role in the development of eukaryotic cells during which the activity of the splicing regulatory machinery changes dynamically. Perturbations in alternative splicing are at the basis of a wide spectrum of human diseases. Recent investigations have shown that alternative splicing plays a role in the homeostasis of lymphoid cells in order to mount an efficient adaptive immune response. However, a detailed characterization of the role of alternative splicing in myeloid cells during an innate immune response is lacking. The characterization of alternative splicing events in myeloid cells will provide a basis to study perturbations that occur in various human disorders such as allergies, autoimmune diseases and cancers. Thus, we performed two transcriptome studies by RNA-sequencing in the aforementioned cells. In the first study, we performed analysis of the human monocytes and various subtypes of macrophages derived from them, both in the THP-1 cell line and in primary human cells. We identified genes that changed mRNA expression and changes in alternative splicing during the differentiation from monocytes to macrophages and during their polarization to various macrophage subsets. We found that alternative splicing plays a major role in the differentiation process and in the polarization of M1 macrophages but likely less in the activation of the M2 macrophages to the various subsets. In the second study, we successfully knocked out by CRISPR the splicing factor PRPF40B in K562 cell line. We rescued the knock out cells by overexpressing PRPF40B as wild-type or as mutated by introducing in the wild-type gene two mutations of the factor found in myelodysplastic syndromes. We found that PRPF40B controls the alternative splicing of thousands of events, hundreds of which are cassette exons, mainly repressing their inclusion, and we proved that it occurs co-transcriptionally. In addition, we found an activation of a hypoxia gene signature in the KO cells thus showing that PRPF40B has a role in inhibiting hypoxia, in contrast to its homologue PRPF40A which instead promotes it. We found that when PRPF40B is mutated, no major changes in gene expression and alternative splicing are occurring thus showing that the two mutations might be just passengers. Unexpectedly, we found a possible contrasting role of PRPF40A and PRPF40B in regulating hypoxia in Acute Myeloid Leukemia which could lead to a potential therapeutic application to decrease chemoresistance.

1 INTRODUCTION

1.1 Constitutive and alternative pre-mRNA splicing

According to the central dogma of biology, in eukaryotic cells gene expression is finely tuned by a set of complex regulatory processes, which begin from nuclear transcription of the precursor messenger RNAs (pre-mRNAs) and terminate in the translation of the mature RNAs into proteins in the cytoplasm (Crick et al., 1970). One of the crucial steps in this whole process is pre-mRNA splicing, which is a two-step transesterification reaction that catalyzes the removal of non-coding intronic sequences as lariat and the joining of the selected coding sequences (exons) (Burge et al., 1999; Keren et al., 2010; Kim et al., 2008) (Figure 1). The 5'start of an intron typically contains an almost invariant GU dinucleotide sequence defined as 5'splice site (5'ss). A nearly invariant AG dinucleotide sequence is present at the 3' end of the intron defined as 3'splice site (3'ss) (Sheth et al. 2006). A branch point sequence (BPS) is found in the middle of the intron to be removed. The BPS contains an adenosine followed by a polypyrimidine tract, a region rich in pyrimidine (C or Us) (Reed and Maniatis, 1988).

Figure 1: Mechanism of pre-mRNA splicing.

Schematic of the two transesterification reactions during the splicing process. Exons are indicated in dark blue while intron in light blue. The first nucleophilic attack occurs from the 2'-OH of the branch point sequence towards the first nucleotide of the intron at the 5'ss forming the lariat intermediate. The second nucleophilic attack occurs from the 3'-OH of the removed intron towards the last nucleotide of the intron at the 3' ss.

By splicing, immature pre-mRNAs are converted to mature mRNAs which are then transported outside the nucleus for protein translation. Splicing can occur in all the domains of life; however the extent of splicing can vary greatly. For instance, eukaryotes splice protein coding RNAs and few non-coding while prokaryotes splice seldom and only non-coding RNAs via self-splicing (Belfort, 1989; Bonen and Vogel, 2001). Splicing was first discovered in 1977, from a work done by Sharp and Roberts laboratories that identified several RNA molecules during the

adenovirus replicative stage which contained noncontiguous sequences from the viral genome (Berget et al., 1977; Chow et al., 1977).

Constitutive pre-mRNA splicing applies to those introns that are always spliced by using the same splice sites, so hence it is not regulated. Alternative pre-mRNA splicing (AS) occurs instead through alternative selections of specific introns and exons. Multiple mRNA transcripts, with different coding potential, are produced from the same gene thus generating multiple protein isoforms, often with different or even opposing functions (Figure 2) (Nilsen and Graveley, 2010). Alternatively, transcripts with identical coding sequence but different localization, stability or microRNA regulation are generated (Black, 2003; House and Lynch, 2008). For instance, the inclusion of an alternatively spliced exon might introduce a premature termination codon (PTC) which results either in the generation of a truncated protein or in the degradation of the mRNA by the non-sense mediated decay (NMD) pathway, an mRNA surveillance mechanism utilized by cells to ensure the translation of properly transcribed mRNAs (Baker and Parker, 2004; Lewis et al., 2003). By acting in cooperation with the NMD pathway, AS can regulate the level of functional mRNAs through the degradation of specific isoforms. Without NMD, a truncated protein can act a dominant-negative inhibitor of the original protein (Seo et al., 2011).

The gene that generates the highest number of transcript variants is the *Drosophila* gene Dscam with potentially more than 38,000 isoforms which is higher than the number of protein-coding genes found in the entire genome, approximately 17,000 (Celotto, 2001). An average human protein-coding gene contains a mean of 8.8 exons with a mean size of 145 nt (Sakharkar and Chow, 2004). The mean intron length is 3,365 nt and the 5' and 3' UTR are 770 and 300 nt, respectively. It has been estimated that more than 90% of human genes are alternatively spliced, which highly increases the proteome diversity (Pan et al., 2008; Nilsen and Graveley, 2010; Wang et al., 2008). AS is therefore an outstanding mechanism which remarkably enhances the coding potential of the genome. AS in humans might allow explaining the discrepancy between the relatively small number of genes presents in a genome and the organismal complexity (Figure 2) (Maniatis and Tasic, 2002).

Constitutive and AS occur in the cell nucleus, in particular domains denominated nuclear speckles which are rich in components of the splicing machinery and poly(A) RNA, and are

carried out by the spliceosome (Wahl et al., 2009). The spliceosome is considered the most complex macromolecular machinery known in eukaryotic cells which consists of five different RNA-protein complexes called small ribonucleoproteins (snRNPs) and other proteins not part of the snRNPs (Figure 3) (Will and Luhrmann, 2011). The small nuclear RNA (snRNA) component of the snRNPs is responsible for the recognition and binding of specific sites on an intron through RNA-RNA or RNA-protein interactions.

The AS events can be organ-specific, tissue-specific or cell type-specific (Mottes and Iverson, 1995; Grosso et al., 2008). In conclusion, AS is a fundamental process needed to produce mature mRNAs that then can be used by the translational machinery to generate various proteins thus increasing their biodiversity (Figure 2).

Figure 2: AS and proteome complexity.

Diagram showing how different protein isoforms are generated by AS from the same pre-mRNA transcript. Introns are removed and exons are joined in different combinations shown here in different colors (Figure from Ivanov and Anderson, 2013).

1.2 The major and the minor spliceosome

In most eukaryotes, two classes of introns exist: the most common type of introns (>99%) are of the U2-type while very rare are U12-type introns. Two different spliceosomes, the major and the minor, catalyze the removal of the U2 and U12 introns respectively (Sharp and Burge, 1997) (Figure 3). The major spliceosome is composed of five snRNPs namely U1, U2, U4, U5 and U6, and other 200 related proteins (Jurica and Moore, 2003). The assembly of the spliceosome

components on the pre-mRNA occurs dynamically in a stepwise manner which enhances the accuracy and the flexibility of this complex machinery (Lynch, 2014).

The first step in the assembly of the major spliceosome involves the binding of the U1 snRNP to the 5'ss and of SF1 to the BPS (Seraphin and Rosbash, 1989). The large subunit of U2AF binds then to the polypyrimidine tract (PPT) whereas the smaller subunit bind to the 3'ss dinucleotide in order to form the early commitment E complex. The formation of this first complex occurs in an ATP-independent manner. Through interaction with U1 and the E complex component U2AF, U2 snRNP is recruited to the BPS where the A complex is formed in an ATP-dependent manner. The formation of the A complex favors the bulging of the adenosine at the branch site, targeting it for the nucleophilic attack of the first transesterification reaction. The U4/U5/U6 tri-snRNP is then recruited to the spliceosome to form the B complex which is followed by the displacement of U1 and U4 away from the spliceosome (Jamison et al., 1992). Further rearrangements give rise to a catalytically active B complex which carries out the first catalytic step generating a free exon and an intron-exon lariat intermediate. The phosphodiester bond at the 5'ss of the intron is modified by the 2'-hydroxyl of an adenosine of the BPS to form a free 5' exon and a branched intron, which is subsequently degraded. The second catalytic step which results in a free intron as a lariat and joined exons is carried out by an activated complex C. As a final step U2, U5 and U6 are recycled back for a new round of splicing (Chen and Manley, 2009; Kornblihtt et al., 2013; Wahl et al., 2009).

The minor spliceosome consists instead of four specific snRNPs, U11, U12, U4atac, and U6atac while U5 snRNP is shared between the two spliceosomes (Tarn and Steitz, 1996). The minor spliceosome recognizes consensus splice site dinucleotide sequence at the 5'ss and at the 3'ss as well as a more conserved BPS. The assembly pathway is similar to the major spliceosome with the exception of the commitment E complex which is absent in the minor spliceosome because U11 and U12 join the spliceosome together as a dimer (Figure 3). The same two step transesterification mechanism is employed for intron removal by both spliceosomes. Removal of U12-type introns appears to be slower than U2-type. It has been hypothesized that a slower rate of splicing could affect the expression of genes containing U12-type introns, considering that the transcripts containing non spliced U12-type introns could be retained in the nucleus (Patel et al., 2002).

Figure 3: Schematic representation of minor and major spliceosome assembly.

On the right splicing of the U12-type introns is shown while on the left of the U2-type of introns.

Spliceosome binding to the pre-mRNA is influenced by a set of auxiliary proteins defined as RNA binding proteins (RBPs) such as SR proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) which are *trans*-acting factors (Wang and Burges, 2008). In addition, *cis*-acting auxiliary elements, referred to as exonic or intronic splicing enhancers or silencers (ESEs/ESSs, ISEs/ISSs) are present on the pre-mRNA sequence (House and Lynch, 2008). The above factors and elements are required due to the degenerate splice site consensus sequences. The positive elements, (ESE and ISE) are usually recognized by SR proteins which are generally viewed as positive splicing regulators and promote exon inclusion by facilitating the binding of U1 snRNP and U2AF in the early step of spliceosome assembly (Long and Caceres, 2009). Negative elements (ESS and ISS), are usually recognized by the hnRNP family members, which instead are viewed as negative splicing regulators since they promote exon skipping either by antagonizing the binding of SR proteins or by other several mechanisms (Figure 4) (Martinez et al., 2007). However this common concept has been recently challenged by reports showing that positive regulators of splicing in certain cases can act as negative regulators and vice versa, depending on the site at which they bind on the pre-mRNA (Witten and Ule, 2011). In general, both SR and hnRNP proteins contain RNA-binding and protein interaction domains. However, many of them bind to the RNA with low specificity to accessible parts of the pre-mRNA. The binding to the RNA is increased by the presence of repetitive sequences on the RNA. Ultimately, the combination of these *trans*-acting factors and *cis*-acting elements is what determines whether an exon should be included or skipped (Barash et al., 2010)

All the RBPs have splicing variants whose functions are poorly studied. In both animals and plants, RBPs and other spliceosomal components are themselves spliced, they can control their own splicing or the splicing of other RBPs (Thomas et al., 2012). In addition, signals from the environments and metabolic cues can result in post-translational modifications (PTMs) of their proteins thus modulating their functions (Stamm et al., 2005).

A combination of multiple weak protein-protein, protein-RNA and RNA-RNA interactions during spliceosome assembly will allow to achieve a high fidelity of exon recognition. A low affinity of interaction can confer various advantages. For instance, small changes in the concentration of regulatory proteins will influence the protein-protein interactions which allow the alternative usage of exons depending on a tissue and/or developmental-specific concentration

of regulatory factors. In addition, phosphorylation of regulatory factors that alter protein-protein interactions can influence splice site selection (Xiao and Manley, 1998).

Figure 4: Schematic representation of *cis*-acting elements and *trans*-acting splicing factors.

Spliceosomal components and auxiliary factors bind to consensus target sequences on the pre-mRNA promoting exon recognition and ligation. This mechanism is also known as "exon definition" and allows the precise localization of exons between all the introns present on the pre-mRNA (adapted from Kornblihtt et al., 2013).

1.3 Intron and exon definition models

Higher vertebrate genes are composed of short exons (170 nt in average) interspersed between much longer introns. Instead, the opposite occurs in lower vertebrates. In the first scenario, when exons are small and introns are large, the splicing machinery most likely will assemble across an exon rather than an intron. It may even take an hour for an intron to be transcribed, thus initial splice site definition occur over the shorter exons. The exon definition model, first proposed by Berget in 1995 (Berget, 1995), allows explaining how the spliceosome is able to recognize exons and introns in a huge variety of pre-mRNA molecules with extremely high precision in higher vertebrates. According to this model, the recognition of the 5'ss by U1 snRNP can be enhanced when U2AF binds to the 3'ss on the same exon and vice versa. In addition, other factors such as the *cis*-acting elements, RNA structure, transcriptional rate, chromatin modifications can

influence the binding of the complex across the exon. In lower vertebrates, such as budding yeast instead, splice sites are paired across introns rather than exons, thus an “intron definition” model is predominant (Figure 5).

Figure 5: Exon and Intron definition models.

Diagrams representing the two different models found in higher (A) and lower vertebrates (B) respectively. Components of the spliceosome are bound to exons depicted as gray cylinders. Straight lines represent introns and arrow interactions between the components of the spliceosome (adapted from De Conti et al., 2013).

1.4 Modes of alternative splicing

Seven main modes of alternative splicing exist (Sammeth et al., 2008) (Figure 6):

1. *Alternative 5'ss*: more than one 5'ss is spliced to the same downstream exon, generating exons with different sizes.
2. *Alternative 3'ss*: more than one 3'ss is spliced to the same upstream exon, generating exons with different sizes.
3. *Mutually exclusive exons*: only one of two or more exons is included.
4. *Exon skipping or cassette exons*: an exon is either included or excluded from the transcript. It is the most common alternative splicing event.
5. *Intron retention*: an intron is retained in the coding region. It is the rarest event.
6. *Alternative promoters*: different promoters associated with different 5' terminal exons are used.

7. *Alternative polyadenylation*: different poly(A) sites associated with different 3' terminal exons are used.

Figure 6: Patterns of AS.

Cylinders represent an exon while straight lines represent an intron. Arrows represent the different combinations by which exons are joined.

Similarly to other processes such as capping and polyadenylation, AS occurs co-transcriptionally and it has been shown that the two events influence each other (Bentley, 2005, Kornblihtt et al., 2004; Maniatis and Reed, 2002). According to the recruitment coupling model, the RNA

Polymerase II (Pol II) carboxy-terminal domain (CTD) and the Mediator complex are responsible for the recruitment of splicing factors to the transcription machinery while the pre-mRNA is still being transcribed (Kornblihtt et al., 2013). In more detail, the CTD of Pol II promotes the binding of Ser/Arg-rich splicing factor 3 (SRSF3), and this promotes inclusion of alternative exons. General transcription factors (GTFs) bound to promoters and specific transcription factors (TFs) bound to transcriptional enhancers associate with the Mediator complex which recruits the negative splicing factor heterogeneous nuclear ribonucleoprotein L (hnRNPL) and inhibits inclusion of alternative exons (Ji and Fu, 2012). Kinetic coupling is instead another mechanism by which the rate of Pol II elongation influences splicing. If the Pol II is proceeding at a slower pace an exon is preferentially included whereas if the Pol II is proceeding at a faster pace an exon is skipped (de la Mata et al., 2003; Kornblihtt et al., 2013) (Figure 7). It is worth noting however, that the aforementioned concept is not always valid but there are cases in which the opposite would occur. If the Pol II proceeds at a slower pace an exon would be excluded and vice versa (Dujardin et al., 2014; Tang et al., 2016). Different factors exist that promote Pol II pausing such as CTCF or factors that enhance the speed such as Zfp326 (Skukla et al, 2011; Liu et al., 2015).

Figure 7: Co-transcriptional splicing.

Schematics representing how AS is coupled to mRNA transcription. As a general rule, for the majority of the alternatively spliced exons, a fast elongation of Pol II results in exon skipping (a) while a slow elongation results in exon inclusion (b and c) (Figure from McCracken et al., 1997).

Besides being regulated by transcription, chromatin structure influences splicing through the effect of histone modifications and nucleosome positioning (Luco et al., 2010; Schor et al., 2012). By either indirect recruitment of splicing factors or by modulating the elongation rate of the Pol II, chromatin structure and histone modifications can influence AS. For instance exon/introns boundaries are marked by regions in which the histone mark trimethylated histone H3 at lysine 36 (H3K36me3) is enriched or depleted (Kolasinska-Zwierz et al., 2009). In addition, given the interaction of histone acetyltransferases and methyltransferases with components of the spliceosome, it appears that chromatin complexes can facilitate the assembly of the spliceosome to the pre-mRNA.

Another connection between chromatin and AS is mediated by Argonaute proteins. Argonaute proteins bound to small RNAs can trigger histone modifications and influence the binding of the spliceosome thus promoting exon inclusion (Batsche and Ameyar-Zazoua, 2015).

Long non coding RNAs (lncRNAs) can introduce epigenetics modifications that impact the speed of transcription and the recruitment of adapters. They can also interact with splicing factors and regulate their availability (Gonzalez et al., 2015).

Another factor that influences splicing is nucleosome positioning. It has been shown that introns are practically almost devoid of nucleosomes, whereas exons are preferentially bound by nucleosomes (Andersson et al., 2009; Nahkuri et al., 2009). The presence of nucleosomes at exons improves exon definition by causing the stalling of the Pol II at these sites, which in turn favor the recruitment of the splicing factors to the nascent pre-mRNA. In agreement to this, it has been reported that alternatively spliced exons, on the contrary to constitutive exons, are highly enriched in nucleosomes and in RNA Pol II density.

In this Ph.D. project we are interested to study the role of AS in the human immune system, in particular in cellular differentiation, thus certain concepts in immunology will be introduced in the next sections.

1.5 Human hematopoiesis

Given the short lived nature of mature human blood cells, hematopoietic stem cells (HSCs) are required in order to replenish them constantly throughout the lifetime of an individual. HSCs are the precursors of all the blood components and reside in the medulla of the bone marrow (Dzierzak, 2002) (Figure 8).

HSCs are maintained in an undifferentiated state and have the ability of both multi-potency and self-renewal. Self-renewal is a process by which stem cells divide, while maintaining the undifferentiated state and the differentiation potential, in order to give rise to other stem cells by asymmetric or symmetric division. Multi-potency is instead the ability to differentiate into any type of mature cells (Seita and Weissman, 2010). Based on their self-renewal capacity, HSCs can be divided into two populations: long term (LT)-HSCs which possess a life-long self-renewal and short term (ST)-HSCs that have a limited self-renewal (Ema et al., 2014).

Upon a particular stimulation, HSCs can differentiate into specialized cells; they then become terminally differentiated and lose their self-renewal capacity. HSCs are maintained in an undifferentiated state inside the so called "stem cell niche". The cells in the niche produce factors such as cytokines and chemokines which maintain the HSCs in a quiescent state but also regulate their differentiation and proliferation kinetics (Morrison and Spradling, 2008).

Inside the niche, during asymmetric division a stem cell produces a daughter cell which is a copy of itself and a second cell that is committed to differentiate. In certain cases this second cell, being a progenitor itself undergoes terminal differentiation, and in other cases the cell could both amplify itself and then differentiate (Dzierkak and Speck, 2008).

According to the currently accepted model of hematopoiesis, at first ST-HSCs give rise to multipotent progenitors (MPPs), which have lost all the self-renewal capacity but can still produce all the mature blood cells. MPP then differentiate into the common lymphoid progenitors (CLPs) and into the common myeloid progenitors (CMPs) (Figure 8). CLPs can generate all types of lymphoid cells whereas CMPs give rise to all the myeloid cells (Akashi et al., 2000, Kondo et al., 1997). B cells, T cells and NK cells belong to the lymphoid lineage, while megakaryocytes, erythrocytes, granulocytes and monocytes belong to the myeloid lineage. Dendritic cells (DCs) are a particular type of cells because they can be generated either from the CMPs or CLPs.

Figure 8: Human hematopoiesis.

Schematic representation of hematopoiesis. (LT-HSC) Long-term Hematopoietic stem cell, (ST-HSC) Short-term Hematopoietic stem cell, (CMP) Common Myeloid Progenitor, (CLP) Common Lymphoid Progenitor, (MEP) Megakaryocyte-Erythroid Progenitor, (GMP) Granulocyte-Macrophage Progenitor (Figure from Larsson and Karlsson, 2005).

1.6 Alternative splicing during normal hematopoiesis and in normal development

Several recent RNA-sequencing (RNA-seq) studies during normal murine and human hematopoiesis have proposed a model in which specific changes in AS occur during the various stages of hematopoiesis. However, very few differentiation stage-specific AS events have been functionally validated. One example is exon 16 of the *EPB41* gene which encodes protein 4.1R (Baklouti et al., 1997). In immature erythroblasts, the exon is skipped while it is included in mature ones. Exon 16 encodes a portion of the spectrin/actin binding domain and its inclusion enhances the affinity of 4.1R protein for spectrin and actin, thereby stabilizing the erythroid membrane under mechanical stress (Yamamoto et al., 2009). Considering that mature

erythrocytes have to pass through blood vessels and be subjected to a mechanical stress, the inclusion of exon 16 confers an important feature to the erythroblasts prior to their release into the blood circulation. Failure to include exon 16 has been shown to generate mechanically unstable red blood cells which would lead to anemia (Conboy, 1993).

Another concept which has emerged from the RNA-seq data of mature granulocytes isolated from murine bone marrow is that during granulopoiesis a pattern of intron retention upon differentiation appears (Wong et al., 2013). Retained introns contain PTCs which will then be exported in the cytoplasm and be subjected to degradation by NMD. A similar pattern was later found in five populations of human erythroid precursors. However, in the latter case, intron-containing transcripts were preferentially retained inside the nucleus and protected from degradation by NMD, in contrast to what would have happened during granulopoiesis (Edwards et al., 2016). In both scenarios, the pattern of intron retention would result in the downregulation of the full length proteins. Similarly to the PTCs introduced during intron retention, during terminal erythropoiesis, SR proteins are downregulated due to the higher expression upon differentiation of the isoforms containing a PTC (Pimentel et al., 2014).

Studies to identify *trans*-acting factors which are responsible for the stage specific switches in AS are currently ongoing. Efforts are much needed considering our lack in the knowledge on the contribution of these splicing factors in hematopoiesis while much is known about the role of the transcription factors.

Examples of splicing factors involved in the normal development are PTBP1 and PTBP2 which can regulate the development of post-mitotic neurons from neuronal precursor cells (NPCs). In undifferentiated cells, PTBP1 levels are high and can repress PTBP2 by increasing its degradation by AS followed by NMD. On the contrary, as cells differentiate, PTBP1 expression becomes silenced which results in the up-regulation of PTBP2, with the latter thus promoting the generation of mature neurons (Vuong et al., 2016; Hamid and Makeyev, 2014).

1.7 Alternative splicing in the immune system: B cells, T cells and DCs

During the past decades, the attention has been focused on the changes in gene expression during an immune response mostly at the transcriptional level (Bunting et al., 2007; Lin et al., 2003; Ollila and Vihinen, 1998). However, recent studies have recognized AS as a novel mechanism of gene regulation which can highly increase the diversity and flexibility of an immune response (Lynch, 2004). Individual cells in the immune system rely on different detection, transduction and effector mechanisms, and respond differently to the changes in the environment. By AS, immune cells can obtain the plasticity needed for the immune response.

Immunological diversity can be obtained in various ways, such as by AS of the major histocompatibility complex (MHC) molecules, by generating alternatively spliced isoforms of the Fc γ receptors, which alters Fc receptor function, turning these molecules in a soluble form. Other examples are by alteration of the T cell receptor (TCR) signaling and by expression of various isoforms of the cytokines and cytokines receptors (Martinez and Lynch, 2013).

The first case of AS reported in the immune system was the differential expression of Immunoglobulin M (IgM) transcripts in naïve B cell upon antigen-induced differentiation generation either a membrane-bound form or a secreted one (Alt et al., 1980).

AS has been recently very well studied in T cells after their activation upon antigen challenge (Lynch, 2004). T cell activation induces a deep change in the transcriptome. Microarray and next generation sequencing (NGS) have been employed to study the changes in AS upon T cell activation, and have revealed AS, together with microRNA, as emerging important post-transcriptional mechanisms that can regulate immune cell function (Grigoryev et al., 2009; Martinez and Lynch, 2013).

More in detail, among the genes that have been validated to undergo AS in T cells, include genes that are known to have an important function in the immune response. The genes include cell surface receptors (CD3, CD28, CD8, CD44, CD45, CTLA-4, and IL-7R α), kinases (FYN, LSK, SYK, MAP4K2, MAP3K7, and MAP2K7), adhesion molecules (ICAM1, PECAM1), phosphatases, adaptor proteins, transcription factors (LEF1, GATA3, and RUNX1), chromatin modifying enzymes (EHMT2) and RBPs (SAM68, CELF2, and AUF1) (Lynch, 2004).

Importantly it has been found that the splicing factor hnRNPLL is a master regulator of alternative splicing involved in T cell activation (Topp et al., 2008).

In addition, a recent study has shown that AS is involved in the response to bacterial challenge in DCs (Rodrigues et al., 2013). In order to elicit an immune response, DCs go through an extensive maturation process which allows them to be converted from antigen-capturing cells to Antigen Presenting Cells (APCs). Acquisition of high mobility, secretion of chemokines, up-regulation of costimulatory molecules, loss of phagocytic receptors are the major changes that occur during maturation and AS plays a role in this process.

Very little was known in the literature about the role of AS in the differentiation of myeloid cells despite many reports showing that AS plays an important role in the activation of cells of lymphoid lineage as previously mentioned (Lynch, 2004; Grigoryev et al., 2009). Thus, in this project we aim at increasing our knowledge of AS in cells of myeloid origin. Concepts about specific myeloid cells will be introduced in the next sections.

1.8 Human Monocytes

Monocytes are myeloid cells which belong to the innate immune system and constitute around 10% of total leukocytes. They are derived from precursor cells called monoblasts, which are originated from HSCs. The main receptor involved in the egress from the bone marrow is CCR2. Monocytes can be divided into classical, non-classical or intermediate. Classical monocytes are characterized by high expression of CD14 and are negative for CD16 ($CD14^{++} CD16^{-}$), non-classical monocytes show low level expression of CD14 and expression of the CD16 receptor ($CD14^{+}CD16^{++}$) and intermediate monocytes possess a high level of CD14 and low level expression of CD16 ($CD14^{++}CD16^{+}$) (Passlick et al., 1989; Ziegler-Heitbrock et al, 2010). Classical monocytes express high level of CCR2 and their major function is phagocytosis, non-classical are involved in the patrolling of vessels and lastly the intermediate ones have a role in inflammation.

Monocytes circulate into the bloodstream for one to three days and then finally migrate to tissues where they differentiate and replenish tissue resident macrophages (Ebert and Florey, 1939). Depending on the particular tissue, macrophages are named differently: microglia in the brain,

Kuppfer cells in the liver, osteoclasts in the bones and Langerhans cells in the skin (Ginhoux and Jund, 2014).

Alternatively, they serve as danger sensors in the circulation considering that in response to infection or injury, they migrate to the site of infection/injury recruited by chemotactic signals sent from macrophages already present at the inflamed site. Once monocytes reach the site of inflammation, they differentiate to either macrophages or DCs, with the latter being responsible of clearing the infection by activating an immune response (Van Furth et al., 1973). Once they differentiate to macrophages and DCs, they acquire particular functions needed for infection clearance such as phagocytosis, antigen presentation and cytokine production. At the site of infection, macrophages and DCs will then elicit an adaptive immune response via the presentation of the antigen to T helper cells (T_h cells).

Monocytes differentiate to macrophages upon activation by Granulocyte Macrophage Colony Stimulating Factor (GM-CSF) or by Macrophage Colony Stimulating Factor (M-CSF) binding to their receptor, which triggers a signaling cascade leading to the activation of various transcription factors, such as PU.1, Egr1, ICSBP/IRF-8, KLF4, MafB and c-Maf (Figure 9). Upon differentiation, monocytes acquire different cell morphology, characterized by an enlarged nucleus, an elongated shape and the presence of vacuoles in the cytoplasm, which are features that typically distinguish macrophages.

Similarly to T helper type 1 (T_H1) and T helper type 2 (T_H2) polarization, macrophages are generally distinguished as a classically activated (M1) macrophages or alternatively activated (M2) macrophages; the former being pro-inflammatory while the latter being anti-inflammatory (Mills et al., 2000). It is known that GM-CSF can induce differentiation to M1 macrophages while M2 macrophages are obtained by stimulation with M-CSF (Mantovani et al., 2004). Macrophages can be polarized to M1 by Interferon gamma ($IFN-\gamma$) alone or in combination with Lipopolysaccharide (LPS), and have a prominent role in the activation of the immune response. Alternatively, they can be polarized to M2 macrophages by a combination of Interleukin-4 (IL-4) and Interleukin-13 (IL-13), which have a role in immunosuppression and wound healing (Nathan et al., 1983; Gordon, 2003). Finally, M2 macrophages can be further subdivided into M2a, M2b and M2c macrophages, each one possessing a distinct chemokine and chemokine receptor profiles and having different functions in the human body (Mantovani et al., 2004) (Figure 9).

M2a macrophages have a role in anti-parasitic response and allergies, M2b in immuno-regulation and lastly M2c in tissue repair.

Besides being involved in the immune response as APCs, macrophages have a role in the removal of dying cells and debris. Furthermore, aged erythrocytes are constantly engulfed and digested by macrophages. Macrophages play a role in iron metabolism since they degrade hemoglobin (Hb) (de Back et al., 2014). Finally macrophages have also a role in wound healing and tissue repair.

Figure 9: Representation of macrophage heterogeneity.

Monocytes can be differentiated to M1 or M0/M2 macrophages by GM-CSF or M-CSF respectively. M0/M2 macrophages can be differentiated to M2a, M2b, M2c by IL-4/IL-13, immune complexes or Interleukine-10 (IL-10). A combination of GM-CSF and IL-4 can differentiate monocytes to DCs (Figure adapted from Gleissner, 2012).

Excessive recruitment of monocytes to the sites of inflammation and the chronic accumulation of macrophages at those sites is a hallmark of inflammatory diseases such as atherosclerosis, rheumatoid arthritis and Crohn's disease (Liu et al., 2014).

Considering that we are interested to study the role of AS in connection with human diseases of the immune system, I will provide background at first on what is known about the role of AS in various human diseases in general, with later on a particular focus on blood cancers known as Myelodysplastic syndromes (MDS) which affect immune cells of the myeloid lineage,.

1.9 Splicing in human diseases

Mutations that occur in either the *cis*-acting elements or in the spliceosomal components often result in abnormal mRNA splicing. *Cis*-acting elements have been found mutated in numerous diseases such as cystic fibrosis, beta-thalassemia, Parkinson's disease, retinis pigmentosa, neurofibromatosis and frontotemporal dementia and many other diseases (Faustino and Cooper, 2003). In the aforementioned diseases, point mutations can disrupt splicing by weakening splice sites or activating cryptic ones, thus generating defective proteins. For example, in β -thalassemia, splice site mutations of the Hemoglobin Subunit Beta (HBB) gene will cause defective production of the β -globin protein (Busslinger et al., 1981). In addition, it has been also found in certain diseases, a part of the genes containing the splice site mutations have reduced transcriptional activity (Vaz-Drago et al., 2015). In addition, an undetermined number of intron mutations may further contribute to splicing misregulation. For instance, a recent study has identified hundreds of intronic single-nucleotide variations (SNVs) associated with autism (Stein et al., 2017). However, it is worth to mention that not always a splicing change has a negative effect. For instance, it was recently found that a splice site variation of the Low-Density Lipoprotein Receptor (LDLR) that lowers non-high density lipoprotein cholesterol can protect against coronary artery disease (Gretarsdottir et al., 2015).

Nuclear level changes of RBPs have been frequently observed in many cancers and in most cases result in the generation of splicing variants affecting cell cycle, apoptosis and migration. Another way by which the level of RBPs is affected is by sequestration (Duan et al., 2014). In the cases of DM1 and DM2 myotonic dystrophies for instance, the expansion of specific repeats in the DNA generates mRNA transcripts which ultimately sequester the splicing factor Muscleblind-like 1

(MBNL1) resulting in its loss-of-function thus leading to downstream disruption of AS not only in muscle tissues but also in the brain (Carpentier et al., 2014). In some neurological diseases, such as Amyotrophic Lateral Sclerosis (ALS), mutated splicing factor aggregates in the cytoplasm and often sequester other spliceosome components thus leading to alteration in splicing profiles (Gerbino et al., 2013).

Cytoplasmic assembly of snRNPs is facilitated by the survival motor neuron complex (SMN). Mutations in the complex can reduce the levels of snRNPs and cause spinal muscular atrophy (SMA), a neurodegenerative disease characterized by destruction of specific motor neurons (Monami et al., 1999). There are two copies of the SMN gene in humans, SMN1 and SMN2. SMN1 contributes to the generation of the majority of the protein while limited amount of full length protein is expressed from SMN2. Having limited amount of protein translated from SMN2 is likely sufficient for normal activity in all the other cells excluding motor neurons. SMA patients carry two non-functional copies of SMN1, and SMN2 cannot compensate for the SMN1 loss, due to a single-nucleotide transition in exon 7 of SMN2 resulting in the exclusion of the exon and the translation of a shorter non-functional protein (Ruggiu et al., 2012).

Somatic mutations in the spliceosome were once thought to be lethal until recently, considering that disruption of the spliceosome components would result in abnormal splicing of thousands of genes and no cell would survive without splicing at all. However, recent findings have shown that mutations in the spliceosome contribute to the pathogenesis of various human cancers and in particular of hematological malignancies (Sveen et al., 2015). It is not yet clear whether a change in AS can cause the disease or it is an indicator for an underlying defect. Potential outcomes of mutations of genes encoding proteins of the spliceosome include exon skipping, intron retention, and cryptic exonic (or intronic) splice site activation with truncated (or elongated) exon. An example of a disease caused by mutations in the spliceosome is the autosomal dominant forms of retinitis pigmentosa, in which PRPF31 and PRPF8 splicing factors have been found mutated (Vithana et al., 2011). PRPF31 is a factor that is crucial for the transition of the spliceosomal complex to the activated state. In addition, recent studies have identified numerous somatic mutations of splicing factors in MDS, a category of disorders of the immune system that affect the production of myeloid cells (Yoshida et al., 2011). Heterozygous and mutually exclusive mutations have been found in SF3B1, SRSF2, U2AF35, ZRSR2, SF3A1, LUC7L2, PRPF40B,

SF1, and U2AF65 splicing factors with SF3B1 being mutated in 20-45% of MDS cases (Malcovati et al., 2011; Yoshida et al., 2011). Mutations in the same splicing factors have been found in other neoplasms such as Chronic Lymphoid Leukemia (CLL), breast, renal and adenoid cystic carcinoma (Ebert and Bernard, 2011; Wang et al., 2011). Interestingly all the splicing factors mutated, excluding PRPF40B and LUC7L2, are involved in 3'ss recognition. PRPF40B and LUC7L2 are most likely the only factor involved in 5'ss recognition via U1. The AS events influenced by these mutated splicing factors are unknown and awaits further investigation.

Besides being interested to study the role of AS in cellular differentiation of monocytes, a second aim of this Ph.D. is to investigate the role of the PRPF40B splicing factor in MDS pathogenesis. Some background on MDS pathogenesis and the mutated splicing factors will be given in the next sections.

1.10 MDS pathogenesis

MDS are a heterogeneous group of hematological pre-malignancies characterized by an inefficient hematopoiesis (Cazzola and Malcovati, 2005). They are a clonal disorder meaning that an aberration amongst a hematopoietic progenitor cell is responsible for the entire disease. They arise from mutations in HSCs in the bone marrow which do not become fully mature blood cells (Welch et al., 2012). A founding driver mutation, typically in a gene involved in RNA splicing or epigenetics confers a selective advantage to the HSCs which clonally expands as a consequence (Papaemmanuil et al., 2013; Yoshida et al., 2011). At first the expansion occurs locally, however at later stages, cells start to migrate and settle to other compartments of the bone marrow. At the very end of the expansion, cells achieve full clonal dominance in the body and all the bone marrow cells are clonally derived. Besides carrying the driver mutation, cells also acquire additional passenger mutations at the beginning of the clonal expansion. In addition, clonal cells can also acquire subclonal driver mutations, such as in genes involved in chromatin modification or regulation of transcription (Walter et al. 2013). As a consequence, the subclones may have further impaired maturation ability and the blast cells number may increase till the point of developing overt Acute Myeloid Leukemia (AML) (Walter et al., 2012). In this scenario, the disease is not anymore monoclonal in the strict sense but it consists in a mosaic of clones with various somatic mutations. Mutually exclusive mutations have been found in more than 25 genes (Figure 10).

Figure 10: Landscape of MDS mutations.

Size of the bubble is proportional to the frequency of the mutations in each gene in MDS patients. Genes are grouped according to their function. Top two categories of mutated genes include pre-mRNA splicing and DNA methylation (Figure from Larsson et al., 2013).

The above mutations are valuable either for prognostic and therapeutic applications. At a certain point after clonal expansion, the disease becomes clinically apparent; however it may or may not require the acquisition of cooperating mutations. Because of the excessive apoptosis, skewed proliferation and defective maturation of the cells, dysplasia in one or more cell lineages occurs. The number of lineages affected (either megakaryocytic or granulocytic or erythroid) can be used as a criteria for classification of patients. The majority of the patients present no symptoms at early stages of the disease while at later stages anemia (low red blood count) appears when diagnosed with MDS which would cause extreme fatigue. Furthermore they can present a low white blood count (neutropenia) and low platelets (thrombocytopenia) which lead to the propensity to infections and bleeding respectively (Germinig et al. 2013).

Diagnosis of MDS is mainly based on visualization of dysplastic features of the cytoplasm or nucleus of cells in the bone marrow following biopsy. Other tests include full blood count, karyotyping, cytogenetics studies and flow cytometry (Germing et al. 2013). For the majority of the patients, there is no therapy that can cure the disease definitely. For younger patients and high risk MDS cases, allogeneic stem-cell transplantation is the only curative therapy. However, there are treatments that can improve the quality of life of the patients and delay the progress to AML. For instance, three chemotherapeutic agents which improve overall survival have been approved by the Food and Drug Administration (FDA): lenalidomide and the demethylating agents 5'-azacitidine and decitabine (Jädersten and Hellström-Lindberg, 2009). The survival rate varies from months to years and depends on many factors such as type of cells affected.

The clear cause of primary MDS has not been confirmed. It may involve exposure to radiations or certain chemicals (Poynter et al., 2017). Secondary MDS results from exposure to chemotherapeutic drugs during cancer treatment. MDS is associated with advanced age since most of the patients are diagnosed at or after 70 years old. A different scenario is seen in Asia, where the disease is more common at younger age (Paydas, 2006; Kuendgen et al., 2007). Children with genetic defects in genes involved in DNA repair can develop MDS. According to the World Health Organization (WHO), incidence of MDS is 2-10 cases per 100,000 people which increases to 50 per 100,000 in people older than 70 years (Neukirchen et al., 2011).

MDS can be distinguished from AML because of various factors. For instance, in AML, HSCs show a blast-like morphology and there is a block in cell differentiation at early stages. On the contrary, in MDS cells seem to differentiate well at the beginning until a terminally differentiated population of cells fails to appear. Thus, in MDS there is an increased cell death while in AML the cell death is reduced. In fact, when MDS progress to AML there is a distinct absence of apoptosis. Another feature that can distinguish MDS from AML is the presence of specific chromosome abnormalities (Bernasconi, 2008).

Diagnostic and prognostic criteria do not include much genetic information because relatively little is known about the genetics of the disease, although cytogenetic abnormalities can be found in around 50% of the MDS cases (such as loss of 5q, 7 or 7q, 20q, and/or Y, and trisomy 8)

(Haase, 2008). In particular, the most common abnormality for half of those cases is the deletion of the long arm of chromosome 5 [del(5q)] which defines a distinct subtype of the disease (Gaballa et al., 2014). This subtype is characterized by transfusion-dependent anemia, <5% blasts in the bone marrow and thrombocytosis. Patients with del(5q) have a more favorable prognosis and rarely progress to AML. However, no gene mapping inside the deleted region was found to be mutated. On the contrary, haploinsufficiency of certain genes belonging to the deleted region was sufficient to recapitulate the phenotype seen in MDS patients (Stoddart et al., 2014). Haploinsufficient genes found in this region are mainly ribosomal protein genes and one of those genes, RPS14, is highly responsible to the defect seen in this MDS subtype especially for the erythroid differentiation (Pellagatti et al., 2008). Disruption of ribosomal biogenesis would cause nuclear stress which would lead to p53 pathway activation in erythroid cells, which would ultimately result in the expression of genes involved in cell cycle arrest and apoptosis. The p53 protein can then keep in check the growth of primitive del(5q) cells thus explaining the better prognosis in those patients. As a further proof of this concept, patients with del(5q) and mutations in TP53 gene have a poor response to therapy and higher propensity to develop AML (Jadersten et al., 2011).

Two other features characteristic of MDS are the presence of DNA damage and shortened telomeres (Colla et al., 2015). Excessive DNA damage is found in HSCs of MDS patients and the inability of these cells to cope with it is directly associated to their genomic instability (Zhou et al., 2013). This excessive damage is commonly observed as oxidative DNA damage. Bone marrow cells of MDS patients show elevated levels of oxidized nucleotides. A recent study has identified a link between shortened telomeres and DNA damage. In a mouse model, progressive erosion of telomeres can cause an accumulation of DNA damage in hematopoietic precursors and impaired differentiation which recapitulates an MDS phenotype (Colla et al., 2015).

According to the World Health Organization (WHO), MDS are classified into different subtypes characterized by various features (Table 1) (Arber et al., 2016).

Name	Abbreviation	Key Feature	Patients, %
Refractory cytopenia, with unilineage, dysplasia	RA	Anemia and erythroid dysplasia	10
	RN	Neutropenia and granulocytic dysplasia	< 1
	RT	Thrombocytopenia and megakaryocytic dysplasia	< 1
Refractory anemia with ring sideroblasts	RARS	≥ 15% ring sideroblasts	5
5q- syndrome	del(5q)	Isolated 5q31 deletion, anemia, hypolobated megakaryocytes	5
Refractory cytopenia with multilineage dysplasia	RCMD	Multilineage dysplasia with > 1 cytopenia with or without ring sideroblasts	20
Refractory anemia with excess blasts, type 1	RAEB-1	5% to 9% blasts	20
Refractory anemia with excess blasts, type 2	RAEB-2	10% to 19% blasts ± Auer rods	20
Unclassifiable	MDS-U	Does not fit other categories	10

Table 1: MDS classification.

WHO classifications of the various MDS subtypes (Figure from Arber et al., 2016).

1.11 MDS and the dysregulated spliceosome

Spliceosomal MDS mutations occur mainly in the splicing complex A (Figure 11), however it is not very clear if they confer a gain or loss of function. Absence of homozygous mutations and the presence of missense rather than nonsense mutations suggest a gain of function. However, in the case of *ZRSR2*, the presence of a frameshift and nonsense mutation allows to hypothesize a loss of function. Both alternative and constitutive splicing can be affected when functional changes in the mutated splicing factors occur, considering that most of the mutated genes are involved in the recognition of the 3'ss at the very early stage of spliceosome assembly.

SF3B1, a subunit of the SF3B complex, helps to facilitate the binding of U2 snRNP to the BPS during A complex formation (Lardelli et al; 2010). SF3B1 also interacts with other spliceosomal proteins such as U2AF2 (or U2AF65), which binds the PPT downstream of the BPS. It is implicated in the pathogenesis of a particular MDS subtype denominated refractory anemia with ring sideroblasts (RARS), being found to be mutated in 60%- 80% of these cases whereas only in 6% in MDS without ring sideroblasts (RS) (Papaemmanuil et al., 2011). The disease is characterized by the appearance of an iron accumulation in the erythroblast's mitochondria thus the name RS. Mutations in *SF3B1* are associated with a favorable prognosis and a lower

likelihood to progress to AML (Malcovati et al, 2011). Different studies have shown that SF3B1 mutant cells use cryptic AG dinucleotides lying 10–30 nt upstream of the canonical AG dinucleotide, thus linking it to an aberrant 3'ss usage, although those studies were not done in HSCs (Darman et al, 2015; DeBoever et al. 2015). It is still unknown how mutations in *SF3B1* lead to the formation of the RS even though it has been suggested a relationship between *SF3B1* mutations and the iron transporter *ABCB7*. RNA-seq data of RARS samples have identified a missplicing event in the *ABCB7* gene which introduces a PTC and targets the transcript to NMD degradation, thus resulting in a marked downregulation of the gene in MDS patients (Dolatshad et al, 2016). SF3B1 also forms a complex with BCLAF1, U2AF, and PRPF8 which together with BRCA1 regulate the AS of several transcripts involved in DNA repair and DNA damage response (DDR), thus possibly providing an explanation about why there is an accumulation of DNA damage in HSCs of MDS patients (Savage et al., 2014). Finally, SF3B1 can co-localize with polycomb proteins which regulate gene expression through epigenetics modifications, thus suggesting a role of SF3B1 outside pre-mRNA processing (Isono et al, 2005).

SRSF2 mutations occur in 20–30% of MDS and ~50% of chronic myelomonocytic leukemia (CMML) patients and are associated with a poor prognosis instead (Wu et al. 2012). Despite not being a permanent component of the spliceosome, it is a member of the SR proteins and it is involved in both constitutive and AS by binding to ESE. A recent study has shown that a P95H mutation in this gene result in a phenotype different from loss of function. The mutation can alter the RNA binding sequence specificity of the gene resulting in an altered recognition of specific ESE motifs and ultimately in the mis-splicing of important genes involved in telomere maintenance, DNA repair, and chromatin remodeling. Importantly, P95H mutation reduces the expression of the histone methyl transferase *EZH2*, the latter being an important regulator of hematopoietic differentiation. In more detail, *SRSF2* mutations cause the inclusion of a poison exon in the *EZH2* transcript thus targeting it to degradation by NMD (Zhang et al., 2015).

U2AF1 (or U2AF35) instead is the small subunit of U2AF and it forms a heterodimer with U2AF2. U2AF1 binds to the 3' AG dinucleotide while U2AF2 binds to the adjacent PPT. Several studies have shown that mutations in *U2AF1* alter 3'ss preference in a sequence-specific manner (Przychodzen et al. 2013; Brooks et al. 2014). A recent study has shown also that they promote splicing causing more exon skipping (Graubert et al., 2012). Transcripts with a defect in

splicing caused by the *U2AF* mutations in MDS patients have a role in cell cycle and splicing control (Okeyo-Owuor et al., 2015). In addition, in *U2AF1* mutant cells similarly to *SRSF2* mutants, several epigenetic regulatory genes such as *DNTM3B* are aberrantly spliced. *U2AF2* mutations are mostly found in more proliferative phenotypes such as high risk MDS, are associated with a poor prognosis and patients with those mutations require more intensive chemotherapeutic treatments (Thol et al. 2012).

ZRSR2 is a component of the U12 snRNP and binds to the 3'ss of U12-type introns. RNA-seq analysis of *ZRSR2* mutant cells versus wild-type (WT) has revealed an accumulation of transcripts with retained U12-type introns. Gene Ontology (GO) analysis has shown that the genes affected are involved in MAPK signaling, cytokine signaling and E2F transcription factors (Madan et al., 2015).

An intriguing hypothesis in a recent review by Naro et al. proposes a model by which depletion or delocalization of splicing factor would result in a formation of a three-stranded nucleic acid structure, composed of a DNA:RNA hybrid and a single-stranded DNA (R-loop) during transcription which would trigger DDR in order to impact AS. Thus, they propose a second mechanism by which AS is affected by the mutated splicing factors in addition to their direct impact on branch point/3'ss selection. If this hypothesis is correct it would be consistent with the accumulation of DNA damage in MDS precursor cells (Naro et al., 2015). In case of the presence of damages sites on the DNA, the transcription machinery would be slowed down and the mRNA transcript would fold back on the template DNA thus forming the R-loop. The R-loop would then activate a non-canonical ATM damage signaling pathway and itself be source of DNA double strand breaks (DSBs) (Aguilera, García-Muse, 2012).

Figure 11: Mutated splicing factors in MDS and other hematological diseases.

(A) Mutated components of the spliceosome and their binding locations on the pre-mRNA are shown. (B) Type of mutations and the percentage occurrence in each blood disorder (Figure adapted from Abdel-Wahab and Levine, 2011; Pellagatti and Boultonwood, 2017).

1.12 PRPF40B and other proteins containing the WW and FF domains

The basic steps in spliceosome assembly are well conserved between yeast and humans (Guthrie, 1991). In yeast, the bridging of the 5' and 3' ends of the intron occur via the interactions between the U1 snRNP to the 5'ss of the intron, and the branch point binding protein (BBP) and Mud2p which cooperatively recognize the BPS and the 3'ss respectively (Berglund et al., 1997). Prp40 is a splicing protein in yeast which can simultaneously bind U1 snRNP and BBP/Mud2p thus helping to define the bridging between the two ends of an intron (Abovich et al., 1997). It contains two WW and four FF domains in its primary sequence. Prp40 can bind BBP directly

through the WW domain whereas the binding to Mud2p occurs indirectly. Prp40 shows a protein structure which is shared with a relatively small number of proteins (Figure 12).

Figure 12: Proteins involved in transcription and splicing.

Schematic representation of the structure of proteins which have a dual role in transcription and splicing and possess both WW and FF domains (WW domain in red and FF in blue). yPrp40, yeast Prp40; hPRPF40A and hPRPF40B, human PRPF40A and PRPF40B, respectively; hTCERG1, human TCERG1 (adapted from Becerra et al; 2016).

The above proteins belong to a special class of factors which might be involved in the coupling of splicing and transcription elongation (Becerra et al; 2016). They include two mammalian orthologues, namely PRPF40A (PRP40 pre-mRNA processing factor 40 homolog A) and PRPF40B (PRP40 pre-mRNA processing factor 40 homolog B). The WW domain of PRPF40A is essential for its localization to nuclear speckles. PRPF40A interacts with SF1 and U2AF and it is present in the spliceosome complex A, B but not C which is consistent with the role of Prp40 at the early step of spliceosome assembly (Sanchez-Alvarez, 2010). Immunoprecipitation experiments have shown that PRPF40A can co-precipitate together with U2 snRNP and other proteins involved in transcription elongation (Makarov et al., 2012). PRPF40B is the least characterized orthologue. Even though it does not possess any RNA recognition motif (RRM), recent reports have shown that it plays a role in AS. In particular, a model has been proposed by which PRPF40B can act as a bridge between the U1 bound to the 5'ss and SF1 bound to the BPS (Padgett, 2012). Very recent findings by Becerra et al., have shown that PRPF40B can bind to SF1 through its WW domain while a weak binding to U2AF is modulated by the FF domain

(Becerra et al., 2016). In addition, it appears that the localization to nuclear speckles of PRPF40B is mediated by the second FF domain contrary to PRPF40A where the WW domain is required. Very importantly, the same group has shown that PRPF40B can control the selection of alternative splice sites of important apoptotic genes such as *Fas* (Becerra et al., 2016). They have proposed a model by which weak exons are more sensitive to PRPF40B alterations given the fact that the presence of weak 3'ss and 5'ss and sequences within the exon near the 3'ss are required for PRPF40B-mediated regulation of AS.

Genome sequencing analysis of samples from MDS patients has revealed several mutations in the PRPF40B gene (Yoshida et al., 2011; Jhanwar, 2015) (Figure 13).

Figure 13: PRPF40B protein structure.

Graph depicting the PRPF40B protein structure with domains in different colors (WW domain in red and FF in blue). Arrows indicate the location of the different mutations found in MDS patients (Red color representing a somatic mutation, black a germline mutation). Number of arrow corresponds to the number of MDS patients in which the mutation was found.

Considering that during my Ph.D. I have found a connection between PRPF40B and the regulation of genes involved in hypoxia, I will thus provide background about this pathway in chapters 1.13 and 1.14.

1.13 Tumor Hypoxia

In malignant solid tumors, cancerous cells consume the nutrients and oxygen supplied by the vascular system of the body. In particular, the oxygen levels in some regions of the tumor is lower than the surrounding healthy tissues, due to malformed blood vessels, leading to a condition known as tumor hypoxia (low O₂) (Semenza, 2000). A reduction in oxygen tension elicits a cellular response which ultimately results in the stabilization of hypoxia-inducible

factors (HIFs) (Figure 14). HIFs are heterodimeric transcription factors which consist of a labile alpha subunit regulated by oxygen and a constitutively expressed stable beta subunit (Adams et al., 2009). The alpha subunit is encoded by *HIF1A*, *EPAS1* (known as HIF2a) and *HIF3A* genes. In normoxic conditions, the alpha subunit of HIF is hydroxylated at specific residues by HIF propyl-hydroxylases (PHDs). The Von Hippel Lindau (VHL) E3 ubiquitin ligase then recognizes those residues thus causing HIFs ubiquitination leading to degradation by the proteasome. However during hypoxia, HIFs degradation is inhibited due to the fact that PHDs are oxygen dependent enzymes (Yu et al., 2001). HIF1A protein accumulates in the nucleus where it binds to specific hypoxia-response elements (HREs) in the DNA thus activating the expression of thousands of genes (Figure 14) (Minet et al., 2001). It has been reported that HIF1A expression is post-transcriptionally regulated since its mRNA expression does not change (Wenger et al., 1997). HIF activation is highly associated with enhanced tumor growth and poor clinical prognosis (Harris, 2002). The genes upregulated by HIFs promote the survival of the cells in low-oxygen conditions. The genes regulated are mainly involved in autophagy, angiogenesis, proliferation, migration, erythropoiesis and metabolism. Other hypoxia related pathways are the PI3K/AKT/mTOR, MAPK and NF- κ B pathways.

In particular, in hypoxic conditions a shift between oxidative phosphorylation to altered glycolysis is observed (Chiche et al; 2010). While normal cells metabolize glucose via a mitochondria dependent pathway and generate ATP via oxidative phosphorylation, in cancerous cells the expression of glycolytic enzymes is upregulated together with an increase in glucose transporters (Macheda et al., 2005). In tumors, glucose uptake is highly increased and it is then metabolized through a glycolytic pathway which allows synthesizing ATP in an oxygen-independent manner (anaerobic glycolysis), resulting in a higher production of lactate as end product which in part explains a higher acid load of cancerous cells (Denko, 2008). In conclusion, HIF1A can induce an adaptive response which ensures that the energy requirement is met thus allowing the cells to survive.

The extracellular pH of tumor is lower than normal cells not only because of lactate production but also for the hypoxic induction of the carbonic anhydrases CA9 and CA12, two membrane bound enzymes involved in the conversion of carbon dioxide to carbonic acid. Uptake of

carbonic acid by specific cellular transporters decreases the intracellular pH (Brahimi-Horn and Pouyssegur, 2007).

In an effort to improve blood supply and ameliorate the hypoxic condition, cancerous cells start to secrete factors which stimulate the generation of new blood vessels, a process known as angiogenesis. Hypoxia can increase the expression of vascular endothelial growth factor (VEGF), a potent stimulator of angiogenesis (Ravi et al, 2000).

However, under severe stress, HIF1A might induce a cell death response by stimulating the expression of known pro-apoptotic proteins such as BNIP3, NIX, Noxa and by promoting cell cycle arrest (Goda et al; 2003). Another mechanism by which HIF1A can induce apoptosis is through the stabilization of the tumor suppressor p53. During DNA damage, p53 can induce programmed cell death (PCD) or causes growth arrest. Hypoxia can induce apoptosis by increasing mitochondrial membrane hyperpermeability which results in cytochrome C release in the cytosol. Reactive Oxygen Species (ROS) formation can also contribute to hypoxia induced apoptosis. Finally, another mechanism which induces apoptosis is through the activation of c-Jun (Greijer and Wall van der, 2004). In conclusion, it is the severity of hypoxia which determines if the cells adapt to it and survive or die by apoptosis.

Hypoxia can increase the invasive behavior of cancer cells by upregulating genes involved in epithelial-to-mesenchymal transition (EMT) thus contributing to metastasis. For instance, in hypoxia-induced EMT, a decrease in epithelial-like genes such as E-cadherin and beta-catenin and an increase of mesenchymal-like genes such as N-cadherin and vimentin are observed. Hypoxia can induce the expression of transforming growth factor beta (TGF- β), a master regulator of EMT genes (Zhang et al., 2013). The decrease in pH can also enhance tumor invasion by stimulating the secretion of metalloproteinases which in turn degrade the extracellular matrix (ECM) thus favoring migration of cancerous cells. Enhanced angiogenesis is also associated with metastasis considering the fact that a permeable vascular bed can facilitate extravasation of cancerous cells from the hypoxic environment (Brown and Giacca, 1998).

Finally, hypoxia can stimulate the expression of erythropoietin (EpO), a hormone which helps the body to adapt to low oxygen by enhancing the production of red blood cells. EpO can

stimulate the growth and prevent apoptosis of red blood cells through induction of Bcl-2 and Bcl-XL (Haase, 2013).

Chronic hypoxia is a condition when low oxygen availability persists for longer time and it is associated with accumulation of DNA mutations, high frequency of DNA breaks and inefficient DNA repair potentially leading to genomic instability (Pireis et al, 2010).

Figure 14: Hypoxia pathways.

Illustration showing the pathways activated during hypoxia. HIF1A is either degraded by the proteasome in normoxia or stabilized during hypoxia and translocated to the nucleus where after forming a complex with other TFs induces the expression of the indicated genes (GeneTex).

1.14 Hypoxia in normal hematopoiesis and in leukemia/MDS

In addition to the protein factors secreted by the stem cell niche, other factors regulate hematopoiesis. For instance, a calcium signaling and hypoxia play an essential role in HSCs biology (Tonelli et al., 2012; Zhang and Sadek, 2014). In particular, the bone marrow is a hypoxic organ due to the limited blood supply and by HSCs actively consuming oxygen. HIF1A mRNA is highly expressed in HSCs and HIF1A is believed to keep them in a quiescent state under hypoxic conditions, partly due to the modulation of cyclin-dependent kinase inhibitor (CDKI) genes (Takubo et al., 2010). In addition it has been shown that HSCs localize in hypoxic regions (Parmar et al. 2007).

It has been hypothesized that the hypoxic environment can limit the supply of ROS to the bone marrow thus protecting HSCs from damage considering that an aberrant ROS production has a negative effect on HSC quiescence, self-renewal and differentiation (Suda et al., 2011). Low oxygen level might influence HSCs indirectly through the modulation of stromal cells. In fact, hypoxia can trigger the secretion of factors known to promote HSCs maintenance from mesenchymal stem cells (MSCs) (Anthony and Link, 2014).

Various types of leukemia could arise from leukemic stem cells (LSCs). LSCs have stem-cell like properties and contribute to tumor development and drug resistance. Recent reports have shown that HIF1A plays an important role in LSC tumorigenicity (Deynoux et al, 2016). Examples of HIF1A-dependent leukemia are chronic myeloid leukemia (CML), T cell acute lymphoblastic leukemia (T-ALL), and AML. However, in addition to the hypoxia-induced stabilization of HIF1A in LSCs, several types of leukemia can stabilize HIF1A even in normoxic condition (Piccoli et al., 2007). For instance, in AML, HIF1A signals in cooperation with Notch1 pathway to promote self-renewal in normoxia. HIF1A can promote disease progression via different mechanism including regulation of metabolism, cell cycle and immune function. Since quiescence is a consequence of HIF1A expression, HIF1A induction can favor chemoresistance (Petit et al., 2016).

In leukemia, hypoxia in the bone marrow can induce an EMT-like phenotype in CML cells for instance by increasing the CXCR4 expression which results in increased HSC chemotaxis or

decreased expression of E-cadherin thus resulting in a lower adhesion to the stroma (Percio et al, 2014).

Recent studies have reported an increased expression of HIF1A in bone marrow cells of MDS patients and have shown that HIF1A can be used as a marker of adverse prognosis. HIF1A expression correlates with blast percentage, hemoglobin count and cytogenetic abnormalities (Tong et al., 2012). It is then believed that therapeutic targeting of HIF1A may benefit many MDS patients. Cytopenias, leukemic blasts and absence of functional reconstitutable HSCs may explain the activation of HIF1A due to low oxygen tension in MDS cells. HIF1A induction might in turn activate proliferation pathways that are important for the transition from MDS to AML. In fact, MDS cells can grow more robustly under hypoxic conditions (Thomson et al., 2007). In addition, low oxygen tension can induce expression of the micro RNA mir-210 which in turn in a feed forward loop promotes cancer cell survival by inhibiting SHIP-1, a tumor suppressor gene (Lee et al., 2012). Another report has also shown the presence of a strong hypoxia gene signature in MSCs of MDS patients which might contribute to their malignant behavior (Medyouf et al., 2014).

1.15 Context of this project

Despite it has been established that AS plays a role in the activation of cells of the lymphoid lineage, conferring to these cells the plasticity needed for an immune response, very few studies on AS were done at the beginning of my thesis in other cells of the immune system, such as the cells of the myeloid lineage. Furthermore most of the studies investigating the role of AS in monocytes have been performed using mouse monocytes whereas there is still no general study reported in the literature involving primary human monocytes.

In addition to this, the finding that various splicing factors are mutated in blood disorders affecting white blood cells, suggests that AS might have a previously unrecognized role in the proper differentiation/activation process of the cells of the myeloid lineage. Thus in this study we are investigating the role of AS in monocytic differentiation and in MDS pathogenesis.

1.16 Aims

Two are the main aims of this Ph.D. project:

1. To characterize the global changes in mRNA expression and AS in human monocytes upon various differentiation stimulations.

A monocytic cell line and primary human monocytes were used and differentiation to different subsets of macrophages was induced by various types of stimulations. Subsequently RNA-seq was employed to identify AS events by comparing terminally differentiated/activated cells to undifferentiated ones. We discovered novel events that play a role in the differentiation/activation process and thus fill the gap in our knowledge of the role of AS upon differentiation of human monocytes.

2. To investigate the role of the splicing factor PRPF40B in AS and in MDS.

I aimed to reveal the AS events regulated by this factor by knocking it out by CRISPR. I performed rescue experiments in knock out cells with WT and mutated PRPF40B and proceed with RNA-seq in order to analyze the transcriptome to find the AS events regulated by PRPF40B. Our main aim is to investigate the AS regulated by the WT gene and in the meanwhile demonstrate a possible link between the mutated gene and MDS pathogenesis.

2 MATERIALS AND METHODS

2.1 Cell culture, differentiation and polarization

THP-1 cell line was obtained from American Type Culture Collection (ATCC) (cat no TIB-202), and maintained between 2×10^5 cells/ml and 1×10^6 cells/ml in Roswell Park Memorial Institute medium (RPMI) 1640 modified medium (ATCC) supplemented with 10% Fetal Bovine Serum (FBS) (Gibco, Life Technologies), 1% Penicillin/streptomycin (Pen/Strep) (Gibco, Life Technologies), 0.05 mM β -mercaptoethanol (Gibco Life Technologies) and incubated at 37 °C with 5% CO₂. For differentiation and polarization of THP-1, cells were seeded in 6 well plates at a concentration of 3×10^5 cells/ml in Dulbecco's Modified Eagle Medium (DMEM) high glucose medium (Hyclone, Thermo Scientific) supplemented with 10% FBS (Gibco, Life Technologies), 1% Pen/Strep (Gibco, Life Technologies). Cells were treated either with Dymethyl Sulphoxide (DMSO, Sigma-Aldrich) as control, 160 nM Phorbol 12-Myristate 13-Acetate (PMA, Sigma-Aldrich), 0.5 μ M 1 α , 25-dihydroxy Vitamin D₃ (VD3) (Calcitriol, Cayman Chemical) for 3 days or a combination of 0.5 μ M 1 α , 25-dihydroxy VD3 (Calcitriol, Cayman Chemical) and 1 μ M Retinoic Acid (Sigma). Differentiation of PMA treated cells was further enhanced by replacing the PMA-containing media with fresh DMEM high glucose medium (Hyclone, Thermo Scientific) supplemented with 10% FBS (Gibco, Life Technologies), 1% Pen/strep (Gibco, Life Technologies) for 5 days (Daigneault et al., 2010). For polarization of THP-1 cells, cells were treated with 160 nM PMA (Sigma-Aldrich) for 3 days and then incubated for 8 h either with a combination of 20 ng/ml IFN- γ (Peprotech) and 100 ng/ml LPS (Sigma-Aldrich) for the pro-inflammatory macrophages, or a combination of 20 ng/ml IL-4 (Peprotech) and 20 ng/ml IL-13 (Peprotech) for the anti-inflammatory macrophages.

2.2 Isolation of primary human macrophages and relative differentiation/polarization methods

For primary monocyte purification, Peripheral Blood Mononuclear Cells (PBMCs) were isolated from four healthy donors by Histopaque-1077 (Sigma-Aldrich) gradient. Subsequently, the samples were incubated with the negative selection microbeads (Pan Monocyte Isolation Kit, cat.no130-096-537, Miltenyi Biotech) to obtain purified monocytes.

Briefly, 60 ml of total blood per donor was drawn into 9 ml Vacuette EDTA coated tubes (cat no 455036, Greiner), transferred to 50 ml falcon tubes and then 1X Phosphate Buffered Saline (PBS) was added to it with a ratio of 1:3. A total of 30 ml of the mixture was poured into filtered Leucosep tubes (Greiner Bio One), followed by centrifugation at 800 g, for 15 min at room temperature and with zero deceleration. Then, the PBMCs layer was harvested and washed with MACS buffer twice. A total of 1×10^7 PBMCs were resuspended in 30 μ l MACS buffer, followed by addition of FcR blocking reagent and mixture were with Biotin-Antibody Cocktail and Anti-Biotin MicroBeads. Mixtures were passed through LS columns standing on magnetic separator, and the flow-through with the unlabeled monocytes was collected. After washing with MACS buffer once, the monocyte sample was collected directly for flow cytometry analysis and total RNA isolation. Purity of monocytes was assessed by CD14 staining.

Monocytes were differentiated to macrophages by culturing the purified monocytes with the different cytokines for 7 days in RPMI 1640 medium (cat.no 30-2001, ATCC) supplemented with 20% FBS (Gibco) and 1% Pen/Strep (Gibco, Life Technologies) at 37 °C and 5% CO₂. A concentration of 100 ng/ml of GM-CSF was used to induce M1 macrophages, and 100 ng/ml of M-CSF to induce M2 macrophages. Every 3 days the culture media was changed with addition of fresh cytokines. On day 7, GM-CSF macrophages were polarized by 100 ng/ml LPS plus 50 ng/ml IFN- γ for 8 h. In turn, M-CSF macrophages were incubated for 8 h with 10 ng/ml of either IL-4, Interleukin 1- β (IL-1 β) or IL-10 for the three polarizations. At the end, differentiated macrophages were washed with 1X PBS and then either detached with Accutase (A6964, Sigma-Aldrich) for flow cytometry analysis or directly lysed for extraction of total RNA.

2.3 Flow cytometry for assessment of macrophage differentiation (cell line and primary cells)

Harvested cells were washed with 1X PBS and were incubated with 2 μ l of 1 mg/ml goat serum (Sigma-Aldrich) for 5 min for Fc receptors blocking on ice prior to antibody staining. For detection of cell surface markers several monoclonal mouse anti-human antibodies were employed (Table 2):

Table 2: List of antibodies for flow cytometry.

Antigen	Fluorochrome	Supplier	Sample Type
CD36	FITC	Biolegend	THP-1
CD11b	PE	Biolegend	THP-1
CD54	PE	Biolegend	THP-1
CD14	APC	Biolegend	THP-1 & Primary cells
CD86	FITC	Miltenyi	THP-1 & Primary cells
CD163	FITC	Miltenyi	Primary cells
CD80	PE	Miltenyi	Primary cells
CD16	PerCP-Vio700	Miltenyi	Primary cells
HLA-DR	PE	Miltenyi	Primary cells
CD206	APC-Vio770	Miltenyi	Primary cells
CD71	VioBlue	Miltenyi	Primary cells
CD206	PE	Biolegend	THP-1
CD33	APC	Biolegend	THP-1
CD68	PE	Biolegend	THP-1

Primary antibodies or the relevant isotypes were incubated with samples for 10 min on ice. Afterwards, cells were washed two times in buffer and resuspended in 300 µl of 1X PBS. Flow cytometric measurements were performed at the FACS facility in SBS using 5 lasers LSR Fortessa X-20 (Becton Dickinson). For unstained cells, Forward (FSC) and Side (SSC) scatters were used to identify the two cell populations, undifferentiated cells and differentiated cells respectively, and to measure the cell size and granularity. Measurement of autofluorescence was taken in the FL-1 channel (excitation at 488 nm wavelength, emission at 530/30 nm wavelength). Data analysis was performed by Cyflogic (v.1.2.1) or alternatively FlowJo (v10.0.7).

2.4 Live cell imaging

THP-1 cells were seeded on tissue-culture treated μ -Dish (81156, Ibidi GmbH, Germany) at a density of $1 \times 10^5/\text{cm}^2$ in DMEM high glucose medium (Hyclone, Thermo Scientific) supplemented with 10% FBS (Gibco, Life Technologies), 1% Pen/strep (Gibco, Life Technologies). Image capture in bright field was performed in a Zeiss Live Imaging Microscope at 37 °C and 5% CO₂, and image was processed by ImageJ software. Experiment was mainly done by Dr. Hongfei Liu but I contributed for a 10% of the work.

2.5 Total RNA extraction and sample preparation for RNA sequencing (RNA-seq)

Total RNA from three replicates of undifferentiated, differentiated and polarized cells was extracted by PureLink Mini RNA kit (Life Technologies) according to the manufacturer's protocol and the precipitated RNA was dissolved in RNase free water. Briefly after resuspending cells in lysis buffer, samples were homogenized and 70% ethanol was added to the homogenized lysate. The lysate was passed through the RNeasy mini-column to allow RNA binding to membrane. The flow through was discarded and Wash Buffer I or II were applied successively to wash the column. RNeasy column was then carefully transferred to a new 1.5 ml collection tube and elution was performed with RNase-free water. RNA quality was assessed by gel electrophoresis and by measuring the absorbance using a Nanodrop Spectrometer (Thermo Scientific). In order to remove genomic DNA contaminations, samples were treated with TurboDNase (Life Technologies) according to the manufacturer's protocol. Briefly, TurboDNase (2 U/ μ l) was added to each sample and incubated at 37 °C for 30 min. Subsequently each reaction was inactivated and RNA resuspended in RNase free water. After DNase treatment, RNA quality was assessed by gel electrophoresis and by measuring the absorbance using a Nanodrop Spectrometer (Thermo Scientific) to ensure no traces of DNA were left. Finally, samples were purified and concentrated by RNA Clean & ConcentratorTM-5 (Zymo Research) according to the manufacturer's protocol. After RNA concentration, RNA quality was assessed by Picogreen assay and RNA Integrity Number (RIN) for each sample was determined by Bioanalyzer 2100 system (Agilent) at the Singapore Centre for Environmental Life Sciences Engineering (SCELSE) facility in NTU.

2.6 Strand-specific RNA-seq library preparation (performed by SCELSE) and RNA-seq

Total RNA was converted into libraries of double stranded cDNA for high-throughput sequencing. For each sample, strand specific libraries were prepared by Illumina Truseq Stranded mRNA sample preparation kit. Each sample was barcoded in a way that all libraries could be pooled together and distributed across various sequencing lanes. Pooled libraries were then sequenced by Illumina HiSeq 2000 DNA sequence analyzer. After cluster generation, 101 bp paired end reads were generated.

2.7 RNA-seq data analysis

The total number of paired end reads received after sequencing was calculated for each sample and Quality Control (QC) on the obtained reads was performed by FastQ tools. Each read has a length of 101 bp. In order to discover Differentially Expressed Genes (DEGs) and the AS events that change during the differentiation and polarization, a RNA-seq bioinformatics pipeline was implemented in the lab as a team work together with my co-supervisor Prof Zheng Jie (Figure 15). Briefly, reads after QC were aligned to the reference human genome or transcriptome by TopHat (Trapnell et al., 2009). Expressed genes and transcripts were assembled by Cufflinks and for each gene Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were calculated (Ghosh and Chan, 2016). Cuffdiff, a subprogram of Cufflinks, was used to identify DEGs. For each gene, fold change was calculated based on the FPKM values of two different comparisons. CummeRbund, another subprogram of Cufflinks, was used to calculate Pearson's correlation between the replicates (Trapnell et al., 2013). MATS, MISO and SpliceTrap were used to obtain the AS events (Katz et al., 2010; Wu et al., 2011; Shen et al., 2014). MATS and MISO use the aligned reads by TopHat to perform the AS analysis while SpliceTrap use its own database to perform the alignment of the reads followed by AS analysis. MATS is the only tool which can handle replicates while the other two do not.

Regarding SpliceTrap, we first mapped the original paired end reads to SpliceTrap database hg19, with low cutoff. We performed postAnalysis subsequently, with junction reads required per junction (-j 2) and low cutoff (-c L). Finally, we compared exon inclusion ratio between each two samples using SpliceChange. As a cutoff we used inclusion level difference ≥ 0.05 and p-value ≤ 0.001 . Briefly, in MISO analysis, first we used the alignment results to estimate isoform

expression levels within each sample and we made pairwise comparisons between samples to detect differentially expressed isoforms. We used Bayes_factor min (BFmin) and Bayes_factor median (BFmedian) as criteria of statistical significance among the replicates ($\text{BFmin} \geq 1$ & $\text{BFmedian} \geq 10$). To run MATS, we adopted the alignment outputs of Tophat and we applied the cut-offs of inclusion level difference ≥ 0.05 and False Discovery Rate (FDR) ≤ 0.05 in order to get the AS events.

Figure 15: Bioinformatics pipeline.

Flow chart describing the bioinformatics pipeline used to perform the RNA-seq data analysis in order to discover DEG and novel AS events.

2.8 cDNA synthesis and RT-PCR

A total of 1 μg of RNA was heated up together with deoxynucleotide (dNTPs) and oligodT at 65 $^{\circ}\text{C}$ in order to disassociate RNA secondary structures. RNA was cooled down and cDNA was

synthesized with M-MuLV Reverse Transcriptase (Biolabs, NEB) at 42 °C for 2 h. At each step, quantity and quality of the RNA and of the cDNA was assessed by Nanodrop Spectrophotometer (Thermo Scientific). PCR was performed by using gene-specific primers and a GoTaqFlexi (Promega), DreamTaq (Thermo Scientific) or alternatively a Prime Star DNA Polymerase. The PCR conditions for the GoTaq Polymerase were as follows: 39 cycles of denaturation at 95 °C for 30 s, annealing at respective temperatures for 40 s and extension at 72 °C at a polymerization speed of 1 min/kb. The PCR conditions for Prime Star Polymerase (Takara) were as follows: 39 cycles of denaturation at 98 °C for 30 s, annealing at 58 °C for 15 s and extension at 72 °C at a polymerization speed of 5 sec/kb. The PCR conditions for DreamTaq Polymerase were as follows: 39 cycles of denaturation at 95 °C for 30 s, annealing at respective temperatures for 50 s and extension at 72 °C at a polymerization speed of 1 min/kb.

PCR products were loaded on an agarose gel containing an agarose percentage adjusted based on the size of the PCR product. PCR bands were visualized by a gel doc system and band intensities were quantified using Quantity One v 4.6.9.

2.9 Semi-quantitative (Semi-Q) RT-PCR

A total of 10 pmol forward primer was 5' end labeled using 10 U T4 PNK (New England Biolabs, USA) and 10.2 pmol γ -32P-ATP (60 μ Ci; Perkin Elmer, USA). Labeled primers (Table 3) were purified by illustra Microspin G-25 column (GE Healthcare Life Sciences, USA) and added into primer mixture containing 100 pmol unlabeled reverse primer and 90 pmol unlabeled forward primer (Table 3). cDNA was amplified by PCR using Go-Taq polymerase (Promega, USA) with the radiolabeled primer mixture for 23-35 cycles. The PCR products were then separated by 8-10% (w/v) native polyacrylamide gel (Acrylamide/Bis 29:1; Biorad, USA) in 1X Tris/Borate/EDTA (TBE) buffer. Radioactive signals from the gel were subsequently exposed to storage phosphor imaging screen which were read by Typhoon Scanner. The analysis of band intensity was done with ImageQuant TL software (GE healthcare life sciences, USA). For gel images, X-ray films (Kodak, USA) were exposed with the gel for 12-48 h in X-ray cassette at -80 °C and then processed by Kodak Model 2000 X-ray film processor.

Table 3: List of primers for the Semi-Q RT-PCR validation.

Gene Name	Forward	Reverse
FAM126A	AGGCTTCATTGCCTCATGGT	CTCCTGACCGTCTGTGGTTC
RAI14	TGCAGGAATTCAAAGCCTTCT	ACCGATTCCCTTTCCCGATAGTG
STAT2	GAAGCTGCACTTGGGAGTGA	ATCTCAGCCAACTGGGTAGG
SAP25	CTGCCCCACCACCTGC	ATACAGCGGCCAGAATGAGG
RCOR3	TGCTAAGCAAGTAAACAGTGC	TCACTTGGCCAACAGTCTTGT
FMNL3	GGTACCATCGAGGACATCATCA	TGGTGGCGAACAACCATAGG
ZNF678	TTCTCTCCAGAGGAGTGGGC	CAAATTGTCAAGGCCCCAGC
AKAP17A	GCTGGAGAAGCTGCAGGCGG	CGCGCAGCTCCCGCTCTTTC
LARP7	AGACATCATTAAAGGAAGCATCAGA	TCTTAATGGTATAACTTCTTCTCCC

2.10 Protocol for CRISPR Knockout (KO) of human *PRPF40B* in K562 cell line

In order to knockout human *PRPF40B*, a modified CRISPR method has been employed as described in the following paragraphs.

2.10.1 Cell Culture

K562 CML cell line was maintained in DMEM (Hyclone, USA) supplemented with 10% FBS (Gibco, USA), and 1% Pen/strep (Gibco, Life Technologies). The cells were incubated at 37 °C and 5% CO₂. CRISPR/Cas9-edited K562 cells were generated and maintained in the same media with the exception of increasing the FBS concentration to 20% in order to favor the survival of the single cells clones.

2.10.2 CRISPR single guide RNA vector design

To generate *PRPF40B* KO in K562, firstly a pair of 20 nt guide sequences each targeting two different introns of *PRPF40B* were designed (Figure 16B). The first forward single guide RNA (sgRNA) was designed targeting intron 4 while the reverse sgRNA targeting intron 8 as shown in the below diagram (Figure 16B).

The sgRNAs were designed in a way to remove a portion of the gene between exon 5 to exon 8, which would result in a frameshift in the mRNA and no protein (Fig 16A).

Figure 16: CRISPR PRPF40B KO design.

Upper panel (A) showing the three isoforms of *PRPF40B* and the deleted common region between exon 5 to exon 8 highlighted in red (UCSC genome browser). Lower panel (B) showing the individual sizes and the total sum in base pairs of the deleted exons. Red arrows indicate the binding regions of the two sgRNAs used for the CRISPR.

The reason why introns were targeted instead of exons is that based on our experience with previous experiments, it was found easier to generate a deletion which will lead to an out-of-frame transcript with this strategy. Previous genomic DNA Sanger sequencing results of deletions using sgRNAs targeting *PRPF40B* exons have shown that the Cas9 cutting site can be either 3 bp or 4 bp upstream of the NGG protospacer adjacent motif (PAM sequence). In the case of *PRPF40B* it was found that during clonal selection, cells with a deletion which will result in an in-frame transcript were preferentially selected in certain cases. Knowing exactly whether the

Cas9 cuts 3 bp or 4 bp upstream of the PAM would be relevant to know if the transcripts would be in or out-of-frame, in case of targeting exons. Thus, by targeting introns, the possibility that a population of cells with the in-frame transcript would be selected was eliminated. The 20 nt sgRNAs excluding the PAM sequence, were cloned individually into the pSpCas9 (BB)-2A-GFP (PX458; Addgene) vector which expresses the WT Cas9 enzyme. A pair of sgRNAs expressed from these plasmids would induce Cas9 mediated DSBs at the two sites, deleting the intervening DNA segment due to precise end joining (Zheng et al., 2014).

All guide sequences (Table 4) were designed using the CRISPR design tool CHOPCHOP (Montague et al., 2014).

Table 4: CRISPR sgRNAs designed online.

sgRNA label	Oligo sequence
FW gRNA PRPF40b Intron 4	GAGCCGTCTGCCGAAGAAAGAGG
FW gRNA PRPF40b Intron 8	AGGGTGAGATGTCCTACGGGTGG

Note: Output sequences from CHOPCHOP are reported with the inclusion of the PAM in red.

To generate the two plasmids containing the sgRNAs, PX458 vector was digested with BbsI at 37 °C for 3 h. A pair of complementary DNA oligonucleotides containing guide sequence without the PAM was ordered from Integrated DNA Technologies (IDT) (Table 5) and was first annealed by heating equal molar amounts of the two oligonucleotides at 37 °C for 30 min followed by 95 °C for 5 min and gradually cooling to room temperature in 1 h. It is of outmost importance not to include in the sgRNAs the PAM nucleotides because by doing so the Cas9 would cut the plasmid as well as the targeted gene based on my own experience. In addition, it is advisable to sequence the DNA region targeted by the sgRNAs to exclude the presence of Single nucleotide polymorphisms (SNPs) which would affect the Cas9 cutting efficiency. The annealed oligonucleotides were then 5' end phosphorylated by T4 polynucleotide kinase (T4 PNK; New England Biolabs, USA) in 1X ligation buffer [50 mM Tris-HCl, 10 mM MgCl₂, 1 mM ATP and 10 mM DTT] and ligated into PX458 vector that was cut by BbsI.

Table 5: CRISPR sgRNAs ordered for cloning.

sgRNA label	Oligo sequence
FW gRNA PRPF40b Intron 4	CACCgGAGCCGTCTGCCGAAGAAAG
RW gRNA PRPF40b Intron 4	AAACCTTTCTTCGGCAGACGGCTC ^c
FW gRNA PRPF40b Intron 8	CACCgAGGGTGAGATGTCCTACGGG
RW gRNA PRPF40b Intron 8	AAACCCCGTAGGACATCTCACCCCT ^c

Note: sgRNA (20 nt) is in black font, BbsI overhang is in blue font. An extra g (shown in black lower case) is added at the 5' of the sgRNA where the 20-nt guide sequence does not begin with g because the U6 RNA polymerase III promoter used to express the sgRNA prefers a guanine (g) nucleotide as the first base of its transcript. All DNA oligonucleotides are in 5' to 3' direction.

Cas9 sgRNA plasmids were transformed into competent DH5 α *E. coli* cells via heat shock. Selection for successfully transformed *E. coli* cells was done using broth agar plates containing 70 μ g/ml of ampicillin (Amp). Colonies were grown in 3 ml of Lysogeny broth (LB) overnight at 37 °C. The plasmids were then extracted using MiniPrep kit (Omega Bio-tek, USA). All purified plasmids were sequenced (1st Base, Singapore) to ensure insertion of correct sgRNA.

2.10.3 Nucleofection

K562 cells were maintained in DMEM with a concentration between 1×10^6 and 2×10^6 cells/ml. When cells reached a maximum concentration of 0.7×10^6 cells/ml, cells were nucleofected. Briefly, 2×10^6 K562 cells were resuspended in 100 μ l SF reagent (Lonza) with 2.5 μ g of both Cas9/sgRNA CRISPR plasmids. Nucleofection was performed according to manufacturer's instruction by using the FF-120 program with Amaxa 4D-Nucleofector (Lonza, Switzerland). After 24 h, cells were sorted for green fluorescent protein (GFP) expression.

2.10.4 Fluorescence-activated cell sorting (FACS)

Post 24 h of nucleofection, cells were resuspended in culture media and sorted for expression of GFP using the BD FACS Aria cell sorter (Becton Dickinson, USA). The entire populations of GFP positive cells were sorted on the first day. Cells were maintained in DMEM in a 12-well plate for an additional day. After 24 h the first FACS sorting, cells were sorted again. FACS

sorting was done for the top ~2 % of GFP positive cells in order to enrich for cells that received high levels of the CRISPR/Cas9 plasmids. The reason why a double sorting strategy was performed instead of one single sorting was to further avoid the possibility that cells without any deletion were selected during later stages of cell culture. A significant loss of GFP was observed after the first sorting which might explain a selection for unedited cells. After the second sorting, cells were maintained either as a “bulk” population in a 12 well plate at first and later on transferred to a 6 well plate or plated in 96 well plates as single cell clones (Figure 17).

Figure 17: CRISPR Workflow.

Schematics of the workflow for the CRISPR KO of *PRPF40B*. Two vectors expressing the sgRNAs were used to co-transfect K562 cells and a double FACS sorting method was adopted to enrich for cells highly expressing the Cas9 protein. Afterwards, single cell clonal selection and PCR genotyping of the selected clones were performed.

2.10.5 Single cell clone selection

Immediately after the second sorting around 200 cells were diluted with 20 ml of DMEM and plated into two 96 well plates. Three days after, cells were counted and plated again in two 96 well plates. Lastly, after six days cells were plated into three 96 plates. The reason why cells were seeded at different days was because we noticed that if bulk cells were first expanded for few days after the FACS sorting, the survival of the single cell clones would increase exponentially. For instance, approximately 10% of the clones survived from the first seeding, which increased to 40 % and 90 % on the 2nd and 3rd seeding respectively. Various clones were subsequently frozen in freezing media containing FBS and 5% DMSO. It was observed, after thawing some of them, that if the frozen clones were not single cells with a complete deleted region, the non-edited cells would have a selective advantage during the recovery stage compared to the cells with a deletion.

2.10.6 Genomic DNA extraction and genotyping CRISPR-edited *PRPF40B*

K562 genomic DNA was extracted by incubating cells with lysis buffer [100 mM NaCl, 100 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 7.5)] at 55 °C for 1 h. DNA was then subjected to ethanol precipitation by adding 2X volume of 100% ethanol and 0.1X volume of 3 M sodium acetate (pH 5.2) and incubated at -20 °C overnight. DNA was centrifuged at 13,500 rpm for 15 min and then washed with 70 % ethanol twice before resuspending the pellets in 150 µl water.

Two different genotyping PCRs were employed (Fig 18).

Figure 18: Genotyping PCRs.

Genotyping PCR schematics adopted for screening of the single cell clones. In red and blue are the binding regions of the two sgRNAs. Green arrows depict the primers used in the PCR. Primer P1 was used together with primer P3 for the first round of screening. Afterwards, a second round

of validation on the same samples was performed by using primer P2 together with primer P3 to further confirm the deleted region.

In the first PCR, a forward primer upstream and a reverse primer downstream of the sgRNA cleavage sites were used. In the second, one forward primer targeting a sequence inside the deleted region was designed and paired with the downstream reverse primer previously designed to target the portion downstream of the deletion. Sequences of the primers are indicated in Table 6. The reason why an additional pair of primers was used is that a bias towards the lower amplified band during PCR reaction was found if the first pair of primers targeting upstream and downstream of the deletion was used. In certain cases, a high percentage of false positive clones were found most likely due to this PCR bias. By using the second pair of primers the bias towards the lower band was excluded because the comparison between the WT and KO samples would imply a band versus no band.

Table 6: List of primers for the two genotyping PCRs.

Genotyping PCR primers label	Oligo sequence
FW deletion Intron4-Intron8(P1)	TCTCAACAAAATACTCTCACTTGAG
RW deletion Intron4-Intron8 (P3)	CCACAGTGTAGAGATCATGGCGACAG
FW inside deletion Intron4-Intron8 (P2)	CTTAGTCTCTAACCATATTCATGAT

Note: For the first genotyping PCR the first pair of primers (P1) and (P3) was used. For the second genotyping PCR a second pair (P2) and (P3) was used.

PRPF40B CRISPR KO clones were genotyped by PCR using DreamTaq polymerase (Thermo Scientific, USA) and PCR conditions in Table 7.

Table 7: PCR conditions for the genotyping PCRs.

Cycling step	Temperature (°C)	Time	
Initial denaturation	95	2 min	
Denaturation	95	30 sec	} 39 cycles
Annealing	56	50 sec	
Extension	72	50 sec	
Final extension	72	5 min	

WT genomic fragment for K562 and truncated genomic fragments for KO clones were resolved by 1.5% (w/v) ethidium bromide-stained agarose gel and visualized by UV light. Positive clones (showing truncated PCR products with primers P1 and P3) were selected and sequenced (1st Base, Singapore) to confirm deletion of the intended genomic region.

2.10.7 CRISPR qRT-PCR

Quantification of PRPF40B transcript was done using Applied Biosystems 7500/ CFX96 Touch™ (Bio-Rad, USA) Real-Time PCR detection system and SYBR Select Mastermix (Applied Biosystems). PCR was set up in 96 well microplates containing 7.5 µL of Master Mix, 2.5 µL of cDNA, and 1 µL of forward and reverse primer for a total of 11 µL reaction volume. Primer sequences used to amplify β-actin, PRPF40B transcripts are shown in Table 8. The conditions for PCR are indicated in Table 9.

Table 8: List of primers for the qRT-PCR to assess PRPF40B mRNA expression level.

qRT-PCRprimers PRPF40b	Oligo sequence
FW qRT PRPF40B	GCCCTATGGAGTGAGCATGTGG
RW qRT PRPF40B	TATAGTAATAAGGTTTGCCTGT

qRT-PCR primers β -actin	Oligo sequence
FW qRT β -actin	CCAGAGGCGTACAGGGATAG
RW qRT β -actin	CCAACCGCGAGAAGATGA

Transcript levels of β -actin were used to normalize PRPF40B endogenous and overexpression transcripts. Ct values were used to calculate relative mRNA expression by the $2^{-\Delta\Delta C_t}$ relative quantification method (Livak and Schmittgen, 2001).

Table 9: Real Time PCR conditions.

	Temperature (°C)	Time	
Initial denaturation	95	3 min	
Denaturation	95	10 sec	} 39 cycles
Annealing	58	30 sec	
Extension	72	20 sec	

2.11 PCR mutagenesis of PRPF40B

PCR mutagenesis was employed in order to generate vectors for the PRPF40B isoform C and the two of the MDS mutations P383L and P540S (Figure 13). The cDNA Open Reading Frame (ORF) clone for the longest isoform of the human *PRPF40B* gene was purchased from GenScript (Clone ID OHu66462). Mutagenesis PCR using KAPA HiFi DNA polymerase (KAPA Biosystems, USA) was applied to introduce deletion, insertion or nucleotide substitutions into the WT *PRPF40B* plasmid. The strategy adopted was to first derive from the longest isoform of 2679 bp, a shorter sequence of *PRPF40B* of 2616 bp, which transcribes an alternative mRNA isoform C and which corresponds to the canonical sequence of the gene (Figure 19).

PCR mutagenesis was carried out by using the primer pair “FW 2616 deletion” and “RW 2616 deletion” (Table 10).

PRPF40B Chromosome 12

Figure 19: PRPF40B UCSC genome browser isoforms.

UCSC genome browser graph showing the 3 different isoforms of *PRPF40B*. Part highlighted in red represents the portion of the gene which was deleted by PCR mutagenesis. The lengths of the isoform are: Isoform (A) 2679 bp, isoform (B) 2541 bp and isoform (C) 2616 bp.

Note: longest isoform (A) was purchased from GenScript. Isoform (C) was derived from (A) by deleting the portion at N terminus highlighted in red. The deleted region does not contain the epitope recognized by the PRPF40B antibodies used for Western blot (WB) which is at the C-terminus. The forward and reverse pcDNA sequencing primers (pcDNA-FW, pcDNA-RW) and one additional primer in the middle of the sequence (FW PRPF40b seq 850 bp) were used to confirm the DNA sequence by Sanger sequencing (1st Base, Singapore) (Table 10).

Table 10: List of primers used to clone and sequence PRPF40B isoform C.

Primer Name	Sequence
FW 2616 deletion	CTCGGATCCGCCACCATGATGCCACCACCC
RW 2616 deletion	GGGTGGTGGCATCATGGTGGCGGATCCGAG
FW PRPF40b seq 850bp	GGAGGAATCAAAGCCAGAACCAGAG
pcDNA-FW	GAGACCCAAGCTGGCTAGCGTT
pcDNA-RW	GAGGCTGATCAGCGGGTTTAAAC

In addition, it was also found the presence of a NAGNAG sequence towards the end of the isoform (C). NAGNAG sequences are tandem alternative 3'ss which can be found across the

genome (Hiller et al., 2004). They consist of two different AGs preceded by any of the 4 nucleotides (N) which are alternatively used leading to the exclusion or insertion of one amino acid. The amino acid missing was then introduced by PCR mutagenesis using the primers indicated (Table 11).

Table 11: List of primers to insert one NAG into the PRPF40B isoform C.

Primer Name	Sequence
FW NAGNAG	CCTACAGGTGCTGGAGCAGACTGAATGCCAGCACCTCC
RW NAGNAG	GGAGGTGCTGGCATTCACTCTGCTCCAGCACCTGTAGG

At last, P383L and P540S mutations were introduced into the canonical WT sequence by PCR mutagenesis (Table 12).

Table 12: List of primers to introduce two MDS mutation into PRPF40B WT.

Primer Name	Sequence
FW P540Smuta	GCCAGCCGGGCTCCACCTCTCTGGACTTATTCAAGTTCTATGTGGAGGAG
RW P540Smuta	CTCCTCCACATAGAACTGAATAAGTCCAGAGAGGTGGAGCCCGGCTGGC
FW P383Lmuta	AGGTCTGGGCTGTGGTCCTTGAGAGGGATCGAAAAGAGGTTTATGATG
RW P383Lmuta	CATCATAAACCTCTTTTCGATCCCTCTCAAGGACCACAGCCCAGACCT

To confirm the correct sequences of the NAGNAG and the mutations introduced, several sequencing primers were designed (Table 13 and Figure 20).

Figure 20: PRPF40B vectors sequencing primers design.

Graph describing the design of the primers used for the sequencing of the plasmids. In black (pcDNA FW and RW) are indicated the two external primers used to check the beginning, end of the sequence and the introduced NAGNAG (adapted from GenScript).

“P383L Primer” indicates the location of the primer used to sequence the P383L mutation while “P383L (1213 bp)” indicates the position of the mutation. Similarly, “P540S Primer” indicates the primer used to sequence the P540S mutation while “P540S (1684 bp)” indicates the position of the mutation. “NAGNAG (2119 bp)” indicates the position of the NAGNAG where we introduced the three nucleotides or one amino acid. Lower part of the graph showing map of the backbone of the vector used for the cloning.

Table 13: List of sequencing primers for the two MDS mutations and PRPF40B isoform C.

Primer Name	Sequence
FW P383L seq	CAACCAGAGTAAAGAGTCCCGCTGG
FW P540S seq	CTCTTCTTCCTGGCCAAGAAGGAGA
pcDNA-FW	GAGACCCAAGCTGGCTAGCGTT
pcDNA-RW	GAGGCTGATCAGCGGGTTTAAAC

Note: Sequencing primer “FW P383L seq” at 451 bp was used to check the correct introduction of the P383L mutation while primer “FW P540S seq” at 1176 bp was used to check the correct the P540S mutation. “pcDNA-FW” was used to check the deletion of the longer isoform while “pcDNA-RW” was used to check the NAGNAG sequence (Figure 20).

In addition, from the sequencing results it was found that the DYK-FLAG tag is at the N-terminus and not at the C-terminus like it is indicated by the plasmid map from the vendor.

PCR conditions for the mutagenesis were as follows: one cycle of denaturation at 95 °C for 2 min, 18 cycles of denaturation at 98 °C for 30 s followed by annealing at 50 °C or 55 °C for 1 min and extension at 72 °C for 5 min, and one final extension for 5 min. Total PCR products were treated with 20 U DpnI (New England Biolabs, USA) to remove methylated PCR templates followed by transformation of chemically competent *E. coli* DH5 α with 4 μ l PCR products.

2.12 Transfection with overexpression plasmids (Rescue experiments)

To restore expression of *PRPF40B* WT and mutants in KO cells, various rescued experiments were performed. K562-*PRPF40B*-KO cells were seeded at a density of 0.2×10^6 cells/ml in a 6 well plate before transfection. For the titration experiment, *PRPF40B* WT vector and pUC19 mock plasmid that is not expressed in eukaryotic cells were mixed in various ratios as indicated (Table 14).

Table 14: Amounts of PRPF40B WT overexpression plasmid and mock used for titration.

PRPF40B	0	25 ng	50 ng	75 ng	100 ng	300 ng	500 ng	1 µg	1.5 µg	2 µg
MOCK	2 µg	2 µg	1.9 µg	1.9 µg	1.9 µg	1.7 µg	1.5 µg	1 µg	500 ng	0

For the time course overexpression, 100 ng of the *PRPF40B* WT vector and 900 ng of mock plasmid were used for a single transfection for each of the different time points. For the overexpression of the *PRPF40B* mutations alone as homozygous-like, 100 ng of the P383L vector or 100 ng of the P540S vector were mixed with 1.9 µg of mock plasmid and transfected separately. Overexpression was considered homozygous-like if only one PRPF40B overexpression plasmid was used. Alternatively, for the overexpression of the heterozygous-like mutations, 1.8 µg of mock was mixed with 100 ng of the *PRPF40B* WT vector and then either with 100 ng of the P383L or P540S vector and transfected separately. Overexpression was considered heterozygous-like if a 1:1 ratio of the *PRPF40B* WT gene and one of the two mutations was used. Plasmid DNA was mixed with 100 µl Opti-MEM (Life Technologies, USA) and XtremeGENE HP transfection reagent (Roche, Switzerland) in a ratio of 1:2 (DNA:reagent). Transfection mixture was then incubated at room temperature for 40 min before it was added to cells in a drop-wise manner.

2.13 Total RNA extraction, sample preparation for RNA sequencing and RNA-seq data analysis

Total RNA from three replicates of K562 WT cells, the three *PRPF40B* KO clones and three replicates for each of the overexpression of PRPF40B WT and its mutations as homozygous-like or heterozygous-like was extracted with a method as previously described in section 2.6. Samples were sent at SCELSE facility for assessment of RIN. cDNA library preparations and RNA sequencing was done by the same facility as previously described in section 2.7. RNA-seq data analysis was performed by employing the same pipeline as previously described in section 2.8 with the exception that MATS only was used for the analysis of the AS events.

2.14 Additional Bioinformatics analysis:

Several R packages were used to perform the various analyses in R studio v1.0.153 (Table 15).

Table 15: List of various types of RNA-seq data analysis and corresponding R packages.

Package	Version	Type of Analysis
plot3d	1.1.1	Principal Component Analysis (PCA)
heatmap2	2.3.1	heatmaps
pheatmap	1.0.8	heatmaps
TCGABiolinks	2.5.7	TCGA data mining
ggplot2(boxplot)	2.2.1	Boxplots
ggplot2(wilcox.test)	2.2.1	Statistical Analysis

Pathways and GO analysis of the DEGs and AS events in THP-1 cells and blood derived macrophages were done both with the Reactome v61 online database (Fabregat et al., 2018) and with the GO online tool DAVID (Dennis et al., 2003). Gene set enrichment analysis of the DEGs and AS events regulated by PRPF40B was done both by the online tool ConsensusPathDB (Max Plank Institute for Molecular Genetics) (Kamburov et al., 2011) and Enrichr (Chen et al., 2013). Protein-Protein Interactions (PPI) network for the AS events regulated by PRPF40B was built with the online database STRING v10.5 (<https://string-db.org/>) with a high confidence score of 0.7 and a kmeans clustering score of 10 (Szklarczyk et al., 2017). A network of the DEGs found in the cell line model and a network of common AS events between the THP-1 cells and primary cells were built using the Cytoscape v3.2.1. plug-in ClueGO v2.1.7 by using a p-value less than 0.05 and a clustering k score of 0.5 (Shannon et al., 2003). Visualization of RNA-seq reads was done by Integrative Genomics Viewer (IGV) software (Thorvaldsdóttir et al., 2013).

Input DNA sequences used to calculate the 5'ss and 3'ss strength respectively were retrieved from the genome with bedtools package and the entropy score for each sequence was obtained with the online tool MaxEntScan (Yeo and Burge, 2004). Additional TCGA data mining was performed mainly with the online tool cBioPortal (Gao et al. 2013, Cerami et al. 2012) and other databases such as Oncogeno, Gent and Gepia (Shin et al., 2011; Tang et al., 2017).

2.15 Primers for the validation of the PRPF40B regulated AS events

Validation of the AS events regulated by PRPF40B was performed by PCR with DreamTaq Polymerase as previously described. Primers were designed online with the NCBI/Primer-Blast tool (Table 16).

Table 16: List of primers for the validation of the AS events regulated by PRPF40B.

Gene Name	Forward Primer	Reverse Primer
NAB1	GCAGGGAAGCCAGAAACTCT	AAGCTAACTCGGAGGCCTTG
WDHD1	TCACCCGTTGTGACTTGTG	TGGCTGCGGCCGCCCT
CTC1	TGAACTTCTGCAGGGTGGTC	TGGGCTCGTCACAATAAGCA
PRKD2	CCCGACTCAGAGAGTCCCC	CGGGCAACAGCACGAAC
SPTAN1	AGAAGAGTCCCCGAGAGGTC	AACCTGCTCCAAGTCTGCTC
MFSD2B	TTGGGAAGAAGACGTCAGCC	GGGTGGAGATGCCCAGG
CAMKK1	CGGGTGTACCAGGAGATTGC	AGGGCACTTCCATGGTTTGA
SCRN2	TTGGAGCACAAGACCCTGTT	CCAGATCCTGCTGTTTCTGCT
TIA1	TTTGTGTTGGTGATCTCAGCCC	ACTCGGGCATCTTTCCAA
FIP1L1	TTAGCAAACCACCTCCGTTT	GCGCCTGGTGGAGGA
SETD4	GCATGCAGAAAGGAAAAGGG	TTCAGACTTGTGGCTCTCTT
INTS6	TGACTACCACCAGTGGACAT	TGGTGTGATTGCAGAGTCAT
DPH7	TTAGGCCGTCTCTTCCTGTA	CCACCGGGATGTATGTCCA
C1orf52	GACGAGCTGTTTAGGAGCGT	TTGAATCGACGTGCCTCTCC
IP6K2	CCGCTTCAATGAGACAACCC	AACAAGGGGCCTTTTACTGG
KIAA1468	TCAGAGGGTTGTTCTGCTC	CCAAAGGCTGGAATTGTGGC
RIPK2	CTCTGGCACTGTGTCGTCC	GGGCAATTTCATGCAGGATGC
ANKRD54	ACCACGTTCTGTCATCACC	GGTACTGGTCATCTGCAGGC
MAP3K8	CCTGGTTCCTGCTTCTGCC	TCAGACTCCTGGCTTTGCAG
WNK1	TTTTGATTTAGAGAGAGATGTCC	CTTGTGGAGCTTGAGGCTGA
ARVCF	CTTTTGAGATGGTGCCCTG	TGCAGGCCATGGTCAATGAT
SUGP2	ACAGGCATCTGGAAGTGCTC	GCAACGGCTAAATGATTAACAACA
NFIC	CGACTTCCAGGAGAGCTTTGT	GCAGGATCCCTTGGTCCTAA
ORC2	GGACAACCAGGATAACCCTCT	GCTTGTGGTCCCTAAATTCAGTT
HFM1	TGAATTGGGAAATGAAGTTTGGG	ATCCGACATCTCAGTCTTCTC
ASIC3	GCAGTACAAGAACTGTGCCC	GGATCTCGAGGATGGTGAGC

ACAA1	CCTGCCTTCAAGAAAGATGGTTTG	CCAACCTCTTCTGCCTTGGACCTC
MPPE1	GGGAGCAGATGCCAACTTAG	TGGCATTTCAGGTCTTAAGGAA
LPCAT4	AGGAGCCTGGAAGAGCTAAC	GAGAGGCCTTGGCTGGAT
WSB1	CTATGCTGTGTTTCAGTCGGA	TGCAGTAGCCAGTAATGCTC
DNASE1	AAGATCGCAGCCTTCAACAT	TATAGCTGTTCCGTCCTCAGT
XRRA1	AATGCCTCAGAACCTTTCCTTC	GCAGGGAGGTAAGGCCATTG
NCALD	AGGCACTGGCTGAATTCTTG	GAAGGCCCATTTTCAGCTTCT
TXLNA	GAGGTGGGGAAGCCATCGG	CCCTTGGAGGGTTCCTTCTCTC
DMTF	AAGTGCCCGGACCCAA	CCTTCCCTTAGCAACCAGCA
NAT1	TAGTTCCTGGTTGCCGGCTG	TTGCAGTTCACATCCAGGCT
MFSD8	AACAAGCCTGGTGCCTCTAC	CCATGTGGTCCTAGAATTTTTGAAT
MARK2	GCCTCCATCCAGAATGGCAA	CGGGGGAAGTTAGTTCGGTC
INF2	CTCGGGGAGAGGACGG	GGCTGTTGTAGCATCAAGGC
MYLIP	TTCTGTGCGGGATAGCGAAG	GCCCTTCAAGTCACGGCTAT
ALAS1	GCCGCCTCTGCAGTCC	TGGCCCCAACTTCCATCATC
AB2	TCACCCTGAAATCCAGCAGC	CATCTGTCCAGCAGCCCTTC
SEC24B	ATCACCCAGTTGAGCCTGTG	TTGGGCATTGGAGAAGCACT
GUK1	TGGCTGCGGCCGCCCT	TCCAGAGCCACAGAGGAGTGAAGGC
KIAA0232	ATGGCACAATGCGAGGAAGA	GGCTTTTTTCCCACCAGGTCT
C2orf47	CGTCCTCGAGGTGGCCT	GGTGTCTCTGCTGCTCTGAAG
C2CD5	GGAGAGTGCAAGTTCTAACTCA	AAGTAGTCTCTCGAATAAAAAACAT
ZNF2	CAGAGAGACACATTCCCTGC	TCCAGCATCACCTCCTTGTA
AFMID	GGAGGATACTGGCAGAGCG	CTTGGATAACCGCTTCTGGAC
NOTUM	CCCCTACTGCTCCAGTGATG	CACGTGATCGTGTCGACGCA

2.16 qRT-PCR *PRPF40B* regulated DEGs validation

Validation of the DEGs regulated by *PRPF40B* was performed by Real Time PCR as previously described in section 2.7.7 with primer listed in Table 17.

Table 17: List of primers for the validation of the DEGs regulated by PRPF40B.

Gene Name	Forward	Reverse
B-actin	CCAGAGGCGTACAGGGATAG	CCAACCGCGAGAAGATGA
ENO2	GAAGAAAAGGCCTGCAACTG	CCAGGTCAGCAATGAATGTG
PGK1	CTGTGGGGGTATTTGAATGG	CTTCCAGGAGCTCCAAACTG
CA9	TAAGCAGCTCCACACCCTCT	TTCATCTGCACAAGGAACG
BNIP3	GCATGAGTCTGGACGGAGTA	TCCTCAGACTGTGAGCTGTT
TGF-β3	GCGCCTCAAGAAGCAGAAGG	TCCTCCAAGTTGCGGAAGCA
COL1A2	GTCCAGCTGGCTCCAAAGGA	AGCTCTGCCATCAGCTCCAG

Note: primers were designed with GenScript Real Time PCR Primer Design online tool. *β -actin* was used for the normalization of the values. Validation of the PRPF40B regulated AS events and DEG was performed by FYP student Resilind Chew under my mentorship.

2.17 Splicing minigenes

To study mechanistically how *PRPF40B* is regulating AS, minigenes for *RIPK2* and *SPTAN1* containing alternative middle exons with shortened flanking introns only were constructed into the UMV vector (Tan and Roca, 2016). In addition, alternative middle exons together with endogenous flanking constitutive exons and their own introns were cloned into the pcDNA3.1+ plasmid. For the former cloning method, *RIPK2*-exon 2 and *SPTAN1*-exon 23 with 300 bp of the upstream and downstream introns respectively were amplified from Human Genomic DNA (Promega, USA) using PCR with primers indicated in Table 18. Fragments were then cloned into a UMV vector with BamHI and EcoRI sites (Figure 21A). For the latter cloning method, three fragments of the *RIPK2* were amplified from Human Genomic DNA (Promega, USA) using PCR with primers indicated in Table 19 (top): (i) exon 1 of *RIPK2* and downstream 300-nucleotide (nt) intronic sequences (ii) exon 2 of *RIPK2* and 300 nt of their upstream and downstream introns (iii) exon 3 of *RIPK2* and 300 nt of its upstream intron. Subsequently, these three fragments were merged into one using overlap PCR with PrimeStar Polymerase (Takara). Alternatively, an entire fragment of *SPTAN1* gene with the flanking constitutive exons and the middle alternative exon (*SPTAN1*-exon 22-23-24) was amplified from Human Genomic DNA

(Promega, USA) using PCR with primers indicated in Table 19 (bottom) and directly cloned into the pcDNA 3.1+ plasmid by using NheI and XhoI restriction sites (Figure 21B).

Table 18: List of primers for the cloning of the UMV minigenes.

Primer Name	Sequence
FW SPTAN1 UMV BamH1	TGCCTGTTGGATCCCAGTAAAATTTCTTTCTAAGGGT
RW SPTAN1 UMV EcoR1	TCCTTCCTGAATTCAAACTATAGAACAGGCCTTCAG

Primer Name	Sequence
FW RIPK2 UMV BamH1	TTGAATACGGATCCAACTTAATCATGTAATTAGTTCA
RW RIPK2 UMV EcoR1	CAGTATGCGAATTCTATCTCTTCCATTTCTACCCATGT

Table 19: List of primers for the cloning of pcDNA3.1+ minigenes.

Primer Name	Sequence
FW RIPK2 (up)	GCCCGGATCCATGAACGGGGAGGCCATCTG
RW RIPK2 (up)	TACACACAAAATTACAGACACTTTATGAACTGGTATTAAT
FW RIPK2 (middle)	ATTAATACCAGTTCATAAAGTGCTGTAATTTGTGTGTA
RW RIPK2 (middle)	ATATCCTGAATATTAAACCTATCAATTCAGTAAACCCAG
FW RIPK2 (down)	CTGGGTTTACTGAAATTGATAGGTTTAAATATTCAGGATAT
RW RIPK2 (down)	AAAAAAGTGAATTCCTTAACATGAAATTCATT

Primer Name	Sequence
FW SPTAN1 pcDNA NheI	TCTTTTTTGCTAGCGATTGGTGGAAAGTGGAAGTGAA
RW SPTAN1 pcDNA XhoI	CTTCTGGAAGTCATCAAACCTCCTCGAGAAGACTGC

Figure 21: Minigenes design.

Schematic describing the cloning of the *RIPK2* and *SPTAN1* alternative middle exons into the UMV vector (A) or into the pcDNA3.1+ vector (B).

2.18 Bacterial transformation and plasmid extraction

A total of 50 μ l of chemically competent *E. coli* DH5 α was mixed with 4 μ l PCR or ligation products and incubated on ice for 30 min followed by a heat shock at 42 $^{\circ}$ C for 45 s. Subsequently, cells were incubated on ice for another 5 min and then recovered in 1 ml of pure LB at 37 $^{\circ}$ C for 1 h with agitation.

Finally, transformed cells were plated on ampicillin-supplemented LB agar plate (70 µg/ml) and grown at 37 °C overnight. Colonies were grown in 5 ml LB with ampicillin (70 µg/ml) overnight and then subjected to miniprep (Omega Bio-Tek, USA). All minigene sequences were confirmed by sequencing (1st Base, Singapore). For the ligation product, plasmids were digested with the respective restriction enzymes as indicated previously and run on an agarose gel to confirm the insertion of specific fragments before submitting for sequencing.

2.19 Inhibition of Pol II mRNA transcription

Non edited K562 WT, *PRPF40B*-KO and rescued cells were seeded at a concentration of 0.5×10^6 cells/ml and incubated with either DMSO, 50 µM 5,6-dichloro-1-β-Dribofuranosylbenzimidazole (DRB) or 1 µM Camptothecin (CPT) for 5 h to inhibit transcription (Sze Jing et al., 2016). After RNA extraction and cDNA synthesis, with methods as previously described in paragraphs 2.5 and 2.8, primers designed in exons flanking the alternative exon 8 of hnRNPD L were used to detect the efficiency of the drugs by RT-PCR (Table 20).

Table 20: List of primers for the assessment of the inhibition of Pol II.

Primer Name	Sequence
hnRNPD L.Ex7.F	GCCAACAGAGCCACTTATGG
hnRNPD L.Ex9.R	ACAGTTGGACACAATGGTGTC

The PCR was carried out with Prime Star Polymerase (Takara) and the conditions were as follows: 39 cycles of denaturation at 98 °C for 30 s, annealing at 58 °C for 15 s and extension at 72 °C at a polymerization speed of 1 kb/min. PCR products were loaded on a 1.5 % agarose gel and visualized by UV light in a gel doc system.

2.20 Western Blotting

Harvested cell were lysed in NP-40 buffer (5M NaCl, 1M TRIS pH 8.0, 100% Triton-X, 100% Glycerol) and 1X proteinase inhibitors (Roche) for 60 min on ice and supernatants were collected after centrifugation at maximum speed for 30 min. Protein concentrations for each

sample were determined by Bradford Assay (Biorad) and 40 µg of protein from each sample were boiled for 10 min with Sodium dodecyl sulfate (SDS) loading buffer before loading into gel. Cell lysates were separated by SDS-PAGE on an 8% gel (Acrylamide/Bis 37.5:1; Biorad, USA) in 1X SDS running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS) and transferred to a Polyvinylidene difluoride (PVDF) membrane (Biorad) in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol] at 100 V for 1.5 h. Membrane was blocked in 5% milk in 1X TBST [144 mM NaCl, 20 mM Tris, 0.05% Tween 20, pH7.6] for 1 h and washed with 1X TBST thrice, each for 5 min. Next, the membrane was incubated with the primary antibody anti-rabbit PRPF40B (Abcam), anti-rabbit SF1 (Thermo), anti-rabbit snRNP70 (U1-70k) (Thermo), or anti-mouse GAPDH (Sigma) diluted in 1X TBST at 4 °C overnight followed by three washes with 1X TBST and incubation with the secondary anti-mouse/rabbit IgG-HRP antibodies (BioLegend, USA/Sigma-Aldrich, USA/Santa cruz biotechnology, USA) diluted in 1X TBST with or without 5% milk (Table 2.12). After three washes, proteins were detected by Western Lightning Plus ECL (Perkin Elmer, USA).

2.21 Sample preparation for LC-MS/MS (Shotgun) and MS analysis

Prior to Mass Spectrometry (MS) analysis, each sample was polymerized as a gel with 40% Acrylamide (Biorad) for 1 h and fixed in 50% Methanol and 12% Acetic Acid for 1 h. Gels were cut in smaller pieces and submitted to the Proteomics Centre (NUS). At the Proteomics Centre samples were further processed. Briefly, they were subjected to column purification, digested with trypsin to generate smaller peptides and finally analyzed by a 5600 Triple ToF machine.

2.23 Measurement of Apoptosis, DNA damage and proliferation

Assessment of Apoptosis, DNA damage and proliferation was done by using the Apoptosis, DNA Damage and Cell Proliferation Kit (BD Biosciences, cat no 562253) according to manufacturers' protocol. Briefly, K562 WT cells and three *PRPF40B* KO clones were seeded at 1×10^6 cells/ml in 6 well plates and pulsed for 4 h with 1 mM Bromodeoxyuridine (BrdU) (Bradford and Clarque, 2011). Cells were harvested, washed with 1X PBS and BD Cytofix/Cytoperm Fixation/Permeabilization Solution was added to permeabilize and fix the cells for 45 min on ice. Cells were washed with 1X PBS and treated with DNase at a concentration of 300 µg/ml for 1 h at 37 °C to expose incorporated BrdU. Cells were then stained

on ice for 30 min with Anti-BrdU PerCP-CyTM5.5 antibodies to assess proliferation (Bradford and Clarke, 2011), Alexa Fluor® 647 Anti-H2AX (pS139) PE antibodies for DNA damage (Huang and Darzynkiewicz, 2006) and Anti-Cleaved PARP (Asp214) antibodies for apoptosis (Kunzmann et al., 2006). Cells were washed with 1X PBS, resuspended in buffer and flow cytometric measurements were performed at the FACS facility in SBS using 5 lasers LSR Fortessa X-20 (Becton Dickinson). Data analysis was performed by Cyflogic (v.1.2.1).

3 RESULTS

Aim 1

In the following section, the results of the transcriptomic analysis of the changes in AS and DEGs during the differentiation and activation of human monocytes will be described.

3.1 Cell surface markers for macrophage differentiation and polarization of THP-1

Considering that the THP-1 cell line is a very good model to study monocyte physiology (Tsuchiya et al., 1980), THP-1 cells were differentiated into macrophages with the different stimulations indicated in chapters 2.1 and 2.2 (Auwerx, 1991). Differentiated cells were polarized to M1-like and M2-like macrophages (Genin et al., 2015). In order to verify the efficiency of our differentiation and polarization protocols, cell surface expression of well-established markers for macrophage differentiation/polarization (Rogler et al., 1998) were assessed by flow cytometry (Figure 22).

3.1.1 Cell surface markers for THP-1 macrophage-like cell differentiation with PMA and VD3

According to the literature, THP-1 cells can be differentiated to macrophage-like cells by phorbol diesters such as PMA, thus THP-1 cells were differentiated to macrophages by PMA for 3 d only (PMA 3d) (Tsuchiya et al., 1982). Alternatively, in order to enhance differentiation, after 3 d with PMA, fresh media without PMA was added and cells were cultured for 5 additional days (PMA 5dr) (Daigneault et al., 2010).

Upon PMA stimulation, cells change morphology and overexpress or down-regulate specific cell surface markers (Daigneault et al., 2010). Thus, we assessed expression of various surface

antigens by flow cytometry (Figure 22). CD14, the co-receptor of the LPS receptor, is a well-established antigen which is employed to define monocyte/ macrophage differentiation and which is expressed on neutrophils too (Zamani et al. 2013) and it is highly up-regulated in our differentiation protocols. We observed also a striking up-regulation of CD54, an adhesion molecule, and of CD11b, an alpha-M beta-2 (α M β 2) integrin (Larsson and Springer, 1990). Regarding CD36, which is a class B scavenger receptor and the major receptor used by macrophages for Low Density Lipoprotein (LDL) uptake (Krieger and Hertz, 1994), we observed an up-regulation at 3 d with PMA, followed by a down-regulation after PMA removal. Conversely, we did not find any expression of CD86, a co-stimulatory molecule for antigen presentation, after 3 d but the marker appeared to be expressed at 5 d after PMA removal (Figure 22). We did not detect any expression both of CD68, a glycoprotein that binds LDL, and of CD206, a mannose receptor, either after 3 d only or after PMA withdrawal (data not shown). PMA withdrawal results in a further upregulation of CD14, CD11b and CD54 while CD36 is downregulated. Lastly, we did not observe any change in CD33 expression, a marker of myeloid lineage, which is highly expressed in all conditions (data not shown). Excluding some exceptions, these results confirm what is reported in the literature for the differentiation of monocytes to macrophages (Rogler et al., 1998, Daigneault et al., 2010; Paulsen et al., 2011; Aldo et al., 2013).

THP-1 cells can be differentiated to macrophage-like cells by VD3 with differences in the state of differentiation compared to the PMA treatment (Schwende et al., 1996). Expression of cell surface markers was assessed by flow cytometry (Figure 22). As reported in the above study, we did not observe any attachment of the cells and any loss in proliferation. We observed an up-regulation of CD14 and of CD11b. However, in contrast to THP-1 treated with PMA, it appears that not all the cells over-express CD11b after treatment with VD3. We did not observe any expression of CD36, CD54 and CD86. In addition, we did not notice any change in the expression of CD33, CD68 and CD206 (data now shown). These results completely reproduce the data reported in the literature (Schwende et al., 1996; Daigneault et al., 2010; Aldo et al., 2013).

3.1.2 Cell surface markers for polarization of M1-like and M2-like macrophages (cell line)

Macrophages can be polarized to classically activated (M1) macrophages, which are pro-inflammatory, or to alternatively activated (M2) macrophages, having an anti-inflammatory phenotype (Martinez et al., 2006). After differentiating THP-1 monocytes to macrophage-like cells by PMA for 3 d, we have polarized them towards the M1-like phenotype with a combination of LPS and IFN- γ or towards the M2-like phenotype with a combination of IL-4 and IL-13 for 8 h.

CD11b appears to be higher in M2-like macrophages compared to M1-like macrophages, while CD54 expression is higher in M1-like. We detected a higher CD36 expression in macrophages polarized to M2-like phenotype compared to M1 (Figure 22). CD33 expression decreased in both phenotypes compared to control (data not shown). Interestingly, we did not observe any expression of CD14 after polarization. CD86 expression was not detected either in M1-like or M2-like macrophages (Figure 23). Similarly we did not detect any surface expression of other markers such as CD80, CD209, CD206, and CD68 (data not shown). Excluding some exceptions, such as CD86 and CD206, overall our results match the reported ones (Park et al., 2007; Kadowaki et al., 2011). We could not identify any specific markers by flow cytometry for the two polarization states. However, reports in the literature suggest assessing the activation by analyzing the gene expression levels of various cytokines and chemokines (Mantovani et al., 2004; Ushach and Zlotnik, 2016). We performed such analysis from our RNA-seq data which will be described in section 3.7.

Figure 22: Flow cytometry results for differentiation/polarization of THP-1 cells.

Cell surface expression of macrophage-like differentiation and polarization markers by flow cytometry after various treatments. Gray shaded area shows IgG isotype controls. Lines with different colors representing surface antigen expressions for the various treatments as indicated. X-axis represents fluorescence intensity while y-axis represents cell count. On the right side, mean fluorescence intensities (MFI) of three replicates were plotted as bar charts.

3.2 Cell surface markers for macrophage differentiation and polarization of primary human blood monocytes

Similarly to what was previously done with the THP-1 cell line, primary human monocytes isolated from blood from various human donors were differentiated to M1 and M2 macrophages and subsequently polarized to M1a, M2a, M2b, M2c as described in the literature (Martinez et al., 2008, Rey-Giraud et al., 2012). Human blood was withdrawn from 4 human donors upon written consent and monocytes were isolated by magnetic negative selection with a method described previously (Figure 23). Negative selection was preferred over positive selection in order to obtain untouched non activated cells and to further enhance the purity of the isolated

monocytes. Purity was assessed by CD14 and CD16 expression (Figure 24). We obtained purity higher than 97% as our quantification of CD14 expression show.

Figure 23: Procedure for isolation of human blood monocytes.

Schematics representing the workflow for the isolation of the primary monocytes from human donors. Hystopaque was used to separate the fresh blood in various layers, PBMCs were collected and CD14⁺ monocytes isolated by negative selection after magnetic separation.

Figure 24: Results for the isolation of primary cells.

Panel A indicates the amount of blood drawn and the quantity of PBMCs and monocytes obtained during the isolation process. Panel B shows the flow cytometry histograms of CD14 and CD16 markers measured after each monocyte isolation. The percentage of the population expressing CD14 was used as a measure to assess the purity of the collected fraction of monocytes. For the CD16 we observed two populations corresponding to those expressing higher level or a lower level of CD16. Note that Donor 2 expresses higher total CD16 levels but still shows the two subpopulations of cells expressing relatively lower or higher CD16, thus we did not exclude this sample in order to preserve some human variability.

3.2.1 Cell surface markers for macrophage differentiation of primary human blood monocytes

Human monocytes were differentiated to M1 macrophages by GM-CSF and to M2 macrophages by M-CSF as reported in the literature (Lacey et al., 2012), and efficiency of differentiation was assessed by flow cytometry (Figure 25).

Figure 25: Flow cytometry results for differentiation/polarization of human monocytes.

Cell surface marker expressions of the various subsets of macrophages obtained by differentiating primary human monocytes from 4 human donors. Gray shaded area shows IgG isotype controls. Lines with different colors representing surface antigen expressions. X-axis represents fluorescence intensity while y-axis represents cell count. On the right side MFI of the 4 human donors were plotted as bar charts.

We observed that the expression of the scavenger receptor CD163 is high in M2 and low in M1, while vice versa the expression of the co-stimulatory molecule CD80 is high in M1 and low in

M2. Monocytes are negative for the macrophage mannose receptor CD206 which is highly induced upon differentiation especially by GM-CSF. Monocytes are also negative for the transferrin receptor CD71 which is highly induced by GM-CSF and not by M-CSF. The expressions of the two co-stimulatory molecules HLA-DR and CD86 are further induced by GM-CSF, while to a lesser extent by M-CSF. M2 macrophages seem to have a higher expression of CD14 and CD16 compared to M1. In general, these results confirm what is reported (Zhang et al., 2013; Samaniego et al., 2014; Hoppstädter et al., 2015).

3.2.2 Cell surface markers for macrophage polarization of primary human blood monocytes

M1 macrophages were exposed to LPS and IFN- γ for 8 h in order to activate them. Activated M1 macrophages were denominated as M1a. In parallel, M2 macrophages were treated for 8 h with IL-4, IL-1 β and IL-10 in order to obtain M2a, M2b and M2c macrophages respectively, as described in the literature (Italiani and Boraschi, 2014). Flow cytometry was used to access the expression of the various cell surface markers upon the various stimulations (Figure 25). The expressions of the co-stimulatory molecules CD80, CD86 and HLA-DR were further induced by LPS confirming what is reported (Zhang et al., 2013). However, regarding the activation of the M2 macrophages, we did not observe any major changes in the cell surface marker expression upon the various treatments. A reason could be that the aforementioned cytokines are weak thus a treatment longer than 8 h could be necessary in order to observe any major effects. However, reports in the literature would suggest assessing the M2 activations by analyzing the gene expression levels of specific cytokines and chemokines rather than relying solely on flow cytometry (Mantovani et al., 2004; Ushach and Zlotnik, 2016). We indeed performed such analysis which will be described in the later section 3.7.

3.3 Live cell images for macrophage differentiation and polarization

Live cell images during the differentiation/polarization process, at different time points, were taken by a Zeiss Live cell imaging microscope for the cell line and primary cells (Figure 26 and 27). We performed such experiments in order to further assess the efficiency of our differentiation/polarization protocols. When monocytes differentiate to macrophages, they undergo strong morphological changes which can be visualized by live cell imaging.

3.3.1 Live cell images for the differentiation and polarization of THP-1 cells

Images of live cells were taken before and after differentiation of THP-1 cells with PMA at different time points. After addition of PMA, cells progressively attach to the plastic surface till day 3 when they are completely attached. However at day 3, cells appear to be still round whereas after 5 days from PMA withdrawal cells become more elongated with a higher cytoplasmic content, a sign of terminal differentiation (Figure 26A). Live cell images were taken before and after differentiation of THP-1 cells with VD3 or a combination of VD3 and Retinoic Acid for 3 days. Cells do not attach to the surface but remain in suspension, thus our results reproduce what is reported (Daigneault et al., 2010). In both conditions, cells are still round and they do not lose proliferation capacity (Figure 26B). Live cell images were taken before and after polarization to M1-like macrophages with a combination of LPS and IFN- γ or to M2-like macrophages with a combination of IL-4 and IL-13. In our polarization protocol, M2-like macrophages appear elongated, whereas M1-like macrophages appear to be rounder (Figure 26C).

Figure 26: THP-1 live cell imaging.

Live cells images of differentiated and polarized THP-1 cells taken at 20X magnification. Images for differentiation by PMA (A), Vitamin D3 and a combination of it and Retinoic Acid (B) and for the polarization to M1-like and M2-like macrophages (C).

3.3.2 Live cell images for the differentiation and polarization of primary cells

Live cell images were taken before and after differentiation of human monocytes to M1 macrophages by GM-CSF or to M2 macrophages by M-CSF. As reported, macrophages differentiated by GM-CSF have an oval shape while M-CSF macrophages have a spindle shape (Rey-Giraud et al., 2012). In addition, pictures were taken after the various polarization treatments (Figure 27). We observed that M1 macrophages activated by LPS/IFN- γ shift shape to spindle-like while vice versa activated M2 macrophages from spindle shape became rounder upon activation.

Figure 27: Primary cells live cell imaging.

Live cells images of human monocytes and macrophages taken a 20X magnification after differentiation with GM-CSF for M1 macrophages and M-CSF for M2 macrophages and of the various stimulations for the polarization of the cells. Number of days in which cells were kept in culture after adding the specific stimuli are indicated.

3.4 Total RNA quality control before RNA-seq

Considering that our differentiation and polarization protocols are very robust and reproducible, we proceeded to extract total RNA from undifferentiated cells and differentiated cells for all the stimulations. After DNase treatment RNA was concentrated, and RNA quality for each sample was assessed before proceeding to RNA-seq. Only samples with an RNA Integrity Number (RIN) equal or higher than 8 were submitted for RNA sequencing (RNA-seq).

3.5 RNA-seq data analysis results

Total number of pair end reads after RNA-seq and number of aligned reads to the human hg19 genome was calculated for each sample (Figure 28).

Sample	Pair-end reads	Aligned pairs	Concordant alignment
Mo r1	87,329,362	80,713,261	92.3%
Mo r2	80,983,890	72,580,577	89.4%
Mo r3	107,522,086	96,439,607	89.5%
Mo r4	80,583,672	73,934,962	90.9%
GM-CSF MΦ r1	82,701,142	76,612,393	92.5%
GM-CSF MΦ r2	118,097,824	105,921,996	89.2%
GM-CSF MΦ r3	61,842,136	56,378,933	90.4%
GM-CSF MΦ r4	106,168,369	97,443,266	90.9%
LPS+IFNγ MΦ r1	89,506,445	82,214,443	91.6%
LPS+IFNγ MΦ r2	90,580,459	69,907,275	76.8%
LPS+IFNγ MΦ r3	61,003,943	55,047,817	90.1%
LPS+IFNγ MΦ r4	82,484,686	75,910,902	91.0%
M-CSF MΦ r1	101,433,727	94,529,210	93.1%
M-CSF MΦ r2	82,973,373	73,877,285	88.6%
M-CSF MΦ r3	70,971,617	64,821,545	91.0%
M-CSF MΦ r4	78,549,770	72,124,878	90.8%
IL4 MΦ r1	80,650,550	75,194,139	93.1%
IL4 MΦ r2	78,271,375	70,626,496	89.7%
IL4 MΦ r3	70,773,508	64,608,239	91.0%
IL4 MΦ r4	82,661,230	76,117,322	91.5%
IL1β MΦ r1	82,014,291	75,870,383	92.3%
IL1β MΦ r2	73,599,383	66,126,586	89.2%
IL1β MΦ r3	83,874,143	77,350,738	91.6%
IL1β MΦ r4	86,912,297	80,521,392	91.8%
IL10 MΦ r1	80,883,431	75,021,190	92.6%
IL10 MΦ r2	87,677,824	78,865,051	89.3%
IL10 MΦ r3	84,354,630	77,959,168	91.9%
IL10 MΦ r4	80,922,918	75,603,349	93.2%
THP-1 r1	40,690,106	37,544,522	90.2%
THP-1 r2	43,461,363	38,054,309	87.3%
THP-1 r3	66,591,098	58,035,938	86.9%
VD3 r1	58,786,248	51,359,222	87.1%
VD3 r2	65,073,093	56,697,087	86.8%
VD3 r3	65,046,128	56,050,703	85.8%
PMA3D r1	42,072,634	38,438,030	90.0%
PMA3D r2	39,051,497	35,142,923	89.8%
PMA3D r3	68,412,077	59,484,062	86.8%
PMA5Dr r1	39,273,995	36,165,751	91.8%
PMA5Dr r2	38,742,446	34,491,767	88.8%
PMA5Dr r3	73,310,657	64,061,954	87.2%
LPS+IFNγ r1	54,284,053	49,497,542	91.0%
LPS+IFNγ r2	34,990,002	31,771,748	90.5%
LPS+IFNγ r3	33,670,544	30,309,713	89.7%
IL4/IL13 r1	43,052,593	39,017,413	90.4%
IL4/IL13 r2	34,211,065	30,978,769	90.3%
IL4/IL13 r3	35,021,490	31,914,357	90.8%
sh2-MBNL1 r1	52,000,116	47,646,967	91.4%
sh2-MBNL1 r2	53,757,395	48,702,931	90.1%
sh2-MBNL1 r3	52,906,285	48,228,450	90.8%

Figure 28: Alignment results of the RNA-seq of cell line and primary cells.

Number of reads aligned with TopHat. Total numbers of paired end reads, with a length of 101 bp, obtained after sequencing and number of reads mapped and unmapped to the reference genome and transcriptome by TopHat. Last column indicates percentage of aligned reads.

3.6 Analysis of the DEG in cell line and primary cells

In order to perform a detailed analysis of the transcriptome, as a team we identified thousands of DEGs in macrophages compared to monocytes in the cell line model (Figure 29 Panel A and B) and in primary human cells. (Figure 29 Panel C and D). Venn diagrams were generated by Dr. Hongfei Liu.

Figure 29: Results of DEGs in cell line and primary cells.

Venn diagrams with numbers of DEGs with fold change ≥ 2 (q-value ≤ 0.05) for the THP-1 cell line (Upper panels A and B) and primary cells (Lower panels C and D).

In addition, common DEGs between cell line and primary cells were identified (Figure 30). Venn diagrams were generated by Dr. Hongfei Liu. Validation of 28 events was performed by qRT-PCR by FYP student Ms. Michelle Wong who under my co-supervision found that the events are in agreement to what was found by RNA-seq (data not shown).

Figure 30: Results of DEGs in common between cell line model and primary cells.

Venn diagrams showing the proportion of DEGs with fold change ≥ 2 (q-value ≤ 0.05) between the indicated conditions in primary cells (pink) and THP-1 cells (blue).

3.7 DEGs analysis of chemokines and cytokines in cell line and primary cells

In order to assess the changes in gene expressions of specific chemokines and cytokines upon macrophage differentiation and polarization, heat maps were generated in R with the package heatmap2, using the FPKM values from Cufflinks (Figure 31).

We observed specific clusters of chemokines and cytokines which were either downregulated or upregulated after macrophage differentiation and polarization. In particular, for the polarization of THP-1 cells to M1-like macrophages by LPS/IFN- γ , a large cluster of chemokines and cytokines was upregulated. However, very few changes upon IL-4/IL-13 treatment of the THP-1 cells were detected. For the primary cells, we observed a large cluster upregulated upon treatment with LPS/IFN- γ (M1a). We finally observed upregulation of chemokines and cytokines specific to M1a, M2a, M2b and M2c macrophages. The result of our chemokines/cytokines profile in part confirms and in part is different from what is reported (Mantovani et al., 2004;

Mantovani et al., 2013; Moghaddam et al., 2018). In general, these results confirm that our differentiation and polarization protocols were indeed successful.

Figure 31: Cytokines/Chemokines DEG in cell line and primary cells. Heat map of the mRNA expression of various cytokines and chemokines in the cell line model (A) and in the primary cells (B). Dark blue color represents upregulation while light blue downregulation. Columns represent the different conditions while rows gene names.

3.8 Principal Component Analysis (PCA) of the DEGs in the cell line and primary cells

In order to further distinguish the various differentiated and polarized cells and to further confirm the validity of our differentiation and polarization protocols, PCA was performed based on the FPKM values of Cufflinks (Figure 32). PCA is a method to assess the distance between samples and to discover which samples are more similar to each other and which not in our case by measuring RNA levels of all genes. A PCA 3D interactive model was generated by using the R package plot3d. Our PCA results show that THP-1 cells differentiated with VD3 are close to undifferentiated cells while PMA treated cells are very much different. The two PMA treatments are close to each other. Treatment with IL-4/IL-13 and LPS/IFN- γ are instead different from PMA treated cells and they go into two different directions. Regarding the primary cells, M1 and M2 macrophages are very different from undifferentiated monocytes while all the M2s are very similar and cluster together. M1a macrophages are instead very different from M1 (Mia et al., 2014). Our PCA results thus support what was observed from the previous results for the live cell imaging, flow cytometry markers and DEG analysis.

Figure 32: PCA in cell line and primary cells.

Principal Component Analysis for the cell line model (A) and primary cells (B). Spheres represent the various conditions plotted in a 3D space.

3.9 GO analysis of the DEGs in the cell line model an in primary cells

A network of the DEGs events in the cell line model was built in Cytoscape to further assess what are their functions. The enriched terms are cell differentiation, cell communication, cell migration, extracellular matrix and inflammatory response (Figure 33).

Figure 33: GO Network of the DEGs during differentiation of THP-1 cells.

Sphere represents a single GO class. Different colors represent clusters of GO classes which belong to the same family. Size of the sphere is proportional to the significance of the GO class.

To investigate more in detail the functions of the various DEGs upon macrophage differentiation and polarization, genes were sorted based on their functions by the Reactome database (Croft et al., 2010), and heat maps of the expression of those genes were generated. Categories were selected based on the relevance to macrophage biology. It was observed that PMA and VD3 treatment downregulates genes involved in cell cycle, DNA repair, DNA replication and chromatin organization, confirming that cells stop to proliferate once they differentiate. LPS

polarization of the PMA treated cells upregulates specific subsets of those genes and downregulation of others, while no major changes were observed in the treatment with IL-4. PMA treatment and LPS but not VD3 results in an upregulation of genes involved in cell-cell communication, ECM organization, vesicle mediated transport, immune system and signal transduction, which are all features of macrophage biology (Figure 34). GM-CSF derived macrophages upregulate genes involved in DNA repair and DNA replication while they downregulate genes involved in immune system. For the other categories, M1 macrophages appear to upregulate some genes and at the same time downregulate others. M1a macrophages upregulate instead genes involved in cell-cell communication, ECM organization, vesicle mediated transport, immune system and signal transduction. Primary monocytes differentiated with M-CSF downregulate genes involved in chromatin organization, vesicle mediated transport, immune system and signal transduction.

Figure 34: GO Reactome results.

Heat maps of the mRNA expression of various genes in the cell line model on top and primary cells on the bottom. Red color represents upregulation while green downregulation. Columns represent the different conditions while rows gene names. Genes were grouped based on their functions found by the Reactome database. A cutoff of a fold change of less than -2 or larger than 2 was used.

3.10 Identification of novel AS events that change upon differentiation and polarization of the cell line and primary cells

In order to discover novel AS events that change during the transition from monocytes to macrophages, MATS, MISO and SpliceTrap were used. I contributed to establish the SpliceTrap analysis. I discovered hundreds of AS events changes in the cell line model and in the primary cells. AS events were sorted based on their Percentage Spliced In (PSI) values. PSI indicates how efficiently sequences of interest are spliced into transcripts. To conclude, as a team we discover by various tools thousands of AS events that change from monocytes to macrophages for the primary cells (Figure 35) and for the THP-1 cells (Figure 36). Venn diagrams were generated by Dr. Hongfei Liu.

Figure 35: Changes in AS in primary cells.

Venn diagrams with the numbers of AS events for differentiation and activation of primary cell as indicated (PSI change ≥ 0.05 ; SpliceTrap p-value ≤ 0.001 , MISO BFmin ≥ 1 & BFmedian ≥ 10 , MATS FDR ≤ 0.05). In parentheses, number of genes affected by AS.

Figure 36: Changes in AS in THP-1.

Venn diagrams with the numbers of AS events (PSI change ≥ 0.05 ; SpliceTrap p-value ≤ 0.001 , MISO BFmin ≥ 1 & BFmedian ≥ 10 , MATS FDR ≤ 0.05) for the indicated treatments of PMA/VD3 differentiation (A) and polarization (B). In parentheses, number of genes affected by AS.

Lastly, Dr. Hongfei Liu found that there are hundreds of AS events in common between the cell line model and the human primary cells (Figure 37).

Figure 37: Common changes in AS between THP-1 and primary cells.

Venn diagrams showing the numbers of changes in AS that are common in primary and THP-1 cells upon analogous stimuli (differentiation in panel A and polarization in panel B).

3.11 Validation of the novel AS events identified by RNA-seq

Due to the high complexity of the RNA-seq data and to the fact that the tools for the analysis of the AS events are not completely accurate, validation of some of the AS event was done by semi-quantitative PCR. Various types of AS events found by different tools during the differentiation of the cell line model were randomly selected and PCR primers into the flanking exons of the AS exon were designed. Validation of a total of 18 AS events was performed by radioactive semi-quantitative RT-PCR, and I contributed to validate some of those AS events, which are all shown here (Figure 38). A clear example of validated case is the inclusion of a cassette exon in the FAM126A gene (Figure 38 top left). The PSI value in the RNA-seq for the THP-1 control sample is close to zero indicating almost total exclusion of the exon which is in agreement with the RT-PCR where only the lower exclusion band was detected. In all the other samples where cells are differentiated, there is an increase in the PSI values detected by RNA-seq which excluding for VD3 is again in agreement with the RT-PCR in which an upper inclusion band starts to appear in addition to the lower exclusion band. Values of the quantification of the inclusion bands by the typhoon scanner are indeed also higher in the differentiated cells compared to control.

Figure 38: Validation of AS in cell line.

On top of bar charts, schematic of the type of AS event is depicted. SE (Skipped exon), A5SS (Alternative 5'ss). At the bottom, scanned images of X-ray films after exposure to the radioactive gels. Bar charts on top of X-ray images show the changes in PSI for the specific AS event found by RNA-seq. Upper band indicates inclusion of the exon while lower band exclusion. Below the X-ray images, values of the quantification of the inclusion bands by the typhoon scanner are indicated.

A network of the common AS events between the cell line model and primary cells was built in Cytoscape to further assess what are the functions of the genes in which the AS events were detected. The GO analysis shows that the genes have mainly intracellular functions (Figure 39).

Figure 39: GO Network of common changes in AS between THP-1 and primary cells.

Network of the common AS events generated in Cytoscape (ClueGO). Different colors represent different GO terms.

Further GO analysis by the online tool DAVID was performed by Dr. Hongfei Liu on the DEGs and AS events detected in the primary cells. It was found that the DEGs and the genes in which AS was detected have different functions (Figure 40).

Figure 40: Results of the GO analysis by DAVID.

Enrichment of GO terms for DEGs and alternatively spliced genes in primary cells.

In addition, Dr. Hongfei Liu found that many splicing factors change expression during the transition from monocytes to macrophages. In an effort to identify which factors could be involved in the differentiation process, it was found by Dr. Hongfei Liu that MBNL1 seems to be required for the proper differentiation of THP-1 cells because by knocking it down, cells do not differentiate as well as WT cells as measured by specific flow cytometry markers.

In conclusion, we identified thousands of DEGs and AS events that change during the transition from monocytes to various subtypes of macrophages, with hundreds of them in common between the cell line model and the primary cells. We also identified MBNL1 as the major splicing factor playing a role in the differentiation process (Liu et al., 2018, in press).

Aim 2

In the following sections the results of the transcriptomic analysis of the *PRPF40B* KO cells and of two selected MDS mutations of the gene will be described.

3.12 Knock out of PRPF40B by CRISPR

To investigate the role of PRPF40B in regulating AS, CRISPR Cas9 was used to knock out the gene by deleting a genomic portion as previously described in section 2.10, which results in a frameshift leading to the transcription of a defective mRNA.

After several trials to KO the gene in THP-1 cells, we decided to knock it out in K562 cells instead because we found that THP-1 cells are much more difficult to genetically manipulate compared to K562. However, the reason is not because K562 have a lower number of chromosomes, in fact it has been reported that they are hypotriploid (Naumann et al., 2001). An explanation is that in THP-1 intracellular pathways such as cGAS against foreign DNA are activated (Diner et al., 2013), thus making them harder to transfect with plasmid DNA. An additional reason for choosing K562 over THP-1 is because there is no MDS cell line publically available, and K562 is the closest cell line which has been used in the literature as MDS model (Dolatshad et al., 2015). Another advantage of using K562 is that it can be differentiated to both erythroid and myeloid cells while THP-1 can only be differentiated to myeloid, thus making K562 cells a better model for MDS (Sutherland et al., 1986). Genotyping PCRs were performed on the genomic DNA extracted from dozens of single cell clones in order to identify those in which the deletion was complete. If the deletion was not introduced only an upper band of more than 1000 bp was observed, while if there was a complete deletion only a lower band of more than 300 bp was observed (Figure 41). If the deletion was partial, different ratios of upper and lower bands were observed. Only clones with a complete deletion for all alleles were selected. However, for certain clones where the PCR showed a complete deletion, it was found that if the cells were kept in culture for longer time, an upper band would reappear. It was hypothesized that this could be due either to the fact that the expanded cells were not coming from a single cell clone and that there was a bias in the PCR in favor of the lower band which gave false positive results. Considering that the lower band would be much more easily amplified compared to a much larger band, a potential positive clone with no upper band detected could be instead a negative clone as there would be still some residual upper band not detected by the PCR. To further confirm the complete deletion in each specific clone, a second round of genotyping PCR on the potential positive clones found in the first round of PCR was performed. The second PCR would consist of an internal primer inside the deletion paired with a downstream primer external

to the deletion (Figure 41). If there was no complete deletion an upper band of 500 bp would be amplified while if there was a complete deletion no band would be amplified because that the internal primer would not bind to the DNA, thus eliminating the bias toward a lower band. Various clones with a complete deletion confirmed with the two genotyping PCRs were obtained and frozen down. Three clones from this batch were selected for the downstream experiments.

Figure 41: Results of the genotyping PCRs.

A panel showing schematic of the two genotyping PCRs. Boxes represent the PRPF40B exons targeted for deletion. Dotted lines represent introns. In red and blue are indicated the binding sites of the two sgRNAs. In green, the binding sites of the two sets of PCR primers. P1 and P3 were used for the first round of PCR while P2 and P3 for the second. B panel showing agarose gel image for the two PCRs. Ladder (L), negative control (-), genomic DNA of K562 WT cells (+), genomic DNA of the three single cells clones are indicated as 1, 2, 3.

Lower bands of 500 bp were gel extracted with QIAquick gel extraction kit (Qiagen) according to manufacturers' protocol. Samples were submitted to 1st Base for Sanger sequencing by using sequencing primer flanking the deleted region. From the sequencing results it was observed a complete deletion and joining of the flanking DNA regions in the genome of all the 3 clones. Deletion was considered to be in all alleles since nucleotide peaks were clearly single and the deletion junction was identical in all clones (Figure 42).

Figure 42: Sequencing results of the PRPF40B KO clones.

Sanger sequencing results of genomic DNA from WT K562 cells and 3 KO clones. On top, lower case letters indicate introns while exons are in upper case. Blue sequence indicates the binding sites of the two sgRNAs while red sequence indicates the NGG PAM site. Red triangles indicate the site where the Cas9 likely introduces a cut. Red arrow shows the point where a junction was formed after deleting the genomic portion in WT cells. Peaks with different colors represent the different nucleotides.

3.13 WB and qRT-PCR results of PRPF40B KO clones

In order to assess the KO of *PRPF40B* by CRISPR, the levels of the mRNA and protein were measured by qRT-PCR and WB respectively (Figure 43).

It was found that upon KO of the gene the mRNA and protein have totally disappeared in each of the 3 single cell clone compared to WT cells. It was thus concluded that the PRPF40B KO was complete.

Figure 43: Results of WB and qRT-PCR of the PRPF40B KO.

Top panel shows protein expression level of PRPF40B in the WT cells and in 3 single cell clones measured by WB. GAPDH was used as loading control. Lower panel shows the mRNA expression of PRPF40B measured by qRT-PCR in the same 3 clones. Values of the qRT-PCR were normalized against β -actin.

3.14 Rescue experiments

In order to restore expression of WT PRPF40B and to overexpress the two MDS mutations of the gene, several plasmids were cloned. Cloned plasmids were sent to 1st Base for sequencing by using sequencing primers as previously described. Sequencing results show correct deletion of the N-terminus portion of PRPF40B isoform A to obtain isoform C. Furthermore, the two mutations were correctly introduced into the WT gene (Figure 44).

Figure 44: Sequencing results of PCR mutagenesis.

Sanger sequencing results of the plasmid used for the overexpression. On top is shown the sequence of the PRPF40B longer isoform A: the part deleted to generate the shorter isoform C is underlined in red while blue denotes the retained part. In green is the start codon while in yellow the FLAG tag sequence. At the bottom, Sanger sequencing results of the two mutations are indicated.

In order to find out which is the best time point to harvest proteins after rescue with the PRPF40B expression plasmids, a time course experiment in which proteins were harvested after 24 h, 48 h, 72 h or 96 h post-transfection was carried out. It was found that 72 h after transfection is the best time point to collect the proteins, because at 72 h PRPF40B accumulated at similar levels than in the WT K562 (Figure 45).

In addition, in order to find the most suitable concentration of plasmid to be used for the rescue experiments, various amounts of PRPF40B WT vector were transfected. Isoform C was employed for all the subsequent overexpression experiments. It was found that 100 ng is a

sufficient amount of plasmid which can result in a protein level similar to the endogenous amount of PRPF40B in K562 non transfected cells (Figure 45).

Lastly, the two plasmids in which PRPF40B was mutated (P383L and P540S) were overexpressed together with the two PRPF40B WT isoforms A and C (Figure 45D). In this case overexpression was performed with both WT isoform A from genescript and the derived isoform C from it in order to double confirm the molecular weight of the protein. It was found that overexpression of 100 ng of each plasmid results in production of an amount of protein similar for all the vectors (Figure 45).

Figure 45: Rescue experiments results.

qRT-PCR and Western Blot results for the overexpression of the WT and mutants. Panels A and B show overexpression with various increasing amounts of PRPF40B WT plasmid. At the bottom, time course experiment on panel C and overexpression with the mutated plasmids on panel D. WT (A) and WT (C) indicate overexpression with PRPF40B isoform A and isoform C respectively. Quantifications of the WB bands by densitometry are indicated below each band. GAPDH was used as loading control.

3.15 RNA-seq analysis of the PRPF40KO and of the overexpression samples

Total mRNA was extracted from 3 replicates of CRISPR non edited K562 WT cells, from the 3 single cell *PRPF40B* KO clones and from the overexpression of the PRPF40B WT gene and of the mutations either alone or mixed in a 1:1 ratio with the WT vector. Samples were submitted to the sequencing facility at SCELSE for RNA QC and only those samples with RIN higher than 8 were further processed for RNA-seq. A similar bioinformatics pipeline from the first aim was employed with the exception that for the analysis of the AS events only MATS was used. The reasons why only MATS was used were because it was found that in our hands it is more accurate than the other two tools and it is the only tool which can handle replicates, which facilitates the analysis. Paired end reads from the various conditions were aligned to the reference human genome hg19 by allowing a maximum of two mismatches. The majority of the reads were successfully aligned to the genome (Figure 46).

Sample	Total Reads	Aligned	% Aligned
k562wt1	50835985	47199209	92.80%
k562wt2	53560741	49543671	92.50%
k562wt3	54434481	51982962	95.50%
KO1-B1	54029223	50644127	93.70%
KO2-B2	52996375	49624932	93.60%
KO3-B8	52123470	48904479	93.80%
B1-Wt	55570498	52433548	94.40%
B1-383	52977955	49665775	93.70%
B1-540	55523931	52473665	94.50%
B2-Wt	54724442	51537953	94.20%
B2-383	53315449	50278358	94.30%
B2-540	54191326	51123714	94.30%
B8-Wt	57841100	54316060	93.90%
B8-383	61609629	57870356	93.90%
B8-540	53366903	49839035	93.40%
B1-WT/383	53063244	49961391	94.20%
B1-WT/540	50305902	46289538	92.00%
B2-WT/383	52711998	49565862	94.00%
B2-WT/540	53153231	49560260	93.20%
B8-WT/383	50715167	47463056	93.60%
B8-WT/540	57342902	53652256	93.60%

Figure 46: Alignment results of the RNA-seq of the PRPF40B KO and MDS mutations.

Statistics for the total reads obtained from RNA-seq. Numbers and percentages of total and aligned reads are indicated. B1, B2, B8 denote the numbers of the 3 single cell clones. WT indicates overexpression of the WT gene, overexpression of the P383L and P540S mutations are abbreviated as 383 and 540 if they were overexpressed as homozygous-like. WT/383 and WT-540 indicate overexpression of the mutation as 1:1 ratio with WT.

Cummerbund was employed to assess the correlation between replicates. It was found that replicates correlate well with each other (Figure 47).

Figure 47: Results of correlation between replicates.

Bar charts showing correlation between the various replicates generated by Cummerbund based of FPKM values.

3.16 Analysis of the DEG upon KO and overexpression

A fold change larger and equal to 0.3 or lesser and equal to -0.3 and a p-value lesser and equal to 0.05 were used to derive significant DEGs. It was found that several genes change expression upon KO of PRPF40B, and some of them were rescued when PRPF40B was reintroduced in the KO cells (Figure 48A). However it was observed that most of the genes rescued by overexpressing the WT gene were also rescued by the mutants. It was observed that the heterozygous-like mutations are rescuing the DEGs regulated by PRPF40B WT less well compared to the homozygous-like mutants (Figure 48B). It was concluded that besides being involved in regulating AS, PRPF40B is regulating directly or indirectly the mRNA expression of a certain subset of genes.

Figure 48: DEGs results of the PRPF40 KO and rescue experiments

Venn diagrams for the DEG regulated by PRPF40B. On top (A), Venn diagram of the DEGs regulated by PRPF40B WT. On the bottom (B), pie charts of the rescued versus not rescued genes when the mutants were overexpressed.

Gene set over-representation analysis for the 408 rescued DEGs was performed by the online tool ConsensusPathDB (Max Plank Institute for Molecular Genetics). It was found that the top two enriched pathways are cholesterol biogenesis and HIF1A signaling (Figure 49).

pathway name	set size	candidates contained	p-value	q-value	pathway source
Cholesterol Biosynthesis	15	10 (66.7%)	1.59e-13	1.15e-10	Wikipathways
HIF-1-alpha transcription factor network	67	17 (25.4%)	2.4e-13	1.15e-10	PID
Cholesterol biosynthesis	24	10 (41.7%)	8.58e-11	2.73e-08	Reactome
superpathway of cholesterol biosynthesis	25	10 (40.0%)	1.4e-10	3.02e-08	HumanCyc
Activation of gene expression by SREBF (SREBP)	19	9 (47.4%)	1.88e-10	3.02e-08	Wikipathways
Activation of gene expression by SREBF (SREBP)	26	10 (38.5%)	2.23e-10	3.02e-08	Reactome
Sterol Regulatory Element-Binding Proteins (SREBP) signalling	68	14 (20.6%)	6.61e-10	3.02e-08	Wikipathways

Figure 49: Results of GO analysis of the DEGs regulated by PRPF40B

Name of the pathways found to be enriched is indicated on the right. In red, number of genes involved in that particular pathway which are present in our data set (Kamburov et al., 2011).

In addition, a heat map of the most important selected genes involved in hypoxia according to the literature was generated by the R package heatmap2 in R. The heatmap identifies a large cluster of genes which are upregulated in the KO and which are then rescued. Other genes are instead downregulated while others barely change expression (Figure 50).

Figure 50: Heat map of DEGs in hypoxia.

Heat map generated with the R package pheatmap of the most important genes known to be involved in hypoxia in the literature. Light blue means downregulation while dark blue upregulation. Red box highlights genes with stronger changes.

It was found by a more in depth analysis of the DEGs regulated by *PRPF40B* that some of those genes are involved in development (HOX genes) and others are accordingly to the literature targets of the transcription factor KLF1 which regulates erythroid differentiation (Pilon et al., 2008) (Figure 51). Thus, *PRPF40B* could be also involved in the regulation of genes which are important for erythroid differentiation.

Figure 51: DEGs with role in development and erythroid differentiation.

Heat map generated with the R package pheatmap for the HOX genes and for the KLF-1 targets. Light blue means downregulation while dark blue upregulation. Red box highlights genes with stronger changes.

3.18 Validation of the DEG regulated by PRPF40B

In order to assess the reliability of the RNA-seq data, validation of selected DEGs involved in hypoxia was performed by qRT-CPR by FYP student Miss Resilind Chew and by me.

We validated genes involved in hypoxia known to regulate different pathways (Figure 52). For instance, BNIP3 regulates autophagy, ENO2 and PGK1 are involved in glycolysis and CA9 regulates pH (Swietach et al., 2007; Yeh et al., 2008; Zhang and Ney, 2009). TGF- β 3 is instead involved in ECM remodeling and COL1A2 in collagen production (Karamichos et al, 2011; Baumann et al., 2016). It was found that excluding TGF- β 3, which shows an opposite trend in the first 3 samples compared to RNA-seq, all the genes validated confirm what was found by RNA-seq (Figure 52). We did not consider TGF- β 3 validated because the RNA-seq shows it is downregulated in the KO cells while the qRT-PCR shows it is upregulated.

Figure 52: qRT-PCR validation results.

qRT-PCR results of 3 technical replicates for the validation of selected DEG involved in hypoxia. In blue are RNA-seq FPKM average values while in black qRT-PCR expression average values for the corresponding gene.

3.19 Analysis of the expression of PRPF40B and its homologue PRPF40A in various human cancers

Due to the finding that PRPF40B represses hypoxia, it was hypothesized that it could act as tumor suppressor. Thus, we checked the mRNA expression of PRPF40B and its homologue PRPF40A in various human cancers according to The Cancer Genome Atlas (TCGA) in cBioportal (Gao et al., 2013, Cerami et al., 2012).

It was found that in AML, *PRPF40B* is at its lowest level (Figure 53A). On the contrary, *PRPF40A* is expressed higher in AML than in any other cancer (Figure 53B).

Figure 53: PRPF40A and PRPF40B mRNA expression in TCGA database.

Expression of PRPF40A (A) and PRPF40B (B) in various human cancers (cBioportal) (Gao et al., 2013, Cerami et al., 2012). Blue dot represents the mRNA expression level in a specific human patient. Expression is based on FPKM values of RNA-seq data. Red dot represents mutation.

We checked the expression of *PRPF40B* also in other human cancer databases such as GEPIA, GENT and ONCOGEO and the expression profile was similar to TCGA (Figure 54) (Shin et al., 2011; Tang et al., 2017).

Figure 54: *PRPF40B* mRNA expression levels in various cancers.

Example of the mRNA expression profile of *PRPF40B* in various types of cancer according to ONCOGEO database (OncoLand, Omicssoft). Colored dots represent different type of tumors.

Interestingly it was thus found that the expression of *PRPF40A* and *PRPF40B* are inversely correlated in AML (Figure 55). A recent report has shown that *PRPF40A* mRNA is induced in lung cancer and its expression correlates with hypoxia markers (Oleksiewicz et al., 2017), which is consistent with the two genes having opposing functions, considering that in this study *PRPF40B* seems to be instead repressing it.

Figure 55: Inverse correlation of PRPF40A and PRPF40B mRNA expression levels in AML patients.

Correlation of the mRNA expression levels of PRPF40B (x-axis) and PRPF40A (y-axis) measured by RNA-seq in AML patients accordingly to TCGA database. Blue dots represent the mRNA expression level of the two genes in a specific human patient. N denotes the size of the cohort. Red line is the intersect representing inverse correlation.

Since PRPF40B regulates the expression of genes involved in hypoxia, we checked the correlation at the mRNA level between HIF1A and PRPF40B in AML patients according to the TCGA database. It was found that the 2 genes are inversely correlated (Figure 56).

Figure 56: Inverse correlation of HIF1A and PRPF40B mRNA expression levels in AML patients.

Correlation of the mRNA expression levels of PRPF40B (x-axis) and HIF1A (y-axis) measured by RNA-seq in AML patients accordingly to TCGA database. Blue dots represent the mRNA expression level of the two genes in a specific human patient. N denotes the size of the cohort. Red line is the intersect representing inverse correlation.

3.20 Assessment of PRPF40A protein level in the KO cells (WB)

Because it was found that the mRNA expressions of *PRPF40A* and *PRPF40B* are inversely correlated, we then measured the protein expression of PRPF40A in the *PRPF40B* KO cells by WB. We did not detect any clear change in PRPF40A expression at the protein level after KO of PRPF40B (Figure 57). This further confirms that the AS changes observed are truly attributable to PRPF40B regulation and not to its homologue.

Figure 57: PRPF40A WB results in the KO cells.

Top panel shows protein expression level of PRPF40A in three replicates of K562 WT cells and in the 3 single cell PRPF40B KO clones measured by WB. GAPDH was used as loading control (bottom).

3.21 Analysis of the AS events upon KO and overexpression

MATS was used in order to discover AS events regulated by PRPF40B and any difference with the mutants. Hundreds of AS events were discovered in all the various conditions (Figure 58).

Condition	KO	Rescue	P383L	P540S	WT-383	WT-540
A3SS	1791	913	1075	964	1024	1036
A5SS	1408	986	965	1004	1005	1032
IR	723	448	507	477	472	440
SE	13359	12395	12964	12261	12904	12956
MXE	16923	18199	18559	19425	18947	17853

Figure 58: RNA-seq data analysis of AS.

Statistics of all the AS events found by MATS. Rows indicate the type of the AS event while columns the various conditions. Cutoff of $FDR < 1$ was applied. Alternative 5'ss (A5SS), Alternative 3'ss (A3SS), Intron Retention (IR), Skipped exon or cassette exon (SE), Mutually exclusive exons (MXE).

Since we mainly focused to study the PRPF40 KO cells, we did a more in depth analysis of all the AS events regulated by PRPF40B. AS events were sorted based on a cutoff of an FDR <1 and a Δ PSI larger or equal to 0.1 or lesser or equal to -0.1. Δ PSI provides the information of the inclusion level of an exon. A negative value means an exon is more excluded while positive more included. We calculated the number of AS events that are rescued after reintroducing PRPF40B expression in KO cells. A total of 1,396 events are regulated by PRPF40B (Figure 59).

Event type	K562 vs KO	KO vs WTov	Common Rescued
SE	2544	2302	793
IR	145	91	54
MXE	1790	1841	317
A5SS	381	257	123
A3SS	376	203	109

Figure 59: PRPF40B regulated AS events.

Statistics of all the AS events found by MATS. Rows indicate the type of the AS event. K562vsKO indicates the events that change upon KO while KO vs WTov those that change upon PRPF40B overexpression in KO cells. Last column indicates rescued events in common between the two comparisons. Cutoff of FDR<1 and $\text{PSI} \leq -0.1$ or ≥ 0.1 was applied. Alternative 5'ss (A5SS), Alternative 3'ss (A3SS), Intron Retention (IR), Skipped exon (SE), Mutually exclusive exons (MXE).

More in detail, it was found that hundreds of cassette exon events (here indicated as SE) changed the level of inclusion or exclusion upon KO and in rescued cells (Figure 59). It was found that 793 out of 4,846 events were rescued by reintroducing the PRPF40B gene in the KO cells. The 793 events belong to 477 unique genes (Figure 60). AFMID, ZNF76, SUGP2, MRPL55 LUC7L2, IP6K2, CTNND1 were found to be the genes with the most number of AS events detected. It was concluded that PRPF40B is truly regulating AS. In addition, upon PRPF40B KO, 459 SE events show more exon inclusion while 333 more exon exclusion.

Figure 60: Cassette exon events regulated by PRPF40B.

Venn diagrams showing the numbers of cassette exon events regulated by PRPF40B. Blue circle indicates the events that change upon KO while the yellow circle the events that change when PRPF40B was overexpressed in the KO cells. Intersection shows the number of events in common between the two comparisons and which were rescued when PRPF40B was reintroduced. Cutoff of $FDR < 1$ and $PSI > 0.1$ or $PSI < -0.1$ was applied. Number of unique genes in which the AS events were found is indicated.

Furthermore, it was found that when PRPF40B was reintroduced as MDS mutant rather than WT, a portion of the 793 AS events were not rescued. Intriguingly, it was observed that the AS events were rescued less efficiently when the mutations were overexpressed when mixed with WT plasmid rather than alone (Figure 61).

Figure 61: Analysis of the AS events regulated by mutations of PRPF40B.

Venn diagrams showing the portions of the cassette exon events which were rescued or not by overexpressing the mutated plasmids either as homozygous-like (left) or heterozygous-like (right).

3.22 Validation of the AS events regulated by PRPF40B

In order to assess the reliability of the RNA-seq data, validation of selected AS events found by MATS was performed by RT-PCR. It was found that excluding some exceptions most of the genes underwent AS changes as found by the RNA-seq. It was found that when the mutations were overexpressed as 1:1 ratio with WT rather than heterozygous-like, the AS events were rescued less efficiently in 8 out of 20 cases (Figure 62).

Figure 62: Validation of the AS events regulated by PRPF40B.

Results of the validation of the AS events. Bar charts show the changes in PSI from the RNA-seq data for one AS event in a specific gene while at the bottom are shown PCR bands from agarose gels for the same AS event. Values of the quantifications for the inclusion bands are shown below the images.

For instance, we detected AS change in exon 23 of the SPTAN1 gene which express the alpha II-spectrin protein (Figure 63). Spectrins are cytoskeletal proteins which are involved in actin crosslinking, cell adhesion, intercellular communication and cell cycle regulation (Machnicka et al., 2014). Alpha II-spectrin is highly expressed in cardiac muscle cells and mutations of the gene are associated with neurological disorders, while Alpha I-spectrin is instead expressed in erythrocytes (Lewis et al., 2008). Accordingly to Uniprot database, the exclusion of exon 23 results in a switch from the canonical isoform 1 to an alternative isoform 3 (Figure 63).

Figure 63: Visualization of SPTAN1 AS.

Example of cassette exon event regulated by PRPF40B visualized with Integrative Genome Viewer (IGV) (Thorvaldsdóttir et al., 2013). Shown is representative of three replicates for the conditions indicated on the left. Grey arrows represent reads from RNA-seq aligned to the human genome. Blue boxes represent exons while blue lines introns. Colors indicate a read whose mate was mapped to a different location in the genome.

Considering recent evidence that AS can affect PPI, due to the fact that AS events significantly occur in regions where a protein interaction domain or a short linear motif is present, a PPI network of the AS events regulated by PRPF40B was built using the STRING database (Yang et al, 2016). Several clusters were identified in the network. In particular, the alternatively spliced genes regulated by PRPF40B are mainly involved in the Akt/MAPK1 cascade and ubiquitination (Figure 64).

Figure 64: PPI network of the AS events regulated by PRPF40B.

STRING database PPI network of the AS event regulated by PRPF40B (Yang et al, 2016). Circles represent the various clusters highlighted with different colors. Proteins are shown as spheres (nodes) and lines between them represent interactions (edges).

The functions of the DEGs and AS events regulated by PRPF40B were compared by Enrichr, an online tool which gives information about the enrichment in GO terms and others categories in a queried gene set (Kuleshow et al., 2016). Similarly to what was found by ConsensusPathDB, DEGs are mainly involved in steroid biogenesis, HIF1 signaling and glycolysis (Figure 65A). It was found instead that AS events are enriched in genes having binding sites for POLR2A, a subunit of Pol II, based on a Chip-seq data in K562 cells. Furthermore, the genes are shown to have enrichment in the H2AFZ histone marker and are involved in the MAPK signaling pathway and in RNA binding (Figure 65B).

Figure 65: Gene set enrichment analysis of the DEGs and AS events regulated by PRPF40B.

Result of the Enrichr analysis for the DEGs and AS events regulated by PRPF40B (Kuleshow et al., 2016). Top 3 categories sorted by p-value for each GO term and various databases are shown.

Blue color represent most enriched term in the correspondent category while orange second and gray third.

Figure 66: Gene set enrichment analysis of the DEGs and AS events regulated by PRPF40B by KEGG database.

Result of the Enrichr analysis for the DEGs and AS events regulated by PRPF40B (Kuleshow et al., 2016) in the KEGG database. Top 10 categories sorted by p-value for each GO term and various databases are shown. Panel A shows the functions of the DEGs while panel B of the AS events.

3.23 Features of the AS events regulated by PRPF40B

Various features of the 793 cassette exon events were assessed (Figures 66, 67, 68 and 69). For instance, we checked the general distribution of the exon lengths in addition to the lengths of the introns flanking those events. Their 5'ss and 3'ss strengths (degree of match to the best motifs) were obtained by the online tool MAXENT. It was found that the cassette exon events regulated by PRPF40B have a weaker 5'ss and 3'ss compared to AS events which are not regulated by PRPF40B (Figure 66). In addition, there were no differences in the lengths of the exons while the flanking upstream introns are shorter (Figure 66). Lastly, we performed motif enrichment analysis (Figure 69).

Figure 67: Features of the cassette events regulated by PRPF40B.

Panel A and B show boxplots of the 5'ss and 3'ss strengths. In panel C and D, boxplots for the intron and exon lengths. P-values of the comparisons are indicated on top of each figure.

Considering there was a recent report claiming that exons with a weaker 5'ss or 3'ss could be more sensitive to changes in the expression of PRPF40B (Becerra et al., 2016), we checked

globally the correlation between the strength of the splice sites of the cassette exon events obtained with MAXENT and their PSI values, obtained by MATS. However, we did not find any correlation between strength and change in the PSI for the 793 rescued events (Figure 67).

Figure 68: Correlation of the PSI and 5'ss strengths of the cassette events regulated by PRPF40B.

Correlation between MATS PSI values on the x-axis and MAXENT 5'ss values on the y-axis. P-value and correlation coefficient are indicated below the graph.

In addition, considering that PRPF40B is involved both in the regulation of AS and transcription according to recent reports, we checked the transcription elongation rates of the 477 unique genes based on a Bromouridine-sequencing (Bru-seq) dataset in K562 cells publically available (Veloso et al., 2014) (Figure 68). Bru-seq is a methodology which maps sequences of nascent

RNA transcripts and determines relative transcription rate of each gene. We observed that the 477 genes in which an AS event was found by MATS have a higher transcription rate than the average rate of the genes in the dataset. However, when we selected AS events which were not regulated by PRPF40B, we also found a similar trend. We thus concluded that a faster elongation rate was not specific to PRPF40B AS regulated events (Figure 68).

TRANSCRIPTION ELONGATION RATE

Figure 69: Transcription elongation rate of the cassette exons events regulated by PRPF40B.

Boxplots showing the distribution of the values of the rate of transcription for the AS in genes regulated by PRPF40B (PRPF40B regulated) in comparison to genes with AS events which are not regulated by it (AS non regulated), and to all the genes expressed by WT K562 cells (K562 Bru-seq). Values of the rate of transcription are taken from a Bru-seq experiment (Veloso et al., 2014). Dots represent outliers. Median value is indicated with a bold black straight line.

3.24 Motifs analysis of the PRPF40B regulated cassette exon events

We performed motifs analysis by DREME, a subprogram of the online web portal MEME Suite v4.12.0 (Bailey et al., 2009), in order to discover which are the most enriched motifs in the sequences of the PRPF40B regulated exons (Figure 69). DREME discovers short, ungapped motifs that are relatively enriched in input sequences compared with shuffled sequences. Exon sequences of the 793 events with 200 bp of the flanking upstream and downstream introns were retrieved from the genome by using bedtools and used as input.

Figure 70: Motifs analysis of the cassette exons events regulated by PRPF40B.

Result of the top 10 motifs found by DREME (Bailey et al., 2009). RC Logo indicates the reverse complement (RC) of the motif. E-value is an estimate of how likely it is that the motif is not just a statistical artifact.

3.25 Splicing minigenes

In order to further validate the AS events regulated by PRPF40B and to investigate how PRPF40B is regulating splicing, experiments with minigenes were performed. Several minigenes for RIPK2 and SPTAN1 with three exons and two introns, and the PRPF40B-regulated cassette exon in the middle, were cloned and transfected as previously described. We selected those two genes because they showed a very obvious change in AS based on our validation (Figure 62). PCR with primers specific to the flanking regions of the middle exon were used in order to assess any AS of the transfected cells compared to WT cells.

3.25.1 UMV minigenes

RIPK2 Exon 2 and SPTAN1 exon 23 with portions of their flanking introns were cloned into the UMV vector as previously described. The UMV vector contains already flanking constitutive exons. Plasmids were transfected in WT, PRPF40B KO and rescued cells. Primers designed into the UMV flanking exons were used for the PCR. For the transfection with RIPK2-EXON2-UMV we detected only one band, while for transfection with SPTAN1-EXON23-UMV we detected two bands which show in WT cells a similar splicing pattern to the validation done previously, but there was no AS change in the KO cells (Figure 70). We concluded that the UMV minigenes do not recapitulate the PRPF40B regulated AS.

Figure 71: Results of UMV minigenes

Agarose gel image of PCR bands after transfection with UMV minigenes.

3.25.2 pCDNA3.1+ minigenes

We aimed to test splicing minigenes with a closer architecture to the endogenous AS event thus RIPK2 Exon 2, SPTAN1 Exon 23 with their endogenous flanking exons were cloned into the pCDNA3.1+ vector as previously described and transfected into K562 WT, *PRPF40B* KO and rescued cells. Primers specific to the pCDNA vector were designed and used for the PCR. In this case we detected two bands for the RIPK2-EXON2 in K562 WT transfected cells which recapitulates the splicing pattern observed from the previous validation. However, no change in AS was detected in the transfected KO cells. Regarding the SPTAN1 transfected cells, we instead observed more skipping of exon 23 in the WT cells which contradicts what was found by the validation (Figure 71). From the experiments with the UMV and pCDNA minigenes, we observed that the splicing patterns found during the validation of the AS events was not recapitulated.

Figure 72: Results of pCDNA minigenes.

Agarose gel image of PCR bands after transfection with pCDNA minigenes.

We hypothesized that other factors such as chromatin state and/or rate of Pol II transcription might be influencing the AS outcome, thus we did not carry out additional experiments with the minigenes.

3.26 RNA Pol II inhibition

In order to prove that there is a connection between the splicing factor PRPF40B and mRNA transcription, a Pol II inhibition assay was performed. DRB was used to stall the polymerase at the initiation stage while CPT was used to inhibit its elongation. As a positive mRNA control, exon 8 of hnRNP DL, which is known to increase upon inhibition of the Pol II, was used (Ip et al., 2011). It was found that there is an increase in inclusion of exon 8 in PRPF40B KO cells at a comparable level of WT K562 cells treated with DRB and CPT and this increase was fully reversed by the rescue experiment (Figure 72). In addition, we still observed an increase in exon 8 inclusion upon DRB treatment of the KO cells.

Figure 73: Pol II inhibition assay.

Result of the Pol II inhibition assay. K562 WT, PRPF40B KO and rescued cells were treated for 5 h with vehicle (DMSO) or drugs to inhibit Pol II (DRB and CPT). Percentage of inclusion indicates the mean values for the inclusion of hnRNP DL exon 8 of 5 replicates. SD indicates the standard deviation.

We checked the AS of hnRNP DL exon 8 in our RNA-seq data and we observed a slightly higher inclusion of that exon upon PRPF40B KO in all the 3 clones (Figure 73).

Figure 74: Visualization of hnRNP DL exon 8.

IGV visualization of hnRNP DL exon 8 in the various replicates (Thorvaldsdóttir et al., 2013). Grey arrows represent reads from RNA-seq aligned to the human genome. Blue boxes represent exons while blue lines introns. Colors indicate a read whose mate was mapped to a different location in the genome.

3.27 Assessment of Apoptosis, DNA damage and proliferation in PRPF40B KO cells

In order to further characterize the PRPF40B KO clones, apoptosis, DNA damage and proliferation were measured by flow cytometry as previously described in section 2.23. Our preliminary results - did not detect any major difference in apoptosis, DNA damage and proliferation when comparing the KO clones to WT cells (Figure 74). In order to make final conclusions about cell cycle, a staining with propidium iodide (PI) could be included in the experiment. In addition, a longer incubation with BrdU could be tried in order to make sure the staining was completely achieved.

Figure 75: Flow cytometry results for the assessment of apoptosis, DNA damage and proliferation in KO cells.

Flow cytometry histograms of a representative of three replicates. Proliferation was assessed by staining with anti-BrdU antibodies, DNA damage by anti-H2AX antibodies and apoptosis by anti-PARP antibodies in K562 WT cells (WT) or 3 KO clones (KO1, KO2, KO3).

4 DISCUSSION AND FUTURE DIRECTIONS

Recent reports have shown that AS plays a role in cell proliferation, apoptosis and in cell lineage differentiation in a tissue specific context (Kazantseva et al., 2013; Kim et al., 2013; Schwerk, and Schulze-Osthoff, 2005). Furthermore, AS is involved in the activation of various cells of the immune system such as T cells, B cells and DCs (Alt et al., 1980; Lynch, 2004, Rodrigues et al., 2013). Very little was known about the function of AS in monocytes when this Ph.D. was started. In more recent times, studies on the role of AS in those cells were published. For instance, a study by de Bruin et al., has shown that the protein Quaking (QKI) is involved in the regulation of AS and in the differentiation of M-CSF human macrophages (deBruin et al., 2016). Another study by Lin et al., has investigated AS in those cells in the context of cardiometabolic diseases (Lin et al., 2016). Alasoo et al., have instead found AS events that change in macrophages differentiated from human induced pluripotent stem cells (IPSDMs) when treated with LPS (Alasoo et al., 2015). Those studies have come to the conclusion that AS plays a role in the differentiation to macrophages in a stimulus-specific fashion. In the past, AS has been investigated by microarray and/or exome arrays (Cuperlovic-Cuif et al., 2006). The advent of RNA-seq technology provides an exceptional opportunity to study in depth changes in AS on a genome-wide scale in a less biased manner (Pan et al., 2008).

For the first aim of this Ph.D., we have investigated the role of AS in the differentiation of monocytes to macrophages. We have employed RNA-seq in order to find novel AS events that change during the differentiation. In addition, we aimed to find any possible splicing factor which could act as a major regulator during that process. For the second aim of my Ph.D., I focused to study the biology of PRPF40B, which is one of the various splicing factors found to be mutated in MDS (Yoshida et al., 2011). I have used the same RNA-seq pipeline implemented for the first aim of this Ph.D. in order to find AS events regulated by PRPF40B. Furthermore, I have investigated the impact on gene expression and AS when two of the MDS mutations were introduced into the WT gene.

4.1 Comparison between differentiation protocols of the cell line model and primary cells

Regarding the first aim of this Ph.D., focused to study the role of AS in the differentiation of human monocytes, we started by performing experiments with a cell line rather than primary

cells because THP-1 can be easily differentiated and it is easier to obtain large amount of differentiated cells. Our experiments are based on an in culture model of differentiation, thus we cannot excluded that culturing cells on a plastic surface would be affecting to a certain extent our results. However, an old report has shown that the changes observed during an in culture differentiation of macrophages are in general of the same kind of an *in vivo* differentiation (Hammerstrom, 1979). Indeed, we opted for experiments in culture since they are more cost-effective and less challenging to perform rather than *in vivo* and we are still able to obtain very useful information. According to what is reported, THP-1 cells can be differentiated to macrophages by different stimuli (Schwende et al., 1996). Classical stimulations are PMA, VD3 or retinoic acid while IL-32 γ is a novel cytokine which has been shown to differentiate monocytes to macrophage-like cells (Netea et al., 2008). These stimulations activate different signaling pathways (Makishima et al., 1999; Tahara et al., 2009). For instance PMA is a small molecule that can cross the cell membrane and activates Protein Kinase C (PKC), which results in induction of downstream signaling pathways. PMA is a compound that is used to mimic diacylglycerol (DAG), a second messenger in our bodies (Zhong et al., 2008). Instead, VD3 can bind to the retinoic acid nuclear receptor whose complex activates transcription (Wu et al., 1996). Furthermore, THP-1 cells differentiated to macrophages by PMA can be polarized to M1-like or M2-like macrophages according to different protocols (Martinez et al., 2006). M1 macrophages are pro-inflammatory, can be stimulated by Th1 cytokines and secrete inflammatory cytokines such as Tumor Necrosis Factor-alpha (TNF- α) and IL-1 β , whereas M2 macrophages are anti-inflammatory, can be stimulated by Th2 cytokines and secrete anti-inflammatory cytokines such as IL-10 (Muraille et al., 2014).

In addition to the classical protocol which consists on a differentiation of THP-1 cells with PMA for 3 days only (PMA3d), we have modified the protocol according to another report (Daigneault et al., 2010) in order to further enhance differentiation by removing PMA after 3 days and resting the cells in fresh media for additional 5 days (PMA5dr). Furthermore, we established a protocol for the differentiation by VD3 and a protocol for the polarization of M1-like and M2-like macrophages. Cells differentiated by PMA for 3 days were then polarized with a combination of LPS/IFN- γ for M1-like and with a combination of IL-4/IL-13 for M2-like macrophages for 8 h. In our methods, we employed a PMA concentration of 100 ng/ml because it is the standard

concentration used in the literature, however more recent reports have suggested to use a lower concentration (Park et al., 2007) that could be used in similar future experiments. In addition, we polarized the cells with a treatment of 8 h only rather than a longer time in order to find the changes in AS that occur during an immediate immune response. In future, experiments in which cells are polarized for longer time could be performed in order to find late changes in AS.

As our analysis of cell surface antigen expression shows (Figure 22), VD3 treatment results in an upregulation of the macrophage markers CD14 and CD11b. CD14 plays a crucial role in immune recognition being the co-receptor of LPS and the receptor for other bacterial cell wall components, heat shock proteins and viruses (Ziegler-Heitbrock and Ulevitch, 1993). In addition, CD14 plays a role in the phagocytosis of apoptotic cells (Antal-Szalmás et al., 1997). CD11b is an adhesion molecule known to be involved in leukocyte migration (Larsson and Springer, 1990). When monocytes become activated, they undergo phenotypical changes with upregulation of adhesion molecules, which allow them to adhere to endothelial cell, pass through the endothelial layer and finally reach the site of inflammation (Cifarelli et al., 2007). PMA treatment results instead in an upregulation of CD36 and CD54 in addition to the two previous markers. We see upregulation of CD36, which is a very important macrophage receptor for LDL (Krieger and Hertz, 1994). The fact that CD54, an adhesion molecule which plays an important role in the formation of an immunological synapse (Duston, 2014), was detected only after PMA treatment and not after VD3 is consistent with the live cell images, because the cells after PMA treatment became in close contact to each other while with VD3 cells stayed in suspension. In addition, CD54 plays an important role in the interaction between M1 macrophages and MSCs (Espagnol et al., 2017). Enhancement of differentiation by the PMA5dr protocol was confirmed by the upregulation of CD86 in that particular condition only. CD86 is a very important co-stimulatory molecule expressed on APCs and necessary for T-cell activation (Subauste et al., 1998). Regarding the polarization of PMA derived macrophages to M1-like and M2-like macrophages, we could not find any specific FACS marker by flow cytometry (Figure 22), and thus we assessed instead the efficiency of the polarization protocols by changes in gene expression of various cytokines and chemokines. We indeed found specific clusters of cytokines and chemokines which change expression upon polarization (Figure 31), which will be discussed later on. This observation confirms the claim that assessment of the polarization protocols could

be done by alternative methods other than flow cytometry, such as ELISA (Enzyme-Linked ImmunoSorbent assay).

Our live cell images clearly show the acquisition of the characteristic macrophage phenotype with an enlarged shape and granular cytoplasm upon differentiation by PMA, while no change in morphology was observed for the differentiation by VD3 (Figure 26). From the live cell images, we observed that cells become rounder when they are polarized to M1-like, while they become more elongated when polarized to M2-like. In addition, we observed upon differentiation, an increase in autofluorescence and phagocytosis both measured by flow cytometry (data not shown), which confirms what is reported (Njoroge et al., 2001).

After establishing differentiation protocols with the cell line model, we moved to perform similar experiments with human primary cells which are more difficult to isolate and culture. Three peripheral blood monocyte subsets have been described, according to their immunophenotypes: classical, non-classical and intermediate (Wong et al., 2012). Each subset seems to have different roles and tissue distribution during inflammation or steady-state conditions, however there are contrasting results in the literature. For instance, it was once recognized that intermediate monocytes are the pro-inflammatory population (Ziegler-Heitbrock, 2007) but a recent report has claimed that the non-classical monocytes and not the intermediate ones should be instead considered pro-inflammatory, and in addition that the classical monocytes are phagocytic with no inflammatory role (Mukherjee et al., 2015). We isolated human monocytes from PBMCs by negative selection from 4 donors (Figure 23), and we obtained fractions with high purity as shown by CD14 and CD16 expression (Figure 24). We used a Pan Monocyte Isolation kit which allowed us to isolate all the three populations of monocytes. We decided to differentiate all the three types of monocytes pooled together to macrophages rather than looking at a single particular population. Indeed, future experiments could be done in which the three populations of monocytes are separated by using cell type specific isolation kits and subsequently each population would be differentiated individually. Our live cell images for the differentiation of human monocytes clearly show the acquisition of a macrophage phenotype (Figure 27). GM-CSF treated monocytes acquire a typical oval shape while M-CSF treated cells an elongated spindle-shape as reported (Rey-Giraud et al., 2012). We observed that after polarization, M1 macrophages switch from an M1 oval shape to spindle like shape while M2 polarized

macrophages switch from a M2 spindle-like shape to a rounder one (Figure 27). By flow cytometry (Figure 25), we observed a high upregulation of the macrophage marker CD206, which is negative in monocytes while it becomes positive upon differentiation. CD206 is a mannose receptor on surface of tissue macrophages and other cells and which plays a very important role in immune homeostasis and phagocytosis of pathogens and/or dying cells (Azad et al., 2014). It is considered to be an M2 marker, however we found it to be higher in GM-CSF derived macrophages rather than in M-CSF. To support our result, a study by Samaniego et al., has shown a similar observation (Samaniego et al., 2014). We found that M1 macrophages are CD80^{high} and CD163^{low} while vice versa the M2s are CD80^{low} and CD163^{high} confirming what is reported (Samaniego et al., 2014). CD163 is a scavenger receptor expressed on M2 macrophages which is involved in the clearance of free Hb from the blood circulation thus protecting tissues from oxidative damage mediated by Hb (Graversen et al., 2002). CD80 is instead an M1 marker and a co-stimulatory molecule necessary for antigen presentation and T cell activation (Balbo et al., 2001). In addition, M1 macrophages seem to express CD14 and CD16 at higher level than the various M2 macrophages while CD71, CD86, HLA-DR are higher in the M1s compared to M2s. HLA-DR is a MHC Class II surface receptor responsible for antigen presentation and in eliciting or suppressing T_h cell responses (Pinet et al., 1995). CD71 is instead a transferrin receptor which mediates the uptake of transferrin-iron complexes and it has been used as erythroid marker (Marsee et al., 2010). We see it upregulated in M1 macrophages but it has been identified on alveolar macrophages too (Taylor et al., 2000). Regarding the polarizations, we found that LPS/IFN- γ activation results in a further increase in CD86 and HLA-DR. LPS can enhance immune responses to antigens by increasing expression of antigen presenting molecules such as HLA-DR and CD86 (Heinzelmann et al., 1997). We did not detect major obvious changes in the polarization to M2a, M2b and M2c. We assessed instead the efficiency of the polarization protocols for the primary cells once again by analyzing changes in gene expression of various cytokines and chemokines (Figure 31), similarly to what we did for the cell line model.

We found that the fractalkine chemokine receptor CX3CR1 is downregulated in both cell line and primary cells upon differentiation confirming what is reported (Panek et al., 2015). CX3CR1 binds to its ligand CX3CL1 and the interaction between the two might control monocyte survival and migration (Landsman et al., 2009). CXCR3 and CCL25 are other two molecules

that in our protocols are downregulated in both cell line and primary cells. CXCR3 is a chemokine receptor which has been hypothesized to play a role in directing macrophage activities (Oghumu et al., 2014), while CCL25 is a cytokine which stimulates macrophage differentiation in rheumatoid arthritis (Schmutz et al., 2010). Toll-like receptor 4 (TLR4) is a receptor present on the cell surface which binds LPS from Gram-negative bacteria (Lu et al., 2008). Upon binding, two different intracellular cascades in macrophages are activated: the MyD88-dependent and MyD88-independent pathway. Activation of the cascades leads to the production of specific cytokines and chemokines. Indeed, in the LPS treatment we see upregulation of CCL2, CCL7, CXCL2, CCL8, and CXCL9, which are MyD88-dependent and of CCL5, CXCL10, CXCL11, and CXCL13 which instead are MyD88-independent confirming what is reported (Chow et al., 1999). However, we found some discrepancies regarding the expression of the aforementioned molecules between cell line and primary cells. Exceptions are CXCL9, 10 and 11 which are downregulated in primary cells while upregulated in THP-1. These chemokines are known to be induced by IFN- γ ; nevertheless we found a study in the literature where GM-CSF derived macrophages are shown to downregulate their expression (Dabritz et al., 2015). We do not know why we observed this discrepancy but we still kept those results because other molecules are perfectly matching what is reported. CXCL16 is a pro-inflammatory chemokine up-regulated in both the cell line model and primary cells induced by LPS (Lehrke et al., 2007). In contrast to the literature, we see upregulation by LPS of CCL17, CCL1, CCL22 and CCL24 which instead are reported to be M2b cytokines (Arango Duque and Descoteaux, 2014). We found that the cytokine CCL26 is specific of IL-4 activation since it is found upregulated both in the cell line and primary cells confirming what is reported (Bao et al., 2012). CCL7 and CCL18 were found to be upregulated by LPS in THP-1 cells while in the primary cells were upregulated by IL-4. We came to the conclusion that our results in part recapitulate and in part are different to what is reported.

4.2 Analysis of the RNA-seq data of the differentiation of monocytes

Together with my co-supervisor Asst. Prof Zheng Jie from the School of Computer Engineering (SCE), as a team we implemented a pipeline for the analysis of DEGs and AS from the RNA-seq data. We have performed an analysis of the expression level of about 8000-10000 human genes and compared the gene expression across different conditions. We have then obtained a list of

DEGs which are up-regulated or down-regulated or whose expression has remained unchanged after differentiation/polarization in the cell line model and in primary cells (Figure 29). We identified hundreds of DEGs which are in common between the cell line model and the primary cells (Figure 30). PCA analysis has captured variabilities between the various conditions based on their gene expression (Figure 32). By a more in depth PCA, we merged the cell line and primary cells into a single 3D plot (data not shown). We found that there is partial overlap of the two thus highlighting the fact that THP-1 cells are indeed a model good enough to study differentiation of human monocytes. In general, these PCA results are in agreement to what we have found previously by flow cytometry, live cell images and cytokine/chemokine profiles. PCA has been used by other studies in the literature to separate the various monocyte and macrophage populations. For instance, a study by Zhang et al., has used PCA to separate M1-like and M2-like macrophages obtained from THP-1 monocytes (Zhang et al., 2017). Similarly, a study by Li et al., has used PCA to separate M1 and M2 macrophages polarized from mouse bone marrow-derived macrophages (BMDM) (Li et al., 2015). In general all those studies have preferred to employ a PCA 3D model over a 2D model since a better separation of the cell populations was achieved with the former. We took the output of Cuffdiff, which give us the information about the fold change of the DEGs and we performed GO analysis by sorting the genes in the Reactome database accordingly to their cellular functions. A heat map for the various categories was generated by using the FPKM values from Cufflinks (Figure 33). We clearly found that for the cell line model, treatment with PMA and to a certain extent with VD3, results in a downregulation of genes involved in cell cycle, DNA repair, DNA replication, and chromatin organization confirming the fact that PMA treated cells stop to proliferate during the differentiation process (Richter et al., 2016). Genes involved in cell-cell communication, ECM organization, vesicle mediated transport, immune system and signal transduction, which are all functions associated with macrophages, were upregulated by PMA but not by VD3. We found for instance upregulation of several metalloproteinases (MMPs) which confirms that macrophages are involved in ECM degradation (Valentin et al., 2009). Regarding the primary cells, we found upregulation of genes involved in DNA repair and DNA replication for the M1 and M1a which are different to what was found in the PMA treatment of THP-1 cells. For the rest of the GO terms for the differentiation to M1 and M2 and activation to M1a, we see some genes that are upregulated while others are downregulated. Regarding the M2 polarization of

primary cells, we see that different clusters of genes involved in cell-cell communication, ECM organization, immune system and vesicle mediated transport are upregulated and are specific for each stimulation, probably confirming that indeed all the M2s have different roles in the immune system as reported (Martinez et al., 2008). For instance, M2a macrophages are involved in allergy, killing of parasites and regulation of Type II inflammation. M2b are instead involved in immuno-regulation and Th2 activation while M2c in tissue remodeling and matrix deposition.

We performed analysis of AS by three tools, namely MISO, MATS and SpliceTrap, and we identified thousands of AS events that changed upon differentiation/polarization in the cell line model (Figure 35) and in the primary cells (Figure 34). We observed that in our hands there is a poor overlap between the three tools due probably to the fact that they perform an AS analysis in different ways and they are built with different programming languages. Based on our communication with other experts in this field, we confirmed that the aforementioned discrepancies between the AS tools for RNA-seq data analysis observed by us, were also encountered by other researchers. We found hundreds of AS events which are in common between the cell line model and the primary cells (Figure 36) which are going towards the same direction thus highlighting the fact that those events are truly occurring during the transition from monocytes to macrophages. We validated selected AS events by radioactive PCR and we found that the validation confirms the AS changes observed from the RNA-seq, we thus concluded that our RNA-seq pipeline was reliable. Furthermore, besides being robust, our methods are also reproducible as it was found by the Pearson's correlation between replicates (data not shown). In addition, the results of the validation has shown that for certain genes the intensities of the bands are different across the various conditions, thus suggesting that there are AS changes which are accompanied by changes in gene expression. In future, such events can be further investigated.

We observed that the genes under transcriptional control are different from those under the control of AS, similarly to what is reported in the literature (Barbosa-Morais et al., 2012; Merkin et al., 2012). For instance, by comparing the GO heat map (Figure 33) we can clearly see that genes upregulated or downregulated have different functions from the genes in which AS events were detected in common between the cell line and the primary cells (Figure 38). This finding indicates that certain pathways are regulated by overall expression via transcription or decay, and others via AS for more subtle changes. Genes that use one mechanism tend not to use others. Downstream analysis of the RNA-seq data have found that various splicing factors change

expression at the mRNA level between the undifferentiated and differentiated cells and that those changes are conserved between the cell line model and the primary cells. We then aimed at identifying which one of those factors is responsible for the changes in AS observed. By knockdown experiments it was later found that MBNL1 is a splicing factor which is truly responsible for the regulation of a fraction of those AS events and for the proper cellular differentiation.

4.3 PRPF40B CRISPR KO

Regarding the second aim of this Ph.D., focused on the biological function of PRPF40B, I have established an efficient PRPF40B gene KO protocol by CRISPR. We decided to perform a KO of the gene by CRISPR rather than using CRISPR to Knock-In its MDS mutations because we are more interested in studying its basic biological function (Figure 77).

Figure 76: CRISPR KO and Knock-In strategies.

Schematics of CRIPR KO on the right and Knock-In on the left. In the CRIPSR KO no functional protein is translated while in the Knock-In a mutated one is generated. Black star indicates mutation. PRPF40B exons are in red while gray lines in between them represent introns.

Another reason is that based also on other experiments carried out in our lab, the Knock-Ins are less efficient and frequently carry secondary mutations.

After several trials in which I used sgRNAs targeting PRPF40B exons, I decided to use instead sgRNAs targeting introns because it was found to be easier to obtain an out-of-frame transcript with the latter strategy. In fact, based on my experience, when exons were targeted, even if the CRISPR sgRNAs were designed in a way to generate an out-of-frame transcript, clones with an in-frame transcript were preferentially selected after nucleofection. This was possible because Cas9 most of the times cuts 3 bp upstream of the PAM, however in certain cases the cut can occur 4 bp upstream. I have selected sgRNAs targeting exons which were supposed to cut 3 bp upstream and thus to generate an out-of-frame transcript, however after single cell clone selection, I found clones in which either one or both sgRNAs were cutting instead 4 bp upstream thus leading to a transcript which was still in-frame. In addition, it is worth noting that we opted for a Cas9 introducing a double cut rather than a single cut nickase system because based on some of my trials with the latter, I could not obtain any single cell clone showing any deletions or insertions (Indels). However, I could not exclude this was due to the detection method employed at that time because during those trials I was not using the Surveyor assay, which is the standard assay to detect indels (Qiu et al., 2004), but I was simply sending PCR bands for Sanger DNA sequencing. Based on the literature, by using a nickase system it is possible to reduce off-target effects of the Cas9 (Zhang et al., 2015), however we continued with the experiments using the double cut strategy because in our hands it was a more efficient method than the nickase to obtain KO clones and we made sure of using a not so high amount of Cas9 plasmid for nucleofection. In addition, by performing later on rescue experiments we made sure that results seen were truly attributable to PRPF40B and not due to artifacts generated from the CRISPR protocol.

A strategy to remove a genomic portion of PRPF40B from intron 4 to intron 8, thus including full exon 5 and 8, was designed as described in Figure 16. I adopted a modified CRISPR KO workflow based on a protocol in the literature (Figure 17) (Ran et al., 2013). It consists at first in the nucleofection of two sgRNAs to introduce two DSBs in two different genomic locations which are then preferentially repaired by the nonhomologous end-joining (NHEJ) repair system thus leading to a deletion junction (Figure 40) (Mali et al., 2013). After nucleofection, a double FACS sorting strategy rather than a single one was preferentially adopted in order to enrich for

cells highly expressing GFP and also the Cas9 protein which is cloned in the same plasmid. One of the reasons is that by isolating cells with a higher Cas9 expression it is possible to increase the DNA cutting efficiency. By sorting cells into fractions with low, medium, and high GFP expression levels, Sigma-Aldrich and others (Duda et al., 2014), have shown that the high GFP population is enriched in genome edits compared to the other two fractions. I managed to obtain 3 clones with a full PRPF40B KO (Figure 39, 40 and 41). In addition, we also concluded that since PRPF40B cells are viable, the gene is not essential for splicing. I did not observe any major change in proliferation in the KO cells.

4.4 Rescue experiments in PRPF40B KO cells

Having obtained a full KO, we proceeded to generate an overexpression vector to reintroduce WT PRPF40B expression in the KO cells. Because we aimed to study the impact of MDS mutations on the gene, we generated vectors also for two selected MDS mutations (Yoshida et al., 2011). We selected the P383L mutation because it is the only somatic mutation between those found, and the P540S mutation because it was found in two MDS patients rather than in one (Figure 13). We performed rescue experiments with the WT gene only or alternatively by overexpressing the two mutations as homozygous-like or mixed in a 1:1 ratio with the WT vector as heterozygous-like. We decided to overexpress the mutations as heterozygous-like in order to closer recapitulate the MDS phenotype because in MDS only one allele is mutated. In addition, we decided to overexpress them as homozygous-like because we were hypothesizing that we could obtain a clearer phenotype by doing so rather than as heterozygous-like (Figure 78). Overexpression efficiencies were assessed by measuring PRPF40B mRNA transcript level and protein level by qRT-PCR and WB respectively (Figure 43).

Figure 77: Overexpression experiments.

Image representing how the rescue experiments were performed. Circle represent plasmid, black star indicates mutation. Double circles indicate the rescue experiments where mutation was mixed with WT at 1:1 ratio.

4.5 RNA-seq data analysis of the PRPF40B KO and its MDS mutations

I applied the same bioinformatics pipeline used for the first aim with the exception that I only used MATS for the analysis of AS. I have performed pathway and GO analysis of the DEGs with the online tools ConsensusPathDB/Enrichr and I came with the conclusion that PRPF40B is regulating the mRNA expression of hundreds of genes (Figure 46). However, I have found that most of the genes which change expression upon KO were rescued by the two MDS mutations as well as by overexpressing the WT gene similarly (Figure 48).

4.6 Role of PRPF40B in regulating hypoxia and cellular differentiation

By performing a pathway and GO analysis of the DEGs regulated by PRPF40B with ConsensusPathDB/Enrichr, I have found that the majority of the DEGs are involved in steroid biogenesis and in the HIF1A signaling pathway (Figure 47). Importantly, I have detected major changes in the regulation of hypoxia as it is shown by a heat map of genes known to be involved in hypoxia (Figure 52). Indeed, I have found that the important transcription factors c-Jun (JUN), c-Fos (FOS) and Early growth response protein 1 (EGR1) are upregulated upon KO. c-Jun and c-

Fos are in fact known to be upregulated in ischemic hypoxia (Gilby et al., 1997). EGR1 is instead another transcription factor known to increase in hypoxia and to be upstream of HIF1A (Sperandio et al., 2009). EGLN1 is a gene which encodes for the enzyme prolyl hydroxylase domain-containing protein 2 (PHD2) which targets HIF1A for degradation during normoxia. It has been reported that its mRNA is upregulated during hypoxia probably due to a feedback mechanism to stop hypoxic responses in reoxygenated cells (D'Angelo et al., 2003), and our results support that claim. We observed also upregulation of its homologue EGLN3 probably having a similar function. Several reports have shown that the micro-RNA miR-210 is upregulated in hypoxia thus strongly linking it with the hypoxia pathway (Camps et al., 2008; Giannakakis et al., 2007; Kulshreshtha et al., 2007). I have found upregulation in the KO cells of the precursor of the aforementioned micro-RNA which is transcribed from the MIR210 gene. The BNIP3/BNIP3L axis is crucial for hypoxia-induced autophagy (Bellot et al., 2009), and we have found those two genes upregulated. CA9 is a transmembrane enzyme which converts carbon dioxide to bicarbonate and H^+ thus increasing the acidification of the extracellular tumor environment during hypoxia. It is known to be upregulated in tumor cells and it has been widely used as a marker of hypoxia (Swietach et al., 2007) and we saw upregulation of its mRNA. We found upregulation of angiopoietin-like protein 4 (ANGPTL4) which is an important protein involved in angiogenesis and tumorigenesis (Le Jan et al., 2003) and known to be induced by hypoxia (Murata et al., 2009). Together with the FYP student Ms. Resilind Chew we validated some of the DEGs involved in hypoxia by qRT-PCR and we found that the validation largely confirms the changes observed from the RNA-seq (Figure 52). We found that TGF- β 3 validation in WT, KO and rescued cells was an exception because RNA-seq data can be noisy. We can conclude that the PRPF40B KO induces a clear hypoxia gene signature.

By a more in depth DEG analysis, I have found that a subset of genes under regulation of KLF1 change expression upon KO (Figure 51). KLF1 is a transcription factor known to regulate erythroid differentiation (Pilon et al., 2008), thus we can infer that PRPF40B regulates the expression of genes which are involved in erythroid differentiation. We could then hypothesize that changes in the gene expression level of PRPF40B would have an impact on the expression of the regulated erythroid genes thus ultimately affecting the proper differentiation to erythrocytes. In order to further characterize the KO cells, I have assessed the apoptosis, DNA

damage and proliferation of the KO cells compared to WT cells by flow cytometry. Preliminary results have not identified any major changes in these processes (Figure 73).

4.7 PRPF40A and PRPF40B roles in hypoxia and connection with AML drug therapy

Because we hypothesized that PRPF40B could act as tumor suppressor, I checked the mRNA expression levels of the gene in various human cancers according to TCGA database, and I have found a striking downregulation of the gene in AML patients compared to other cancers (Figure 52). Other databases have shown a similar pattern of expression (Figure 53). In addition, it was found from the literature that an upregulation of the mRNA expression of the PRPF40B homologue PRPF40A in lung cancer is positively correlated with hypoxia (Oleksiewicz et al., 2017). Based on the aforementioned study and on our results, we can thus deduce that both an increase in the level of the mRNA expression of PRPF40A and a PRPF40B KO could promote the induction of a hypoxia gene signature. I checked the mRNA expression level of PRPF40A across various cancers in the same database and intriguingly I have found that its expression is higher in AML than in any other cancer (Figure 53). Indeed, I have found an inverse correlation between the two genes in AML patients at the mRNA level (Figure 55). I could not find any overlapping hypoxia markers between our data set and the study by Oleksiewicz et al. because their dataset is based only on qRT-PCR and not on RNA-seq. I have found by RNA-seq data analysis that in our K562 WT cells the expression of PRPF40A is higher than the expression of PRPF40B. We could thus hypothesize that PRPF40A and PRPF40B might have opposing roles in the regulation of hypoxia in AML or in other cancers. There are several reasons why these two genes could affect hypoxia at the level of gene expression. For instance, both could regulate the splicing of a transcription factor which controls hypoxia; however I could not identify any AS changes in such factor in our RNA-seq data. Another hypothesis could be that the splicing factors are introducing epigenetics modifications which ultimately lead to changes in gene expression (Zhou et al., 2011). We obtained the RNA-seq results from K562 which is a CML cell line, thus future investigations on the role of PRPF40A and PRPF40B in AS and hypoxia should be performed in AML cell lines. I have measured expression of PRPF40A in the KO cells by WB and I have found that there was no clear difference in the PRPF40A protein level in the KO clones compared to WT cells, thus PRPF40A cannot compensate for the loss of PRPF40B (Figure 57). Lastly, because I have found that PRPF40B regulates hypoxia, I have

checked the correlation at the mRNA level in AML patients between PRPF40B and the major regulator of hypoxia HIF1A. Indeed I have found that there is an inverse correlation between the two genes (Figure 56). It does seem that when PRPF40B is lowly expressed, HIF1A instead is highly expressed which might lead to induction of hypoxia. We hypothesized that these findings if confirmed with experiments in primary cells or in mice could have a significant clinical impact, considering that a high hypoxia contributes to chemoresistance in AML blasts (Deynoux et al., 2016). We could envision as therapeutic strategy to increase expression of PRPF40B while at the same time reducing the expression of PRPF40A, by combining it with current drug treatments for AML patients which would possibly reduce hypoxia and chemoresistance (Figure 79).

Figure 78: Roles of PRPF40A and PRPF40B in AML.

Diagram showing the effects on hypoxia when PRPF40A and PRPF40B have different level of expression and relative possible consequences on cancer drug treatment.

Additional studies, such as knocking out PRPF40A in K562 WT cells as a single KO or in PRPF40B-KO cells as double KO, followed by RNA-seq and other functional studies in mouse models could be performed in future, in order to study in depth the relationship between the two genes in regulating hypoxia and AS.

4.8 Regulation of AS by PRPF40B

I have detected by MATS thousands of AS events in the various conditions (Figure 58). Since we are more interested in the PRPF40B KO cells rather than the two MDS mutations considering that no major changes at the level of gene expression were detected upon overexpression of the

mutations, I focused on comparing the AS events between K562 WT, *PRPF40B* KO and rescued cells with the WT gene. I have found thousands of AS event that change upon KO and hundreds that are rescued by reintroducing expression of WT PRPF40B (Figure 58). We then mainly focused on the cassette exon events and we identified hundreds of cassette exon events that are rescued by PRPF40B (Figure 59). Considering that the majority of those events show more exon inclusion upon KO, it was concluded that PRPF40B can act mainly as a repressor of AS. Together with FYP student Resilind Chew, we validated by RT-PCR some of those PRPF40B regulated AS events, and we found that in general the validation confirms the changes in AS observed from the RNA-seq (Figures 62 and 63). Our results provide a clear proof that PRPF40B is indeed involved in regulating AS and those changes are truly attributable to PRPF40B, because the protein expression of its homologue PRPF40A does not clearly change in the KO.

Intriguingly, from the validation it was also observed that if the MDS mutations were overexpressed as heterozygous-like, some of the AS were rescued less efficiently than when the mutations were overexpressed as homozygous-like (Figure 62). Validation confirms what was found by a more in depth analysis of the 793 cassette exon events, in which it was compared which portion of those genes is rescued also by the two mutations either as homozygous-like or heterozygous-like and which not (Figure 61). It is worth noting that the aforementioned observation applies in part also to DEGs. For instance, in the case of TGF- β 3, we see a high upregulation when the mutations were overexpressed as heterozygous-like but much less when homozygous-like. These results could in part explain why in MDS only one allele is mutated and not two. There could be the possibility that the mutated proteins are forming some toxic dimers together with the WT proteins, which could be trapped in the cytosol and not being translocated into the nucleus to perform their functions. However we cannot exclude that this phenomenon could be due to slight differences in transfection efficiency. Thus, probably this observation should be confirmed by further experiments in which the mutations are introduced into the endogenous gene by a CRISPR Knock-In, either mutating one allele or both and then comparing the two. It is worth noting that there is a similar example of this concept in the literature, where in Alzheimer's disease, the A2V mutation in the amyloid precursor protein (APP) protects from the disease when it is in its heterozygous form while cause an early disease dementia in its homozygous form (Nguyen et al., 2016). Thus, there could be some opposing difference in the

phenotype if the MDS mutations are heterozygous or homozygous. Structural and biochemical studies could be performed in future investigations.

A PPI network of the 477 unique genes in which the 793 AS events were detected was generated by STRING PPI database and clusters of genes with similar function were identified. It was found that the majority of those genes are involved in the Akt1/MAPK signaling cascade (Figure 64). Reports in the literature have shown that there is an association between Akt1 activation and hypoxia which are influencing each other (Alvarez et al., 2001; Dai et al., 2008). We could hypothesize that changes in AS in genes involved in the Akt1/MAPK cascade upon PRPF40B KO would lead to activation of hypoxia, but this notion has to be tested by future experiments.

Gene enrichment analysis by Enrichr has confirmed that the top three pathways in the set of DEGs regulated by PRPF40B are steroid biogenesis, glycolysis and HIF1 signaling. Importantly, it has shown that the AS events are in genes which have binding sites for the subunit of Pol II according to a Chip-seq data publicly available in the same cell line used in this study, thus potentially sustaining our hypothesis that the regulation of the AS by PRPF40B occurs co-transcriptionally. In addition, the analysis has shown that many genes having AS are involved in the MAPK signaling confirming what was found by the STRING database. We observed that the functions of the DEGs are completely different from the functions of the alternatively spliced genes.

I have checked various features of those 793 cassette exons. For instance, I have found that the most occurring motif in those exons is an A rich motif (Figure 68). A family of proteins which is found to bind those motifs are the poly(A)-binding protein (PABPs), which are RBPs known to recognize stretches of A in the poly(A) tail of the mRNAs (Gray et al., 2015). We could hypothesize that PRPF40B might recognize those motifs indirectly via one or multiple PABPs, with the latter facilitating PRPF40B interaction with mRNAs, considering that PRPF40B has no putative RRM. By mass spec, I have detected in the co-IP pull down sample with anti-PRPF40B antibodies a few PABPs, which if confirmed could provide evidence for this hypothesis. A similar example can be found in the study of Ran-binding protein 9 (RANBP9) which regulates AS by forming a complex in the nucleus with specific poly(A) binding proteins (PABPs) (Bao et al., 2014).

In addition, I have found that the PRPF40B regulated cassette exon events have a weaker 5'ss and 3'ss compared to constitutive background exons and AS events which are not regulated by it (Figure 65 top). This supports our conclusion that those events are truly alternatively spliced events because it is known that alternatively spliced events in general have weak 5'ss and 3'ss. Furthermore, I have observed that the p-value for the 5'ss comparison is lower than the p-value of the 3'ss, which could in part reflect that PRPF40B is involved more in the recognition of the 5'ss rather than the 3'ss (Abovich et al., 1997). I could not find any major changes in the lengths of those cassette exon events and of their flanking introns (Figure 65 bottom). I have shown that there is no correlation between PSI changes and 5'ss strengths (Figure 66), which is not consistent with the hypothesis proposed in a recent report claiming that exons with weaker 5'ss and 3'ss are more sensitive to changes in PRPF40B expression (Becerra et al., 2016). For instance, if the hypothesis is true, we would expect that exons with a very weak 5'ss tend to have a large change in PSI upon KO, but I could not find such inverse correlation. The aforementioned claim is based on the study of sequences from a single AS event, while our results are more comprehensive because they take into consideration hundreds of AS events.

4.9 Role of PRPF40B in co-transcriptional splicing

Minigenes have been commonly used in the literature to identify *cis*-elements and *trans*-acting factors that regulate AS (Cooper, 2005). We designed various minigenes in order to further validate the AS changes found by the RNA-seq (Figure 9). Those minigenes would have been employed to study how mechanistically is PRPF40B regulating AS, for instance by introducing deletion/mutations in *cis*-elements of the minigenes and then observing if there are any alterations in the AS. However, in our experiments with the minigenes we could not recapitulate the changes in AS previously observed (Figures 69 and 70), thus we could not utilize these vectors for follow-up studies. We hypothesized that the reason why we could not recapitulate the AS changes was mainly due to the fact that the regulation of the AS events under PRPF40B control occurs in a co-transcriptional way and it is thus influenced by speed of the Pol II in its chromatin context. In order to prove that PRPF40B is involved in co-transcriptional splicing as it has been hypothesized (Becerra et al., 2016), I performed a Pol II inhibition drug assay and I have shown that indeed there is a connection between the splicing factor and the speed of transcription (Figure 71). A model in which the regulation of AS by PRPF40B is connected to

the speed of Pol II could be then proposed. If PRPF40B is present, Pol II would proceed at a faster speed thus resulting in hnRNPDL exon 8 skipping while if PRPF40B is absent, Pol II would proceed at slower pace thus giving the time for exon 8 to be included (Figure 80). In addition, DRB also further increased exon 8 inclusion in KO cells, indicating that other factors also contribute to this function, such as possibly PRPF40A.

Figure 79: Model of the role of PRPF40B in co-transcriptional splicing.

Diagram showing possible connection between PRPF40B and transcription based on results of Pol II inhibition assay. On the left, PRPF40B maybe bound to other splicing factors would increase the speed of Pol II thus resulting in exon skipping. On the right, exon inclusion is favored by slow Pol II and absence of PRPF40B. Transcription factors might also be involved in facilitating interaction of the splicing factors to the CTD of Pol II.

Indeed, there are exceptions to the above model. As it has been reported in the literature, in certain cases a fast Pol II would result in exon inclusion rather than skipping and vice versa (Dujardin et al., 2014; Tang et al., 2016). In fact, as it has been stated previously when PRPF40B was KO, we observed more exon skipping rather than inclusion in the minor percentage of the cases in our RNA-seq. I have observed by IGV visualization, a higher number of RNA-seq reads in introns preceding the alternatively spliced middle exons of the AS events regulated by PRPF40B in the KO samples(data not shown), which could suggest a stalling of the Pol II when

PRPF40B is knocked out. I am currently testing the changes in AS for other validated PRPF40B AS events upon drug treatment in order to further test the above concept.

In vivo and *in vitro* splicing and transcription assays could be performed in future in order to further elucidate the mechanisms by which splicing and transcription are coordinated. For instance, real-time live-cell imaging of co-transcriptional splicing is now more feasible (Schmidt et al., 2011). Technical advances in *in vitro* transcription-splicing coupled systems are making these assays a promising tool to study co-transcriptional splicing (Yu et al., 2010). Chromatin Immunoprecipitation (ChIP) is another technique which could be used but it is still an indirect measure of co-transcriptional splicing (Nielsen, 2005).

5 CONCLUSION

Regarding the first aim of the Ph.D., we successfully obtained various types of macrophages derived from differentiating and polarizing monocytes, both in a cell line model and in primary human cells. We did extensive assessment of the differentiation/polarization protocols as our results for the flow cytometry, live cell imaging, PCA and DEG have shown. We established a reliable bioinformatics pipeline and we performed a transcriptomic analysis by RNA-seq of the monocytes and macrophages. We identified thousands of DEGs and AS events that changed upon differentiation and polarization and we found hundreds of events in common between the cell line model and the primary cells. We observed that there is a poor overlap between the tools used to detect AS and that the genes under transcriptional control are different from those under AS. Our study of the transcriptome of the monocytes and macrophages can thus provide the basis for future detailed characterization of these important myeloid cells in connection with any alteration seen in immunodeficiencies, autoimmunity, atherosclerosis and cancer.

Regarding the second aim of this Ph.D., I successfully generated a PRPF40B KO by CRISPR in a CML cell line. I performed a transcriptomic analysis by RNA-seq of the WT cells, the KO clones and the samples in which PRPF40B was overexpressed as WT or with two MDS mutations. I identified hundreds of DEGs and AS events that are regulated by PRPF40B. By DEG analysis I have found that PRPF40B could act as tumor suppressor by inhibiting hypoxia while at the same time acting as a splicing factor by mainly repressing AS. I discovered a possible link between PRPF40B and its homologue PRPF40A in regulating hypoxia in AML which could lead to a potential therapeutic strategy. I could not identify any major changes in DEGs and AS events when two MDS mutations were overexpressed which lead us to conclude that they could be passenger mutations. Lastly, I proved that PRPF40B is involved in co-transcriptional splicing. In conclusion, I have generated data on PRPF40B's biology which can be used in future by the lab to investigate the overlapping roles with its homologue in regulating AS and hypoxia in AML.

REFERENCES

- Abdel-Wahab O., Levine R. (2011) The Spliceosome as an Indicted Conspirator in Myeloid Malignancies. *Cancer Cell* **4**:420–422.
- Abovich N., Rosbash M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**:403–412.
- Adams J.M., Difazio L.T., Rolandelli R.H., Luján J.J., Haskó G., Csóka B., Selmeczy Z., Németh Z.H. (2009) HIF-1: a key mediator in hypoxia. *Acta Physiol. Hung.* **96**:19–28.
- Aguilera A., García-Muse T. (2012) R Loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**:115–124.
- Akashi K., Traver D., Miyamoto T., Weissman I.L. (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**:193–197.
- Alasoo K., Martinez F.O., Hale C., Gordon S., Powrie F., Dougan G. (2015) Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription. *Sci. Rep.* **5**:12524.10.1038/srep12524.
- Alt F.W., Bothwell A.L.M., Knapp M., Siden E., Mather E., Koshland M., Baltimore D. (1980) Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNA's that differ at their 3' ends. *Cell* **20**:293–301.
- Alvarez-Tejado M., Naranjo-Suarez S., Jimenez C., Carrera A.C., Landazuri M.O., Del Peso L. (2001) Hypoxia induces the activation of the phosphatidylinositol 3-kinase/Akt cell survival pathway in PC12 cells: protective role in apoptosis. *J. Biol. Chem.* **25**:22368–22374.
- Andersson R., Enroth S., Rada-Iglesias A., Wadelius C., Komorowski J. (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**:1732–1741.
- Anthony B.A., Link D.C. (2014) Regulation of hematopoietic stem cells by bone marrow stromal cells. *Trends Immunol.* **35**:32–37.

Arber D.A., Orazi A., Hasserjian R., Thiele J., Borowitz M.J., Le Beau M.M. (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**:2391–2405.

Auwerx J. (1991) The human leukemia cell line, THP-1: a multifaceted model for the study of monocyte-macrophage differentiation. *Experientia* **47**:22–31.

Azad A.K., Rajaram M.V., Schlesinger L.S. (2014) Exploitation of the macrophage mannose receptor (CD206) in infectious disease diagnostics and therapeutics. *J. Cytol. Mol. Biol.* **10**:1 pii: 1000003.

Bachegowda L., Gligich O., Mantzaris I., Schinke C., Wyville D., Carrillo T., Braunschweig I., Steidl U., Verma A. (2013) Signal transduction inhibitors in treatment of myelodysplastic syndromes. *J. Hematol. Oncol.* **6**:50.

Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J.Y., Li W.W., Noble W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**:W202–W208.

Bao J., Tang C., Li J., Zhang Y., Bhetwal B.P., Zheng H., Yan W. (2014) RAN-binding protein 9 is involved in alternative splicing and is critical for male germ cell development and male fertility. *PLoS Genet.* **10**, e1004825.

Barash Y., Calarco J.A., Gao W., Pan Q., Wang X., Shai O., Blencowe B.J., Frey B.J. (2010) Deciphering the splicing code. *Nature* **465**:53–59.

Becerra S., Montes M., Hernández-Munain C., Suñé C. (2015) Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing in vivo. *RNA* **21**:438–457.

Baker K.E., Parker R. (2004). Nonsense-mediated mRNA decay: Terminating erroneous gene expression. *Curr. Opin. Cell Biol.* **3**:293–299.

Baklouti F., Huang S.C., Vulliamy T.J., Delaunay J., Benz E.J. (1997) Organization of the human protein 4.1 genomic locus: new insights into the tissue-specific alternative splicing of the pre-mRNA. *Genomics* **39**:289–302.

Balbo P., Silvestri M., Rossi G.A., Crimi E., Burastero S.E. (2001) Differential role of CD80 and CD86 on alveolar macrophages in the presentation of allergen to T lymphocytes in asthma. *Clin. Exp. All.* **31**:625–636.

Batsche E., Ameyar-Zazoua M. (2015) The influence of Argonaute proteins on alternative RNA splicing. *Wiley Interdiscip. Rev. RNA* **6**:141–156.

Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J.Y., Li W.W., Noble WS. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**:W202–W208.

Becerra S., Andrés-León E., Prieto-Sánchez S., Hernández-Munain C., Suñé C. (2016) Prp40 and early events in splice site definition. *Wiley Interdiscip. Rev. RNA* **7**:17–32.

Belfort M. (1989) Bacteriophage introns: Parasites within parasites? *Trends Genet.* **5**:209–216.

Bellot G., Garcia-Medina R., Gounon P. (2009) Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains. *Mol. Cell Biol.* **10**:2570–2581.

Bentley D.L. (2005) Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell. Biol.* **17**:251–256.

Berglund J.A., Chua K., Abovich N., Reed R., Rosbash M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**:781–787.

Berges C., Naujokat C., Tinapp S., Wieczorek H., Höh A., Sadeghi M., Opelz G., Daniel V. (2005) A cell line model for the differentiation of human dendritic cells. *Biochem. Biophys. Res. Commun.* **333**:896–907.

Berget S.M., Moore C., Sharp P.A. (1977) Spliced segments at 5' terminus of adenovirus 2 late messenger-RNA. *Proc. Natl. Acad. Sci. U. S. A.* **8**:3171–3175.

Berget S.M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**:2411–2414.

Bernasconi P. (2008) Molecular pathways in myelodysplastic syndromes and acute myeloid leukemia: relationships and distinctions—a review. *Brit. J. Haemat.* **142**:695–708.

- Black D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**:291–336.
- Bonen L., Vogel J. (2001) The ins and outs of group II introns. *Trends Genet.* **17**:322–331.
- Bradford J.A., Clarke S.T. (2011) Dual-pulse labeling using 5-ethynyl-2'-deoxyuridine (Edu) and 5-bromo-2'-deoxyuridine (BrdU) in flow cytometry. *Curr. Protoc. Cytom.* **55**:7.38.1–7.38.15.
- Brahimi-Horn M.C., Pouyssegur J. (2007) Hypoxia in cancer cell metabolism and pH regulation. *Essays Biochem.* **43**:165–178.
- Brooks A.N., Choi P.S., de Waal L., Sharifnia T., Imielinski M., Saksena G., Pedamallu C.S., Sivachenko A., Rosenberg M., Chmielecki J. (2014) A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS One* **9**: e87361.
- Brown J.M., Giaccia A.J. (1998) The unique physiology of solid tumors: Opportunities (and problems) for cancer therapy. *Cancer Res.* **58**:1408–1416.
- Bunting K., Rao S., Hardy K., Woltring D., Denyer G.S., Wang J., Gerondakis S., Shannon M.F. (2007) Genome-wide analysis of gene expression in T cells to identify targets of the NF-kappa B transcription factor c-Rel. *J. Immunol.* **178**:7097–7109.
- Burge C.B., Tuschl T., Sharp P.A. (1999) The RNA World, 2nd Ed.: The Nature of Modern RNA Suggests a Prebiotic. *RNA World* **37**:525–560.
- Busslinger M., Moschonas N., Flavell R.A. (1981) Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell* **27**:289–298.
- Camps C., Buffa F.M., Colella S., Moore J., Sotiriou C., Sheldon H., Harris A.L., Gleadle J.M., Ragoussis J. (2008) hsa-miR-210 Is Induced by Hypoxia and Is an Independent Prognostic Factor in Breast Cancer. *Clin. Cancer Res.* **14**:1340–1348.
- Carpentier C., Ghanem D., Fernandez-Gomez F.J., Jumeau F., Philippe J.V., Freyermuth F. (2014) Tau exon 2 responsive elements deregulated in myotonic dystrophy type I are proximal to

exon 2 and synergistically regulated by MBNL1 and MBNL2. *Biochim. Biophys. Acta Mol. Basis Dis.* **1842**:654–664.

Cazzola M., Malcovati L. (2005) Myelodysplastic syndromes—coping with ineffective hematopoiesis. *N. Engl. J. Med.* **6**:536-538.

Celotto A.M., Graveley B.R. (2001) Alternative splicing of the *Drosophila Dscam* pre-mRNA is both temporally and spatially regulated. *Genetics* **159**:599–608.

Cerami E., Gao J., Dogrusoz U., Gross B.E., Sumer S.O., Aksoy B.A., Jacobsen A., Byrne C.J., Heuer M.L., Larsson E., Antipin Y., Reva B., Goldberg A.P. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**:401–404.

Crisci A., Raleff F., Bagdiul I., Raabe M., Urlaub H., Rain J.C., Kramer A. (2015) Mammalian splicing factor SF1 interacts with SURP domains of U2 snRNP-associated proteins. *Nucleic Acids Res.* **43**:10456–10473.

Chen M., Manley J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**:741–754.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* **14**:128-142.

Chiche J., Brahimi-Horn M.C., Pouyssegur J. (2010) Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer. *J. Cell Mol. Med.* **14**:771–794.

Chow L.T., Gelinas R.E., Broker T.R., Roberts R.J. (1977) An amazing sequence arrangement at the 5'ends of adenovirus 2 messenger RNA. *Cell* **12**:1–8.

Colla S., Ong D.S.T., Ogoti Y., Marchesini M., Mistry N.A., Clise-Dwyer K., Ang S.A., Storti P., Viale A., Giuliani N. (2015) Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome. *Cancer Cell.* **27**:644–657.

Conboy J.G. (1993) Structure, function, and molecular genetics of erythroid membrane skeletal protein 4.1 in normal and abnormal red blood cells. *Semin. Hematol.* **30**:58–73.

Cooper T.A. (2005) Use of minigene systems to dissect alternative splicing elements. *Methods* **37**:331–340.

Crick F. (1970) Central dogma of molecular biology. *Nature* **227**: 561–563.

Cuperlovic-Cuif M., Belacel N., Culf A.S., Ouellette R.J. (2006) Microarray analysis of alternative splicing. *OMICS: J. Integr. Biol.* **10**:344–357.

Daebritz J., Weinlage T., Varga G., Wirth T., Walscheid K., Brockhausen A., Schwarzmaier D., Brueckner M., Ross M., Bettenworth D., Roth J., Ehrchen J.M., Foell D. (2015) Reprogramming of monocytes by GM-CSF contributes to regulatory immune functions during intestinal inflammation. *J. Immunol.* **194**:2424–2438.

Dai T., Zheng H., Fu G.S. (2008) Hypoxia confers protection against apoptosis via the PI3K/Akt pathway in endothelial progenitor cells. *Acta Pharmacol. Sin.* **29**:1425–1431.

Daigneault M., Preston J.A., Marriott H.M., Whyte M.K., Dockrell D.H. (2010) The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages. *PLoS One* **5**: e8668.

D'Angelo G., Duplan E., Boyer N., Vigne P., Frelin C. (2003) Hypoxia up-regulates prolyl hydroxylase activity: a feedback mechanism that limits HIF-1 responses during reoxygenation. *J. Biol. Chem.* **278**:38183–38187.

Darman R.B., Seiler M., Agrawal A.A., Lim K.H., Peng S., Aird D., Bailey S.L., Bhavsar E.B., Chan B., Colla S., Corson L., Feala J., Fekkes P., Ichikawa K., Keaney G.F., Lee L., Kumar P., Kunii K., MacKenzie C., Matijevic M., Mizui Y., Myint K., Park E.S., Puyang X., Selvaraj A., Thomas M.P., Tsai J., Wang J.Y., Warmuth M., Yang H., Zhu P., Garcia-Manero G., Furman R.R., Yu L., Smith P.G., Buonamici S. (2015) Cancer-Associated SF3B1 hotspot mutations induce cryptic 3' Splice site selection through use of a different branch point. *Cell Rep.* **13**:1033–1045.

de Back D.Z., Kostova E.B., van Kraaij M., van den Berg T.K., van Bruggen R. (2014) Of macrophages and red blood cells; a complex love story. *Front. Physiol.* **5**: article 9.

- DeBoever C., Ghia E.M., Shepard P.J., Rassenti L., Barrett C.L., Jepsen K., Jamieson C.H., Carson D., Kipps T.J., Frazer K.A. (2015) Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput. Biol.* **11**:e1004105.
- de la Mata M., Alonso C.R., Kadener S., Fededa J.P., Blaustein M.A., Pelisch F., Cramer P., Bentley D., Kornblihtt A.R. (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell.* **12**:525–532.
- De Conti L., Baralle M., Buratti E. (2013) Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **4**:49–60.
- Denko N.C. (2008) Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat. Rev. Cancer* **8**:705–713.
- Dennis G. Jr., Sherman B.T., Hosack D.A., Yang J., Gao W., Lane H.C., Lempicki R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**:P3.
- Deynoux M., Sunter N., Herault O., Mazurier F. (2016) Hypoxia and Hypoxia-Inducible Factors in Leukemias. *Front. Oncol.* **6**:41.
- Diner E.J., Burdette D.L., Wilson S.C., Monroe K.M., Kellenberger C.A., Hyodo M., Hayakawa Y., Hammond M.C., Vance R.E. (2013) The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3**:1355–1361.
- Dolatshad H., Pellagatti A., Fernandez-Mercado M., Yip B.H., Malcovati L., Attwood M., Przychodzen B., Sahgal N., Kanapin A.A., Lockstone H., Scifo L., Vandenberghe P., Papaemmanuil E. (2015) Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* **5**:1092–1103.
- Dolatshad H., Pellagatti A., Liberante F.G., Llorian M., Repapi E., Steeples V. (2016) Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia* **12**:2322–2331.

- Duana R., Sharmab S., Xiaa Q., Garberb K., Jinb P. (2014) Towards Understanding RNA-Mediated Neurological Disorders *J. Genet. and Genom.* **9**:473–484.
- Dustin ML. (2014) What counts in the immunological synapse? *Mol. Cell.* **2**:255–62.
- Dvinge H., Kim E., Abdel-Wahab O., Bradley R.K. (2016) RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer.* **16**:413–430.
- Dzierzak E. (2002) Hematopoietic stem cells and their precursors: developmental diversity and lineage relationships. *Immunol. Rev.* **187**:126–138.
- Dzierzak E., Speck N.A. (2008) Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat. Immunol.* **9**:129–136.
- Ebert B., Bernard O.A. (2011) Mutations in RNA splicing machinery in human cancers. *N. Engl. J. Med.* **365**:2534–2535.
- Ebert R. H., Florey H. W. (1939) The extravascular development of the monocyte observed in vivo. *Brit. J. Exp. Pathol.* **20**:342–356.
- Ema H., Morita Y. and Suda T. (2014) Heterogeneity and hierarchy of hematopoietic stem cells. *Exp. Hematol.* **42**:74–82.
- Espagnolle N., Balguerie A., Arnaud E., Sensebe L., Varin A. (2017) CD54-Mediated Interaction with Pro-inflammatory Macrophages Increases the Immunosuppressive Function of Human Mesenchymal Stromal Cells. *Stem Cell Rep.* **8**:961– 976.
- Fabregat A., Jupe S., Matthews L., Sidiropoulos K., Gillespie M., Garapati P. (2018) The Reactome pathway knowledgebase. *Nucleic Acids Res.* **46**:D649–Dd55.
- Fajtova M., Kovarikova A., Svec P., Kankuri E., Sedlak J. (2013) Immunophenotypic profile of nucleated erythroid progenitors during maturation in regenerating bone marrow. *Leuk. Lymph.* **54**:2523–2530.
- Faustino N.A., Cooper T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.* **17**:419–437.

- Gaballa M.R., Besa E.C. (2014) Myelodysplastic syndromes with 5q deletion: pathophysiology and role of lenalidomide. *Ann. Hematol.* **93**:723–733.
- Gao J., Aksoy B.A., Dogrusoz U., Dresdner G., Gross B., Sumer S.O., Sun Y., Jacobsen A., Sinha R., Larsson E., Cerami E., Sander C., Schultz N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**:11.
- Genin M., Clement F., Fattaccioli A., Raes M., Michiels C. (2015) M1 and M2 macrophages derived from THP-1 cells differentially modulate the response of cancer cells to etoposide. *BMC Cancer.* **15**:577–591.
- Germing U., Kobbe G., Haas R., Gattermann N. (2013) Myelodysplastic syndromes: diagnosis, prognosis, and treatment. *Dtsch. Ärztebl. Int.* **46**:783–790.
- Giannakakis A., Sandaltzopoulos R., Greshock J., Liang S., Huang J., Hasegawa K., Li C., O'Brien-Jenkins A., Katsaros D., Weber B.L. (2008) miR-210 links hypoxia with cell cycle regulation and is deleted in human epithelial ovarian cancer. *Cancer Biol. & Ther.* **7**:255–264.
- Gilby K.L., Armstrong J.N., Currie R.W., Currie R.W., Robertson H.A. (1997) The effects of hypoxia-ischemia on expression of c-Fos, c-Jun and Hsp 70 in the young rat hippocampus. *Brain Res. Mol. Brain Res.* **48**:87–96.
- Ginhoux F., Jung S. (2014) Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat. Rev. Immunol.* **14**:392–404.
- Gleissner C.A. (2012) Macrophage phenotype modulation by CXCL4 in atherosclerosis. *Front. Physiol.* **3**:1–7.
- Goda N., Ryan H.E., Khadivi B., McNulty W., Rickert R.C., Johnson R.S. (2003) Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia. *Mol. Cell. Biol.* **23**:359–369.
- Gonzalez I., Munita R., Agirre E., Dittmer T.A., Gysling K., Misteli T. (2015) A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat. Struct. Mol. Biol.* **22**:370–376.

- Gordon S. (2003) Alternative activation of macrophages. *Nat. Rev. Immunol.* **3**:23–35.
- Graubert T.A., Shen D., Ding L., Okeyo-Owuor T., Lunn C.L., Shao J. (2012) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* **44**:53–57.
- Graversen J.H., Madsen M., Moestrup S.K. (2002) CD163: a signal receptor scavenging haptoglobin-hemoglobin complexes from plasma. *Int. J. Biochem. Cell. Biol.* **34**:309–314.
- Gray N.K., Hrabáľková L., Scanlon J.P., Smith R.W. (2015) Poly(A)-binding proteins and mRNA localization: who rules the roost?. *Biochem. Soc. Trans.* **6**:1277–1284.
- Greijer A.E., Wall E. van der (2004) The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J. Clin. Pathol.* **57**:1009–1014.
- Gretarsdottir S., Helgason H., Helgadóttir A., Sigurdsson A., Thorleifsson G., Magnusdottir A. (2015) A Splice Region Variant in LDLR Lowers Non-high Density Lipoprotein Cholesterol and Protects against Coronary Artery Disease. *PLoS Genet.* **11**(9):e1005379.
- Grosso A.R., Gomes A.Q., Barbosa-Morais N.L., Caldeira S., Thorne N.P. (2008) Tissue-specific splicing factor gene expression signatures. *Nucleic. Acids Res.* **36**:4823–4832.
- Grigoryev Y.A., Kurian S.M., Nakorchevskiy A.A., Burke J.P., Campbell D., Head S.R., Deng J., Kantor A.B., Yates J.R. 3rd, Salomon D.R. (2009) Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. *PLoS One* **4** (11):e7906.
- Guthrie C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* **253**:157–163.
- Haase V. H. (2013) Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Rev.* **27**:41–53.
- Haase D. (2008) Cytogenetic features in myelodysplastic syndromes. *Ann. Hematol.* **87**:515–526.
- Hamid F.M., Makeyev E.V. (2014) Regulation of mRNA abundance by polypyrimidine tract-binding protein-controlled alternate 5' splice site choice. *PLoS Genet.* **10**:e1004771.

- Hammerstrøm J. (1979) Human macrophage differentiation in vivo and in vitro. A comparison of human peritoneal macrophages and monocytes. *Acta Pathol. Microbiol. Scand. C.* **87C**:113–120.
- Harris A.L. (2002) Hypoxia - a key regulatory factor in tumour growth. *Nat. Rev. Cancer.* **2**:38–47.
- Heinzelmann M., Mercer-Jones M.A., Gardner S.A., Wilson M.A., Polk H.C. (1997) Bacterial cell wall products increase monocyte HLA-DR and ICAM-1 without affecting lymphocyte CD18 expression. *Cell Immunol.* **176**:127–134.
- Hietakangas V., Poukkula M., Heiskanen K.M., Karvinen J.T., Sistonen L., Eriksson J.E. (2003) Erythroid differentiation sensitizes K562 leukemia cells to TRAIL-induced apoptosis by downregulation of c-FLIP. *Mol. Cell. Biol.* **23**:1278–1291.
- Hiller M., Huse K., Szafranski K., Jahn N., Hampe J., Schreiber S., Backofen R., Platzer M. (2004) Widespread occurrence of alternative splicing at nagnag acceptors contributes to proteome plasticity. *Nat. Genet.* **36**:1255–1257.
- Hoppstadter J., Seif M., Dembek A. (2015) M2 polarization enhances silica nanoparticle uptake by macrophages. *Front Pharmacol.* **6**:55-67.
- House A.E., Lynch K.W. (2008) Regulation of Alternative Splicing: More than just the ABCs. *J. Biol. Chem.* **283**:1217–1221.
- Huang X., Darzynkiewicz Z. (2006) Cytometric assessment of histone H2AX phosphorylation: a reporter of DNA damage. *Methods Mol. Biol.* **314**:73–80.
- Ip J. Y., Schmidt D., Pan Q., Ramani A.K., Fraser A.G., Odom D.T., Blencowe B.J. (2011) Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* **21**:390–401.
- Isono K., Mizutani-Koseki Y., Komori T., Schmidt-Zachmann M.S., Koseki H. (2005) Mammalian polycomb-mediated repression of hox genes requires the essential spliceosomal protein Sf3b1. *Genes Dev.* **19**:536–541.

- Italiani P., Boraschi D. (2014) From monocytes to M1/M2 macrophages: phenotypical vs. functional differentiation. *Front. Immunol.* **5**:514–536.
- Ivanov P., Anderson P. (2013) Post-transcriptional regulatory networks in immunity. *Immunol. Rev.* **253**:253–272.
- Jädersten M., Hellström-Lindberg E. (2009) Myelodysplastic syndromes: biology and treatment. *J. Intern. Med.* **3**:307–328.
- Jadersten M., Saft L., Smith A., Kulasekararaj A., Pomplun S., Gohring G., Hedlund A., Hast R., Schlegelberger B., Porwit A., Hellstrom-Lindberg E., Mufti G.J. (2011) TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. *J. Clin. Oncol.* **29**:1971–1979.
- Jamison S.F., Crow A., Garcia-Blanco M.A. (1992) The spliceosome assembly pathway in mammalian extracts. *Mol. Cell. Biol.* **12**:4279–4287.
- Jhanwar S.C. (2015) Genetic and epigenetic pathways in myelodysplastic syndromes: A brief overview. *Adv. Biol. Regul.* **58**:28–37.
- Ji X., Fu X.D. (2012) The mediator couples transcription and splicing. *Mol. Cell.* **4**:433–437.
- Jurica M.S., Moore M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**:5–14.
- Kadowaki T., Inagawa H., Kohchi C., Hirashima M., Soma G-I. (2011) Functional Characterization of Lipopolysaccharide derived from Symbiotic Bacteria in Rice as a Macrophage-activating Substance. *Anticancer Res.* **31**:2467–2476.
- Kamburov A., Pentchev K., Galicka H., Wierling C., Lehrach H., Herwig R. (2011) ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* **39**:D712–D717.
- Kao H.Y., Siliciano P.G. (1996) Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell Biol.* **16**:960–967.

- Katz Y., Wang E.T., Airolidi E.M., Burge C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**:1009–1015.
- Kazantseva J., Kivil A., Tints K., Kazantseva A., Neuman T., Palm K. (2013) Alternative splicing targeting the hTAF4–TAFH domain of TAF4 represses proliferation and accelerates chondrogenic differentiation of human mesenchymal stem cells. *PLoS One* **8**:e74799.
- Keren H., Lev-Maor G., Ast G. (2010) Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11**:345–355.
- Kim E., Goren A., Ast G. (2008) Alternative splicing: current perspectives. *Bioessays* **30**:38–47.
- Kim K.K., Nam J., Mukoyama Y.S., Kawamoto S. (2013) Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. *J. Cell Biol.* **200**:443–458.
- Kolasinska-Zwierz P., Down T., Latorre I., Liu T., Liu X.S., Ahringer J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**:376–381.
- Kondo M., Weissman I. L., Akashi K. (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**:661–672.
- Kornblihtt A.R., Schor I.E., Allo M., Dujardin G., Petrillo E., Munoz M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**:153–165.
- Kornblihtt A.R., de la Mata M., Fededa J.P., Munoz M.J., Nogues G. (2004) Multiple links between transcription and splicing. *RNA* **10**:1489–1498.
- Kuendgen A., Matsuda A., Germing U. (2007) Differences in epidemiology of MDS between Western and Eastern countries: ethnic differences or environmental influence? *Leuk. Res.* **31**:103–107.
- Kuleshov M.V., Jones M.R., Rouillard A.D., Fernandez N.F., Duan Q., Wang Z., Koplev S., Jenkins S.L., Jagodnik K.M., Lachmann A., McDermott M.G., Monteiro C.D., Gundersen G.W., Ma'ayan A. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**:W90–7.

- Kulshreshtha R., Ferracin M., Wojcik S.E., Garzon R. (2007) A MicroRNA Signature of Hypoxia. *Mol. Cell. Biol.* **27**:1859–1867.
- Kunzmann A., Liu D., Annett K., Malaisé M., Thaa B., Hyland P., Barnett Y., Bürkle A. (2006) Flow-cytometric assessment of cellular poly(ADP-ribosyl)ation capacity in peripheral blood lymphocytes. *Immun. Ageing* **19**:3–8.
- Landsman L. (2009) CX3CR1 is required for monocyte homeostasis and atherogenesis by promoting cell survival. *Blood* **113**:963–972.
- Lardelli R.M., Thompson J.X., Yates J.R. III, Stevens S.W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**:516–528.
- Larsson J., and Karlsson S. (2005) The role of Smad signaling in hematopoiesis. *Oncogene* **24**:5676–5692.
- Larsson C.A., Cote G., Quintás-Cardama A. (2013) The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol. Cancer Res.* **11**:815–827.
- Lee D.W., Futami M., Carroll M., Feng Y., Wang Z., Fernandez M. (2012) Loss of SHIP-1 protein expression in high-risk myelodysplastic syndromes is associated with miR-210 and miR-155. *Oncogene* **31**:4085–4094.
- Le Jan S., Amy C., Cazes A., Monnot C., Lamande N., Favier J. (2003) Angiopoietin-like 4 is a proangiogenic factor produced during ischemia and in conventional renal cell carcinoma. *Am. J. Pathol.* **5**:1521–1529.
- Lehrke M., Millington S.C., Lefterova M. (2007) CXCL16 is a marker of inflammation, atherosclerosis, and acute coronary syndromes in humans. *J. Am. Coll. Cardiol.* **49**:442–451.
- Lewis B.P., Green R.E., Brenner S.E. (2003) Evidence for the wide-spread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci.* **7**:189–192.

- Li L., Ng D.S., Mah W.C., Almeida F.F., Rahmat S.A., Rao V.K., Leow S.C., Laudisi F., Peh M.T., Goh A.M., Lim J.S., Wright G.D., Mortellaro A. (2015) A unique role for p53 in the regulation of M2 macrophage polarization. *Cell Death Differ.* **22**:1081–1093.
- Lin J.S., Lai E.M. (2017) Protein-Protein Interactions: Co-Immunoprecipitation. *Methods Mol Biol.* **1615**:211–219.
- Lin Z., Fillmore G.C., Um T.H., Elenitoba-Johnson K.S., Lim M.S. (2003) Comparative microarray analysis of gene expression during activation of human peripheral blood T cells and leukemic Jurkat T cells. *Lab. Invest.* **83**:765–776.
- Liu Y.C., Zou X.B., Chai Y.F., Yao Y.M. (2014) Macrophage polarization in inflammatory diseases. *Int. J. Biol. Sci.* **10**:520–529.
- Lynch K.W. (2004) Consequences of regulated pre-mRNA splicing in the immune system. *Nat. Rev. Immunol.* **4**:931–940.
- Lynch K.W. (2014) Mechanisms of spliceosomal assembly. *Methods Mol. Biol.* **1126**: 35–43.
- Long J.C., Caceres J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**:15–27.
- Lu Y.C., Yeh W.C., Ohashi P.S. (2008) LPS/TLR4 signal transduction pathway. *Cytokine.* **2**:145–151.
- Luco R.F., Pan Q., Tominaga K., Blencowe B.J., Pereira-Smith O.M., Misteli T. (2010) Regulation of alternative splicing by histone modifications. *Science* **327**:996–1000.
- Macheda M.L., Rogers S., Best J.D. (2005) Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *J. Cell Physiol.* **3**:654–662.
- Machnicka B., Czogalla A., Hryniewicz-Jankowska A., Boguslawska D.M., Grochowalska R., Heger E. (2014) Spectrins: a structural platform for stabilization and activation of membrane channels, receptors and transporters. *Biochim. Biophys. Acta.* **2**:620–634.

- Makarov E.M., Owen N., Bottrill A., Makarova O.V. (2012) Functional mammalian spliceosomal complex E contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Res.* **40**:2639–2652.
- Malcovati L., Papaemmanuil E., Bowen D.T., Boulwood J., Della Porta M.G., Pascutto C. (2011) Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118**:6239–6246.
- Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**:823–826.
- Maniatis T., Reed R. (2002) An extensive network of coupling among gene expression machines. *Nature* **416**:499–506.
- Mantovani A., Sica A., Sozzani S., Allavena P., Vecchi A., Locati M. (2004) The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* **25**:677–686.
- Marsee D. K., Pinkus G. S., Yu H. (2010) CD71 (transferrin receptor) an effective marker for erythroid precursors in bone marrow biopsy specimens. *Am. J. Clin. Path.* **3**:429–435.
- Martinez F.O., Sica A., Mantovani A., Locati M. (2008) Macrophage activation and polarization. *Front. Biosci.* **13**:453–461.
- Martinez N.M., Lynch K.W. (2013) Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immuno. Rev.* **253**:216–236.
- Martinez-Contreras R., Cloutier P., Shkreta L., Fiset J.F., Revil T., Chabot B. (2007) hnRNP proteins and splicing control. *Adv. Exp. Med Biol.* **623**:123–147.
- Madan V., Kanojia D., Li J., Okamoto R., Sato-Otsubo A., Kohlmann A., Sanada M., Grossmann V., Sundaresan J., Shiraishi Y., Miyano S., Thol F., Ganser A., Yang H., Haferlach T., Ogawa S., Koefler H.P. (2015) Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* **6**:6042–6055.

- Makishima M., Shudo K., Honma Y. (1999) Greater synergism of retinoic acid receptor (RAR) agonists with vitamin D3 than that of retinoid X receptor (RXR) agonists with regard to growth inhibition and differentiation induction in monoblastic leukemia cells. *Biochem. Pharmacol.* **57**:521–529.
- Malcovati L., Papaemmanuil E., Bowen D.T., Boulton J., Della Porta M.G., Pascutto C. (2011) Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood*. **118**:6239–6246.
- Maniatis T., Tasic B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*. **6894**:236–243.
- Mantovani A., Biswas S. K., Galdiero M. R., Sica A., Locati M. (2013) Macrophage plasticity and polarization in tissue repair and remodelling. *J. Pathol.* **2**:176–185. Matera G.A., Wang Z. (2014) A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**:108–121.
- Medyouf H., Mossner M., Jann J.C., Nolte F., Raffel S., Herrmann C., Lier A., Eisen C., Nowak V., Zens B. (2014) Myelodysplastic cells in patients reprogram mesenchymal stromal cells to establish a transplantable stem cell niche disease unit. *Cell Stem Cell*. **14**:824–837.
- Mia S., Warnecke A., Zhang X.M., Malmstrom V., Harris R.A. (2014) An optimized protocol for human M2 macrophages using M-CSF and IL-4/IL-10/TGF-beta yields a dominant immunosuppressive phenotype. *Scand. J. Immunol.* **79**:305–14.
- Mills C.D., Kincaid K., Alt J.M., Heilman M.J., Hill A.M. (2000) M-1/M-2 macrophages and the Th1/Th2 paradigm. *J. Immunol.* **164**:6166–6173.
- Moghaddam A.S., Mohammadian S., Vazini H., Taghadosi M., Esmaili S.A., Mardani F. (2018) Macrophage plasticity, polarization and function in health and disease. *J. Cell Physiol.* **epub**:1-16.
- Montague T.G., Cruz J.M., Gagnon J.A., Church G.M., Eivind V. (2014) CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**:W401–W407.

Monani U.R., Lorson C.L., Parsons D.W., Prior T.W., Androphy E.J., Burghes A.H., McPherson J.D. (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* **8**:1177–1183.

Morrison S.J., Spradling A.C. (2008) Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**:598–611.

Mottes J.R., Iverson L.E. (1995) Tissue-specific alternative splicing of hybrid Shaker/lacZ genes correlates with kinetic differences in Shaker K⁺ currents in vivo. *Neuron*. **14**:613–623.

Mukherjee R., Kanti Barman P., Kumar Thatoi P. (2015) Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematosus. *Sci. Rep.* **5**:13886-13990.

Murata M., Yudo K., Nakamura H., Chiba J., Okamoto K., Suematsu N. (2009) Hypoxia upregulates the expression of angiopoietin-like-4 in human articular chondrocytes: role of angiopoietin-like-4 in the expression of matrix metalloproteinases and cartilage degradation. *J. Orthop. Res.* **1**:50–57.

Nahkuri S., Taft R.J., Mattick J.S. (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**:3420–3424.

Naro C., Bielli P., Pagliarini V., Sette C. (2015) The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. *Front. Genet.* **6**:142-152.

Nathan C. F., Murray H.W., Wiebe M.E., Rubin B.Y. (1983) Identification of interferon- γ as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. *J. Exp. Med.* **158**:670–689.

Naumann S., Reutzel D., Speicher M., Decker H.-J. (2001) Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. *Leuk. Res.* **25**:313–322. Netea M.G., Lewis E.C., Azam T., Joosten L.A., Jaekal J., Bae S.Y., Dinarello

- C.A., Kim S.H. (2008) Interleukin-32 induces the differentiation of monocytes into macrophage-like cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**:3515–3520.
- Neukirchen J., Schoonen W.M., Strupp C. (2011) Incidence and prevalence of myelodysplastic syndromes: Data from the Düsseldorf MDS-registry. *Leuk. Res.* **35**:1591–1596.
- Nguyen P.H., Sterpone F., Campanera J.M., Nasica-Labouze J., Derreumaux, P. (2016) Impact of the A2V mutation on the Heterozygous and Homozygous A β 1-40 Dimer Structures from Atomistic Simulations. *ACS Chem. Neur.* **6**:823–832.
- Nilsen T.W., Graveley B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**:457–463.
- Okeyo-Owuor T., White B.S., Chatrikhi R., Mohan D.R., Kim S., Griffith M., Ding L., Ketkar-Kulkarni S., Hundal J., Laird K.M., Kielkopf C.L., Ley T.J., Walter M.J. (2015) U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* **29**:909–917.
- Oghumu S., Varikuti S., Terrazas C., Kotov D., Nasser M. W., Powell C. A., Ganju R. K., Satoskar A. R. (2014) CXCR3 deficiency enhances tumor progression by promoting macrophage M2 polarization in a murine breast cancer model. *Immunology* **143**:109–119.
- Ollila J., Vihinen M. (2007) Immunological system biology: gene expression analysis of B-cell development in Ramos B-cells. *Mol. Immunol.* **44**:3537–3551.
- Padgett R.A. (2012) New connections between splicing and human disease. *Trends Genet.* **28**:147–154.
- Pan Q., Shai O., Lee L.J., Frey, B.J., Blencowe B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**:1413–1415.
- Panek C.A., Ramos M.V., Mejias M.P., Abrey-Recalde M.J., Fernandez-Brando R.J., Gori M.S. (2015) Differential expression of the fractalkine chemokine receptor (CX3CR1) in human monocytes during differentiation. *Cell Mol. Immunol.* **6**:669–680.

Papaemmanuil E., Cazzola M., Boultonwood J. (2011) Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **15**:1384-1395.

Papaemmanuil E., Gerstung M., Malcovati L. (2013) Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**:3616–3627.

Parmar K., Mauch P., Vergilio J.A., Sackstein R., Down J.D. (2007) Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. *Proc. Natl. Acad. Sci. U. S. A.* **13**:5431–5437.

Park E.K., Jung H.S., Yang H.I., Yoo M.C., Kim C., Kim K.S. (2007) Optimized THP-1 differentiation is required for the detection of responses to weak stimuli. *Inflamm. Res.* **1**:45–50.

Park K.-Y., Li W.A., Platt M.O. (2012) Patient specific proteolytic activity of monocyte-derived macrophages and osteoclasts predicted with temporal kinase activation states during differentiation. *Integr. Biol.* **4**:1459–1469.

Passlick B., Flieger D., Ziegler-Heitbrock H.W. (1989) Identification and characterization of a novel monocyte subpopulation in human peripheral blood. *Blood* **74**:2527–2534.

Patel A.A., McCarthy M., Steitz J.A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* **21**:3804–3815.

Paulsen K., Tauber S., Timm J. (2011) The cannabinoid receptors agonist WIN55212-2 inhibits macrophageal differentiation and alters expression and phosphorylation of cell cycle control proteins. *Cell Commun. Signal* **9**:33–40.

Paydas S. (2006) Young age MDS: differences between Western and Eastern countries. *Leuk. Res.* **30**:362.

Pellagatti A., Hellstrom-Lindberg E., Giagounidis A., Perry J., Malcovati L., Della Porta M.G., Jadersten M., Killick S., Fidler C., Cazzola M., Wainscoat J.S., Boultonwood J. (2008) Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. *Brit. J. Haemat.* **142**:57–64.

- Pellagatti A., Boulton J. (2017) Splicing factor gene mutations in the myelodysplastic syndromes: impact on disease phenotype and therapeutic applications. *Adv. Biol. Regul.* **63**:59–70.
- Percio S., Coltella N., Grisanti S., Bernardi R., Pattini L. (2014) A HIF-1 network reveals characteristics of epithelial-mesenchymal transition in acute promyelocytic leukemia. *Genome Med.* **6**:84–96.
- Petit C., Gouel F., Dubus I., Heuclin C., Roget K., Vannier J.P. (2016) Hypoxia promotes chemoresistance in acute lymphoblastic leukemia cell lines by modulating death signaling pathways. *BMC Cancer.* **16**:746–756.
- Piccoli C., D'Aprile A., Ripoli M., Scrima R., Boffoli D., Tabilio A., Capitanio N. (2007) The hypoxia-inducible factor is stabilized in circulating hematopoietic stem cells under normoxic conditions. *FEBS Let.* **16**:3111–3120.
- Pilon A.M., Arcasoy M.O., Dressman H.K., Vayda S.E., Maksimova Y.D., Sangerman J.I. (2008) Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol. Cell. Biol.* **24**:7394–7401.
- Pimentel H., Parra M., Gee S., Ghanem D., An X., Li J., Mohandas N., Pachter L., Conboy J.G. (2014) A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **42**:4031–4042.
- Pinet V., Vergelli M., Martin R., Bakke O., Long E. O. (1995) Antigen presentation mediated by recycling of surface HLA-DR molecules. *Nature* **375**:603–606.
- Pires I.M., Bencokova Z., Milani M., Folkes L.K., Li J.L., Stratford M.R., Harris A.L., Hammond E.M. (2010) Effects of acute versus chronic hypoxia on DNA damage responses and genomic instability. *Cancer Res.* **70**:925–935.
- Porta C., Rimoldi M., Raes G., Brys L., Ghezzi P., Di Liberto D. (2009) Tolerance and M2 (alternative) macrophage polarization are related processes orchestrated by p50 nuclear factor κ B. *Proc. Natl Acad. Sci. U.S.A.* **106**:14978–14983.

Poynter J.N., Richardson M., Roesler M., Blair C.K., Hirsch B., Nguyen P., Cioc A., Cerhan J.R., Warlick E. (2017) Chemical exposures and risk of acute myeloid leukemia and myelodysplastic syndromes in a population-based study. *Int. J. Cancer* **1**:23–33.

Przychodzen B., Jerez A., Guinta K., Sekeres M.A., Padgett R., Maciejewski J.P., Makishima H. (2013) Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* **122**:999–1006.

Qiu P., Shandilya H., D'Alessio J.M., O'Connor K., Durocher J., Gerard G.F. (2004) Mutation detection using Surveyor nuclease. *Biotechniques* **36**:702–707.

Rajan P., Elliott D.J., Robson C.N. and Hing Y. L. (2009) Alternative splicing and biological heterogeneity in prostate cancer. *Nat. Rev. Uro.* **6**:454–460.

Ran F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**:2281–2308.

Ravi R., Mookerjee B., Bhujwala Z.M. Sutter C.H., Artemov D., Zeng Q., Dillehay L.E., Madan A., Semenza G.L., Bedi A. (2000) Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. *Genes Dev.* **14**:34–44.

Reed R., Maniatis T. (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev.* **2**:1268–1276.

Rey-Giraud F., Hafner M., Ries C.H. (2012) In vitro generation of monocyte-derived macrophages under serum-free conditions improves their tumor promoting functions. *PLoS One.* **7** (8):e42656.

Richter E., Ventz K., Harms M., Mostertz J., Hochgrafe F. (2016) Induction of macrophage function in human THP-1 cells is associated with rewiring of MAPK signaling and activation of MAP3K7 (TAK1) protein kinase. *Front. Cell Dev. Biol.* **4**:21.

Rodrigues R., Grosso A.R., Moita L. (2013) Genome-Wide Analysis of Alternative Splicing during Dendritic Cell Response to a Bacterial Challenge. *PLoS One* **8** (4):e61975.

Ruggiu M., McGovern V.L., Lotti F., Saieva L., Li D.K., Kariya S., Monani U.R., Burghes A.H., Pellizzoni L. (2012) A role for SMN exon 7 splicing in the selective vulnerability of motor neurons in spinal muscular atrophy. *Mol. Cell. Biol.* **32**:126–138.

Samaniego R., Palacios B.S., Domiguez-Soto A., Vidal C., Salas A., Matsuyama T., Sanchez-Torres C., de la Torre I., Miranda-Carus M.E., Sanchez-Mateos P. (2014) Macrophage uptake and accumulation of folates are polarization-dependent in vitro and in vivo and are regulated by activin A. *J. Leukoc. Biol.* **5**:797–808.

Sakharkar M.K., Chow V.T., Kanguane P. (2004) Distributions of exons and introns in the human genome. *In Silico Biol.* **4**:387–393.

Sammeth M., Foissac S., Guigo. (2008) A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.* **4** (8): e1000147.

Sánchez-Álvarez M., Montes M., Sánchez-Hernández N., Hernández-Munain C., Suñé C. (2010) Differential effects of sumoylation on transcription and alternative splicing by transcription elongation regulator 1 -(TCERG1). *J. Biol. Chem.* **285**:15220–15233.

Savage K.I., Gorski J.J., Barros E.M., Irwin G.W., Manti L., Powell A.J., Pellagatti A., Lukashchuk N., McCance D.J., McCluggage W.G. (2014) Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. *Mol. Cell.* **54**:445–459.

Schmutz C., Cartwright A., Williams H., Haworth O., Williams J.H., Filer A., Salmon M., Buckley C.D., Middleton J. (2010) Monocytes/macrophages express chemokine receptor CCR9 in rheumatoid arthritis and CCL25 stimulates their differentiation. *Arthritis Res. Ther.* **12**:R161.

Schor I.E., Lleres D., Risso G.J., Pawellek A., Ule J., Lamond A.I., Kornblihtt A.R. (2012) Perturbation of chromatin structure globally affects localization and recruitment of splicing factors. *PLoS One* **7**:e48084.

Schwende H., Fitzke E., Ambs P., Dieter P. (1996) Differences in the state of differentiation of THP-1 cells induced by phorbol ester and 1,25-dihydroxyvitamin D3. *J. Leukoc. Biol.* **59**:555–561.

- Schwerk C., and Schulze-Osthoff K. (2005) Regulation of apoptosis by alternative pre-mRNA splicing. *Mol. Cell* **19**:1–13.
- Seita J., Weissman I.L. (2010) Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**:640–653.
- Semenza GL. (2000) HIF-1 and human disease: one highly involved factor. *Genes Dev.* **16**:1983–1991.
- Seo P. J., Kim M. J., Ryu J. Y., Jeong E. Y., Park C. M. (2011) Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* **2**:303–311.
- Seraphin B., Rosbash M. (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**:349–358.
- Sharp P.A., Burge C.B. (1997) Classification of Introns: U2-Type or U12-Type. *Cell* **7**:875–879.
- Shen S., Park J.W., Huang J., Dittmar K.A., Lu Z.X., Zhou Q., Carstens R.P., Xing Y. (2012) MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* **40**:61.
- Soares-Schanoski A., Gomez-Pina V., del Fresno C., Rodriguez-Rojas A., Garcia F., Glaría A., Sánchez M., Vallejo-Cremades M.T., Baos R., Fuentes-Prior P. (2007) 6-Methylprednisolone down-regulates IRAK-M in human and murine osteoclasts and boosts bone-resorbing activity: a putative mechanism for corticoid-induced osteoporosis. *J. Leukoc. Biol.* **82**:700–709.
- Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**:2498–2504.
- Sheth N., Roca X., Hastings M.L., Roeder T., Krainer A.R., Sachidanandam R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**:3955–3967.

- Shin G., Kang T.W., Yang S., Baek S.J., Jeong Y.S., Kim S.Y. (2011) GENT: Gene expression database of normal and tumor tissues. *Cancer Inform.* **10**:149–157.
- Srivastava A., Shinn A.S., Lam T.T., Lee P.J., Mannam P. (2016) SILAC based protein profiling data of MKK3 knockout mouse embryonic fibroblasts. *Data Brief.* **7**:418–422.
- Sperandio S., Fortin J., Sasik R., Robitaille L., Corbeil J., de Belle I. (2009) The transcription factor *egr1* regulates the *hif-1alpha* gene during hypoxia. *Mol. Carcinog.* **48**:38–44.
- Stamm S., BenAri S., Rafalska I., Tang Y., Zhang Z., Toiber D., Thanaraj T.A., Soreq H. (2005) Function of alternative splicing. *Gene* **344**:1–20.
- Stein J.L., Parikshak N.N., Geschwind D.H. (2013) Rare inherited variation in autism: beginning to see the forest and a few trees. *Neuron* **77**:209–211.
- Stoddart A., Fernald A.A., Wang J., Davis E.M., Karrison T., Anastasi J. (2014) Haploinsufficiency of *del(5q)* genes, *Egr1* and *Apc*, cooperate with *Tp53* loss to induce acute myeloid leukemia in mice. *Blood* **123**:1069–1078.
- Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., Paulovich A., Pomeroy S.L., Golub T.R., Lander E.S., Mesirov J.P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**:15545–15550.
- Suda T., Takubo K., Semenza G.L. (2011) Metabolic Regulation of Hematopoietic Stem Cells in the Hypoxic Niche. *Cell Stem Cell.* **9**:298–310.
- Sutherland J.A., Turner A.R., Mannoni P., McGann L.E., Turc J.M. (1986) Differentiation of K562 leukemia cells along erythroid, macrophage, and megakaryocyte lineages. *J. Biol. Response Mod.* **5**:250–262.
- Sveen A., Kilpinen S., Ruusulehto A., Lothe R.A., Skotheim R.I. (2015) Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **19**:2413–2427.

Szkarczyk D., Morris J.H., Cook H., Kuhn M., Wyder S., Simonovic M. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**:D362–D368.

Swietach P., Vaughan-Jones R.D., Harris A.L. (2007) Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer Metastasis Rev.* **26**:299–310.

Tabilio A., Pelicci P.G., Vinci G., Mannoni P., Civin C.I., Vainchenker W., Testa U., Lipinski M., Rochant H., Breton-Gorius J. (1983) Myeloid and megakaryocytic properties of K-562 cell lines. *Cancer Res.* **43**:4569–4574.

Tahara E., Kadara H., Lacroix L., Lotan D., Lotan R. (2009) Activation of protein kinase C by phorbol 12-myristate 13-acetate suppresses the growth of lung cancer cells through KLF6 induction. *Cancer Biol. Ther.* **8**:801–807.

Takubo K., Goda N., Yamada W., Iriuchishima H., Ikeda E., Kubota Y. (2010) Regulation of the HIF-1 α level is essential for hematopoietic stem cells. *Cell Stem Cell.* **3**:391–402.

Tan J., Roca X. (2016) Informational suppression to probe RNA:RNA interactions in the context of ribonucleoproteins: U1 and 5' splice-site base-pairing. *Methods Mol. Biol.* **1421**:243–268.

Tang S.J., Luo S., Ho J.X., Ly P.T., Goh E., Roca X. (2016) Characterization of the Regulation of CD46 RNA Alternative Splicing. *J. Biol. Chem.* **291**:14311–14323.

Tang Z., Li C., Kang B., Gao G., Li C., Zhang Z. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**:W98–W102.

Tarn, W.Y., Steitz, J.A. (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**:801–811.

Taylor M.L., Noble P.W., White B., Wise R., Liu M.C., Bochner B.S. (2000) Extensive surface phenotyping of alveolar macrophages in interstitial lung disease. *Clin. Immunol.* **94**:33–41.

Thol F., Kade S., Schlarmann C. (2012) Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* **119**:3578–3584.

- Thomas J., Palusa S.G., Prasad K.V., Ali G.S., Surabhi G.K. (2012) Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. *Plant J.* **72**:935–946.
- Thompson J.E., Conlon J.P., Yang X., Sanchez P.V., Carroll M. (2007) Enhanced growth of myelodysplastic colonies in hypoxic conditions. *Exp. Hemat.* **35**:21–31.
- Thorvaldsdóttir H., Robinson J.T., Mesirov J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform. Oxford Univ. Press* **14**, 178–192.
- Tonelli F.M., Santos A.K., Gomes D.A., da Silva S.L., Gomes K.N., Ladeira L.O., Resende R.R. (2012) Stem cells and calcium signaling. *Adv. Exp. Med. Biol.* **740**:891–916.
- Tong H., Hu C., Zhuang Z., Wang L., Jin J. (2012) Hypoxia-inducible factor-1 α expression indicates poor prognosis in myelodysplastic syndromes. *Leu. & Lymph.* **53**:2412–2418.
- Topp J.D., Jackson J., Melton A.A., and Lynch K.W. (2008) A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA* **14**:2038–2049.
- Trapnell C., Pachter L., Salzberg S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**:1105–1111.
- Tsuchiya S., Yamabe M., Yamaguchi Y., Kobayashi Y., Konno T., Tada K. (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int. J. Cancer* **26**:171–176.
- Tsuchiya S., Kobayashi Y., Goto Y., Okumura H., Nakae S., Konno T., Tada K. (1982) Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. *Cancer Res.* **42**:1530–1536.
- Tuomisto T.T., Riekkinen M.S., Viita H., Levonen A.L., Yla-Herttuala S. (2005) Analysis of gene and protein expression during monocyte-macrophage differentiation and cholesterol loading—cDNA and protein array study. *Atherosclerosis* **180**:283–291.

Valentin J.E., Stewart-Akers A.M., Gilbert T.W., Badylak S. F. (2009) Macrophage participation in the degradation and remodelling of extracellular matrix scaffolds. *Tissue Eng. Part A* **15**:1687–1694.

Van Furth R., Diesselhoff Den Dulk M., Mattie H. (1973) Quantitative study on the production and kinetics of mononuclear phagocytes during an acute inflammatory reaction. *J. Exp. Med.* **138**:1314–1330.

Veloso A., Kirkconnell K.S., Magnuson B., Biewen B., Paulsen M.T., Wilson T.E., Ljungman M. (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**:896–905.

Vithana E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Magthteh M., Ebenezer N.D., Willis C., Moore A.T., Bird A.C., Hunt D.M., Bhattacharya S. S. (2011) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8**:375-381.

Vuong J.K., Lin C.H., Zhang M., Chen L., Black D.L., Zheng S. (2016) PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell* **10**:2766-2775.

Wahl M. C., Will C. L. and Luhrmann R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**:701–718.

Wang E.T., Sandberg R., Luo S., Khrebtkova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**:470–476.

Wang L., Lawrence M.S., Wan Y., Stojanov P., Sougnez C., Stevenson K., Werner L., Sivachenko A., De Luca D.S., Zhang L., Zhang W., Vartanov A.R., Fernandes S.M., Goldstein N.R., Folco E.G., Cibulskis K., Tesar B., Sievers Q.L., Shefler E., Gabriel S., Hacohen N., Reed R., Meyerson M., Golub T.R., Lander E.S., Neuberger D., Brown J.R., Getz G., Wu C.J. (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**:2497–2506.

Wang Z., Burge C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**:802–813.

Walter M.J., Shen D., Ding L. (2012) Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**:1090–1098.

Walter M., Shen D., Shao J., Ding L., White B., Kandoth C. (2013) Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* **27**:1275–1282.

Welch J.S., Ley T.J., Link D.C., Miller C.A., Larson D.E., Koboldt D.C., Wartman L.D., Lamprecht T.L., Liu F., Xia J. (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* **2**:264–278.

Wenger R.H., Kvietikova I., Rolfs A., Gassmann M., Marti H.H. (1997) Hypoxia-inducible factor-1 α is regulated at the post-mRNA level. *Kidney Int.* **51**:560–563.

Wiese S., Reidegeld K.A., Meyer H.E., Warscheid B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**:340–350.

Will C. L., Luhrmann R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**: a003707.

Witten J.T., Ule J. (2011) Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27**:89–97.

Wong J.J., Ritchie W., Ebner O.A., Selbach M., Wong J.W., Huang Y. (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**:583–595.

Wong K.L., Yeap W.H., Tai J.J.Y., Ong S.M., Dang T.M., Wong S.C. (2012) The three human monocyte subsets: implications for health and disease. *Immunol. Res.* **53**:41–57.

Wu J., Akerman M., Sun S., McCombie W.R., Krainer A.R., Zhang M.Q. (2011) SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics* **27**:3010–3016.

Wu S.J., Kuo Y.Y., Hou H.A. (2012) The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood* **120**:3106–3111.

Xiao S.H., Manley J.L. (1998) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.* **11**:334–344.

Yamamoto M.L., Clark T.A., Gee S.L., Kang J.A., Schweitzer A.C., Wickrema A., Conboy J.G. (2009) Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. *Blood* **14**:3363–3370.

Yeo G., Burge C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**:377–394.

Yingling J., Lahn M., List A. F., Bitzer M., Verma A. (2009) Dysregulation of TGF-Beta Stimulated Smad Signaling Is Seen in Myelodysplasia and Points to the Potential Therapeutic Efficacy of TGF-Beta Receptor I Kinase Inhibition in Low Grade Disease. *Blood* **114**(22),737.

Yoshida K., Sanada M., Shiraishi Y., Nowak D., Nagata Y., Yamamoto R., Sato Y., Sato-Otsubo A., Kon A., Nagasaki M., Chalkidis G., Suzuki Y., Shiosaka M., Kawahata R., Yamaguchi T., Otsu M., Obara N., Sakata-Yanagimoto M., Ishiyama K., Mori H., Nolte F., Hofmann W.K., Miyawaki S., Sugano S., Haerlach C., Koeffler H.P., Shih L.Y., Haerlach T., Chiba S., Nakauchi H., Miyano S., Ogawa S. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**:64–69.

Yu F., White S.B., Zhao Q., Lee F.S. (2001) HIF-1 α binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc. Natl. Acad. Sci. U. S. A.* **98**:9630–9635.

Yu Y., Das R., Folco E.G. (2010) A model in vitro system for co-transcriptional splicing. *Nucleic Acids Res.* **21**:7570–7578.

Zamani F., Zare Shahneh F., Aghebati-Maleki L., Baradaran B. (2013) Induction of CD14 expression and differentiation to monocytes or mature macrophages in promyelocytic cell lines: new approach. *Adv. Pharm. Bul.* **2**:329–332.

Zhang L., Huang G., Li X., Zhang Y., Jiang Y., Shen J., Liu J., Wang Q., Zhu J., Feng X., Dong J., Qian C. (2013) Hypoxia induces epithelial-mesenchymal transition via activation of SNAI1 by hypoxia-inducible factor -1 α in hepatocellular carcinoma. *BMC Cancer* **13**:1–9.

Zhang X.H., Tee L.Y., Wang X.G., Huang Q.S., Yang S.H. (2015) Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. Nucleic Acids*. **4**:e264.

Zhang Y., Choksi S., Chen K., Pobeziinskaya Y., Linnoila I., Liu Z.G. (2013) ROS play a critical role in the differentiation of alternatively activated macrophages and the occurrence of tumor-associated macrophages. *Cell Res*. **23**:898–914.

Zhang C.C., Sadek H.A. (2014) Hypoxia and metabolic properties of hematopoietic stem cells. *Antioxid. Redox Signal*. **20**:1891–1901.

Zhang J., Lieu Y.K., Ali A.M., Penson A., Reggio K.S., Rabadan R., Raza A., Mukherjee S., Manley J.L. (2015) Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **112**:E4726–E4734.

Zhang C., Wang Y., Wang F., Wang W., Lu Y., Xu Y., Wang K., Shen H., Yang P., Li S., Qin X, Yu H. (2017) Quantitative profiling of glycerophospholipids during mouse and human macrophage differentiation using targeted mass spectrometry. *Sci. Rep.* **412** doi:10.1038/s41598-017-00341-2.

Zhou H.L., Hinman M.N., Barron V.A., Geng C., Zhou G., Luo G., Siegel R.E., Lou H. (2011) Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **108**:E627–E635.

Zhou T., Hasty P., Walter C.A., Bishop A.J., Scott L.M., Rebel V.I. (2013) Myelodysplastic syndrome: An inability to appropriately respond to damaged DNA? *Exp. Hematol.* **41**:665–674.

Zhou L., McMahon C., Bhagat T., Alencar C., Yu Y., Fazzari M., Sohal D. P., Gundabolu K., Ng C., Mo Y., Wickrema A., Kong G., Friedman E., Sokol L., Burnette P. K., Steidl U., Yan L.,

Ziegler-Heitbrock L. (2007) The CD14⁺ CD16⁺ blood monocytes: their role in infection and inflammation. *J. Leukoc. Biol.* **81**:584–592.