Pathophysiology and Treatment of Myelodysplasia with del(5q)

ERICA BELLO

St Cross College



A thesis submitted to the Medical Sciences Division of the University of Oxford in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Michaelmas Term 2017

Nuffield Division of Clinical Laboratory Sciences
Radcliffe Department of Medicine
University of Oxford

Abstract

Pathophysiology and treatment of myelodysplasia with del(5q)

Erica Bello D.Phil. Medical Sciences

St Cross College Michaelmas Term 2017

The myelodysplastic syndrome (MDS) with deletion of chromosome 5q [del(5q)] is a distinct subtype of MDS characterised by refractory hypoplastic anaemia and the del(5q) as the sole karyotypic abnormality. There is a clinical need for novel treatments for MDS with del(5q). A more complete understanding of the pathophysiology of del(5q) MDS would likely impact on development of new therapies and improve patients outcome. The aim of this D.Phil. project was to investigate various aspects of the pathophysiology of MDS with del(5q) and potential therapeutic approaches for this MDS subtype. Mutations of the *CSNK1A1* gene, which maps to the commonly deleted region on chromosome 5q, were found to be rare events in a large cohort of del(5q) MDS patients. Moreover, del(5q) MDS patients with *CSNK1A1* mutation showed significant up-regulation of genes involved in the Wnt/β-catenin pathway. These data support the notion that *CSNK1A1* is an important gene in the pathogenesis of MDS with del(5q). The mechanism of L-leucine-mediated enhancement of erythropoiesis in the MDS with del(5q) was also investigated. L-leucine was found to increase proliferation and differentiation of cultured erythroblasts from MDS del(5q) patients through the activation of the mTOR pathway and increased translation of mTOR-regulated 5'Terminal Oligo Pyrimidine tract (TOP) mRNAs. Lastly, in a proof-of–principle study, the CRISPR/Cas9 system was used to correct a *TP53* mutation in the chronic myeloid leukaemia cell line K562. Correction of the *TP53* mutation by homology-directed repair (HDR) was obtained but corrected K562 cells also harboured additional genetic changes in the *TP53* locus. This study provided new insights into the pathogenesis of MDS with del(5q), and into the mode of action of the translation enhancer L-leucine in this disorder.

Acknowledgements

I would like to thank my first supervisor Professor Jacqueline Boultwood for the opportunity of conducting my D.Phil. project in her lab. I would like to express my gratitude to my second supervisor Dr Andrea Pellagatti for his support and guidance throughout the project and for his critical reading of this thesis. I would like to thank the members of the group for their advice and Bloodwise for providing the funding for this project.

My gratitude goes to the outstanding scientists I collaborated with at the MRC Weatherall Institute of Molecular Medicine (WIMM). I would like to thank Professor Tudor Fulga and all the members of his group for their advice and for sharing ideas which helped shaping this project. In particular, I would like to thank Dr Qianxin Wu for being an incredibly inspiring scientist and an immense source of knowledge. I thank her for her patience, support, guidance and encouragement throughout my D.Phil. project. I am grateful to Professor Tom Milne and his lab for always offering a welcoming environment and for their personal and scientific support. I would like to thank the Computational Biology Research Group at the WIMM and Ross Thorne (Milne group) for their help and advice on the analysis of my data. I would like to thank the friends I made during my time in Oxford and in particular Gemma Moir-Meyer for her endless encouragement and support. I would like to thank Dr Jon Kerry, for being the best partner I could have met and for offering me advice, encouragement and motivation throughout these four years. I am also grateful for his help on the analysis of my data and for providing insightful comments on this thesis. Finally, I would like to thank my family, my parents Antonella and Luciano and my brother Francesco, for always being interested and enthusiastic about my work and for being encouraging and supportive throughout my D.Phil. project.

Table of Contents

Abstract **1**

Acknowledgements **2**

Table of Contents **3**

List of Figures **11**

List of Tables **13**

List of abbreviations **15**

1. Chapter One - Introduction **19**

1.1. The myelodysplastic syndromes **19**

1.2. The molecular pathogenesis of MDS **21**

1.3. Treatment of MDS **24**

1.4. MDS with isolated del(5q) and 5q- syndrome **26**

1.5. The molecular pathogenesis of 5q- syndrome and MDS with del(5q) **26**

1.5.1. The commonly deleted region 26

1.5.2. Haploinsufficiency of *RPS14* 27

1.5.3. Role of p53 in pathogenesis of MDS with del(5q) 29

1.5.3.1. The tumour protein 53 29

1.5.3.2. Activation of p53 in MDS with del(5q) 29

1.5.4. *TP53* mutations 31

1.5.5. The role of *CSNK1A1* in the pathogenesis of MDS with del(5q) 32

1.5.5.1. The Wnt/β catenin pathway 32

1.5.5.2. *CSNK1A1* haploinsufficiency in MDS with del(5q) 33

1.5.5.3. *CSNK1A1* mutations in 5q- syndrome and MDS with del(5q) 34

1.5.5.4. *CSNK1A1* and the mechanism of action of lenalidomide 35
1.5.6. Other important genetic events in MDS with isolated del(5q)..... 35
1.6. Therapeutic approaches for MDS with isolated del(5q) 36
1.6.1. Lenalidomide..... 36
1.6.2. Cenersen..... 37
1.6.3. L-leucine 37
1.6.3.1. Evidence from animal and cellular models of MDS with del(5q) and other ribosomopathies 37
1.6.3.2. The mTOR pathway in regulation of mRNA translation..... 38
1.6.3.3. Current knowledge of the mechanisms of action of L-leucine 42
1.7. The CRISPR/Cas9 system..... 44
1.7.1. Genome editing for modelling of MDS..... 45
1.8. Aims of this D.Phil. project..... 45
2. Chapter Two – Materials and Methods 46
2.1. Patients samples..... 46
2.1.1. CSNK1A1 study..... 46
2.1.2. L-leucine study..... 54
2.2. Cell culture 55
2.2.1. Differentiation of mononuclear cells along the erythroid lineage..... 55
2.2.2. Cell lines 55
2.2.3. CRISPR-edited single cells..... 56
2.3. L-leucine treatment..... 57
2.4. Genomic DNA extraction 57
2.4.1. *CSNK1A1* genotyping..... 57
2.4.2. DNA extraction from cell lines 57

2.4.3. DNA extraction from CRISPR-edited single cells 57
2.5. Polymerase Chain Reaction (PCR) 57
2.6. Sanger sequencing 64
2.7. RNA extraction 64
2.7.1. RNA extraction (standard method) 64
2.7.2. Extraction of total RNA and ribosome-bound RNA (for polysome profiling) 64
2.8. DNase I treatment 65
2.9. Reverse transcription 65
2.10. RT-qPCR 66
2.10.1. TaqMan RT-qPCR reactions 66
2.10.2. SYBR Green RT-qPCR reactions 67
2.11. GEP and data analysis 69
2.12. Preparation of cell extracts 69
2.12.1. Cell extracts for Western Blotting 69
2.12.2. Cell extracts for sandwich ELISA 70
2.12.3. Cell extracts for polysome profiling 70
2.13. Protein quantification 70
2.14. Western blotting 71
2.15. Sandwich ELISA 71
2.16. Ribosome profiling 72
2.17. Polysome Profiling 73
2.17.1. Cycloheximide treatment 73
2.17.2. Preparation of cell extracts 73
2.17.3. EDTA treatment of cell extracts 73

2.17.4. Isolation of polysome-mRNA complexes 74
2.17.5. Isolation of total RNA and polysome-bound RNA 74
2.17.6. DNase I treatment of total and polysome-bound RNA 74
2.17.7. Assessment of RNA quality and quantity 74
2.17.8. RNA sequencing 75
2.17.8.1. Library preparation 75
2.17.8.1.1. Ovation Human Blood RNA-Seq Library System 75
2.17.8.1.2. SmartSeq2 76
2.17.8.2. Sequencing of libraries 76
2.17.8.3. Analysis of RNA sequencing data 77
2.17.8.3.1. Trimming of *fastq* files 77
2.17.8.3.2. Mapping of trimmed *fastq* files 77
2.17.8.3.3. Quality control 78
2.17.8.3.4. Counting of reads mapping to genes 78
2.17.8.3.5. Calculation of transcripts per million 78
2.17.8.3.6. Differential gene expression 79
2.17.8.3.7. Gene Set Association Analysis 79
2.17.8.3.8. Calculation of translation efficiency 79
2.17.8.3.9. Analysis of 5' TOP and 5'TOP-like motifs 80
2.17.8.3.10. Plotting of RNA-seq data 80
2.18. Flow cytometry 81
2.18.1. Staining of cell surface markers 81
2.18.2. Apoptosis analysis 81
2.18.3. Intracellular staining of alpha and beta globin 82
2.19. CRISPR design 82

2.20. Cloning of sgRNAs into the pSpCas9(BB)-2A-GFP vector..... 86

2.21. Cell transfection of CRISPR/Cas9 system..... 87

2.22. Fluorescence activated cell sorting..... 88

2.23. CRISPR-transfected cells culture and harvesting..... 89

2.24. High Resolution Melting Analysis..... 89

2.25. Restriction digest assay..... 90

2.26. Off-target sites analysis..... 90

2.27. TOPO-TA cloning..... 91

2.28. Cell growth assay..... 92

2.29. Targeted NGS sequencing of the CRISPR-edited locus..... 92

2.29.1. Library preparation..... 92

2.29.2. Sequencing..... 104

2.29.3. Analysis of raw DNA sequencing data..... 105

2.29.3.1. Quality control..... 105

2.29.3.2. Trimming *fastq* files..... 105

2.29.3.3. Mapping *fastq* files..... 105

2.29.3.4. Counting CRISPR-edited reads..... 106

2.29.3.5. Analysis of additional mutations in CRISPR-edited reads..... 107

2.30. Statistical analysis..... 107

2.30.1. Flow cytometry and ELISA..... 107

2.30.2. Analysis of translation efficiency..... 108

3. Chapter 3 – Analysis of *CSNK1A1* mutations and gene expression in myelodysplastic syndromes with del(5q)..... 109

3.1. Introduction..... 109

3.2. *CSNK1A1* mutations are rare events in MDS with del(5q)..... 110

3.3. Haploinsufficiency of *CSNK1A1* is associated with increased expression of Wnt/β-catenin effector genes 113

3.4. *CSNK1A1* mutations in del(5q) MDS lead to increased expression of genes involved in Wnt signalling 114

3.5. Discussion 117

4. Chapter Four - An investigation into the mode of action of the translation enhancer L-leucine in the MDS with del(5q) 122

4.1. Introduction 122

4.2. Results 124

4.2.1. L-leucine increases erythroid differentiation 124

4.2.2. L-leucine increases phosphorylation of mTOR targets 132

4.2.3. Optimisation of polysome profiling technique in cultured erythroblasts from patients with del(5q) MDS 134

4.2.4. Expression of genes in the CDR of 5q- syndrome is decreased in cultured erythroblasts from del(5q) MDS patients 143

4.2.5. MDS with del(5q) is associated with decreased expression of ribosomal genes 148

4.2.6. Erythroblasts from patients with del(5q) MDS show increased expression of HSC genes and decreased expression of markers of erythroid differentiation..... 154

4.2.7. MDS with del(5q) is characterised by decreased translation efficiency of ribosomal and erythroid genes 155

4.2.8. L-leucine treatment affects mRNA translation but not mRNA transcription in cultured erythroblasts from del(5q) MDS patients 158

4.2.9. L-leucine treatment increases mRNA translation in cultured erythroblasts from del(5q) MDS patients 162

4.2.10. L-leucine treatment of cultured erythroblasts from del(5q) MDS patients increases translation of 5'TOP mRNAs 165
4.2.11. Translation of 5'TOP mRNAs is decreased in MDS with del(5q) and increased by L-leucine treatment 170
4.3. Discussion 174
5. Chapter Five – Correction of a *TP53* mutation using the CRISPR/Cas9 system 180
5.1. Introduction 180
5.2. Results 181
5.2.1. Validation of *TP53* mutations in K562 cells 181
5.2.2. CRISPR design 184
5.2.3. Evidence of HDR in K562 cells transfected with the CRISPR/Cas9 system 184
5.2.4. *TP53* mutation correction and detection of additional mutations 189
5.2.5. Restoration of *TP53* protein expression in a corrected clone with an additional mutation 192
5.2.6. Investigation of HDR-mediated modification of the *TP53* locus in K562 cells 194
5.2.7. Investigation of HDR-mediated modification of *TP53* and *ASXL1* using CRISPR-introduced barcodes 203
5.3. Discussion 212
6. Chapter Six – General Discussion 217
References 222
Appendix 242

A: Custom script for analysis of RNA-seq data from polysome profiling experiments 242

B: R script for analysis of differential gene expression using DESeq2 246

C: R script to calculate translation efficiency from read counts..... 248

D: Custom script for analysis of DNA-seq data from CRISPR-edited K562 cells 251

List of Figures

Figure 1.1.	The role of *RPS14* haploinsufficiency in p53 activation and apoptosis of erythroid progenitors of 5q- syndrome patients	28
Figure 1.2.	The role of β-catenin in the Wnt signalling pathway	33
Figure 1.3.	Overview of mTORC1-mediated regulation of mRNA translation	40
Figure 1.4.	The CRISPR/Cas9 system for genome editing of mammalian cells	44
Figure 2.1.	Two-step PCR method for preparation of dual-indexed libraries for DNA sequencing	93
Figure 3.1.	*CSNK1A1* mutations in del(5q) MDS patients	112
Figure 3.2.	Expression of Wnt targets in del(5q) MDS patients with and without *CSNK1A1* mutations	116
Figure 4.1.	Percentage of early erythroid progenitors in samples from patients with MDS with del(5q) and healthy controls.	126
Figure 4.2.	Impaired erythroid differentiation of erythroblasts from patients with MDS with del(5q).	128
Figure 4.3.	Reduced globin levels in del(5q) MDS patient erythroblasts are increased following L-leucine treatment	131
Figure 4.4.	Reduced phosphorylation of mTOR targets in patients is increased following L-leucine treatment of patient-derived erythroblasts.	133
Figure 4.5.	Ribosome Profiling using the TruSeq Ribo Profile Library Prep Kit (Illumina) didn't produce sufficient DNA yield for NGS	136
Figure 4.6.	Optimisation of Ribosome Profiling for analysis of erythroblasts from del(5q) MDS patients	139
Figure 4.7.	GSAA analysis of gene expression in patients with MDS with del(5q)	144
Figure 4.8.	DESeq2 analysis of gene expression in del(5q) MDS patients	150
Figure 4.9.	Ribosomal genes show decreased translation efficient in MDS del(5q) patients.	157
Figure 4.10.	L-leucine treatment affects mRNA translation but not transcription	159
Figure 4.11.	L-leucine increases mRNA translation	163
Figure 4.12.	L-leucine treatment increases translation of 5' TOP and 5' TOP-like mRNAs	167

Figure 4.13.	Translation of 5'TOP and 5'TOP-like mRNAs is decreased in MDS del(5q) and increased by L-leucine treatment	172
Figure 4.14.	Model of the mechanism of L-leucine mediated enhancement of erythroid differentiation in MDS with del(5q)	178
Figure 5.1.	Design of the CRISPR/Cas9 system in K562 cells	182
Figure 5.2.	Restriction digest assay	185
Figure 5.3.	Transfection with ssODN containing the T7 primer binding site	187
Figure 5.4.	Analysis of Colony 1 with correction of the *TP53* mutation	193
Figure 5.5.	HDR-mediated modification of the *TP53* locus using a ssODN containing the T7 sequence	196
Figure 5.6.	HDR-mediated modification of *TP53* using a ssODN containing a barcode of silent mutations	200
Figure 5.7.	Analysis of cellular apoptosis in K562 cells transfected with wild type or mutant ssODNs	201
Figure 5.8.	Transfection of K562 cells with a mix of two ssODNs for HDR-mediated editing	204
Figure 5.9.	Transfection of KBM5 cells with a mix of two ssODNs for HDR-mediated editing of *ASXL1*	207
Figure 5.10.	Variation in the percentage of mutation-corrected cells	210
Figure 6.1.	Contribution of the work presented in this thesis to del(5q) MDS research.	221

List of Tables

Table 2.1.	MDS subtype and karyotype of del(5q) MDS patients screened for *CSNK1A* mutations	47
Table 2.2.	MDS subtype and karyotype of del(5q) MDS patients analysed in the L-leucine study	54
Table 2.3.	Samples from healthy controls analysed in the L-leucine study	54
Table 2.4.	PCR primers	59
Table 2.5.	TaqMan assays (Applied Biosystems)	66
Table 2.6.	Cycling conditions of TaqMan RT-qPCR reactions	67
Table 2.7.	Primers and quantification method of SYBR Green RT-qPCR reactions	68
Table 2.8.	Cycling conditions of SYBR Green RT-qPCR reactions	82
Table 2.9.	Single-guide RNAs (sgRNAs)	83
Table 2.10.	Single-stranded oligodeoxynucleotides (ssODNs)	84
Table 2.11.	List of top five off-target sites for sgRNA 2 sequenced	91
Table 2.12.	Primers for the first PCR for preparation of DNA-seq libraries	95
Table 2.13.	Primers for the second PCR for preparation of DNA-seq libraries	99
Table 2.14.	Cycling conditions of PCR reactions for preparation of DNA-seq libraries	103
Table 2.15.	Barcodes for analysis of CRISPR-edited sequences for *TP53*	106
Table 2.16.	Barcodes for analysis of CRISPR-edited sequences of *ASXL1*	106
Table 3.1.	*CSNK1A1* mutations identified in del(5q) MDS patients	111
Table 3.2.	The average fold change in expression of *CSNK1A1* and *CCND1* measured by GEP	114
Table 3.3.	Cell cycle related gene sets identified by GSEA.	117
Table 4.1.	Karyotypes of del(5q) MDS patients analysed by polysome profiling	148
Table 4.2.	Ribosomal and translation-related gene sets identified by GSAASeqSP analysis	149
Table 4.3.	Expression of ribosomal genes in del(5q) MDS patients	152
Table 4.4.	Top 20 differentially translated known 5'TOP mRNAs in L-leucine treated erythroblasts from del(5q) MDS patients	169
Table 5.1.	Percentage of viable colonies per 96-well plate at day 12 post transfection	189

Table 5.2. Detection of c.406_407insC *TP53* mutation correction in colonies grown from single cells transfected with the CRISPR/Cas9 system 190

Table 5.3. Percentage of viable colonies per 96-well plate at day 14 post transfection in a second experiment 191

Table 5.4. Detection of c.405_406insC *TP53* mutation correction in colonies grown from single cells in a second experiment 191

List of abbreviations

5’ TOP	5' Terminal Oligopyrimidine
AML	Acute Myeloid Leukaemia
BC	Blast Crisis
bp	Basepair
BWA	Burrows-Wheeler Aligner
Cas9	CRISPR-Associated Protein 9
cDNA	Complementary Deoxyribonucleic Acid
CDR	Commonly Deleted Region
CK1α	Casein Kinase 1 Alpha
CML	Chronic Myeloid Leukaemia
CMML	Chronic Myelomonocytic Leukaemia
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4',6-Diamidin-2-Phenylindole, Dihydrochloride
DBA	Diamond–Blackfan Anaemia
DBA	Diamond–Blackfan anaemia
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
DNA-Seq	DNA Sequencing
EDTA	Ethylenediaminetetraacetic Acid
EIF	Eukaryotic Translation Initiation Factor
EPO	Erythropoietin
Exo-I	Exonuclease I
FAB	French American British
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum

FC	Fold Change
FDR	False Discovery Rate
FITC	Fluorescein Isothiocyanate
FMO	Fluorescence Minus One
gDNA	Genomic DNA
GEP	Gene Expression Profiling
GFP	Green Fluorescent Protein
GO	Gene Ontology
GSAA	Gene Set Association Analysis
GSEA	Gene Set Enrichment Analysis
HDR	Homology Directed Repair
hiPSC	Human Induced Pluripotent Stem Cell
HRMA	High Resolution Melting Analysis
HSC	Haematopoietic Stem Cell
IL-3	Interleukin 3
IMDM	Iscove's Modified Dulbecco's Medium
InDA-C	Insert Dependent Adaptor Cleavage
inDels	Insertions And Deletions
MDS	Myelodysplastic Syndrome
MDS-EB	MDS With Excess Blasts
MDS-MLD	MDS With Multilineage Dysplasia
MDS-RS	MDS With Ring Sideroblasts
MDS-SLD	MDS With Single Lineage Dysplasia
MDS-U	MDS-unclassifiable
MEF	Mouse Embryonic Fibroblast
mRNA	Messenger RNA
MSigDB	Molecular Signatures Database
mTOR	Mammalian Target Of Rapamycin

mTOR	Mammalian Target of Rapamycin
mTORC1	Mtor Complex 1
MUT	Mutated
nanoCAGE	nano Cap Analysis Of Gene Expression
NGS	Next Generation Sequencing
NHEJ	Non Homologous End Joining
nt	Nucleotide
NTC	No Template Controls
PAGE	Polyacrylamide Gel Electrophoresis
PAM	Protospacer Adjacent Motif
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PE	Phycoerythrin
PI	Propidium Iodide
PMSF	Phenylmethylsulfonyl Fluoride
PNK	Polynucleotide Kinase
Pol	Polymerase
PolyPhen	Polymorphism Phenotyping
RA	Refractory Anaemia
RAEB	Refractory Anaemia With Excess Of Blasts
RAEB-t	Refractory Anaemia With Excess Of Blasts In Transformation
RBC	Red Blood Cells
RBR	Ribosome-bound RNA
RBS	Roberts Syndrome
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing

RP	Ribosomal Protein
RPF	Ribosome-Protected Mrna Fragments
Rrna	Ribosomal RNA
RS	Ring Sideroblasts
rSAP	Recombinant Shrimp Alkaline Phosphatase
RT	Reverse Transcription
RT-qPCR	Real Time Quantitative Polymerase Chain Reaction
SAM	Sequence Alignment Map
SCF	Stem Cell Factor
SFEM	Serum-Free Expansion Medium
sgRNA	Single Guide RNA
shRNA	Short Hairpin RNA
SIFT	Sorting Intolerant From Tolerant
SNP	Single Nucleotide Polymorphisms
ssODN	Single-Stranded Oligodeoxynucleotides
STAR	Spliced Transcripts Alignment To A Reference
TE	Translation Efficiency
t-MDS	therapy-related MDS
TPM	Transcripts Per Million
TSS	Transcription Start Site
WHO	World Health Organisation
WT	Wild Type

1. Chapter One - Introduction

1.1. The myelodysplastic syndromes

The myelodysplastic syndromes (MDS) are a group of clonal disorders of the haematopoietic stem cell characterized by ineffective haematopoiesis and frequent progression to acute myeloid leukaemia (Bejar and Steensma, 2014, Gangat *et al.*, 2016, Pellagatti and Boultwood, 2015a). Abnormal maturation of bone marrow stem cells is manifested in peripheral blood cytopenias, bone marrow hypercellularity and morphologic dysplasia of bone marrow hematopoietic cells (List, 2008). MDS patients may suffer from anaemia, neutropenia, thrombocytopenia and bone marrow dysplasia which lead to the occurrence of recurrent infections, symptomatic anaemia and bleeding (Bejar and Steensma, 2014). The most prevalent causes of death in MDS patients are bone marrow failure and progression to aggressive AML, which occurs in around 30% of patients (Gangat *et al.*, 2016). The incidence of MDS in the United Kingdom is 4 cases per 100,000 per year and 30 cases per 100,000 per year among patients above 70 years of age (Bejar and Steensma, 2014). The median age at diagnosis is 71 years and diagnosis of MDS is based on the presence of peripheral blood cytopenias, peripheral blood and bone marrow dysplasia, increased marrow blasts and detection of clonal cytogenetic abnormalities (Bejar, 2015). MDS was first classified in 1982 according to the French American British (FAB) criteria, which divided MDS into five subtypes: refractory anaemia (RA), RA with ring sideroblasts (RARS), RA with excess of blasts (RAEB) and RA with excess of blasts in transformation (RAEB-t) and chronic myelomonocytic leukaemia (CMML). RA is characterised by the absence of blasts in the peripheral blood, the presence of less than 5% of blasts in the bone marrow and by anaemia which is resistant to iron treatment (List, 2008). Patients with RARS have the same characteristics of patients with RA but they also present with bone marrow

erythroblasts called ring sideroblasts, characterised by iron-loaded mitochondria which appear as a ring around the erythroblasts nucleus when visualised by Prussian blue staining (Cazzola and Invernizzi, 2011). In the FAB classification of MDS RAEB is defined as RA with less than 5% blasts in the peripheral blood and between 5% and 20% bone marrow blasts (List, 2008). RA and RARS are defined as low-grade MDS subtypes characterised by a longer median survival while RAEB, RAEB-t and CMML are classified as high-grade MDS with a poorer prognosis (Gangat *et al.*, 2016). The World Health Organisation (WHO) classification of myeloid neoplasms and acute leukaemia, first published in 2001 and revised in 2008 and 2016, proposed changes to the FAB classification of MDS. In the 2018 WHO classification, CMML was described as a separate myeloid disorder and the characteristics defining RAEB-t (21%–30% myeloblasts) were included in the definition of AML (Bejar and Steensma, 2014). RAEB was divided into RAEB-1 (MDS-EB-1 in the 2016 WHO classification), which is characterised by less than 5% blasts in peripheral blood and 5% to 9% bone marrow blasts, and RAEB-1 (MDS-EB-2 in the 2016 WHO classification), with 5%–19% blasts in peripheral blood and 10%–19% bone marrow blasts (Gangat *et al.*, 2016, Arber *et al.*, 2016). In the 2016 WHO classification, MDS is divided into six subtypes: MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts (MDS-RS) with either single or multilineage dysplasia, MDS with isolated del(5q), MDS with excess blasts (MDS-EB), subdivided into MDS-EB-1 and MDS-EB-2, and MDS-unclassifiable (MDS-U) (Arber *et al.*, 2016). MDS-SLD is characterised by dysplasia of one bone marrow lineage and associated blood cytopenias (refractory anaemia, refractory neutropenia or refractory thrombocytopenia) and the presence of RS as less than 15% of erythroblasts. MDS-MLD patients have the same characteristics of patients with MDS-SLD but dysplasia of more than one lineage in the bone marrow (Arber *et al.*, 2016). The assessment of prognosis in MDS is based on the Revised

International Prognostic Scoring System (IPSS-R), which uses bone marrow blast percentage, karyotype, and number of cytopenias to estimate survival and risk of AML transformation, assigning patients to five risk categories based on a scoring system (Greenberg *et al.*, 2012). The main cause of MDS development is the age-related accumulation of DNA mutations in haematopoietic stem cells (Papaemmanuil *et al.*, 2013, Haferlach *et al.*, 2014, Sperling *et al.*, 2017). Exposure to chemicals such as benzene, insecticides, and pesticides might also contribute to MDS development (Gangat *et al.*, 2016). Moreover, therapy-related MDS (t-MDS) is caused by exposure to radiation or chemotherapy and is characterised by a poor prognosis (Abou Zahr *et al.*, 2017).

1.2. The molecular pathogenesis of MDS

Several studies have identified numerous genes recurrently mutated in MDS, which underlie important mechanisms in MDS pathogenesis and could help evaluation of disease prognosis (Pellagatti and Boultwood, 2015a). Recurrently mutated genes in MDS include genes coding for splicing factors, epigenetic regulators, transcription factors and genes involved in signal transduction (Pellagatti and Boultwood, 2015a). Mutations in splicing factor genes are frequent in MDS (Boultwood *et al.*, 2014) and in a study by Yoshida *et al.* (2011) mutations in *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* have been identified in 57% of MDS patients analysed (130/228 patients). Mutations in these genes are almost often gain of function mutations which affect pre-mRNA processing and splicing of several mRNA targets (Boultwood *et al.*, 2014). *SF3B1* mutations are highly recurrent in MDS and they are associated with the presence of bone marrow ring sideroblasts (Kennedy and Ebert, 2017). *SF3B1* mutations are frequently found in patients with RARS and RCMD with RS (RCMD-RS) (Malcovati *et al.*, 2011, Papaemmanuil *et al.*, 2011) and the presence of *SF3B1*

mutations has been included in the diagnostic criteria of MDS with RS in the revised version of the WHO classification of myeloid neoplasms and acute leukaemia (Arber *et al.*, 2016).

MDS is also characterised by mutations in epigenetic regulators, which are involved in control of DNA methylation, such as *DNMT3A*, *IDH1*/*IDH2*,and *TET2* and of histone function, such as *EZH2*, *ASXL1* and *UTX* (Pellagatti and Boultwood, 2015a). Mutations in these genes cause epigenetic abnormalities such as the hypermethylation and silencing of genes that control proliferation, adhesion and other phenotypes characteristic of MDS (Pellagatti and Boultwood, 2015a). The drugs azacitidine and decitabine, commonly used for MDS treatment, are hypomethylating agents which increase expression of genes silenced by promoter hypermethylation (Pellagatti and Boultwood, 2015a).

ASXL1 encodes a protein involved in regulation of gene repression by DNA methylation, via recruitment of the polycomb repressive complex 2 (PRC2) and binding of the deubiquitinase BAP1 (Abdel-Wahab *et al.*, 2012). Mutations of *ASXL1* are found in 11 to 15% of MDS patients, they are recurrent in patients with AML and CMML and they are frequently observed in patients with MDS and CMML that transform to AML (Boultwood *et al.*, 2010b, Gelsi-Boyer *et al.*, 2010). *ASXL1* mutations are often nonsense mutations which result is lack of ASXL1 expression (Pellagatti and Boultwood, 2015a). Mutations of *ASXL1* have been found to cause loss of PCR2-mediated methylation and repression of genes important in leukaemic transformation such as *HOXA* genes (Abdel-Wahab *et al.*, 2012). *ASXL1* mutations in MDS have been correlated with poor prognosis and they have been found to collaborate with mutations in other genes, such as *NRAS*, to cause myeloid leukaemogenesis (Pellagatti and Boultwood, 2015a, Abdel-Wahab *et al.*, 2012). Deletion of *Asxl1* in HSCs in mice results in multilineage cytopenias, bone marrow dysplasia and increased numbers of HSCs (Abdel-Wahab *et al.*, 2013). Moreover, HSCs with knockdown of *ASXL1* show impaired myeloid differentiation (Davies *et al.*, 2013), suggesting that

ASXL1 plays an important role in pathogenesis of MDS. However, studies using *Asxl1* knockout mice showed that mice with loss of *ASXL1* presented with a mild phenotype with differentiation defects in the myeloid and lymphoid progenitors and did not progress to a more severe MDS or AML-like disease state (Fisher *et al.*, 2010). Furthermore, mutations of *ASXL1* are found in healthy elderly individuals which exhibit clonal haematopoiesis but show normal blood counts, indicating that *ASXL1* mutations are likely to cause clonal haematopoietic expansion in MDS but unlikely to be oncogenic causative events in leukaemia (Zink *et al.*, 2017, Genovese *et al.*, 2014). A knock-in mouse model of a highly recurrent mutation in exon 12 of *ASXL1* has been recently generated (Hsu *et al.*, 2017). HSCs from heterozygous *Asxl1* mutant mice showed increased proliferation capacity *in vitro* (Hsu *et al.*, 2017). However, *Asxl1* mutant mice did not develop blood disorders, suggesting that *Asxl1* mutation alone was not sufficient for leukaemogenesis (Hsu *et al.*, 2017).

Mutations in mediators of cell signalling, such as *NRAS* and *JAK2* and transcription factors, such as *RUNX1* and *TP53* are also frequent in MDS (Pellagatti and Boultwood, 2015a).

The current model of pathogenesis of MDS involves the acquisition of initiating mutations, such as *ASXL1* and *TET2* mutations, which lead to asymptomatic clonal expansion of HSCs (Kennedy and Ebert, 2017). Clonal expansion is followed by occurrence of further mutations in the expanded clone, such as *TP53* mutations, which cause development of the MDS or secondary AML phenotype (Kennedy and Ebert, 2017).

Studies using gene expression profiling have identified deregulated genes in MDS which play a role in disease pathogenesis (Pellagatti *et al.*, 2006, Pellagatti and Boultwood, 2015a). Several studies have identified gene signatures in MDS which were found to correlate with patient prognosis, suggesting that GEP (Gene Expression Profiling) could aid prognostic evaluation of MDS patients (Pellagatti and Boultwood, 2015a). The accuracy of GEP in

prediction of clinical subtypes and progression in MDS and acute and chronic leukaemia was tested in a study by Haferlach *et al.* (2010). GEP successfully predicted leukaemic subtypes with 93% accuracy, while accuracy of classification of MDS patients was of only 50% (Haferlach *et al.*, 2010). These data suggest that compared to AML, gene expression in MDS is highly heterogeneous (Haferlach *et al.*, 2010). In a later study, a classification model for discrimination of MDS patients with rapid transformation to AML from patients with stable disease has been developed (Mills *et al.*, 2009). A study by Pellagatti *et al.* (2013) identified a gene signature based on the expression of 20 genes which outperformed other predictors of patient prognosis in MDS. The interconnections between gene mutations, GEP and patient outcome in MDS were also investigated (Gerstung *et al.*, 2015). It was found that the transcriptome was the most powerful predictor of outcome in MDS and, therefore, the accuracy of outcome prediction in MDS may be improved by incorporation of GEP into MDS prognostic scoring system (Gerstung *et al.*, 2015).

Nearly 50% of patients with *de novo* MDS and 80% of patients with therapy-related MDS harbour cytogenetic abnormalities (Pellagatti and Boultwood, 2015a). Deletion 5q, monosomy 7 and deletion 7q are the most common cytogenetic abnormalities found in MDS patients as well as trisomy 8 and deletion 20q (Gangat *et al.*, 2016). Cytogenetic abnormalities in MDS have been correlated with prognosis and are reflected in the IPSS-R system which defines five cytogenetic groups in MDS (Greenberg *et al.*, 2012). Moreover, interstitial deletions of chromosome 5q define one of the MDS subtypes according to the MDS WHO classification and are correlated with good prognosis (Arber *et al.*, 2016).

1.3. Treatment of MDS

While a few other therapeutic approaches have been developed, the only curative treatment for MDS is currently allogeneic haematopoietic stem cell transplant (HSCT) (Bejar and

Steensma, 2014, Zeidan *et al.*, 2013). In most cases, supportive care, such as antibiotic treatment for infections and blood transfusions, is provided to control disease symptoms and improve quality of life (Bejar and Steensma, 2014). Erythropoiesis stimulating agents, such as erythropoietin (EPO) and granulocyte colony-stimulating factor (G-CSF), have shown some efficacy in amelioration of anaemia, which occurs in over 90% of MDS patients (Gangat *et al.*, 2016). Three drugs, azacitidine, decitabine and lenalidomide, are used for the treatment of MDS but they have shown limited effect on patient survival (Gangat *et al.*, 2016, Zeidan *et al.*, 2013). Azacitidine and decitabine are DNA hypomethylating agents used for the treatment of high-risk MDS and the can also be utilised in low-risk MDS when treatment with other drugs has failed (Bejar and Steensma, 2014). These compounds remove methylation marks and reverse transcriptional inhibition of hypermethylated genes involved in control of cell proliferation and DNA-repair in MDS (Boultwood and Wainscoat, 2007, Issa, 2010, Itzykson and Fenaux, 2014). The thalidomide-analog lenalidomide is currently the preferred treatment for MDS with isolated del(5q) (Zeidan *et al.*, 2013). Immunosuppressive agents such as cyclosporin have also shown some efficacy in the treatment of low-risk MDS. These therapies are known to target the immune activation underlying MDS but their exact mechanism of action requires further investigation (Glenthoj *et al.*, 2016, Zeidan *et al.*, 2013). Finally, patients with high-risk MDS can also be treated with chemotherapeutic agents normally used in AML treatment, such as daunorubicin and cytarabine, as well as receiving HSCT. Overall, however, the response to disease-modifying therapies in MDS is low and not durable and supportive care remains the main approach in the management of MDS patients (Bejar and Steensma, 2014). Moreover, given the age of onset of the disease, HSCT is not always a feasible approach for the cure of elderly patients. Therefore, new therapeutic approaches and curative treatments are needed.

1.4. MDS with isolated del(5q) and 5q- syndrome

Interstitial deletion within the long arm of chromosome 5 [del(5q)] is found in 10-20% of patients with MDS (Bernasconi *et al.*, 2005, Zeidan *et al.*, 2013). MDS with del(5q) is a pathologically distinct subtype of low-risk MDS characterised by refractory hypoplastic anaemia (Komrokji *et al.*, 2013, Zeidan *et al.*, 2013). The 5q− syndrome is defined by the occurrence of refractory anaemia, the presence of a bone marrow blast count of less than 5% and the del(5q) as the sole karyotypic abnormality (Boultwood *et al.*, 2002). Del(5q) MDS patients, instead, harbour del(5q) deletions but may also present with excess blast counts and one additional cytogenetic changes besides the del(5q), with that abnormality being any change but monosomy 7 or del(7q) (Arber *et al.*, 2016). 5q- syndrome patients are around 5% of patients with MDS (Gangat *et al.*, 2016). The clinical features commonly found in patients with the 5q- syndrome are macrocytic anaemia, hypolobulated megakaryocytes, a normal or high platelet count and a good prognosis (Giagounidis *et al.*, 2004, Pellagatti and Boultwood, 2015b). Approximately 10% of patients transforms to AML (Boultwood *et al.*, 2010a, Giagounidis *et al.*, 2004).

1.5. The molecular pathogenesis of 5q- syndrome and MDS with del(5q)

1.5.1. The commonly deleted region

The 5q- syndrome was first described by Van den Berghe *et al.* (1974) and mapping of the commonly deleted region (CDR) of the 5q- syndrome has provided important insights into the molecular pathogenesis of this MDS subtype (Boultwood *et al.*, 2002, Boultwood *et al.*, 1994). The CDR at 5q, which characterises both 5q- syndrome and del(5q) MDS, comprises a ~1.5Mb interval at 5q32- q33 flanked by the DNA marker D5S413 and the *GLRA1* gene (Boultwood *et al.*, 2002, Jaju *et al.*, 1998). Thirty three genes expressed in haematopoietic stem cells (HSCs) are located in the CDR of 5q- syndrome and loss of these genes has been

found affect cellular processes important for normal haematopoiesis (Boultwood *et al.*, 1994). Genes in CDR of 5q- syndrome include the tumour suppressor gene *SPARC*, the ribosomal protein S14 gene (*RPS14)*, encoding a component of the 40S ribosomal subunit, the tumour suppressor casein kinase 1 alpha 1 (*CSNK1A1*) and several microRNA genes (Boultwood *et al.*, 2002, Boultwood *et al.*, 2007).

1.5.2. Haploinsufficiency of *RPS14*

Eukaryotic ribosomes are composed of two subunits, the small 40S subunit and the large 60S subunit, consisting of four ribosomal RNA (rRNA) species and approximately 80 ribosomal proteins (RPs) (de la Cruz *et al.*, 2015). *RPS14* is located on chromosome 5q33.1 and encodes the ribosomal protein S14, a component of the 40S ribosomal subunit (de la Cruz *et al.*, 2015, Kressler *et al.*, 2010). A study using RNA-mediated interference (RNAi) in HSCs showed that haploinsufficiency of the ribosomal protein gene *RPS14* underlies the erythroid defect found in the 5q- syndrome (Ebert *et al.*, 2008). HSCs with knockdown of *RPS14* to haploinsufficient levels exhibit a block in erythroid differentiation with relative preservation of megakaryocytic differentiation (Ebert *et al.*, 2008). The erythroid defect could be rescued by overexpression of *RPS14* in HSCs derived from 5q- syndrome patients (Ebert *et al.*, 2008). Moreover, *RPS14* knockdown experiments in the leukaemia cell line TF-1 showed that *RPS14* haploinsufficiency leads to a defect in pre-rRNA processing and decreased formation of the 40S ribosomal subunit (Ebert *et al.*, 2008). Short hairpin RNA (shRNA)-mediated knockdown of *RPS14* in HSCs results in a significant decrease of total protein production (Yip *et al.*, 2012) Patients with the 5q− syndrome also show decreased expression of multiple genes involved in ribosome biogenesis and control of mRNA translation (Pellagatti *et al.*, 2008). Taken together evidence from these studies suggests that the 5q- syndrome is a ribosomopathy, a disorder characterised by aberrant ribosome biogenesis (Narla and Ebert, 2010). Erythroid progenitor cells are particularly sensitive to

this defect in ribosome biogenesis and to the impairment of protein production, possibly as a consequence of the rapid cell division and massive translational demand for globin synthesis found in these cells (Dutt *et al.*, 2011, Yip *et al.*, 2012). Therefore, *RPS14* haploinsufficiency and impaired ribosome biogenesis in erythroid progenitors of patients with 5q- syndrome and MDS with del(5q) are considered important causes of the anaemia observed in these patients (Dutt *et al.*, 2011, Pellagatti and Boultwood, 2015b, Yip *et al.*, 2012) (Figure 1.1).

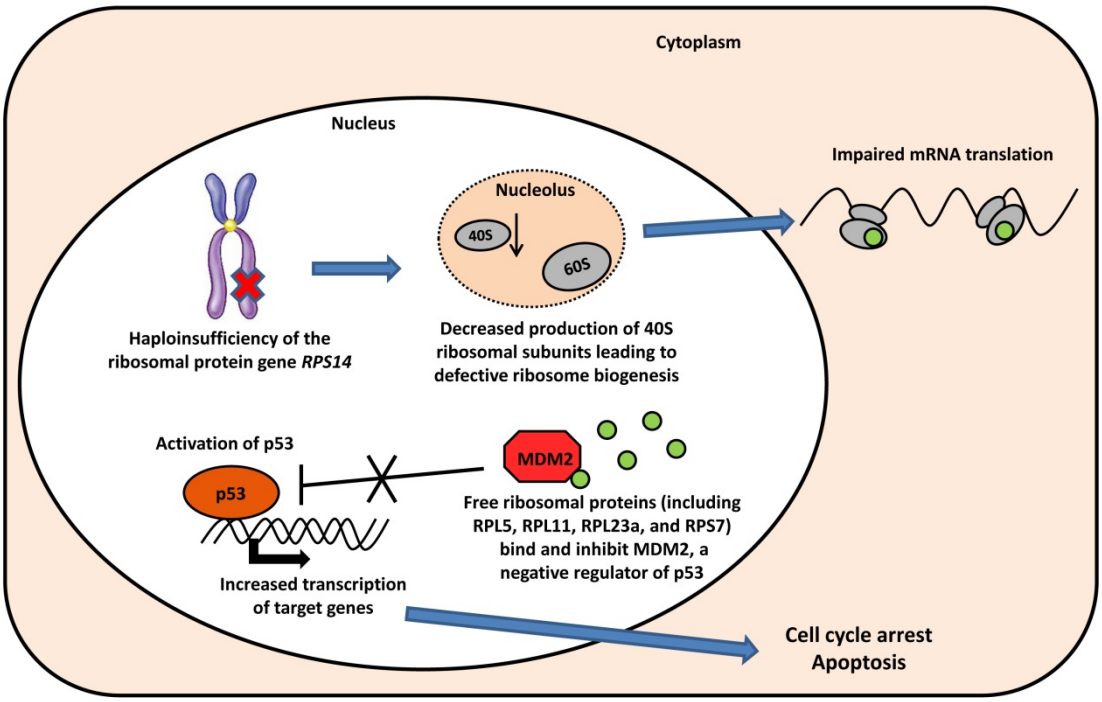


Figure 1.1. The role of RPS14 haploinsufficiency in p53 activation and apoptosis of erythroid progenitors of 5q- syndrome patients. Figure from (Pellagatti and Boultwood, 2015a).

1.5.3. Role of p53 in pathogenesis of MDS with del(5q)

1.5.3.1. The tumour protein 53

The tumour protein 53 (p53) regulates a variety of cellular processes, such as induction of apoptotic pathways in response to DNA damage and cellular stress, regulation of cell growth arrest and regulation of cell cycle progression (Vousden and Lane, 2007). *TP53* mutations are amongst the most recurrent mutations in cancer (Muller and Vousden, 2013). Loss-of-function mutations of *TP53* cause promotion of cell cycle progression, decreased apoptosis, decreased induction of senescence and increased cell proliferation (Bieging *et al.*, 2014). This evidence initially pointed towards a role for *TP53* as a tumour suppressor gene in cancer. However, gain-of-function *TP53* mutations have been found to have an oncogenic role in a variety of cancers (Muller and Vousden, 2013, Bieging *et al.*, 2014, Freed-Pastor and Prives, 2012). Mutant forms of p53 can induce upregulated expression of genes involved in cancer progression, such as *NFKB2*, which regulates anti-apoptotic pathways, and growth factor *EGFR*, which promotes cell division (Freed-Pastor and Prives, 2012). Moreover, mutant forms of p53 can bind to several proteins, such as p53 target proteins p63 and p73, inhibit their activity of inducers of apoptotic pathways (Bieging *et al.*, 2014). Therefore *TP53* is considered to be a tumour suppressor gene with oncogenic functions (Soussi and Wiman, 2015, Vousden and Prives, 2009).

1.5.3.2. Activation of p53 in MDS with del(5q)

The molecular pathogenesis of the 5q− syndrome is also characterised by a p53-dependent mechanism (Figure 1.1). In a study by Barlow *et al.* (2010) a mouse model of the 5q- syndrome was generated by deletion of the *Cd74–Nid67* region, which includes *Rps14*. Mice harbouring this deletion presented with phenotypes associated with 5q- syndrome such as

macrocytic anaemia and hypolobulated megakaryocytes in the bone marrow (Barlow *et al.*, 2010). Moreover, mice with deletion of the *Cd74–Nid67* region showed accumulation of p53 protein, increased apoptosis and defective development of bone marrow cells (Barlow *et al.*, 2010). Evidence that p53 activation plays an important role in the pathogenesis of the 5q- syndrome was provided by the fact that crossing of these mice with p53-deficient mice rescued the deficit in development of bone marrow progenitor cells (Barlow *et al.*, 2010). Another mouse model of 5q- syndrome was generated by conditional heterozygous knockout of *Rps14* (Schneider *et al.*, 2014a). *Rps14*-deficient mice were found to exhibit reduced haemoglobin levels, lower red blood cell counts as well as an erythroid differentiation defect, with the presence of hypolobulated megakaryocytes in the bone marrow (Schneider *et al.*, 2014a). These data show that *Rps14* is sufficient to cause the erythroid defect which characterises 5q- syndrome. Moreover, the erythroid defect could be rescued by heterozygous knockout of p53 in *Rps14* -/+ mice, confirming that p53 activation plays a fundamental role in the pathogenesis of 5q- syndrome (Schneider *et al.*, 2014a). Strong expression of p53 has also been found in erythroid cells from 5q- syndrome patients in a study using immunohistochemistry techniques and the p53 pathway was found to be significantly upregulated in a gene expression profiling study on HSCs of 5q- syndrome patients (Pellagatti *et al.*, 2010b). Other studies have confirmed that erythroid precursors derived from patients with 5q- syndrome show accumulation of p53 (Dutt *et al.*, 2011, Wei *et al.*, 2013). Moreover, shRNA-mediated knockdown of *RPS14* in human bone marrow HSCs resulted in increased levels of p53 (Caceres *et al.*, 2013, Dutt *et al.*, 2011) and p53 targets p21 (Dutt *et al.*, 2011) and PUMA (Caceres *et al.*, 2013). P53 activation resulted in cell cycle arrest and apoptosis of *RPS14*-deficient erythroid precursors, which decreased following treatment with the p53 inhibitor pifithrin-α (Dutt *et al.*, 2011). Evidence from several studies suggests that defective ribosome biogenesis leads to p53 activation (Narla

and Ebert, 2010, Narla *et al.*, 2011, Zhou *et al.*, 2012). Several RPs, such as RPL5, RPL11, RPL23a and RPL7 have been shown to bind the p53 inhibitor protein mouse double minute 2 homolog (MDM2) and inhibit MDM2 E3 ubiquitin ligase activity toward p53, resulting in p53 activation (Dai and Lu, 2004, Dai *et al.*, 2004, Zhou *et al.*, 2012, Chen *et al.*, 2007). Overexpression of RPL11, RPL5 or RPL23 in HeLa cells induces activation of p53 and p21 and results in cell cycle arrest and apoptosis (Dai and Lu, 2004, Dai *et al.*, 2004). Ablation of RPS14 by siRNA has been found to increase levels of RPL11 and RPL5 and induce p53 activation in the lung cancer cell line H1299 (Zhou *et al.*, 2013). Consistent with these findings, *RPS14* knockdown in HSCs results in increased levels of RPL11 and p53 activation (Dutt *et al.*, 2011). Taken together this evidence demonstrates that defective ribosome biogenesis and ribosomal stress results in activation of the p53 pathway and leads to the erythroid defect in MDS with del(5q) and 5q- syndrome (Narla *et al.*, 2011, Pellagatti and Boultwood, 2015b) (Figure 1.1).

1.5.4. *TP53* mutations

TP53 mutations have been identified in 18% of patients with early MDS and del(5q) (Jädersten *et al.*, 2011) and have been correlated with poor prognosis and increased risk of AML development (Sebaa *et al.*, 2012, Boultwood *et al.*, 2010a). Moreover, evidence suggests that *TP53* mutations in del(5q) MDS confer resistance to treatment with the thalidomide analogue lenalidomide (Jädersten *et al.*, 2009, Mallo *et al.*, 2013). Inactivation of p53 may be one of the molecular mechanisms underlying clonal progression of 5q-syndrome to AML (Jädersten *et al.*, 2011). Therefore, information on the mutational status of TP53 may help refine prognosis in MDS del(5q) patients (Jädersten *et al.*, 2011).

1.5.5. The role of *CSNK1A1* in the pathogenesis of MDS with del(5q)

1.5.5.1. The Wnt/β catenin pathway

The *CSNK1A1* gene maps to the CDR of the 5q- syndrome at 5q32 (Boultwood *et al.*, 2002) and it encodes the serine/threonine kinase Casein Kinase 1 alpha (CK1α), which has a regulatory role in several cellular processes, including the Wnt/β-catenin pathway, p53 signalling, cell cycle control and apoptosis (Knippschild *et al.*, 2014, Schittek and Sinnberg, 2014). In the Wnt/β-catenin pathway, CK1α is part of the β-catenin destruction complex, which includes Axin, Adenomatous Polyposis Coli (APC), Glycogen Synthase Kinase 3 beta (GSK3β) and several other CK1 isoforms (Cheong and Virshup, 2011). In the absence of the Wnt ligand, CK1α interacts with and phosphorylates Axin, APC and β-catenin (at Ser-45), priming β-catenin for further phosphorylation by GSK3β and subsequent ubiquitination and degradation (Knippschild *et al.*, 2014) (Figure 1.2A). Therefore, down-regulation of CK1α leads to an accumulation of cytoplasmic β-catenin (Schittek and Sinnberg, 2014). In the presence of the Wnt ligand, the β-catenin destruction complex is inhibited and cytoplasmic levels of β-catenin rise, leading to β-catenin translocation into the nucleus (Knippschild *et al.*, 2014) (Figure 1.2B). β-catenin then associates with transcription factors and activates transcription of genes involved in cell cycle progression, cell proliferation and differentiation processes (Schittek and Sinnberg, 2014) (Figure 1.2B). *CCND1* (encoding cyclin D1) is an important transcriptional target of β-catenin, involved in regulation of the G1/S cell cycle checkpoint. Growth inhibitory signals inhibit cyclin D through p21 and p27 accumulation, thereby leading to entry into cellular quiescence (Alao, 2007).

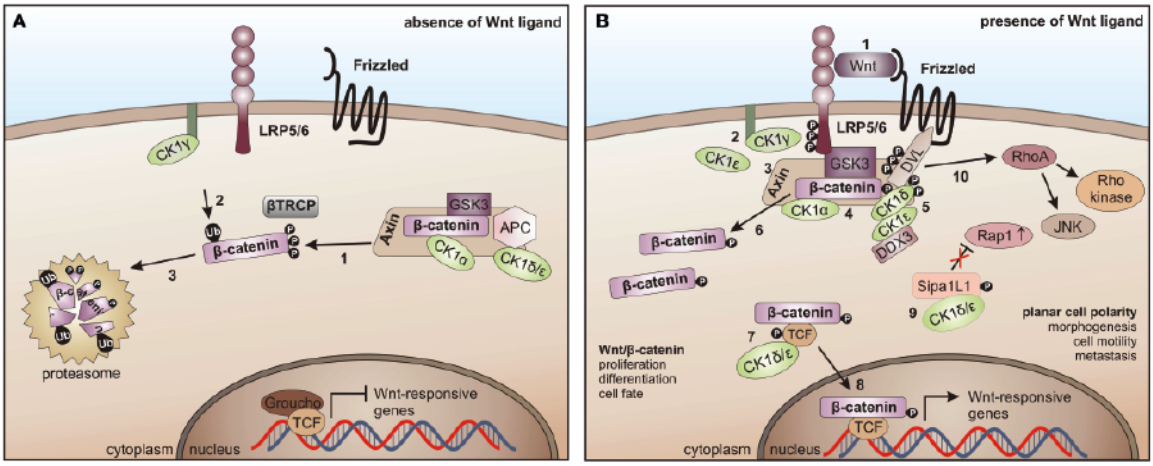


Figure 1.2. The role of β-catenin in the Wnt signalling pathway. A) In the absence of Wnt signalling, β-catenin is ubiquitinated and degraded. B) When the Wnt pathway is activated, β-catenin accumulates and drives transcription of genes involved in promotion of cell proliferation and differentiation. Figure from (Knippschild *et al.*, 2014).

Mutations in various components of the Wnt pathway have been found in a number of human cancers (Schittek and Sinnberg, 2014) and the overexpression of *CCND1* has been linked to the development and progression of several cancers, including those of the breast and colon (Lin *et al.*, 2000, Tetsu and McCormick, 1999). CK1α also plays a regulatory role in the p53 pathway through the interaction with the p53 inhibitor proteins MDM2 and MDMX (Schittek and Sinnberg, 2014). The activation of p53 is triggered by knockdown or downregulation of CK1α in the intestinal epithelium of mice (Elyada *et al.*, 2011), in human colon cancers (Fariña Sarasqueta *et al.*, 2013) and in murine models of acute myeloid leukaemia (Järås *et al.*, 2014). CK1α has been found to phosphorylate MDMX (Ser-289) and cooperate with MDMX to inhibit p53 (Wu *et al.*, 2012). Furthermore, knockdown of CK1α reduces the interaction of MDM2 with p53, thus increasing p53 activity (Huart *et al.*, 2009).

1.5.5.2. *CSNK1A1* haploinsufficiency in MDS with del(5q)

CSNK1A1 was found to be haploinsufficient in the $CD34^+$ cells of 5q- syndrome patients in a gene expression profiling study (Boultwood *et al.*, 2007). *CSNK1A1* expression was found

to be reduced by approximately 50% showing that in this setting the remaining copy of the gene does not compensate the loss of the other allele. *CSNK1A1* has been shown to be a tumour suppressor gene which controls cell proliferation through regulation of β-catenin and the p53 pathway in melanoma and colon cancer (Elyada *et al.*, 2011, Sinnberg *et al.*, 2010). In mice, homozygous deletion of *Csnk1a1* in HSCs results in accumulation of p53 and p21, bone marrow failure and mice lethality, suggesting that *Csnk1a1* function is essential for haematopoiesis (Schneider *et al.*, 2014b). Instead, heterozygous inactivation of *Csnk1a1* leads to β-catenin activation and expansion of murine HSCs, suggesting that *CSNK1A1* haploinsufficiency may be the mechanism underlying the initial clonal expansion in patients with the 5q- syndrome (Schneider *et al.*, 2014b). Moreover, inhibition of CK1α in *CSNK1A1*-haploinsufficient murine myeloid progenitor cells causes targeted depletion of *CSNK1A1*-haploinsufficient cells (Schneider *et al.*, 2014b), suggesting an important role of *CSNK1A1* in the pathogenesis of 5q- syndrome and MDS with del(5q).

1.5.5.3. *CSNK1A1* mutations in 5q- syndrome and MDS with del(5q)

Screening of the remaining copy of all 40 genes within the CDR by Sanger sequencing did not identify any mutations in a cohort of ten 5q- syndrome patients (Boultwood *et al.*, 2007). However, in a study by Schneider *et al*. (2014), mutations of *CSNK1A1* were identified using whole-exome sequencing in 2 of 19 del(5q) cases and a further *CSNK1A1* mutation was found in an additional cohort of 22 MDS cases with isolated del(5q) using next-generation targeted sequencing. *CSNK1A1* mutations were reported as rare events in del(5q) MDS, with an overall frequency of approximately 7% (Schneider *et al.*, 2014b). Since then, a number of studies have reported mutation in *CSNK1A1* in cohorts of patients with 5q- syndrome and

MDS with del(5q) (Graubert *et al.*, 2012, Heuser *et al.*, 2015, Hosono *et al.*, 2017, Meggendorfer *et al.*, 2017, Smith *et al.*, 2015).

1.5.5.4. *CSNK1A1* and the mechanism of action of lenalidomide

It has been reported that lenalidomide, a drug widely used to treat del(5q) MDS (List *et al.*, 2006b), induces the ubiquitination and consequent degradation of CK1α by the E3 ubiquitin ligase CUL4–RBX1–DDB1–CRBN (CRL4CRBN), and that haploinsufficiency of *CSNK1A1* increases lenalidomide sensitivity in del(5q) hematopoietic cells (Fink *et al.*, 2014, Kronke *et al.*, 2015). Lenalidomide modulates the CRL4-CRBN complex and induces ubiquitination of CK1α by CRL4-CRBN, resulting in CK1α degradation (Fink *et al.*, 2014, Kronke *et al.*, 2015, Hollenbach *et al.*, 2014). Binding of CRL4-CRBN to CK1α is dependent on the presence of lenalidomide (Petzold *et al.*, 2016), and studies using *CSNK1A1* shRNA-mediated knockdown in $CD34^+$ cells showed that reduced levels of *CSNK1A1* sensitise haematopoietic cells to lenalidomide (Kronke *et al.*, 2015). Consistent with the role of CK1α in the regulation of the Wnt and p53 pathways, increased levels of β-catenin were observed after lenalidomide treatment of KG-1 cells, together with an increase in expression of the p53 target p21 and increased levels of apoptosis (Kronke *et al.*, 2015). Gain of function mutations of *CSNK1A1* might play a role in the mechanism of resistance to lenalidomide observed in del(5q) MDS patients with *CSNK1A1* mutations (Smith *et al.*, 2015). This hypothesis is supported by the fact that overexpression of *CSNK1A1* in $CD34^+$ cells from del(5q) MDS patients reduces lenalidomide sensitivity (Kronke *et al.*, 2015).

1.5.6. Other important genetic events in MDS with isolated del(5q)

Evidence from several studies suggest that *RPS14* haploinsufficiency underlies the pathogenesis of MDS with isolated del(5q) and cause the erythroid defect and anaemia characteristic of this disease. However, loss of other genes located in the CDR may

contribute to disease pathogenesis (Pellagatti and Boultwood, 2015b). Haploinsufficiency of the miRNA genes miR-145 and miR146a may cause the development of hypolobulated megakaryocytes and thrombocytosis found in patients with del(5q) MDS (Starczynowski *et al.*, 2010). Moreover, haploinsufficiency of other genes located in the CDR, such as SPARC may contribute to expansion of the haploinsufficient clone (Pellagatti *et al.*, 2007). Moreover, a portion of patients with del(5q) MDS and 5q- syndrome harbour various genetic mutations that may define disease development and patient prognosis (Pellagatti and Boultwood, 2015b). In a recent study by Meggendorfer *et al.* (2017) a cohort of 123 patients with MDS with del(5q) was analysed and seven genes were found to have a mutation frequency higher than 5%. These genes were *SF3B1*, *DNMT3A*, *TP53*, *TET2*, *CSNK1A1*, *ASXL1* and *JAK2*.

1.6. Therapeutic approaches for MDS with isolated del(5q)

1.6.1. Lenalidomide

Lenalidomide is currently the most commonly used treatment for MDS with isolated del(5q) (Zeidan *et al.*, 2013). Lenalidomide was shown to be highly effective in the treatment of MDS with del(5q) and 5q- syndrome and it is currently the most commonly used drug for treatment of these MDS subtypes (Zeidan *et al.*, 2013). Treatment of del(5q) MDS patients with lenalidomide results in transfusion independence, suppression of the 5q- clone and improvement in haemoglobin levels (List *et al.*, 2006a, Giagounidis *et al.*, 2014). However, not all patients respond to lenalidomide therapy and approximately 50% of del(5q) MDS patients acquire resistance to lenalidomide within two years (Bejar and Steensma, 2014). Some but not all studies have also raised safety concerns for the use of lenalidomide in del(5q) MDS patients, due to an increased risk of disease progression to AML (Boultwood

et al., 2010b, Jädersten, 2010). Therefore there is a clinical need for novel treatments for MDS with the del(5q).

1.6.2. Cenersen

The pathogenesis of 5q- syndrome and MDS with del(5q) is characterised by p53 activation resulting in increased apoptosis of HSCs (Pellagatti and Boultwood, 2015b). It has been shown that down-regulation of p53 by the antisense oligonucleotide cenersen enhances erythropoiesis in del(5q) MDS (Caceres *et al.*, 2013). Treatment of cultured erythroblasts derived from CD34+ cells of del(5q) MDS patients with cenersen increases cell proliferation and erythroid differentiation (Caceres *et al.*, 2013). Safety issues concerning abrogation of the tumour suppressor p53 should be considered but cenersen could be used as a therapeutic approach since p53 function is only transiently abrogated (Pellagatti and Boultwood, 2015b).

1.6.3. L-leucine

1.6.3.1. Evidence from animal and cellular models of MDS with del(5q) and other ribosomopathies

The 5q- syndrome and MDS with del(5q) are characterised by a defect in ribosome biogenesis (Ebert *et al.*, 2008) and decreased protein production (Yip *et al.*, 2012). Like in other ribosomopathies, defective ribosome biogenesis may lead to decreased mRNA translation, which could constitute a potential target for therapeutic intervention (Pellagatti and Boultwood, 2015b). Treatment with the translation enhancer L-leucine has been shown to have some efficacy in ribosomopathies, such as Diamond–Blackfan anaemia (DBA), as well as in 5q- syndrome. DBA is a disorder characterized by erythroid failure, congenital anomalies, a predisposition to cancer and haploinsufficiency of ribosomal genes, such as

RPS19 (Ito *et al.*, 2010). In a study by Pospisilova *et al.* (2007) a DBA patient was treated with L-leucine and showed a marked improvement in anaemia and reached transfusion independence (Pospisilova *et al.*, 2007). Roberts syndrome (RBS) is a developmental disorder caused by mutations in the cohesin acetyltransferase ESCO2 characterised by decreased mRNA translation (Xu *et al.*, 2013). L-leucine treatment of cellular and zebrafish models of RBS has been found to result in increased cell division and improved development (Xu *et al.*, 2016, Xu *et al.*, 2013). Moreover, studies using zebrafish models of DBA, with *Rps19* haploinsufficiency (Payne *et al.*, 2012, Zhang *et al.*, 2014), and MDS with del(5q), with *Rps14* deficiency (Payne *et al.*, 2012, Ear *et al.*, 2016), found that L-leucine treatment improved haemoglobinization and red cell counts. A study by Jaako *et al.* (2012) using a mouse model of DBA also showed that haemoglobin levels and number of erythrocytes improved as a consequence of treatment with L-leucine. *Ex vivo* treatment with L-leucine of erythroblasts with *RPS14* knockdown to haploinsufficient levels was found to increase cell proliferation, erythroid differentiation and mRNA translation (Yip *et al.*, 2012). In a recent study three MDS patients with isolated del(5q) were treated with L-leucine and showed no adverse effects during treatment (Steensma and Ebert, 2013). However, none of the patients showed improvements in peripheral blood cytopenia or transfusion requirements (Steensma and Ebert, 2013). Clinical trials for the therapeutic use of L-leucine in DBA are underway in the US and in Russia (NCT02386267 and NCT01362595, www.clinicaltrials.gov) (Pellagatti and Boultwood, 2015b).

1.6.3.2. The mTOR pathway in regulation of mRNA translation

L-leucine is a branched chain amino acid and it is both a substrate for protein synthesis and a nutrient that regulates protein metabolism (Dodd and Tee, 2012). L-leucine regulates mRNA translation, ribosome biogenesis and protein synthesis through the activation of the

Mammalian Target of Rapamycin (mTOR) pathway (Boultwood *et al.*, 2013). The mTOR signalling pathway regulates cell growth and proliferation in response to signals from nutrients, growth factors and cellular energy status (Hay and Sonenberg, 2004). Two main mTOR signalling proteins, mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2), are phosphorylated and activated in response to signalling from cellular receptors for nutrients, such as glucose, amino acids and growth factors (Hay and Sonenberg, 2004). mTORC2 is involved in regulation of cell metabolism and cytoskeletal organisation, while mTORC1 controls mRNA translation, protein synthesis and cell growth (Saxton and Sabatini, 2017). For regulation of protein synthesis, activated mTORC1 phosphorylates and activates two key targets, ribosomal protein S6 kinase beta-1 (S6K1) and eukaryotic initiation factor 4E-binding protein 1 (4EBP1) (Sengupta *et al.*, 2010) as well as proteins involved in translation initiation (Figure 1.3). Activated S6K1 and 4EBP1 phosphorylate translation initiation factors, translation elongation factors and the ribosomal protein 6 (RPS6) and promote mRNA translation (Fonseca *et al.*, 2014). In the absence of mTOR signalling, 4EBP1 binds and inhibits the initiation factor eIF4E and prevents formation of the translation initiation complex eIF4F (Fonseca *et al.*, 2014). When mTORC1 is activated and phosphorylates 4EBP1, eIF4E is released and assembles on the 5' cap of mRNA with other initiation factors to start mRNA translation (Gingras *et al.*, 2004). S6K1 is instead involved in phosphorylation and activation of RPS6 and translation elongation factors, such as eEF2K (Fonseca *et al.*, 2014). RPS6 is a component of the 40Sribosomal subunit and phosphorylation of RPS6 by S6K1 promotes ribosome assembly and mRNA translation elongation (Gingras *et al.*, 2004).

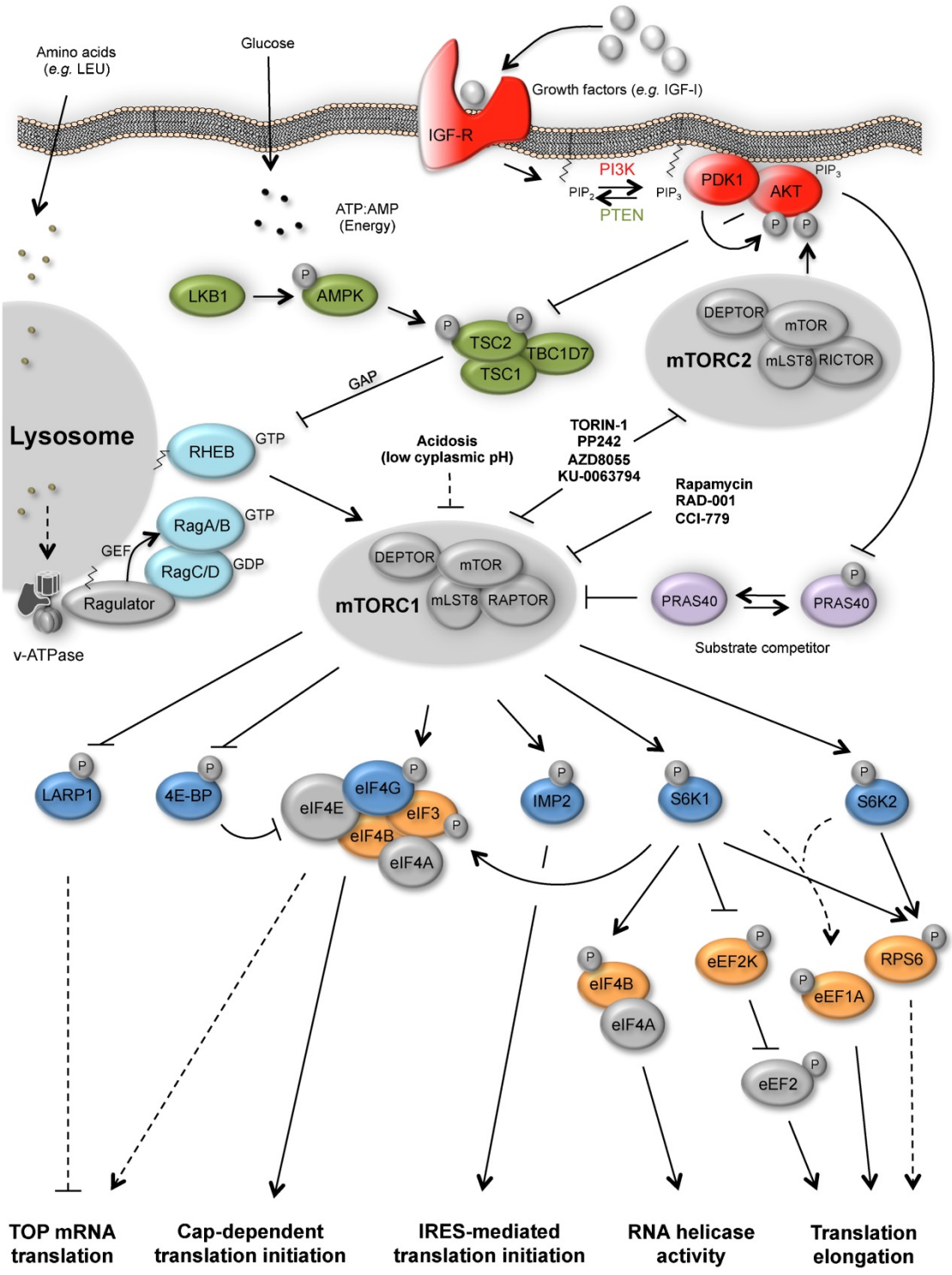


Figure 1.3. Overview of mTORC1-mediated regulation of mRNA translation. Modified from (Fonseca *et al*., 2014).

The mTOR pathway has been found to specifically up-regulate the translation of mRNA transcripts with established 5' terminal oligopyrimidine tracts (5'TOP motifs), many of which code for ribosomal proteins and translation factors (Meyuhas and Kahan, Fonseca *et*

al., 2014, Thoreen *et al.*, 2012). 5'TOP mRNAs are transcripts harbouring an uninterrupted stretch of 4 to 14 pyrimidines immediately adjacent to the transcription start site (TSS) (Thoreen *et al.*, 2012). Similar transcripts called 5'TOP-like mRNAs harbour a stretch of at least five pyrimidines within four nucleotides of the TSS and have also been found to be regulated by mTORC1 (Thoreen *et al.*, 2012, Fonseca *et al.*, 2014, Meyuhas and Kahan, 2015). The traditional concept that mTOR signalling is involved in regulation of translation of 5'TOP-mRNAs (Meyuhas and Kahan, 2015)only has been challenged by recent studies suggesting that the spectrum of mTOR-regulated mRNAs may not be limited to 5'TOP and 5'TOP-like transcripts. Studies on the breast cancer cell line MCF-7 identified other non-5'TOP mRNAs as being regulated by mTOR activity (Gandin *et al.*, 2016, Larsson *et al.*, 2012). The study by Larsson *et al.* (2012) used the polysome profiling technique for analysis of mRNA translation. Polysome profiling involves isolation of ribosome-bound RNA (RBR) and measurement of translation efficiency of transcripts as the ratio of abundance of actively translated RBR and total RNA (Chasse *et al.*, 2017). In contrast with these findings, studies using ribosome profiling on mouse embryonic fibroblasts (MEFs) (Thoreen *et al.*, 2012) and on the prostate cancer cell line PC3 (Hsieh *et al.*, 2012) showed that mRNAs regulated by the mTOR pathway almost exclusively harbour either 5'TOP or 5'TOP-like motifs. Unlike polysome profiling, ribosome profiling is based on deep sequencing of ribosome protected fragments (RPFs) and determines the position of the ribosome on the mRNA transcript (Ingolia *et al.*, 2012). It has been hypothesised that the contradicting results of these studies are due to the ribosome and polysome profiling techniques having different sensitivity in the identification of changes in mRNAs with varying abundance (Masvidal *et al.*, 2017).

The exact mechanism of regulation of translation of 5'TOP mRNAs by the mTOR pathway is still unclear. Several studies have shown that RNA-binding proteins such as the T-cell intracellular antigen-1 (TIA-1), TIA-1-related (TIAR) (Damgaard and Lykke-Andersen,

2011) and the lupus autoantigen (La)-related protein 1 (LARP1) (Tcherkezian *et al.*, 2014, Gentilella *et al.*, 2017, Hong *et al.*, 2017) are regulated by mTORC1 and might mediate mTORC1 promotion of translation of 5'TOP mRNAs (Fonseca *et al.*, 2014). LARP1 is phosphorylated by mTORC1 and acts as a molecular switch for translation initiation of mTOR-regulated mRNAs (Hong *et al.*, 2017) (Figure 1.3). In a recent study it was found that the RNA-binding protein LARP1 stabilises 5'TOP mRNAs and knock-down of LARP1 in human adult bone marrow CD34+ cells causes a reduction in 5'TOP mRNA levels and induction of p53 (Gentilella *et al.*, 2017). LARP1 was found to bind the 40S ribosome subunit and the disruption of the LARP1-40S complex lead to stabilization of p53, underlying a potentially important mechanism in the pathophysiology of 5q- syndrome (Gentilella *et al.*, 2017).

1.6.3.3. Current knowledge of the mechanisms of action of L-leucine

L-leucine activates the mTOR signalling pathway (Dodd and Tee, 2012). L-leucine enters the cells via the SLC7A5-SLC3A2 transporter and intracellular levels of L-leucine are sensed by the leucyl-tRNA synthase protein (Dodd and Tee, 2012). Leucyl-tRNA synthase has been found to bind to Rag GTPases (Rag proteins A/B and C/D) and activate the RHEB protein (Boultwood *et al.*, 2013). Sestrin2 has been also found to play a role in mediating signalling to mTORC1 in response to L-leucine (Wolfson *et al.*, 2016). RHEB in turn activates mTORC1, which then phosphorylates target proteins S6K1 and 4EBP1 (Boultwood *et al.*, 2013). Treatment with L-leucine has been shown to increase phosphorylation of S6K1 and 4EBP1 in cellular and animal models of 5q- syndrome and DBA (Narla *et al.*, 2014, Payne *et al.*, 2012, Yip *et al.*, 2013, Ear *et al.*, 2016). *RPS14* deficient erythroblasts and cultured erythroblasts from del(5q) MDS patients treated with L-

leucine show increased level of phosphorylation of S6K1 and 4EBP1 (Yip *et al.*, 2013). In line with the role of mTOR in regulation of mRNA translation, L-leucine treatment has been shown to increase protein production in *RPS14* deficient erythroblasts (Yip *et al.*, 2013) and in a zebrafish model of DBA (Zhang *et al.*, 2014). Roberts Syndrome (RBR) is characterised by decreased mTORC1 signalling and L-leucine treatment of RBR cells has been found to partially rescue mTOR function and translational efficiency of ribosomal subunits and translation initiation factors (Xu *et al.*, 2016). The majority of 5'TOP mRNAs encode ribosomal subunits and translation-related proteins and 5'TOP mRNAs are translationally repressed following amino acid withdrawal and conversely derepressed upon amino acid re-feeding of cells (Tang *et al.*, 2001). Therefore it has been hypothesised that L-leucine treatment might affect translation of 5'TOP mRNAs and production of ribosomal and translation-related proteins in 5q- syndrome (Boultwood *et al.*, 2013). An increase in haemoglobin levels might also underlie the mechanisms of action of L-leucine-mediated enhancement of erythropoiesis in the 5q- syndrome (Payne *et al.*, 2012, Ear *et al.*, 2016, Zhang *et al.*, 2014). Five of the six *globin* transcripts have been shown to fit the definition of 5' TOP-like mRNAs in a study by Chung *et al.* (2015). Polysome profiling was performed in MEL cells treated with the mTORC1 inhibitor torin-1 and it was found that *α/β-globin* mRNAs shifted to lighter polysome fractions in torin-1-treated cells compared to untreated (Chung *et al.*, 2015). This observation suggests that *α/β-globin* transcripts are direct mTORC1 translational targets (Chung *et al.*, 2015). Moreover, this study found that mTORC1 serves as a sensor for L-leucine via the receptor LAT3 (SLC43A1), and regulates haemoglobin translation through the 4EBP1 activation (Chung *et al.*, 2015). It has been hypothesised that L-leucine activation of mTORC1 and the possible increase in translation of 5'TOP mRNAs may be the mechanism of L-leucine enhancement of erythroid progenitor

cell growth and differentiation in 5q- syndrome and MDS with del(5q) (Boultwood *et al.*, 2013).

1.7. The CRISPR/Cas9 system

The clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) (CRISPR/Cas9) system is a versatile tool for genome engineering in various species, and can be used to correct gene mutations in eukaryotic cells (Ran *et al.*, 2013, Mali *et al.*, 2013, Schwank *et al.*, 2013, Wang *et al.*, 2013, Xie *et al.*, 2014). The system utilizes the endonuclease Cas9, which is targeted to a specific genomic site by a 20-nucleotide synthetic guide RNA (sgRNA) and creates a DNA double-strand break (DSB). The target site must lie immediately 5′ end of a protospacer adjacent motif (PAM) sequence in the form of NGG. The site-specific DNA DSB generated by Cas9 is repaired either by non-homologous end-joining (NHEJ), or by homology-directed repair (HDR) using a single-stranded oligodeoxynucleotide (ssODN) as repair template (Ran *et al.*, 2013) (Figure 1.4).

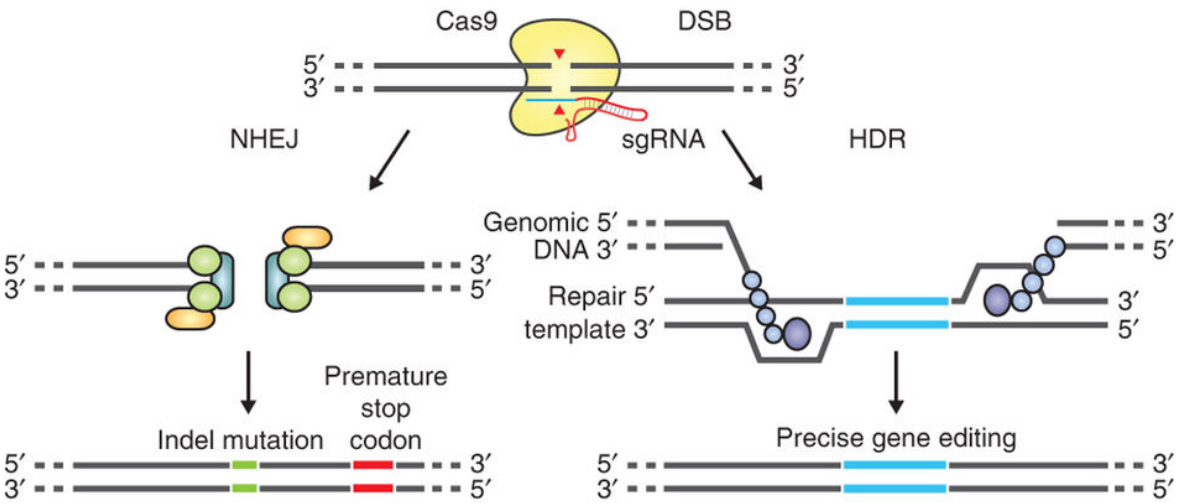


Figure 1.4. The CRISPR/Cas9 system for genome editing of mammalian cells. Figure from (Ran *et al.*, 2013)

1.7.1. Genome editing for modelling of MDS

The CRISPR/Cas9 system has been employed for the functional investigation of mutations in a wide variety of pathologies, including haematological disorders (Pellagatti *et al.*, 2015). In particular, several studies have demonstrated the use of the CRISPR/Cas9 system to generate cellular models of MDS which can be used to characterise disease phenotypes linked to recurrent mutations (Kotini *et al.*, 2015, Kotini *et al.*, 2017).

1.8. Aims of this D.Phil. project

The aim of this D.Phil. project was to gain insights into the pathogenesis of MDS with del(5q) and to investigate possible therapeutic approaches. A more complete understanding of the pathophysiology of del(5q) MDS would likely impact on development of new therapies to improve patients outcome. A large cohort of patients with del(5q) MDS and 5q-syndrome were screened to investigate the occurrence of mutations in *CSNK1A1* and their effects on gene expression. The mechanism of action of L-leucine, a new therapeutic agent, in amelioration of the anaemia in del(5q) MDS was also investigated. Moreover, a proof-of-principle study was carried out to test the use of the CRISPR/Cas9 system for correction of a *TP53* mutation in the CML cell line K562. Genome editing using CRISPR/Cas9 could be used for generating cellular models of MDS with del(5q).

2. Chapter Two – Materials and Methods

2.1. Patients samples

This study was approved by the ethics committee (Oxford 06/Q1606/110) and human bone marrow samples (Table 2.1 and 2.2) were obtained with appropriate consent and in compliance with the Human Tissue Act.

Bone marrow mononuclear cells from healthy donors were purchased from Lonza (Table 2.3) and "ethically obtained with documented, legal permission for research use" (https://www.lonza.com/products-services/bio-research/primary-cells/human-cells-and-media.aspx).

Information on the diagnosis, karyotype and age of patients (Table 2.1, 2.2 and 2.3) was provided by the supplier.

2.1.1. CSNK1A1 study

Bone marrow samples and peripheral blood neutrophils were isolated from 104 MDS patients (Table 2.1) and 17 healthy controls using Histopaque (Sigma-Aldrich) density gradient centrifugation. The isolation was performed by previous members of the group.

Table 2.1. MDS subtype and karyotype of del(5q) MDS patients screened for *CSNK1A* mutations.

Patient ID	Subtype	Karyotype
Patient 1	RA (5q- syndrome)	46,XY,del(5)(q13q33)[30]
Patient 2	RAEB1	46,XY,del(5)(q14q34)[2]/47,XY,del(5)(q14q34),+21[20]
Patient 3	RA	46,XX,del(5)(q13q33)[30]
Patient 4	RA	46,XY,del(5)(q21q34)[5]/46,XY,inv (q13q25),del(5)(q21q34)[25]
Patient 5	RA	46,XX,t(1;3)(p33;p14),del(5)(q14q34)[21]/46,XX[4]
Patient 6	RAEB1	46,XX,del(5)(q14q34)[20]/46,XX[5]
Patient 7	RA	46,XX,del(5)(q14q34),inv(9)(p11q13)c[30]
Patient 8	RARS	46,XX,del(5)(q14q34)[30]
Patient 9	RAEB	92<4n>,XXYY,del(5)(q14q33)[30]
Patient 10	RAEB	46,XX,del(5)(q15q33)[20]
Patient 11	RA	46,XX,t(2;11)(?;?),del5(q13q33)[18]/46,XX[2]
Patient 12	RAEB(RS)	46,XX,del(5)(q13q33),t(6;12)(q13;p12)[2]/44,XX,-7,-22[3]/47,XX,del(5)(q13q33),t(6;12)(q13;p12),+mar[15]/46,XX[3]
Patient 13	RA (5q- syndrome)	46,XY,del(5)(q13q33)[30]
Patient 14	RA (5q- syndrome)	46,XX,del(5)(q13q33)[26]/46,XX[4]

Patient 15	RAEB	46,XX,del(5)(q13q33)[20]/46,XX[10]
Patient 16	CMML	46,XY,del(5)(q13q33),del(13)(q12q22)[20]
Patient 17	RAEB	46,XX,del(5)(q13q33),del(11)(q23)[30]
Patient 18	RA	46,XX,del(5)(q13q33),+8[30]
Patient 19	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 20	RA (5q- syndrome)	46,XX,del(5)(q14q33)[27]/46,XX[3]
Patient 21	RA	46,XX,del(5)(q13q31)[18]/48,XX,del(5)(q13q31),idic(21)(q22),+2mar[2]/46,XX[1]
Patient 22	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 23	RA (5q- syndrome)	46,XY,del(5)(q?)[30]
Patient 24	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 25	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 26	RA (5q- syndrome)	46,XX,del(5)(q13q33)[5]/46,XX[7]
Patient 27	RA (5q- syndrome)	46,XX,del(5)(q22q35)[30]
Patient 28	RA	46,XX,del(5)(q13q33),del(11)(q22)[3]/46,XX[27]
Patient 29	RA (5q- syndrome)	46,XX,del(5)(q33q34)[30]
Patient 30	RAEB	43 ~45,XY,del(5)(q31),der(7)t(7;12)(q22;q1?3),-12,-13,-19,del(20)(q1?3)[cp4]
Patient 31	RA (5q- syndrome)	46,XX,del(5)(q31q33)[8]/46,XX[31]

Patient 32	RAEB	46,XX,del(5)(q13q33)[12]/46,XX[3]
Patient 33	RA (5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 34	RAEB	45,XX,del(5)(q13q33),-7,del(12)(p11p12)[12]/45,XX,del(5)(q13q33),-7,t(8;13)(q21;q31)[2]/46,XX[1]
Patient 35	RA (5q- syndrome)	46,XX,del(5)(q13q33)[26]/46,XX[4]
Patient 36	RA (5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 37	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 38	RA (5q- syndrome)	46,XX,del(5)(q13q33)[30]
Patient 39	RA (5q- syndrome)	46,XX,del(5)(q14q33)[30]
Patient 40	RA	46,Y,der(X)t(X;12)(p22;q21),del(5)(q14;q33),del(12)(p11q13)[27]/46,XY[3]
Patient 41	RA (5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 42	RA (5q- syndrome)	46,XX,del(5)(q31q33)[30]
Patient 43	RA	46,XX,del(5)(q?)[30]
Patient 44	RAEB-1	46,XX,del(5)(q13q33)[17]/46,XX[3]
Patient 45	RA (5q- syndrome)	46,XY,del(5)(q?)[25]
Patient 46	RA (5q- syndrome)	46,XY[51]
Patient 47	RAEB-1	46,XY,del(5)(q11q35)[23]/47,XY,+1,del(5)(q11q35)[3]/ 46,XY,del(5)(q11q35),add(9)(q34)[1]/46,del(5)(q15q35orq22q35)[1]/46,XY[2]

Patient 48	RCMD-RS	46,XY,del(5)(q14q34),del(17)(p12),1dmin[10]
Patient 49	MDS del(5q)	46,XX,del(5)(q13q33)[25]
Patient 50	MDS del(5q)	46,XX,del(5)(q13q33)[30]
Patient 51	MDS/AML	46,XX,del(5)(q13q33)[18]/46,XX[2]
Patient 52	RAEB1	46,XX,del(20)(q11.2)[10]/46,XX,add(5)(q22),del(20)(q11.2)[4]/46,XX,del(5)q13q33)[6]
Patient 53	RAEB2	46,XX,del(5)(q13q31)[17]/46,XX[3]
Patient 54	RA (5q- syndrome)	46,XX,del(5)(q13q31)[4]/46,XX[16]
Patient 55	RAEB	46,XY,del(5)(q13q31)[9]/46,XY[6]
Patient 56	MDS del(5q)	46,XY,del(5)(q13q31)[2]/46,XY[18]
Patient 57	RA(5q- syndrome)	46,XY,del(5)(q15q31)[10]/46,XY[10]
Patient 58	RA(5q- syndrome)	46,XX,del(5)(q13q33)[18]/46,XX[2]
Patient 59	RA(5q- syndrome)	46,XX,del(5)(q13q33)[19]/46,XX[4]
Patient 60	RAEB2	46,XX,del(5)(q13q33)[2]/47,XX,del(1)(p3?4),del(5)(q13q33),+mar[16]/46,XX[2]
Patient 61	RA(5q- syndrome)	46,XX,del(5)(q13q33)[17]/46,XX[3]
Patient 62	RA(5q- syndrome)	46,XX,del(5)(q22q35)[19]/46,XX[1]
Patient 63	RA(5q- syndrome)	46,XY,del(5)(q13q33)[2]/46,XY[19]
Patient 64	RA(5q- syndrome)	46,XY,del(5)(q13q33)[20]

Patient 65	RAEB1	46,XX,del(5)(q31q35)[7]/46,XX[6]
Patient 66	MDS del(5q)	46,XX,del(5)(q13q33)[6]/46,XX[14]
Patient 67	RAEB1	46,XX,del(5)(q13q33)[3]/46,XX[17]
Patient 68	MDS	46,XX,del(5)(q13q31)[15]/46,XX[8]
Patient 69	MDS	46,XX,del(5)(q14q34) [9] / 46,XX [13]
Patient 70	MDS	46,XX,del(5)(q13q33)[14]/46,XX[8]
Patient 71	RA	47 XX,del(5)(q?),+8[20]
Patient 72	RA	46,XX,del(5)(q?),del(7)(q?)[30]
Patient 73	RAEB	46,XY,del(5)(q?)[20]
Patient 74	RA(5q- syndrome)	46,XY,del(5)(q?)[30]
Patient 75	RA(5q- syndrome)	46,XY,del(5)(q?)[30]
Patient 76	RA(5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 77	RA(5q- syndrome)	46,XX,del(5)(q?)[20]
Patient 78	RA(5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 79	RA(5q- syndrome)	46,XX,del(5)(q?)[25]
Patient 80	RA	47,XX,del(5)(q?),+8[25]
Patient 81	RA	46,XX,t(2,11)(p16;q23),del(5)(q?)[30]

Patient 82	RA	46,XY,del(5)(q?)[25]
Patient 83	RA(5q- syndrome)	46,XX,del(5)(q?)[25]
Patient 84	RA(5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 85	RA(5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 86	RA(5q- syndrome)	46,XX,del(5)(q?)[30]
Patient 87	RA(5q- syndrome)	46,XX,del(5)(q?)[20]
Patient 88	RAEB2	46,XY,del(5)(q?)[30]
Patient 89	RA	47,XX,del(5)(q?),del(7)(q?),+8[30]
Patient 90	RA(5q- syndrome)	46,XY,del(5)(q?)[20]
Patient 91	RA	47,XX,del(5)(q?),+8[30]
Patient 92	RAEB2	46,XX,del(5)(q?),del(11)(q?)[30]
Patient 93	MDS	46,XX,del(5)(q?)[30]
Patient 94	MDS	46,XX,del(5)(q13q32)[30]
Patient 95	MDS	45,XY,der(1)(q?),-2,der (q?),-5,-7,+2mar[6]/46,XY[24]
Patient 96	RAEB	46,XX,del(5)(q?)[30]
Patient 97	MDS/AML	45,XY,de**r**(1)(p?),-2,der (q?),-5,-7,t(9;22)(q34;q11),der(17)(?),+2mar[30]
Patient 98	MDS	46,XY,del(5)(q13q33)[8]/46,XY[22]

Patient 99	MDS	46,XX,del(5)(q13q33)[30]
Patient 100	MDS	46,XX,del(5)(q13q33)[20]/46,XX[10]
Patient 101	RAEB	46,XY,del(5)(q13q33),+8,-20,-22,+mar[27]/46,XY[3]
Patient 102	MDS/AML	46,XY,del(5)(q13q33),+8,-20,-22,+mar[20]/46,XY,del(5)(q13q31),+8,add(11)(q23),-20,-22,+mar[10]
Patient 103	MDS	45,XX,der(1)(p?),del(5)(q13q33),-7,-21,-22,+2mar[26]/47,XX,del(5)(q13q33),-7,-17,+3mar[4]
Patient 104	MDS/AML	48 ~50,XX,del(5)(q13q33),del(7)(q11q21),i(12)(q10),+2 ~4mar[41]/46,XX,del(5)(q13q33)[8]/46,XX[1]

2.1.2. L-leucine study

Bone marrow mononuclear cells from 5q- patients were provided by Prof Sophie Raynaud (Laboratoire d'Oncohematolgie, Hopital Pasteur, Nice, France) and Professor Rajko Kušec (Table 2.2).

Table 2.2. MDS subtype and karyotype of del(5q) MDS patients analysed in the L-leucine study.

Patient ID	Subtype	Karyotype
MDS1	RA (5q- syndrome)	46,XX,del(5)(q13q33) [19] / 46,XX [1]
MDS2	RA (5q- syndrome)	46,XX,del(5)(q22q35) [19] / 46,XX [1]
MDS3	RAEB	46,XX,del(5)(q13q33) [20]
MDS4	RA (5q- syndrome)	46,XX,del(5)(q13q33) [16]/ 46,XX [6]
MDS5	MDS del(5q)	46,XX,del(5)(q13q33) [13] / 46,XX [6], an additional mitosis carries trisomy 8
MDS6	RA (5q- syndrome)	46,XX,del(5)(q13q33) [6] / 46,XX [14]
MDS7	RA (5q- syndrome)	46,XY,del(5)(q13q33) [20]
MDS8	Information not available	46,XY,del(5)(q?)

Bone marrow mononuclear cells from healthy controls were purchased from Lonza (Cat number 2M-125C) (Table 2.3).

Table 2.3. Samples from healthy controls analysed in the L-leucine study.

Sample	**Cell type**	**Batch**	**Age**	**Sex**	**Race**
Healthy Control 1	BM mononuclear	080172C	21 Y	Male	Black
Healthy Control 2	BM mononuclear	080176B	35 Y	Female	Caucasian
Healthy Control 3	BM mononuclear	080183B	20 Y	Male	Black
Healthy Control 4	BM mononuclear	080219B	21 Y	Male	Black
Healthy Control 5	BM mononuclear	080264B	20 Y	Male	Black

Healthy Control 6	BM mononuclear	081291B	23 Y	Male	Black
Healthy Control 7	BM mononuclear	081296A	23 Y	Male	Caucasian
Healthy Control 8	BM mononuclear	081286A	25 Y	Female	Black

2.2. Cell culture

2.2.1. Differentiation of mononuclear cells along the erythroid lineage

Bone marrow mononuclear cells from patients and healthy controls were cultured for 11 days in StemSpan™ Serum-Free Expansion Medium (SFEM) media (Stemcell Technologies) supplemented with 0.5 U/mL recombinant human erythropoietin (EPO) (Stemcell Technologies), 100 ng/mL stem cell factor (SCF) (Miltenyi Biotec), 10 ng/mL interleukin 3 (IL-3) (Miltenyi Biotec), 40 µg/mL Lipid Mixture 1 (Contains non-animal derived fatty acids (2 µg/ml arachidonic and 10 µg/ml each linoleic, linolenic, myristic, oleic, palmitic and stearic), 0.22 mg/ml cholesterol from New Zealand sheep′s wool, 2.2 mg/ml Tween-80, 70 µg/ml tocopherol acetate and 100 mg/ml Pluronic F-68 solubilized in cell culture water) (Sigma- Aldrich), 2 mM L-glutamine (Sigma- Aldrich), and 1% penicillin-streptomycin solution (Sigma Aldrich). Cells were passaged every 2 or 3 days. On day 7, the concentration of EPO was increased to 3 U/ml.

2.2.2. Cell lines

K562 cells were purchased from ATCC while MOLM-13 cells were previously stored in the lab. K562 and MOLM-13 cells were cultured in RPMI 1640 media (Life Technologies) containing 10% heat-inactivated foetal bovine serum (FBS) (Life Technologies), 1% L-glutamine (Sigma-Aldrich), 100 units/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich). For one transfection of the CRISPR-Cas9 system for editing of *TP53* and single cell sorting, K562 cells were cultured in Iscove's modified Dulbecco's medium (IMDM)

(Life Technologies) containing 20% heat-inactivated foetal bovine serum (FBS) (Life Technologies) and 1% L-glutamine (Sigma-Aldrich).

KBM5 were kindly provided by Dr. Bing Z. Carter (Department of Leukemia, Anderson Cancer Center). KBM5 cells were maintained in IMDM (Life Technologies) containing 10% heat-inactivated FBS (Life Technologies), 1% L-glutamine (Sigma-Aldrich), 100 units/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich).

293T cells were purchased from ATCC and cultured in Dulbecco's Modified Eagle's medium (DMEM) (Life Technologies) supplemented with 10% FBS (Life Technologies), 100 units/mL penicillin and 100 µg/mL streptomycin (Sigma- Aldrich).

Suspension cell lines K562 and KBM5 were seeded at 2 x 10^5 or 5 x 10^5 and passaged every 2 to 3 days.

The adherent cell line 293T cells were passaged at 90% confluency to 1:8 dilution. Adherent cells were treated with 0.05% (w/v) Trypsin - 0.53 mM Ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich) before passaging.

All cells were maintained at 37° C and 5% CO_2.

2.2.3. CRISPR-edited single cells

CRISPR-edited K562 cells were FACS sorted into 96-well plates and cultured in RPMI 1640 media supplemented with 20% heat-inactivated FBS (Life Technologies), 0.5% penicillin and 0.5% streptomycin (Sigma-Aldrich). Media was carefully aspirated from the top of each well and refreshed every 3-4 days.

2.3. L-leucine treatment

Bone marrow mononuclear cells from 5q- patients and healthy controls were treated with either L-leucine (Sigma Aldrich) or D-leucine (Sigma Aldrich) at a concentration of 600 µg/ml at day 9 of the culture (as described in Yip *et al.* (2013). Culture media containing either L-leucine or D-leucine was refreshed every 2-3 days.

2.4. Genomic DNA extraction

2.4.1. *CSNK1A1* genotyping

DNA was isolated using phenol-chloroform extraction from peripheral blood neutrophils or from bone marrow samples of MDS patients. The extraction was performed by Mrs Jacqueline Shaw.

2.4.2. DNA extraction from cell lines

DNeasy Blood & Tissue Kit (Qiagen) was utilised for DNA extraction from cell lines according to the manufacturer's protocol.

2.4.3. DNA extraction from CRISPR-edited single cells

Genomic DNA was extracted from colonies derived from single CRISPR-edited K562 cells using 20µl of QuickExtract DNA Extraction Solution 1.0 (Epicentre) according to the manufacturer's protocol.

2.5. Polymerase Chain Reaction (PCR)

PCR was performed using various polymerases, depending on the application. PCR reactions were prepared according to the manufacturer's recommendations for each kit. Primers utilised, amount of gDNA and PCR cycling conditions used are described in Table

2.4. For genotyping of CRISPR-edited single cells, 2µl of cell lysate (see section 2.4.3) was used for PCR amplification.

Table 2.4. PCR primers. FWD=forward primer, REV=reverse primer.

Region	Purpose	Primers (FWD top and REV bottom)	Polymerase	Cycling conditions	gDNA
CSNK1A1 ex 3	Genotyping	TCCTTTTGTTTCGTTAGGTGGT AAGGTTAAATAGTGATGCACAGGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 51°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
CSNK1A1 ex 4	Genotyping	GCCAAAGGACACAGCAGGTA CAGCAAATTCAACTTACTATGGC	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 51°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng

TP53 ex 3-4	Genotyping	GTGGGAAGCGAAAATTCCAT GCCAGGCATTGAAGTCTCAT	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 60°C 30s, 72°C 1.5 min and 72°C 10 min	10ng
TP53 ex 5-6	Genotyping	TGTTCACTTGTGCCCTGACT TTAACCCCTCCTCCCAGAGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 56°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
TP53 ex 7	Genotyping	GAGCTTGCAGTGAGCTGAGA GGGATGTGATGAGAGGTGGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 61°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
TP53 ex 8-9	Genotyping	GACAAGGGTGGTTGGGAGTA GCCCCAATTGCAGGTAAAAC	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 56°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
TP53	HRMA sgRNA1	CTCCCCTGCCCTCAACAA G GCCTCACAACCTCCGTCATG	Hotshot Diamond PCR master mix (Clent Lifescience)	95ºC 5 min, 45 cycles at 95ºC 20 s, 60ºC 30 s, 72ºC 30 s then 95ºC 30 s and 25ºC 30 s	10ng
TP53		TTGTTTCTTTGCTGCCGTC		95ºC 5 min, 45 cycles at 95ºC 20 s, 59ºC 30 s, 72ºC	10ng

	HRMA sgRNA2 and sgRNA3	TGGAATCAACCCACAGCTG	Hotshot Diamond PCR master mix (Clent Lifescience)	30 s then 95ºC 30 s and 25ºC 30 s	
TP53	Restriction digest assay	AAAGAAAAGCTCCTGAGGTGTAGACG CAGCTGCTCACCATCGCTATCT	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 61°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
TP53	Amplification of edited locus (WT ssODN – region downstream of T7 sequence)	GCCATACGACTCACTATAGGGAAGA TTAACCCCTCCTCCCAGAGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 53°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng
TP53	Amplification of edited locus (WT ssODN – region upstream of T7 sequence)	TTAGGAGGCTGAGGTGGGAA CTTCCCTATAGTGAGTCGTATGGC	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 60°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng
TP53	Amplification of edited locus (mutated ssODN – region downstream of T7 sequence)	TGTAATACGACTCACTATAGGGATCC TTAACCCCTCCTCCCAGAGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 55°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng

TP53	Amplification of edited locus (mutated ssODN – region upstream of T7 sequence)	TTAGGAGGCTGAGGTGGGAA GGATCCCTATAGTGAGTCGTATTACA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 60°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng
TP53	Amplification of edited locus (region downstream of silent barcode)	TCCTGAGGTGTAGACGCCAAC TTGGACTGGACACGTTTTAGCA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 54°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng
TP53	Amplification of edited locus (region upstream of silent barcode)	TGCTAAAACGTGTCCAGTCCAA GACAACCACCCTTAACCCCTC	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 40 cycles at 95°C 30 s, 54°C 30 s, 72°C 1.5 min and 72°C 10 min	50ng
PRKCA	gRNA2 off-target site	AAAGGCAGATCAGAAACCACAGGA CCATTCATCCTCAGCTGCCATAAT	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 60°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
ASAH1	gRNA2 off-target site	AAGGCATGTAGCCTAAGAGGGAAT TGACCTATCAAGCCTCAGTGATCA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 58°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng

ADA2	gRNA2 off-target site	TGATCAGCTCTGATGACCCAGC CAGCACAGAGATCAGAGAGAGGAGA	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 60°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
LZTS3	gRNA2 off-target site	TGAGTCCACAGAAATCTGATCGACC TCAGAAACTGGGCGAGGAGAGT	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 60°C 30 s, 72°C 1.5 min and 72°C 10 min	10ng
SPTBN1	gRNA2 off-target site	CGCCGCCTCTGGAAGTTCTT GGGAAGCCTGGTTGTTCATTTC	Maxima Hot start Taq DNA polymerase Master Mix (Thermo Scientific)	95°C 4 min, 35 cycles at 95°C 30 s, 60°C 30s, 72°C 1.5 min and 72°C 10 min	10ng
ASXL1	Genotyping	CAGTGGTGATGGTGGTGAGG GGGGAGTTGGGAGGCATCT	Phusion High-Fidelity DNA Polymerase (NEB)	98°C 30 s, 30 cycles at 98°C 10 s, 62°C 30 s, 72°C 30 s and 72°C 7 min	50ng

2.6. Sanger sequencing

PCR products were run on a 1.5% agarose gel to confirm the presence of a single band and treated with Exonuclease I (Exo) (NEB) and Shrimp Alkaline Phosphatase (rSAP) (NEB). 7.5µl of PCR product was mixed with 0.1µl of Exo and 1µl of rSAP and incubated at 37°C for 30 minutes and at 80°C for 20 minutes. PCR products were diluted to 1ng/µl per 100bp sequenced.

Plasmid DNA was diluted to 100ng/µl.

Sanger sequencing was performed at Source Bioscience (Oxford, UK).

2.7. RNA extraction

2.7.1. RNA extraction (standard method)

RNA was extracted from cells using TRIzol Reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. For the *CSNK1A1* study, RNA extraction from bone marrow samples and peripheral blood neutrophils from MDS patients was performed by Dr Andrea Pellagatti.

2.7.2. Extraction of total RNA and ribosome-bound RNA (for polysome profiling)

The RNA Clean and Concentrator-5 Kit (Zymo Research) was used to purify total RNA and ribosome-bound RNA (RBR). Depending on the amount of cell extract available in each experiment, samples were diluted to a minimum of 50µl and total RNA or RBR was extracted according to the manufacturer's protocol.

2.8. DNase I treatment

Total RNA and RBR, used for polysome profiling experiments, was treated with DNase I (Invitrogen) prior to either reverse transcription or library preparation for RNA-seq. A maximum of 1µg RNA was added to 1µl 10X DNase I Reaction Buffer (Invitrogen), 1µl DNase I 1 U/µl (Invitrogen) and DEPC-treated water to 10µl. The sample was incubated at room temperature for 15 minutes. To inactivate the DNase I enzyme, 1µl of 25mM EDTA solution (Invitrogen) was added to the reaction and the sample was heated at 65℃ for 10 minutes.

For use in RT-qPCR assays, DNAse-treated RNA was reverse transcribed prior to amplification.

For use in library preparation for RNA sequencing, RNA was purified using Agencourt RNAClean XP beads (Beckman Coulter) to ensure removal of the enzyme and sample purity. Beads purification of RNA was carried out in 96-well round-bottom plates (Greiner Bio-One) using the Ambion RNA 96 magnetic stand (Life Technologies) according to the manufacturer's protocol.

2.9. Reverse transcription

A maximum of 1µg RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), according to the manufacturer's instructions. 1µl of RNase inhibitor was added to a 20µl reaction. Reverse transcription was performed using random hexamers. cDNA was diluted 1:5 or 1:10 depending on the sample prior to quantification using a NanoDrop (Thermo Fisher Scientific).

2.10. RT-qPCR

Real Time Quantitative Polymerase Chain Reaction (RT-qPCR) reactions were performed in triplicate on a Roche LightCycler 96 instrument (Roche Diagnostics). No template controls (NCT) were included in triplicate for each gene expression assay.

2.10.1. TaqMan RT-qPCR reactions

RT-qPCR amplifications reactions were prepared using LightCycler 480 Probes Master Mix (Roche). 10µl of 2X LightCycler 480 Probes Master was added to 3µl Nuclease-free water, 1µl 20X TaqMan gene expression assay (Applied Biosystems) and 6µl of cDNA. Pre-developed TaqMan gene expression assays (Applied Biosystems) used are outlined in Table 2.5 and cycling conditions in Table 2.6. The *β2-microglobulin* gene (*B2M*) was used to normalize for differences in input cDNA and gene expression was calculated using the ΔΔCT method for relative quantification. Gene expression was measured as fold difference relative to a reference sample (e.g. wild-type cell line or healthy control).

The following calculation was used:

$\Delta CT = CT(Gene) - CT(B2M)$

$\Delta\Delta CT = \Delta CT(sample) - \Delta CT(reference)$

Fold difference in expression = $2^{-\Delta\Delta CT}$

Table 2.5. TaqMan assays (Applied Biosystems).

Gene	TaqMan assay	Quantification method
TP53	H01034249_m1	ΔΔCT method for relative quantification
CDKN1A	Hs00355782_m1	ΔΔCT method for relative quantification
BBC3	Hs0048075_m1	ΔΔCT method for relative quantification
BAX	Hs00414514_m1	ΔΔCT method for relative quantification
B2M	Hs00430330_m1	ΔΔCT method for relative quantification
CSNK1A1	Hs00793391_m1	ΔΔCT method for relative quantification

CCND1	Hs00765553_m1	ΔΔCT method for relative quantification
RPS14	Hs00735285_m1	ΔΔCT method for relative quantification
RPS20	Hs00828752_gH	ΔΔCT method for relative quantification
RPS18	Hs01375212_g1	ΔΔCT method for relative quantification

Table 2.6. Cycling conditions of TaqMan RT-qPCR reactions.

Temperature (℃)	**Time (sec)**	**Cycles**	**Step**
95	300	1	Initial denaturation
95	10	40	Denaturation
60	30		Annealing
72	1		Extension
40	10	1	Cooling

2.10.2. SYBR Green RT-qPCR reactions

RT-qPCR amplification reactions were prepared using LightCycler SYBR Green I Master Mix (Roche). 10µl of 2X LightCycler SYBR Green I Master was added to 1µl of forward and 1µl of reverse primer (10µM), 3µl of Nuclease-free water and 5µl of cDNA. Amplification efficiency of primers designed was tested by setting up RT-qPCR reactions with a serial dilution of cDNA (100ng, 10ng, 1ng, 0.1ng) to generate a standard curve. Only primers with good amplification efficiency and yielding a single amplification product, as assessed by melting curve, were utilised (Table 2.7). RT-qPCR cycling conditions are outlined in Table 2.8. Gene expression was measured using either the ΔΔCT method for relative quantification (see section 2.10.1) or the standard curve method for absolute quantification. For the ΔΔCT method for relative quantification, CT values were normalised to the *Actin Beta* gene *(ACTB)*. For the absolute quantification, a standard curve was

generated using a control sample serially diluted 2-fold from 1ng to 7.8pg. The amount of cDNA in each sample was calculated using the equation of the standard curve and the percent of cDNA in each sample was calculated as nanograms of cDNA per nanograms of input sample, measured using a NanoDrop (Thermo Fisher Scientific).

Table 2.7. Primers and quantification method of SYBR Green RT-qPCR reactions.

Region	**Forward primer**	**Reverse primer**	**Quantification method**
ACTB	ATTGGCAATGAGCGGTTC	GGATGCCACAGGACTCCAT	ΔΔCT method for relative quantification
Alu repeat	GAGGCTGAGGCAGGAGAATCG	GTCGCCCAGGCTGGAGTG	Standard curve method for absolute quantification
RNA5S	GGCCATACCACCCTGAACGC	CAGCACCCGGTATTCCCAGG	Standard curve method for absolute quantification
RNA18S	CTCAACACGGGAAACCTCAC	CGCTCCACCAACTAAGAACG	Standard curve method for absolute quantification
RNA28S	GGGTGGTAAACTCCATCTAAGG	GCCCTCTTGAACTCTCTCTTC	Standard curve method for absolute quantification
B2M	CTCCGTGGCCTTAGCTGTG	TTTGGAGTACGCTGGATAGCCT	Standard curve method for absolute quantification
ACTB	ATTGGCAATGAGCGGTTC	GGATGCCACAGGACTCCAT	Standard curve method for absolute quantification
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG	Standard curve method for absolute quantification

Table 2.8. Cycling conditions of SYBR Green RT-qPCR reactions.

Temperature (°C)	Time (sec)	Cycles	Step
95	30	1	Initial denaturation
95	15	40	Denaturation
60	60		Annealing and extension
95	10	Continuous	Melting
65	60		
97	1		

2.11. GEP and data analysis

Gene expression profiling (GEP) data on $CD34^+$ cells from a cohort of MDS patients with del(5q) and healthy controls were obtained from a previously published dataset (Pellagatti *et al.*, 2010a). The platform Affymetrix GeneChip Human Genome U133 Plus 2.0 (47 000 transcripts) (Affymetrix, Santa Clara, CA, USA) was utilised to produce the data. Analysis of gene set up- or down-regulation was performed using Gene Set Enrichment Analysis (GSEA) as previously described (Papaemmanuil *et al.*, 2011). Dr Andrea Pellagatti performed the experiments and analysed the data generated.

2.12. Preparation of cell extracts

2.12.1. Cell extracts for Western Blotting

Cells were lysed using Pierce RIPA Buffer (Thermo Fisher Scientific) supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) according to the manufacturer's protocol. 1ml of RIPA Buffer was used for 5 x 10^6 cells.

2.12.2. Cell extracts for sandwich ELISA

Total cell lysates were prepared using Cell Lysis Buffer (Cell Signaling) according to manufacturer's instruction. Lysates were vortexed 10 times and placed on ice instead of being sonicated. 400µl of 1X Cell Lysis Buffer plus 1mM Phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) was used for 10^7 cells maximum.

2.12.3. Cell extracts for polysome profiling

Cell extracts were prepared according to the TruSeq RiboProfile (Mammalian) Kit protocol (Illumina). 200µl of Mammalian Lysis Buffer was used for 200,000 cells. Briefly, 1ml of Mammalian Lysis Buffer contains:

- 200µl 5x Mammalian Polysome Buffer (Illumina)
- 100µl 10% Triton X-100 (Illumina)
- 10µl 100mM DTT (Illumina)
- 10µl DNase I (1U/µl) (Illumina)
- 10µl cycloheximide (10 mg/µl) (Sigma-Aldrich)
- 10µl 10% NP-40
- 660µl Nuclease-free water

Cell pellets were resuspended in Mammalian Lysis Buffer and passed through a 25-gauge needle 10 times in order to lyse the cells completely. The lysate was then incubated for 10 minutes on ice with periodic inversions and clarified by centrifugation for 10 minutes at 20,000 x g at 4°C. The supernatant was transferred to a chilled tube and snap frozen or immediately used for RNA extraction.

2.13. Protein quantification

Prior to Western blotting and sandwich ELISA assays, the amount of protein in each cell extract was determined using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific).

2.14. Western blotting

Equal amounts (10µg) of extracted proteins were separated by SDS-PAGE using the Invitrogen NuPage Novex 4–12% Bis-Tris Gels. Samples were then transferred onto a nitrocellulose membrane, which was incubated in blocking buffer, containing 5% non-fat dry milk (Blotto, Santa Cruz Biotechnology) in Tris-buffered saline with Tween-20 detergent (TBST) (Pierce Biotechnology). The membrane was then washed in TBST and incubated with anti-p53-HRP (SC-126, D-10, Santa Cruz Biotechnology) diluted 1:1000 in blocking solution for one hour at room temperature. The membrane was washed repeatedly with TBST and developed using ECL detection reagents (Bio-rad), according to the manufacturer's instruction. The membrane was developed after exposure to an Amersham Hyperfilm ECL (GE Healthcare) for 1 minute. β-actin was used as a loading control. The membrane was stripped using Restore Stripping Buffer (Thermo Fisher Scientific) and then incubated with anti-β-actin -HRP (AC15, Abcam) diluted 1:20,000 in blocking solution for one hour at room temperature. The membrane was then washed in TBST and developed as described above.

2.15. Sandwich ELISA

Cell extracts were diluted to 25µg/ml and 100µl of each lysate was added to the appropriate well. Sandwich ELISA was performed to measure the phosphorylation levels of mTORC1 targets 4EBP1(PathScan Phospho-4EBP1 (Thr37/Thr46) Sandwich ELISA Kit), S6K1 (PathScan Phospho-p70 S6 Kinase (Thr389)) and RPS6 (PathScan Phospho-S6 Ribosomal Protein (Ser235/236)) according to manufacturer's instructions (Cell Signaling Technology). The absorbance (OD450 nm) was measured on a SpectraMax Plus 384 Microplate Reader (Molecular Devices) and it is proportional to the levels of phosphorylated protein in the cell extract.

2.16. Ribosome profiling

The TruSeq Ribo Profile (Mammalian) kit (Illumina) was utilised according to the manufacturer's protocol. The kit includes reagents for ribosome footprinting as well as library preparation for RNA sequencing of total RNA and Ribosome-protected Fragments (RPFs). The following steps were carried out:

- Cell extract preparation
- Ribosome footprinting with TruSeq Ribo Profile Nuclease (Illumina)
- Purification of RPFs with Illustra MiscroSpin S-400 HR Columns (GE Healthcare)
- Extraction of total RNA and RNA from RPFs
- Ribosomal RNA (rRNA) depletion
- PAGE purification of RPFs
- TruSeq library preparation for Next Generation Sequencing (NGS):
    - Fragmentation and End Repair
    - 3' Adapter Ligation
    - Reverse Transcription
    - PAGE purification of cDNA
    - cDNA circularization
    - PCR amplification

The TruSeq Ribo Profile (Mammalian) kit (Illumina) protocol suggests to start with a culture of cell density between 5 and 50 million cells/ml but the exact cell number required is not indicated and the cell type is not specified. Therefore, the protocol was tested using 0.2 $x10^6$ and 0.5$x10^6$ cultured erythroblasts derived from bone marrow mononuclear cells of healthy controls (Lonza) (cultured as described in section 2.2.1), in order to identify the smallest possible amount of starting material to utilise. RNA quantity and quality were assessed by

Agilent RNA 6000 Pico Kit (Agilent Technologies) after extraction of total RNA and RPFs and after PAGE purification of RPFs. 9, 11 or 13 cycles of final PCR amplification were performed, in order to try to obtain sufficient amounts of library for RNA sequencing. Integrity and size of TruSeq libraries were measured using the Agilent High Sensitivity DNA Assay (Agilent Technologies) on the Agilent 2100 Bioanalyser System (Agilent Technologies).

2.17. Polysome Profiling

2.17.1. Cycloheximide treatment

Cycloheximide (Sigma Aldrich) was added to the culture media at a final concentration of 0.1 mg/ml. Cells were incubated in media containing cycloheximide for 10 minutes at 37℃ (Su *et al.*, 2016).

2.17.2. Preparation of cell extracts

Between 200,000 and 1 x 10^6 cells were used for each experiment. Cells were rinsed in 10ml of ice-cold Phosphate-Buffered Saline (PBS) (Life Technologies) containing cycloheximide (Sigma-Aldrich) at 0.1 mg/ml. Cells were pelleted and PBS was removed. Cell extracts were prepared as described in section 2.12.3. For each sample, the cell extract was aliquotted into two tubes, one for the isolation of polysome-bound mRNA and a second one for extraction of total mRNA.

2.17.3. EDTA treatment of cell extracts

In a control experiment, cell extracts from K562 cells were treated with ethylenediaminetetraacetic acid (EDTA, Ambion Molecular biology grade, pH 8.0, Thermo Fisher Scientific) at either 20mM or 30mM for 20 minutes at 4℃.

2.17.4. Isolation of polysome-mRNA complexes

Complexes of mRNA associated with one or more ribosome (polysome) were isolated from cell extracts by size-exclusion chromatography using Illustra MicroSpin S400 HR columns (GE Healthcare) (Su *et al.*, 2016). The columns were vortexed and centrifuged at 735 xg for 1 minute. Before centrifuging, columns were checked for the presence of air bubbles which could disrupt the resin. Between 25µl and 100µl of cell extract was slowly added to the centre of the column, which was then centrifuged at 735 xg for 2 minutes. The purified fraction containing polysome-mRNA complexes was collected at the bottom of the tube.

2.17.5. Isolation of total RNA and polysome-bound RNA

RBR was extracted from purified polysome-mRNA fractions while total mRNA was isolated from cell extracts as described in section 2.7.2.

2.17.6. DNase I treatment of total and polysome-bound RNA

Total mRNA and RBR were treated with DNase I and beads purified as described in section 2.8.

2.17.7. Assessment of RNA quality and quantity

RNA quality and quantity were assessed by Agilent RNA 6000 Pico Kit (Agilent Technologies) on the Agilent 2100 Bioanalyser System (Agilent Technologies). RNA integrity number (RIN) for each sample was calculated using the Agilent 2100 Expert Software (Agilent Technologies) and only samples with a RIN larger than 7 were utilised for NGS.

2.17.8. RNA sequencing

2.17.8.1. Library preparation

2.17.8.1.1. Ovation Human Blood RNA-Seq Library System

RNA was isolated from cultured erythroblasts as described in section 2.17.5 and quantified as in section 2.17.7. 100ng RNA was used for each sample to prepare libraries for RNA sequencing with the Ovation Human Blood RNA-Seq Library System (Nugen), according to the manufacturer's protocol. This low-input method for library preparation enriches libraries for non-rRNA and non-globin sequences, allowing for more efficient sequencing of RNA from human blood cells. The protocol comprises the following steps:

- First strand cDNA synthesis with random priming (without integrated DNase treatment)
- Second strand cDNA synthesis with nucleotide analogue
- cDNA fragmentation
- End repair
- Ligation with nucleotide analogue-marked adaptors
- Strand selection
- Insert Dependent Adaptor Cleavage (InDA-C)
- PCR amplification
- Purification of libraries using Agencourt RNAClean XP Purification Beads (Beckman Coulter)

Eight barcoded libraries were produced with this method and sequenced on a single flow cell lane. Integrity and size of libraries were measured using the Agilent High Sensitivity DNA Assay (Agilent Technologies) on the Agilent 2100 Bioanalyser System (Agilent Technologies).

Preparation of libraries for the same samples was repeated a second time using the Ovation Human Blood RNA-Seq Library System (Nugen) protocol described above except for the first strand cDNA synthesis step, which was carried out with integrated DNase treatment. DNase treatment was performed as outlined in the manufacturer's instructions (Nugen) using 1µl of HL-dsDNase (ArcticZymes) for 100ng RNA and incubating the sample at 37℃ for 10 minutes and 55℃ for 10 minutes.

2.17.8.1.2. SmartSeq2

Between 10pg and 100pg DNase I-treated RNA (see section 2.17.6) was used for RNA sequencing. Library preparation using the Smart-Seq2 protocol (Picelli *et al.*, 2014) was performed at the High-Throughput Genomics Centre (Wellcome Trust Centre for Human Genetics, University of Oxford). Smart-Seq2 library preparation was performed according to the protocol from (Picelli *et al.*, 2014)and adaptors were added to cDNA using the Nextera XT DNA sample preparation kit (Illumina).

2.17.8.2. Sequencing of libraries

Libraries for total RNA samples and RBR were sequenced at the High-Throughput Genomics Centre (Wellcome Trust Centre for Human Genetics, University of Oxford) on a HiSeq4000 Sequencing System (Illumina) using 75bp paired-end chemistry. Barcoded libraries were mixed and sequenced on a single lane of a flow cell. Illumina PhiX Control was added at approximately 1%. Cluster detection on Illumina sequencing platforms is optimized for libraries with a balanced representation of A, T, G, and C nucleotides. Illumina PhiX Control is a library made from PhiX virus DNA which is balanced and is often used as a control for cluster generation, sequencing and alignment during sequencing runs.

2.17.8.3. Analysis of RNA sequencing data

Sequencing data for gene expression and polysome profiling were analysed by running custom Bash scripts in a Linux terminal. Representative scripts for mapping, differential gene expression and calculation of translation efficiency (TE) are outlined in Appendix A, B and C. These custom Bash scripts include the commands outlined below for trimming of *fastq* files, mapping of trimmed *fastq* files, quality control, counting of reads mapping to genes, calculating transcripts per million, analysis of differential genes expression and calculating mRNA translation efficiency.

2.17.8.3.1. Trimming of *fastq* files

Sequencing reads were trimmed using Trim Galore (The Babraham Institute, https://github.com/FelixKrueger/TrimGalore), based on sequencing quality as well as presence of portion of the adapter sequences. The following command was used:

trim_galore -q 20 --paired -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACAACCAGATCTCGTATGCCGTCTTCTGCTTG -a2 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT fastqFileR1 fastqFilesR2

2.17.8.3.2. Mapping of trimmed *fastq* files

Fastq files were mapped to the *Homo sapiens* GRCh37/hg19 genome using the Spliced Transcripts Alignment to a Reference (STAR) software (Dobin *et al.*, 2013) with the following command:

STAR --genomeDir /databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/STAR --sjdbGTFfile /databank/igenomes/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf --outSAMtype BAM Unsorted --sjdbOverhang 74 --readFilesIn $i"_1_val_1.fq.gz"

$i"_2_val_2.fq.gz" --readFilesCommand zcat --outFilterScoreMinOverLread 0.44 --outFilterMatchNminOverLread 0.44 --outSAMattributes All --outFileNamePrefix $i"_" --runThreadN 6

2.17.8.3.3. Quality control

Quality control of mapped data was performed using RNAseQC (DeLuca *et al.*, 2012) on *fastq* files to analyse read metrics (such as read length, mapping and duplicate reads) and coverage (mean coverage and 5'/3' bias). Moreover, *bigwig* files were made in order to visualise each mapped BAM file using the UCSC Genome Brower (https://genome.ucsc.edu/) and verify the absence of DNA contamination.

2.17.8.3.4. Counting of reads mapping to genes

Sequencing reads mapping to genes were counted to estimate transcript abundance using featureCounts (Liao *et al.*, 2014) with the following command:

featureCounts -T 5 -a /databank/igenomes/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf -t exon -g gene_id -o featured_counted.txt -p -s 0 Aligned.sortedByCoord.out.bam 2> fc_exon.txt

2.17.8.3.5. Calculation of transcripts per million

Analysis of gene expression was initially performed on total RNA fractions using the Kallisto software (Bray *et al.*, 2016) (https://pachterlab.github.io/kallisto/) to estimate transcript abundance as transcripts per million (TPM). TPMs are a measurement of the estimated fraction of transcripts made up by a given isoform or gene. This value can be multiplied by 10^6 to obtain the number of transcripts per million. TPMs are calculated by diving the number of sequencing reads mapping to a transcript by the transcript length and multiplying that value by 10^6 (Conesa *et al.*, 2016). The following command was used:

kallisto quant -i indexed_reference_geneome.idx -o kallisto_output_file -b 100 R1_fastq R2_fastq

2.17.8.3.6. Differential gene expression

Analysis of differential gene expression was also performed using software DESeq2 (version 1.14.1) (Love *et al.*, 2014) using the custom R (version 3.3.1) script in Appendix B.

2.17.8.3.7. Gene Set Association Analysis

The software GSAA-seqSP (Xiong *et al.*, 2014) from the GSAA platform (version GSAA_2.0, http://gsaa.unc.edu/) was used to conduct GSEA for RNA-seq expression data with default parameters (Signal2Noise setting). Curated gene sets were retrieved from the Molecular Signatures Database (MSigDB version 6, http://www.broadinstitute.org/gsea/msigdb/).

2.17.8.3.8. Calculation of translation efficiency

Translation efficiency (TE) was calculated for each measured transcript using the custom script in Appendix C. Read counts per gene were calculated using featureCounts (section 2.17.8.3.4) and normalised using DEseq2 (section 2.17.8.3.4) for RBR and total RNA per sample. The TE of each gene was calculated as the ratio between the normalised read count per gene in the RBR fraction and the read count for the same gene in the total RNA fraction. Firstly, only the D-leucine treated samples from patients and controls were analysed: the TE was calculated per gene and the average TE per gene across controls and the average TE per gene across patients were generated. 1 was added to all TE values and the fold change (FC) for each gene was calculated by dividing the TE of each gene in patients by its TE in control samples. Finally the $\log_2$ of the FC value per gene was calculated and plotted. For the analysis of TE in L-leucine and D-leucine treated samples, the analysis was conducted on

patients and controls separately. The TE for each gene was calculated for each sample, 1 was added to the TE values and the FC in TE of each gene was calculated as the ratio between the TE for that gene in L-leucine treated samples and the TE in D-leucine treated samples. The $\log_2$ of the average FC for each gene across patients was calculated and plotted. The $\log_2$ FC for controls was displayed separately.

2.17.8.3.9. Analysis of 5' TOP and 5'TOP-like motifs

mRNAs were classified into 5'TOP and 5' TOP-like using a python script developed by Dr Jon Kerry (MRC Weatherall Institute of Molecular Medicine, Oxford) (https://github.com/jbkerry/NGS/blob/master/5top_mrna.py). The script classifies the transcripts based on the following definitions (Thoreen *et al.*, 2012):

-5'TOP mRNA if the transcript contains a stretch of 4 to 14 pyrimidines immediately upstream the transcription start site (TSS)

-5' TOP-like mRNA if the transcript contains at least 5 pyrimidines within 4 nucleotides from the TSS

The TSS used in the analysis were obtained from the UCSC database (hg19, https://genome.ucsc.edu/).

2.17.8.3.10. Plotting of RNA-seq data

Histograms and scattered plots were generated using the R package ggplot2 (http://ggplot2.org/). The DESeq2 software was used to generate MA-plots and Principal Component Analysis (PCA) while heatmaps were plotted in R (version 3.3.1) using the pheatmap package (https://cran.r-project.org/web/packages/pheatmap/index.html).

2.18. Flow cytometry

Data acquisition was performed on a BD LSRII instrument (BD Bioscience; Franklin Lakes, NJ, USA) and the data were analysed using FlowJo software version 10.

BD CompBeads Set Anti-Mouse Ig (Becton Dickinson UK) stained with each antibody were used for compensation. Unstained live cells and live and dead cells stained with either DAPI or eFluor 780 were used to set the gating for the live cells population. Fluorescence minus one (FMO) controls were used to determine the lowest level for which the antibody binding was considered negative and set the gating for the positive and negative populations for each fluorochrome.

2.18.1. Staining of cell surface markers

Erythroid differentiation was measured by staining of erythroid cell surface markers CD36, CD71 and CD235a at day 11 of culture. Cells were washed in PBS (Life Technologies) and incubated for 30 minutes on ice with PE anti-human CD36 (cat. 336206, Biolegend) anti-CD71 FITC (cat. 11-0719-42, eBioscience) and anti-CD235a APC (cat. 17-9987-42, eBioscience) antibodies. Cells were washed in PBS (Life Technologies), before staining with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) (1µg/ml, Sigma Aldrich).

2.18.2. Apoptosis analysis

Cells were washed in Annexin V buffer (BD Bioscience) and stained with Annexin V-FITC antibody for 20 minutes in the dark at room temperature according to the manufacturer instructions. Cells were then washed in PBS and 1 µg/ml Propidium Iodide (PI) (BD Bioscience) was added immediately before flow cytometry analysis.

2.18.3. Intracellular staining of alpha and beta globin

Mononuclear cells cells were harvested, washed once in PBS and stained with the Fixable Viability dye eFluor 780 (eBioscience) for 30 min. After PBS washing, cells were fixed and permeabilized using the FOXP3 Fix/Perm Buffer Set (eBioscience), according to the manufacturer's recommendations. The Anti-Hemoglobin antibody FITC (ab19361, Abcam) was diluted 1/200 and added to the cells. Cells were incubated for 30 minutes in the dark at room temperature. The Hemoglobin β Antibody (37-8) PE (sc-21757, Santa Cruz Biotechnology) dilutes 1/50 was also added to the cells and incubated for 15 min at 4˚C in the dark. Cells were then washed twice in PBS and and resuspended in Perm buffer (eBioscience) before flow cytometry analysis.

2.19. CRISPR design

To target Cas9 to genomic loci of interest, the sgRNAs listed in Table 2.9 were used. Three sgRNAs to target Cas9 to the *TP53* locus were designed using a publicly available online CRISPR Design Tool (http://crispr.mit.edu/) (Ran *et al.*, 2013). Cas9 was targeted to the *ASXL1* locus using a sgRNA previously designed and tested [gRNA1 from (Valletta *et al.*, 2015)].

Table 2.9. Single-guide RNAs (sgRNAs).

Region	gRNA	PAM sequence
TP53 sgRNA1	CTTAGTTGGGTGTCGACGTG	AGG
TP53 sgRNA2	GTTTTGTAGAACAACTCCCG	AGG
TP53 sgRNA3	TTGTTCTACAAAACGGGTTG	TGG
ASXL1	TCTAAATGGGGAGCACACTC	AGG

Single-stranded oligodeoxynucleotides (ssODN) were also designed to achieve editing of the locus of interest by homology-directed repair (HDR) mechanisms (Ran *et al.*, 2013) (Table 2.10). Silent nucleotide changes were introduced in the ssODN sequences in order to avoid Cas9-mediated cleavage of the CRISPR-edited site and in order to introduce barcodes to distinguish HDR (Ran *et al.*, 2013). In one experiment (Chapter 5 section 5.2.7) K562 and KBM5 cells were transfected with ssODNs for HDR-mediated editing of *TP53* or *ASXL1* mixed in equal ratios and the pX458 plasmid containing the sgRNA targeting the specific locus. The pX458 plasmid expressing Cas9 only (not containing any sgRNA) was used as a control. Each cell line was transfected twice using different ssODNs mixes and the same pX458 vector.

Table 2.10. Single-stranded oligodeoxynucleotides (ssODNs). ssODNs repair templates contained silent mutations (in bold), non-silent mutations (in red) and N nucleotides (in green) (being either a C, a G, a T or an A mixed in equal ratio).

Purpose	**ssODN sequence**	**length**
TP53 mutation correction	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCT**A**AACAAA**A**ATGTTTTGCC AACTT**T**GCCAAGACCTGCCCTGTGCAA**C**T**A**TGGGTTGATTCCACACCCCCGCCCGGCACCCGC GTCCGCGCCATGG	139
TP53 mutated with T7 sequence	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAA**A**ATGTTTTGCC AACTT**T**GTAATACGACTCACTATAGGG**AT**CCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTC CACACCCCCGCCCGGCACCCGCGTCCGCGCCATGG	161
TP53 wild type with T7 sequence	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAA**A**ATGTTTTGCC AACTT**T**GCCATACGACTCACTATAGGGAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCCACA CCCCCGCCCGGCACCCGCGTCCGCGCCATGG	157
TP53 mutated with barcodes of silent mutations	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAA**A**ATGTTTTGCC AACTT**T**GCTAAAACGTGTCCAGTCCAACTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGC GTCCGCGCCATGG	158
TP53 wild type with barcodes of silent mutations	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAA**A**ATGTTTTGCC AACTT**T**GCTAAAACGTGTCCAGTCCAA**T**CTGTGGGTTGATTCCACACCCCCGCCCGGCACCCG CGTCCGCGCCATGG	160
WT *TP53* ssODN	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAA**A**ATGTTTTGCC AACTT**T**GCCAAAACCTG**T**CCTGT**C**CAGCT**C**TGGGTTGATTCCACACCCCCGCCCGGCACCCGC GTCCGCGCCATGG	139

Mutant *TP53* ssODN	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAAATGTTTTGCC CAACTTGCTAAGACGGTGCCCAGTGCAACTGTGGGTTGATTCCACACCCCCGCCCGGCACCCG CGTCCGCGCCATGG	140
TP53 ssODN N (with N base)	ACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAAATGTTTTGCC AACTTGCTAAAACGGTGNCCAGTCCAACTCTGGGTTGATTCCACACCCCCGCCCGGCACCCGC GTCCGCGCCATGG	139
WT *ASXL1* ssODN	AGCGAACACAACTACTGCCGCCTTATCCTCTAAATGGGGAGCACACTCAGGCTGGAACTGCC ATGTCAAGGGCTAGGAGAGAGGACCTGC	90
Mutant *ASXL1* ssODN	AGCGAACACAACTACTGCCGCCTTATCCTCTAAATGGGGAGCACACTCAGGCTTGAACTGCC ATGTCAAGGGCTAGGAGAGAGGACCTGC	90

2.20. Cloning of sgRNAs into the pSpCas9(BB)-2A-GFP vector

Two complementary oligos (Sigma Aldrich) were designed in order to clone each sgRNA into the pX458 [pSpCas9(BB)-2A-GFP vector; Addgene plasmid ID: 48138, (Ran *et al.*, 2013)] expressing the Cas9 nuclease and Green Fluorescent Protein (GFP). Two overhanging sequences were added to each pair of sgRNA-specific oligos as shown below.

5'- CACCGNNNNNNNNNNNNNNNNNNN-3'

3'-CNNNNNNNNNNNNNNNNNNNCAAA -5'

pX458 plasmid was digested using *BbsI* (Thermo Fisher Scientific) restriction enzyme at 37℃ for 5 minutes and at 65℃ for 10 minutes.. 1µg of plasmid was mixed with 2µl of 10X FastDigest Buffer (Thermo Fisher Scientific), 1µl FastDigest BpiI (BbsI) (10 U/L, Thermo Fisher Scientific) and 15.75µl water. The digested plasmid was then run on a 1.5% agarose gel and gel extracted using the QIAquick Gel Extraction Kit (Qiagen). The two complementary oligos were phosphorylated, denatured and annealed. 1µl of each of the oligos was mixed with 1µl of 10X T4 DNA Ligase Reaction Buffer (NEB), 0.5µl T4 Polynucleotide Kinase (PNK) (NEB) and 6.5µl of water and incubated at 37℃ for 30 minutes, 95˚C for 5 minutes and ramp cooled to 25˚C at 5℃/minute to allow annealing. The double-stranded annealed oligo was diluted 1:250 and ligated with the linearized pX458 using T4 DNA Ligase (NEB). 50ng of linearized plasmid was added to 1µl of annealed oligo (1:250), 1.1µl of 10X T4 DNA Ligase Reaction Buffer (NEB), 4.6µl of water and 1µl T4 DNA Ligase (NEB) and incubated at room temperature for 10 min and 65℃ for 10 minutes. Competent cells (One Shot TOP10 Chemically Competent *E. coli*, Invitrogen) were thawed on ice for 30 minutes. 2µl of ligated plasmid was added and bacterial cells were incubated on ice for 30 minutes. Bacterial cells were transformed by heat-shock at 42˚C for 45 seconds

followed by immediate incubation on ice for 2 minutes. 200µl of pre-warmed SOC media (Invitrogen) was added to the cell vial and cells were placed in a shaking incubator for 30 minutes at 37˚C. 80µL of the bacterial culture was streaked on pre-warmed agar plates containing 50µg/mL of selection antibiotic (Ampicillin solution, Sigma Aldrich) and plates were incubated at 37˚C overnight. Single colonies were picked and incubated overnight in 3ml LB Broth (Thermo Fisher Scientific) supplemented with 50µg/ml Ampicillin (Sigma Aldrich) at 37˚C while shaking. DNA was extracted from bacterial cultures using the Miniprep kit (Qiagen) and the correct insertion of the sgRNA sequences was confirmed by Sanger sequencing, using the custom designed primer 5'-GGGGCGGAGCCTATGGAAAA-3' (see section 2.6). Bacterial colonies containing the vector with the correct insertion were then expanded in 150ml of LB Broth (Thermo Fisher Scientific) supplemented with 50µg/ml Ampicillin (Sigma Aldrich) at 37˚C overnight. Plasmid DNA was extracted using Maxiprep kit (Qiagen) and the correct insertion of the sgRNA sequences was verified using Sanger sequencing.

PX458 plasmid containing sgRNA targeting *ASXL1* was previously used in the study by Valletta *et al.* (2015) and it was a kind gift of Dr Simona Valletta (MRC Weatherall Institute of Molecular Medicine, University of Oxford). Bacterial colonies were recovered from glycerol stocks and expanded prior to plasmid DNA extraction using the Maxiprep kit (Qiagen).

2.21. Cell transfection of CRISPR/Cas9 system

Cells were transfected using the Nucleofector 2b Device (Lonza) and Amaxa Cell Line Nucleofector Kit V (Lonza) was utilised for K562 and KBM5 cells. $1x10^6$ cells were centrifuged at 200 x g for 10 minutes at room temperature. The supernatant was removed and cells were resuspended in 100µl of Nucleofector Solution (Lonza). The cell suspension

was then combined with plasmid DNA and/or ssODNs depending on the experiment and gently mixed. The sample was then transferred to a cuvette and inserted in the Nucleofector Device. K562 cells were nucleofected using the Nucleofector Program T-003 while KBM5 with Program T-016. After nucleofection, 500µl of warm media was added to the cuvette and cells were transferred to 12-well culture plates with 2ml media per well and incubated at 37℃, 5% CO_2.

The protocol was adapted from the manufacturer's transfection protocol for each cell line (Amaxa) and the amount of pX458 plasmid transfected was optimised. Both cell lines were transfected either 10µg, 5µg or 2.5µg of plasmid and cell viability was measured after 24 and 48 hours as well as production of GFP (assessed by fluorescent microscopy). K562 cells were transfected with 10µg of pX458 plasmid (either empty or containing one of the sgRNAs targeting *TP53*) and 2.5µl of 100µM ssODN or a mix of three ssODNs. KBM5 cells were transfected with 10µg of pX458 plasmid (either empty or containing the sgRNA targeting *ASXL1*) and 2.5µl of 100µM mix of two ssODNs.

2.22. Fluorescence activated cell sorting

Fluorescence activated cell sorting (FACS) of transfected cells was performed either by Dr Helen Ferry (Flow Sorting Facility, Experimental Medicine Division NDM, University of Oxford) or by Dr Kevin Clark (Flow Cytometry Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford). CRISPR/Cas9 transfected cells were sorted 42 hours after transfection into GFP positive and negative cells. In one experiment, GFP positive or GFP negative K562 cells were sorted into single cells in 96-well plates (repeated twice). Prior to FACS, cells were washed in PBS (Life Technologies), stained with DAPI (1 µg/ml) and resuspended in PBS at a concentration of 3 x 10^5 cells/ml. Sorting was performed

using a BD FACSAria instrument (BD Bioscience). A GFP negative, DAPI-stained control sample was used for gating of GFP negative and positive cell populations.

2.23. CRISPR-transfected cells culture and harvesting

Transfected GFP positive K562 cells were cultured for 96 hours before harvesting in order to perform the restriction digest assay described below or amplify HDR-modified template using the T7 sequence or a barcode of silent mutations.In order to investigate the timing of HDR-mediated modification of *TP53*, K562 cells transfected with ssODNs containing the T7 sequence were harvested every 6 hours after transfection for 42 hours (without sorting of GFP cells). In one experiment, CRISPR-transfected K562 cells were FACS sorted into single cells (section 2.22) and cultured for 12-14 days before harvesting and DNA extraction. K562 and KBM5 cells transfected with a mix of ssODNs (Chapter 5 section 5.2.7) were cultured for 12 days and cells were harvested at five timepoints (48 hours, 96 hours, 6 days, 8 days and 12 days after transfection) in order to extract genomic DNA.

2.24. High Resolution Melting Analysis

High Resolution Melting Analysis (HRMA) was performed on DNA extracted from K562 cells transfected with either empty pX458 plasmid (Addgene) or pX458 plasmid with sgRNA 1, 2 or 3. GFP positive transfected K562 cells were FACS sorted 48 hours after transfection (see section 2.22). DNA was amplified by PCR as described in section 2.5, in the following reaction containing LCGreen plus dye (BioFire Diagnostics):

5 µl 2x Hotshot Diamond PCR mix (Clent Lifescience)

0.2 µl 10 uM forward primer (200 nM final concentration)

0.2 µl 10 uM reverse primer (200 nM final concentration)

1 µl 10x LC Green

2.6 µl Nuclease-free water

1 µl DNA (total 10µg)

Each reaction was repeated three times. HRMA was performed in a Real-Time PCR Detection System machine (Bio-rad) using the following cycling program: 95°C 30 sec, 65°C 2 seconds, high resolution melting 65°C to 95°C in 0.2°C increments every 2 seconds. Melting curves of DNA from K562 cells trasfected with pX458 containing the three sgRNAs were compared to the melting curve generated from DNA of K562 cells transfected with the empty plasmid.

2.25. Restriction digest assay

K562 cells were transfected with the CRISPR/Cas9 system as described in section 2.21 and GFP positive cells were FACS sorted (see section 2.22). Genomic DNA was extracted from transfected and untransfected K562 cells (see section 2.4.2) 96 hours after transfection and PCR amplified (see section 2.5). The PCR product was then column purified using the Wizard SV Gel and PCR Clean-up System (Promega). 10µl PCR product was digested with 1µl BglI restriction enzyme (NEB), 7µl of nuclease-free water and 2µl of 10X NEBuffer 3 (NEB) for 3 hours at 37˚C. PCR products were then separated by electrophoresis on a 2% agarose gel.

2.26. Off-target sites analysis

To investigate potential off-target cleavage by Cas9:sgRNA in *TP53* mutation corrected Colony 1, we performed targeted Sanger sequencing of the top five exonic off-targets sites obtained from the CRISPR Design Tool (http://crispr.mit.edu/) for sgRNA 2 (Table 2.11). sgRNA 2 was predicted to have 402 possible off-target sites, of which 11 were located within

exons. Sanger sequencing was performed as described in section 2.6 following PCR amplification of sites of potential off-target activity (see section 2.5).

Table 2.11. List of top five off-target sites for sgRNA 2 sequenced. Genomic locus of each off-target site, off-target hit score, number of mismatches to the sgRNA-target sequence and primers used for PCR amplification are listed. Off-target hit scores from the CRISPR Design Tool are calculated by taking into account the number of mismatches to the sgRNA-target sequence and their position within the sequence.

UCSC gene	**locus**	**Off-target hit score**	**Number of mismatches [position]**
NM_002737	chr17:-64804111	0.4	4MMs [8:9:10:12]
NM_001127505	chr8:+17916335	0.3	4MMs [1:6:8:19]
NM_017424	chr22:+17662338	0.2	4MMs [1:4:8:14]
NM_014731	chr20:+3144508	0.1	4MMs [4:6:16:20]
NM_003128	chr2:-54856834	0.1	4MMs [2:3:14:15]

2.27. TOPO-TA cloning

DNA from K562 cells transfected with ssODNs containing barcodes of silent mutations was PCR amplified with primers binding the barcodes (Chapter 5 section 5.2.6, Figure 5.6) as described in section 2.5. PCR products were run on a 2% agarose gel, PCR purified using the QIAquick Gel Extraction Kit (Qiagen) and cloned into TOPO-TA vectors (Thermo Fisher Scientific). 4 µl of PCR product were added to 1 µl of Salt Buffer (Thermo Fisher Scientific) and 1 µl of TOPO-TA vector (Thermo Fisher Scientific). The reaction was incubated at room temperature for 20 minute and transferred on ice. Competent cells (One Shot TOP10 Chemically Competent *E. coli*, Invitrogen) were thawed on ice for 30 minutes. 2µl of TOPO-TA vector was added to bacterial cells and incubated on ice for 30 minutes. Bacterial cells transformation was performed by heat-shock at 42˚C for 30 seconds followed

by immediate incubation on ice for 2 minutes. 250µl of pre-warmed SOC media (Invitrogen) was added to the cell vial and cells were placed in a shaking incubator for 1 hour at 37˚C. 20µL of the bacterial culture was streaked on pre-warmed agar plates containing 50µg/mL of Ampicillin solution (Sigma Aldrich) and plates were incubated at 37˚C overnight. Single colonies were picked and incubated overnight in 3ml LB Broth (Thermo Fisher Scientific) supplemented with 50µg/ml Ampicillin (Sigma Aldrich) at 37˚C while shaking. DNA was extracted from bacterial cultures using the Miniprep kit (Qiagen) and HDR-modified DNA templates were analysed by Sanger sequencing, using
M13 forward (5´-GTAAAACGACGGCCAG-3´) and M13 reverse (5´-CAGGAAACAGCTATGAC-3') primers (Thermo Fisher Scientific).

2.28. Cell growth assay

Cells were cultured at an initial density of $5x10^4$/mL, and viable cell counts were determined by trypan blue exclusion for 7 consecutive days. Trypan Blue Solution 0.4% (Thermo Fisher Scientific) was used.

2.29. Targeted NGS sequencing of the CRISPR-edited locus

2.29.1. Library preparation

DNA was extracted from K562 cells and KBM5 cells transfected with a mix of ssODNs (section 2.21). DNA concentration was measured using a NanoDrop (Thermo Fisher Scientific). DNA was also extracted from untransfected cells and sequenced as a control. Libraries for DNA sequencing were prepared according to a customised protocol from (Shalem *et al.*, 2014)adapted by Dr Qianxin Wu (MRC Weatherall Institute of Molecular Medicine, Oxford). Illumina adaptors sequences were added to DNA amplicons in a two-

step PCR reaction using custom designed primers containing portions of Illumina adaptors sequences (Figure 2.1).

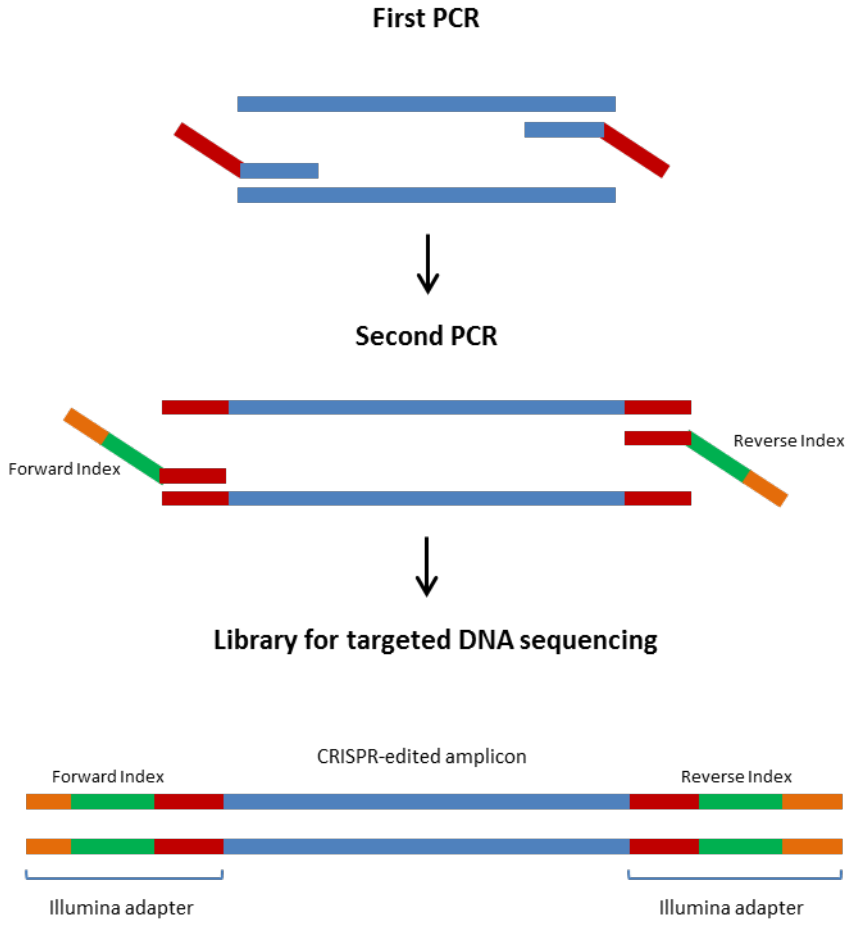


Figure 2.1. Two-step PCR method for preparation of dual-indexed libraries for DNA sequencing.

Primers for the first PCR contained the primer binding site on the DNA as well as part of the Illumina adapter sequences which served as binding sites for the primers for the second PCR (Table 2.12). For the *TP53* library only, primers for the first PCR also included barcodes, which are sequences of variable length to increase library complexity and to allow multiplexing of samples with the same index. Primers for the second PCR were used to add the remaining portions of the adapter sequences and they contained Illumina TruSeq index sequences for multiplexing of different biological samples (Table 2.13). The *TP53* library

was single indexed while the library for sequencing of the *ASXL1* amplicon was dual indexed.

Table 2.12. Primers for the first PCR for preparation of DNA-seq libraries.

Library	Sample	First PCR Forward Primer	First PCR Reverse Primer
TP53	Cas9 only 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCATGATCGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATCTCACAACCTCCGTCATGT
	Transfection 1 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCGTTACCACACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGACTCACAACCTCCGTCATGT
	Transfection 2 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATTCCTTGGTCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGCTCACAACCTCCGTCATGT
	Cas9 only 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTGATCGATAACGCATTCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCCTCACAACCTCCGTCATGT
	Transfection 1 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCGATACAGGTATCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCTCACAACCTCCGTCATGT
	Transfection 2 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTACGATCGATAGGTAAGGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCACAACCTCCGTCATGT
	Cas9 only 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAACAATGGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATCGTCTCACAACCTCCGTCATGT

	Transfection 1 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTATACTGTATCCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATCGCTCACAACCTCCGTCATGT
	Transfection 2 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTGATAGGTCGCACACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATCCTCACAACCTCCGTCATGT
	Cas9 only 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAAGTAGAGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATCTCACAACCTCCGTCATGT
	Transfection 1 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTATACACGATCCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGACTCACAACCTCCGTCATGT
	Transfection 2 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTGATCGCGCGGTCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGCTCACAACCTCCGTCATGT
	Cas9 only 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCATGATCGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCCTCACAACCTCCGTCATGT
	Transfection 1 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCGTTACCACACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCTCACAACCTCCGTCATGT
	Transfection 2 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATTCCTTGGTCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCACAACCTCCGTCATGT
	ssODN mix 1	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAAGTAGAGCTCCTCTGTCTCCTTCCTCTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

	ssODN mix 2	CACTCTTTCCCTACACGACGCTCTTCCGATCTATACACGATCCTCCTCTGTCTCCTTCCTCTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	A375 cells untransfected	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCATGATCGCACTTGTGCCCTGACTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCACAACCTCCGTCATGT
ASXL1	Cas9 only 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 1 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 2 48h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Cas9 only 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 1 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 2 96h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Cas9 only 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 1 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT

	Transfection 2 6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Cas9 only 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 1 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 2 8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Cas9 only 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 1 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	Transfection 2 12d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT
	ssODN mix 1	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACACAACTACTGCCGCCTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTCCTCTCTCCTAGCCCTT
	ssODN mix 2	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACACAACTACTGCCGCCTT	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTCCTCTCTCCTAGCCCTT
	KBM5 cells untransfected	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTGATGGTGGTGAGG	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGAGTTGGGAGGCATCT

Table 2.13. Primers for the second PCR for preparation of DNA-seq libraries.

Library	Sample	Second PCR Forward Primer	Second PCR Reverse Primer
TP53	Cas9 only 48h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTGACCAGTG ACTGGAGTTCAGACGTGTG
	Transfection 1 48h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTGACCAGTG ACTGGAGTTCAGACGTGTG
	Transfection 2 48h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTGACCAGTG ACTGGAGTTCAGACGTGTG
	Cas9 only 96h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATCACTGTGTGA CTGGAGTTCAGACGTGTG
	Transfection 1 96h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATCACTGTGTGA CTGGAGTTCAGACGTGTG
	Transfection 2 96h	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATCACTGTGTGA CTGGAGTTCAGACGTGTG
	Cas9 only 6d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATATTGGCGTG ACTGGAGTTCAGACGTGTG
	Transfection 1 6d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATATTGGCGTG ACTGGAGTTCAGACGTGTG
	Transfection 2 6d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATATTGGCGTG ACTGGAGTTCAGACGTGTG
	Cas9 only 8d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATGATCTGGTG ACTGGAGTTCAGACGTGTG

	Transfection 1 8d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATGATCTGGTG ACTGGAGTTCAGACGTGTG
	Transfection 2 8d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATGATCTGGTG ACTGGAGTTCAGACGTGTG
	Cas9 only 12d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTACAAGGTG ACTGGAGTTCAGACGTGTG
	Transfection 1 12d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTACAAGGTG ACTGGAGTTCAGACGTGTG
	Transfection 2 12d	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATTACAAGGTG ACTGGAGTTCAGACGTGTG
	ssODN mix 1	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATCGTGATGTG ACTGGAGTTCAGACGTGTG
	ssODN mix 2	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATCGTGATGTG ACTGGAGTTCAGACGTGTG
	K562 cells untransfected	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATGCCTAAGTG ACTGGAGTTCAGACGTGTG
ASXL1	Cas9 only 48h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGT GACTGGAGTTCAGACGTGT
	Transfection 1 48h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATGCGCATTAGT GACTGGAGTTCAGACGTGT
	Transfection 2 48h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATCATAGCCGG TGACTGGAGTTCAGACGTGT

	Cas9 only 96h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATTTCGCGGAGT GACTGGAGTTCAGACGTGT
	Transfection 1 96h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATGCGCGAGAG TGACTGGAGTTCAGACGTGT
	Transfection 2 96h	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATCTATCGCTGT GACTGGAGTTCAGACGTGT
	Cas9 only 6d	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATCGAGTAATG TGACTGGAGTTCAGACGTGT
	Transfection 1 6d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGT GACTGGAGTTCAGACGTGT
	Transfection 2 6d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATAATGAGCGG TGACTGGAGTTCAGACGTGT
	Cas9 only 8d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATGGAATCTCGT GACTGGAGTTCAGACGTGT
	Transfection 1 8d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATTTCTGAATGT GACTGGAGTTCAGACGTGT
	Transfection 2 8d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATACGAATTCGT GACTGGAGTTCAGACGTGT
	Cas9 only 12d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGT GACTGGAGTTCAGACGTGT
	Transfection 1 12d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATGCGCATTAGT GACTGGAGTTCAGACGTGT

	Transfection 2 12d	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATCATAGCCGG TGACTGGAGTTCAGACGTGT
	ssODN mix 1	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATTTCGCGGAGT GACTGGAGTTCAGACGTGT
	ssODN mix 2	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATGCGCGAGAG TGACTGGAGTTCAGACGTGT
	KBM5 cells untransfected	AATGATACGGCGACCACCGAGATCTACAC CCTATCCTACACTCTTTCCCTACACGAC	CAAGCAGAAGACGGCATACGAGATCTATCGCTGT GACTGGAGTTCAGACGTGT

The first PCR was performed using 100ng of genomic DNA from CRISPR-transfected cells, amplified for 20 cycles. 15µl of amplicon from the first PCR reaction was added to the second PCR reaction (in a 30µl reaction) and amplified for 15 cycles. For each sample, separate 10 ul reactions were performed and the resulting amplicons were combined. Different DNA polymerases and PCR cycling conditions were utilised depending on the CRISPR-edited amplicon sequenced (Table 2.14).

Table 2.14. Cycling conditions of PCR reactions for preparation of DNA-seq libraries.

DNA Sequencing Library	**First PCR**		**Second PCR**	
	Polymerase	**Cycling condition**	**Polymerase**	**Cycling condition**
TP53	Phusion High-Fidelity DNA Polymerase (NEB), 5X Phusion HF Buffer	98°C 30 s, 20 cycles at 98°C 10 s, 57°C 15 s, 72°C 15 s and 72°C 5 min	Phusion High-Fidelity DNA Polymerase (NEB), 5X Phusion GC Buffer	98°C 30 s, 15 cycles at 98°C 10 s, 62°C 15 s, 72°C 15 s and 72°C 5 min
ssODN mix *TP53*	Phusion High-Fidelity DNA Polymerase (NEB), 5X Phusion HF Buffer	98°C 30 s, 15 cycles at 98°C 10 s, 53°C 15 s, 72°C 15 s and 72°C 5 min	Phusion High-Fidelity DNA Polymerase (NEB), 5X Phusion GC Buffer	98°C 30 s, 15 cycles at 98°C 10 s, 62°C 15 s, 72°C 15 s and 72°C 5 min
ASXL1	Kapa HiFi HotStart (Kapa Biosystems), 5X KAPA HiFi Fidelity Buffer	95°C 3 min, 20 cycles at 98°C 20 s, 67°C 15 s, 72°C 15 s and 72°C 1 min	Kapa HiFi HotStart (Kapa Biosystems), 5X KAPA HiFi Fidelity Buffer	95°C 3 min, 20 cycles at 98°C 20 s, 62°C 15 s, 72°C 15 s and 72°C 1 min
ssODN mix *ASXL1*	Kapa HiFi HotStart (Kapa Biosystems), 5X KAPA HiFi Fidelity Buffer	95°C 3 min, 15 cycles at 98°C 20 s, 63°C 15 s, 72°C 15 s and 72°C 1 min	Kapa HiFi HotStart (Kapa Biosystems), 5X KAPA HiFi Fidelity Buffer	95°C 3 min, 20 cycles at 98°C 20 s, 62°C 15 s, 72°C 15 s and 72°C 1 min

In order to control for errors in transfection and mixing of different ssODNs, ssODNs mixes were also sequenced. 100ng of ssODN mix was added to the first PCR reaction and amplified for 15 cycles and 6.6µl of amplicon from the first PCR amplicon was added to the second PCR (total 10µl reaction) and amplified for 15 cycles.

For each sample, resulting amplicons from the second PCR were run on a 2% agarose gel at 110V for 1 hour, gel purified using the QIAquick Gel Extraction Kit (Qiagen), column purified with the MiniElute PCR Purification Kit (Qiagen) and cleaned up using Agencourt AMPure XP beads (Beckman Coulter). Libraries were then quantified using Qubit dsDNA HS Assay (Invitrogen) according to the manufacturer's protocol and barcoded libraries were pulled together at a final concentration of 16nM. Pulled libraries were quantified a second time using KAPA Library Quantification Kit for Illumina platforms (Kapa Biosystems) and further diluted to 4nM. The integrity and size of the libraries were measured using the High Sensitivity D1000 ScreenTape System (Agilent Technologies) on a 4200 TapeStation System (Agilent Technologies), according to the manufacturer's instructions.

2.29.2. Sequencing

Targeted sequencing of the CRISPR-edited *TP53* amplicon was performed on a MiSeq (Illumina) using a MiSeq Reagent Kit v2 (300 cycles, 150bp paired-end sequencing chemistry). CRISPR-edited *ASXL1* targeted sequencing was performed at the Centre for Genome Engineering (Seoul National University, Seoul, Corea) in collaboration with Dr Omer Habib (Professor Jin-Soo Kim's group). A MiniSeq (Illumina) was utilised, libraries were sequenced with 150bp paired-end chemistry and PhiX DNA was added at 50%.

2.29.3. Analysis of raw DNA sequencing data

Custom Bash scripts were written for the analysis of the DNA sequencing data from CRISPR-edited K562 cells and KBM5 cells. The same analysis pipeline was applied for the analysis of both amplicons and a representative Bash script is outlined in Appendix D. All scripts were run in a Linux terminal.

2.29.3.1. Quality control

Fastq files were analysed using the FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) software for quality control.

2.29.3.2. Trimming *fastq* files

Sequencing reads were trimmed using the package Sickle (https://github.com/najoshi/sickle) in order to remove bases with low sequencing quality. The following command was used:

sickle pe -f R1.fastq -r R2.fastq -t sanger -o R1_trim.fastq.gz -p R2_trim.fastq.gz -s R2_singles.fastq.gz -q 28 -l 50 -n

2.29.3.3. Mapping *fastq* files

Trimmed *fastq* files were then mapped to the *Homo sapiens* GRCh37/hg19 genome using Burrows-Wheeler Aligner (BWA) MEM (http://bio-bwa.sourceforge.net/) with the following command:

bwa mem

/databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/BWAIndex/genome.fa

R1_trim.fastq.gz R2_trim.fastq.gz > mapped.BWA.sam

Sequence Alignment Map (SAM) files of reads mapping to each amplicon analysed by targeted sequencing were also extracted from the outputted SAM file. The regions extracted were chr17:7578399-7578637 for *TP53* and chr20:31022540-31022729 for *ASXL1*.

2.29.3.4. Counting CRISPR-edited reads

The SAM file was then split into different SAM files containing the barcodes corresponding to each ssODN template transfected and reads in each file were counted using samtools view (Li, 2011) (http://samtools.sourceforge.net/). Barcodes used to isolate edited sequencing reads for *TP53* are outlined in Table 2.15 and barcodes used for *ASXL1* reads are in Table 2.16.

Table 2.15. Barcodes for analysis of CRISPR-edited sequences of *TP53*.

Sample	Barcode
WT *TP53* ssODN and *TP53* ssODN "N"	G C C A A C T T
Mutant *TP53* ssODN	G C C C A A C T T
WT *TP53* ssODN	A C C T G T
TP53 ssODN "N" = A	A C G T G A
TP53 ssODN "N" = C	A C G T G C
TP53 ssODN "N" = T	A C G T G T
TP53 ssODN "N" = G	A C G T G G

Table 2.16. Barcodes for analysis of CRISPR-edited sequences of *ASXL1*.

Sample	Barcode
WT *ASXL1* ssODN	T C A G G C T G
Mutant *ASXL1* ssODN	T C A G G C T T

2.29.3.5. Analysis of additional mutations in CRISPR-edited reads

Samtools mpileup (Li, 2011) (http://samtools.sourceforge.net/) was utilised to create variant calling files (*vcf*) and characterise mutations as single nucleotide polymorphisms (SNPs) or small insertions and deletions (INDELs) with the *Homo sapiens* GRCh37/hg19 genome. Reads were then classified as either containing only silent mutations included in the transfected ssODNs or containing additional mutations, using a script developed by Dr Jon Kerry (MRC Weatherall Institute of Molecular Medicine, Oxford) (https://github.com/jbkerry/NGS/blob/master/crispr_tools.py). This script incorporates ENSEMBL Variant Effect Predictor (McLaren *et al.*, 2016) (VEP, http://www.ensembl.org/info/docs/tools/vep/index.html) to classify reads as follows:

-Deleterious to protein function if the SIFT score was lower than 0.05 or the PolyPhen score was between 0.15 and 1

-Non-deleterious to protein function if the SIFT score was higher or equal to 0.05 or the PolyPhen score was lower than 0.05

2.30. Statistical analysis

2.30.1. Flow cytometry and ELISA

Statistical analysis was performed using GraphPad Prism Software (GraphPad Software). Flow cytometry data for expression of erythroid cells surface markers and intracellular staining of alpha and beta globin were analysed using a two-way ANOVA to compare the difference in protein phosphorylation between 5q- patients and healthy controls and between L-leucine and D-leucine treated samples. Two-way ANOVA was also used for the analysis of data from ELISA experiments. A p value <0.05 was considered significant.

2.30.2. Analysis of translation efficiency

Statistical analysis was performed using custom R scripts (version 3.3.1). Translation efficiency (TE) and log_2FC of translation efficiency were calculated as described in section 2.17.8.3.8 . A z score was calculated by dividing the difference between the TE for one gene and the mean TE of all genes, by the standard deviation. Genes with z scores lower than -1.5 or higher than 1.5 were considered differentially translated.

A two-tailed Mann–Whitney U test was used to analyse the difference in TE of all genes between 5q- patients and healthy controls and between L-leucine and D-leucine treated samples. A p value <0.05 was considered significant.

A two-tailed Mann–Whitney U test was also used to analyse the difference between the log_2FC of the TE of all genes and the log_2FC TE of 5'TOP and 5'TOP-like mRNAs. This statistical analysis was performed using custom R scripts (version 3.3.1) and significance was determined by a p value <0.05.

3. Chapter 3 – Analysis of *CSNK1A1* mutations and gene expression in myelodysplastic syndromes with del(5q)

3.1. Introduction

The *CSNK1A1* gene maps to the commonly deleted region (CDR) of the 5q- syndrome at 5q32 (Boultwood *et al.*, 2002, Boultwood *et al.*, 2007) and it encodes the serine/threonine kinase Casein Kinase 1 alpha (CK1α), which has a regulatory role in several cellular processes, including the Wnt/β-catenin pathway (Knippschild *et al.*, 2014, Schittek and Sinnberg, 2014). In the Wnt/β-catenin pathway, CK1α is part of the β-catenin destruction complex (Cheong and Virshup, 2011) and primes β-catenin for ubiquitination and degradation (Knippschild *et al.*, 2014). Therefore, down-regulation of CK1α leads to an accumulation of cytoplasmic β-catenin (Schittek and Sinnberg, 2014). In the presence of the Wnt ligand, β-catenin translocates into the nucleus (Knippschild *et al.*, 2014) where it activates transcription of genes involved in cell cycle progression, cell proliferation and differentiation processes (Schittek and Sinnberg, 2014). *CCND1* (encoding cyclin D1) is an important transcriptional target of β-catenin, involved in regulation of the G1/S cell cycle checkpoint (Alao, 2007). *CSNK1A1* is a tumour suppressor gene which controls cell proliferation through regulation of β-catenin and the p53 pathway in cancer (Elyada *et al.*, 2011, Sinnberg *et al.*, 2010). It has been reported that lenalidomide induces the ubiquitination and consequent degradation of CK1α by the E3 ubiquitin ligase CUL4–RBX1–DDB1–CRBN (CRL4CRBN), and that haploinsufficiency of *CSNK1A1* increases lenalidomide sensitivity in del(5q) hematopoietic cells (Fink *et al.*, 2014, Kronke *et al.*, 2015). *CSNK1A1* was found to be haploinsufficient in the $CD34^+$ cells of 5q- syndrome patients in a gene expression profiling study (Boultwood *et al.*, 2007). Screening of the remaining copy of the genes within the CDR by Sanger sequencing did not identify any

mutations in a cohort of ten 5q- syndrome patients (Boultwood *et al.*, 2007). However, in a study by Schneider *et al*. (2014), mutations of *CSNK1A1* were identified using whole-exome sequencing with an overall frequency of approximately 7%.

We analysed a cohort of MDS patients with del(5q) to investigate whether, as suggested by previous studies, *CSNK1A1* plays a central role in the pathophysiology of this disease. Firstly, the patient cohort was screened for *CSNK1A1* mutations. Secondly, the impact of *CSNK1A1* haploinsufficiency and mutation on the expression of β-catenin-related genes was investigated in the $CD34^+$ cells from MDS patients with del(5q). The results of this study were published in the British Journal of Haematology in 2015 (Bello *et al.*, 2015).

3.2. *CSNK1A1* mutations are rare events in MDS with del(5q)

Sanger sequencing was performed to determine the frequency of *CSNK1A1* mutations in a cohort of 104 cases of MDS with del(5q) (Methods section 2.1.1), 45 of which had 5q-syndrome (Methods Table 2.1). Mutations of *CSNK1A1* were previously identified in del(5q) MDS in exon 3 (Schneider *et al.*, 2014b) and in exon 4 (Graubert *et al.*, 2012, Woll *et al.*, 2014). Therefore, this investigation focused on the analysis of the sequences of these two exons of *CSNK1A1*. Missense mutations of *CSNK1A1* were identified in five del(5q) MDS cases in the cohort (frequency 4.8%) (Table 3.1, Figure 3.1A). Three of the mutations identified were located in exon 3 (involving amino acid 98) whereas two were located in exon 4 (amino acids 134 and 140). Unlike the amino acid mutations E98 and D140, the mutation at 134 (H134L) had not been previously reported. Sorting intolerant from tolerant (SIFT) and polymorphism phenotyping (PolyPhen) are two software designed to predict whether an amino acid substitution is likely to have a deleterious effect on protein function, based on the sequence homology of related proteins throughout evolution (Adzhubei *et al.*, 2010, Ng and Henikoff, 2003, Flanagan *et al.*, 2010). Moreover, PolyPhen predictions are

also based on the location of the amino acid substitution within the structure of the protein and its domains. PolyPhen-2 scores are from 0 (benign effect of amino acid change) to 1 (deleterious). SIFT scores instead go from 1 to 0, with 1 describing a non-deleterious amino acid change (Flanagan *et al.*, 2010). The *CSNK1A1* mutations identified were analysed using the PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and SIFT (http://sift.jcvi.org/) online tools, in order to predict the effect of the mutations on protein function. All *CSNK1A1* mutations, including the newly identified H134L mutation, were reported as damaging by PolyPhen-2 and SIFT analysis (Table 3.1), with the mutation affecting amino acid H134 having the highest Polyphen-2 score amongst the *CSNK1A1* mutations identified. All mutations were located in the protein kinase domain of CK1α and one mutation affected one of the protein catalytic residues and an ATP and ADP binding site (D140) (Figure 3.1B).

Table 3.1. *CSNK1A1* mutations identified in del(5q) MDS patients. Details of patient's diagnosis, karyotype, PolyPhen-2 and SIFT prediction of the effect of the mutations on protein function.

Patient ID	Diagnosis	Karyotype	*CSNK1A1* mutation	PolyPhen-2 prediction/score	SIFT prediction/score
Patient 05	RA	46,XX,t(1;3)(p33;p14),del(5)(q14q34)[21]/46XX[4]	c.401A>T p.H134L	Probably damaging/1	Damaging/0
Patient 07	RA	46,XX,del(5)(q14q34),inv(9)(p11q13)c[30]	c.293A>G p.E98G	Probably damaging/0.999	Damaging/0
Patient 14	RA (5q- syndrome)	46,XX,del(5)(q13q33)[26]/46,XX[4]	c.419A>C p.D140A	Possibly damaging/0.877	Damaging/0
Patient 36	RA (5q- syndrome)	46,XX, del(5)(q?)[30]	c.292G>A p.E98K	Probably damaging/0.999	Damaging/0
Patient 72	RA	46,XX,del(5)(q?),del(7)(q?)[30]	c.292G>A p.E98K	Probably damaging/0.999	Damaging/0

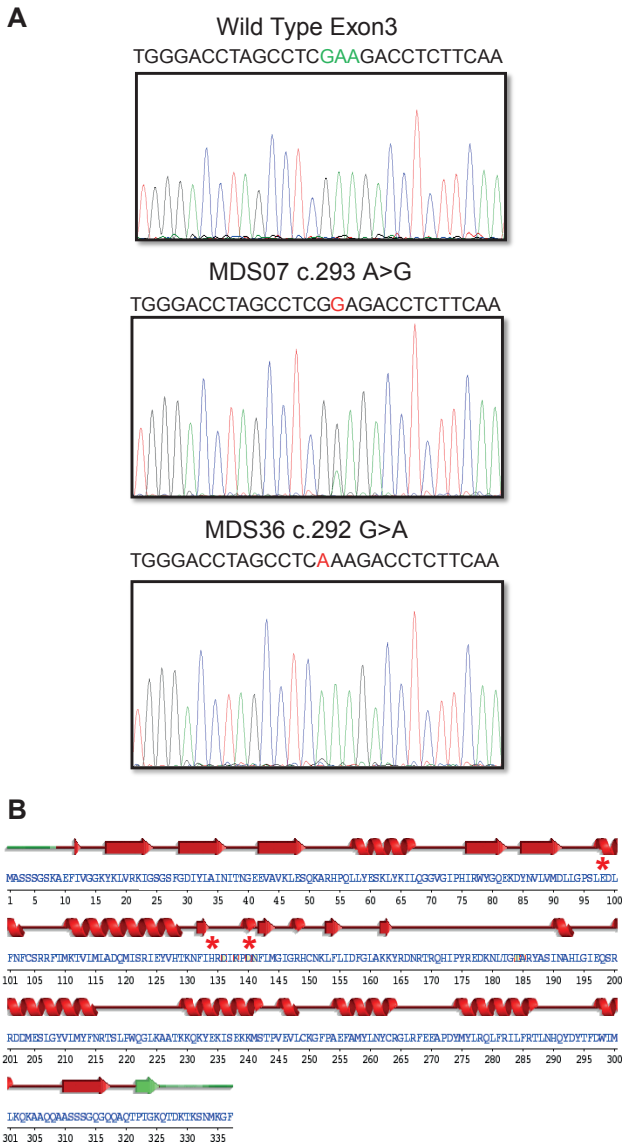


Figure 3.1. *CSNK1A1* mutations in del(5q) MDS patients. A) Representative Sanger sequencing traces from del(5q) MDS patients harbouring *CSNK1A1* mutations in exon 3 of the gene. The top panel shows the wild-type sequence (wild-type codon in green). The bottom traces are from two del(5q) MDS patients with *CSNK1A1* mutations highlighted in red. **B)** Experimentally determined secondary structure of the protein CK1α (UniProt accession number P48729_KC1A_HUMAN) (Petzold *et al.*, 2016). Helices represent alpha helices and arrows represent beta sheets. Red asterisks indicate the location of the mutations identified.

Next generation sequencing data from a previous study conducted in our group (Fernandez-Mercado *et al.*, 2013), using a panel targeting 25 genes mutated in various myeloid malignancies, were available for 4 out of 5 patients with *CSNK1A1* mutations and for 37 patients without mutation of *CSNK1A1*. The association of *CSNK1A1* mutations with other myeloid gene mutations in the panel was investigated. The additional mutations found in the cases with *CSNK1A1* mutations were a *RUNX1* mutation in one patient and a *U2AF1* mutation in another patient but there was no specific association of *CSNK1A1* mutations in the cohort with other myeloid gene mutations. All five patients harbouring *CSNK1A1* mutations had RA and two of them had the 5q- syndrome (Table 3.1). One of the cases with the *CSNK1A1* mutation E98K [Patient 72, carrying a del(5q) and a del(7q), Table 3.1] received lenalidomide and did not respond to treatment.

3.3. Haploinsufficiency of *CSNK1A1* is associated with increased expression of Wnt/β-catenin effector genes

Gene expression profiling (GEP) data from a previous study on MDS $CD34^+$ cells (Pellagatti *et al.*, 2010a) were re-analysed (analysis performed by Dr Andrea Pellagatti) to determine whether reduced expression of *CSNK1A1* in patients with del(5q) is associated with increased expression of the major downstream effectors of the Wnt/β-catenin signalling pathway. Gene expression profiling confirmed that patients with 5q- show haploinsufficient levels of *CSNK1A1* (Table 3.2). Moreover, expression levels of the Wnt/β-catenin target gene *CCND1* were found to be increased by approximately 50% in patients with 5q- (Table 3.2). cDNA was available for three patients with del (5q) included in the GEP study. RT-qPCR analysis of *CSNK1A1* and *CCND1* expression in these three patients confirmed the findings of the GEP experiments (Figure 3.2A). These data show

that haploinsufficiency of *CSNK1A1* is associated with increased expression of the Wnt/β-catenin transcriptional target *CCND1* in the HSC of MDS patients with del(5q).

Table 3.2. The average fold change in expression of *CSNK1A1* and *CCND1* measured by GEP. Fold changes were calculated by dividing the median expression of *CSNK1A1* and *CCND1* in patients by the median expression of these genes in healthy controls. MDS del(5q) patients n=30, 5q- syndrome patients n=16 and healthy controls n=17.

	LogFC *CSNK1A1*	**LogFC *CCND1***
MDS Del(5q)	0.58	1.42
5q- syndrome	0.56	1.58

3.4. *CSNK1A1* mutations in del(5q) MDS lead to increased expression of genes involved in Wnt signalling

GEP data were available for four patients with del(5q) with known mutational status of *CSNK1A1*: two patients harbouring a *CSNK1A1* mutation and two patients wild-type for *CSNK1A1*. Gene Set Enrichment Analysis (GSEA) was performed (analysis performed by Dr Andrea Pellagatti) to compare the gene expression profiles of the two patients with mutation of *CSNK1A1* with those of the two patients without mutations of this gene, in order to determine whether coordinated upregulation of pathways and processes associated with Wnt/β-catenin function could be observed. The output of GSEA includes a q-value which represents the false discovery rate (FDR) adjusted p-value for each gene set (q-values lower than 0.25 are considered significant) (Subramanian *et al.*, 2005). The "REACTOME_SIGNALING_BY_WNT" gene set was found to be significantly up-regulated ($q<0.001$) in the patients with mutation of *CSNK1A1* compared with the patients without mutations of this gene (Table 3.3 and Figure 3.2B, 3.2C). Moreover, GSEA identified a number of cell cycle related gene sets significantly upregulated ($q<0.01$) in

patients with *CSNK1A1* mutations compared with the patients without *CSNK1A1* mutations (Table 3.3). This observation is in line with the role of the Wnt signalling pathway in transcriptional regulation of genes which control cell cycle progression, such as *CCND1* (Alao, 2007, Stolz and Bastians, 2015). These data suggest that *CSNK1A1* mutations in del(5q) MDS may lead to an increase in the expression of genes regulated by the Wnt signalling.

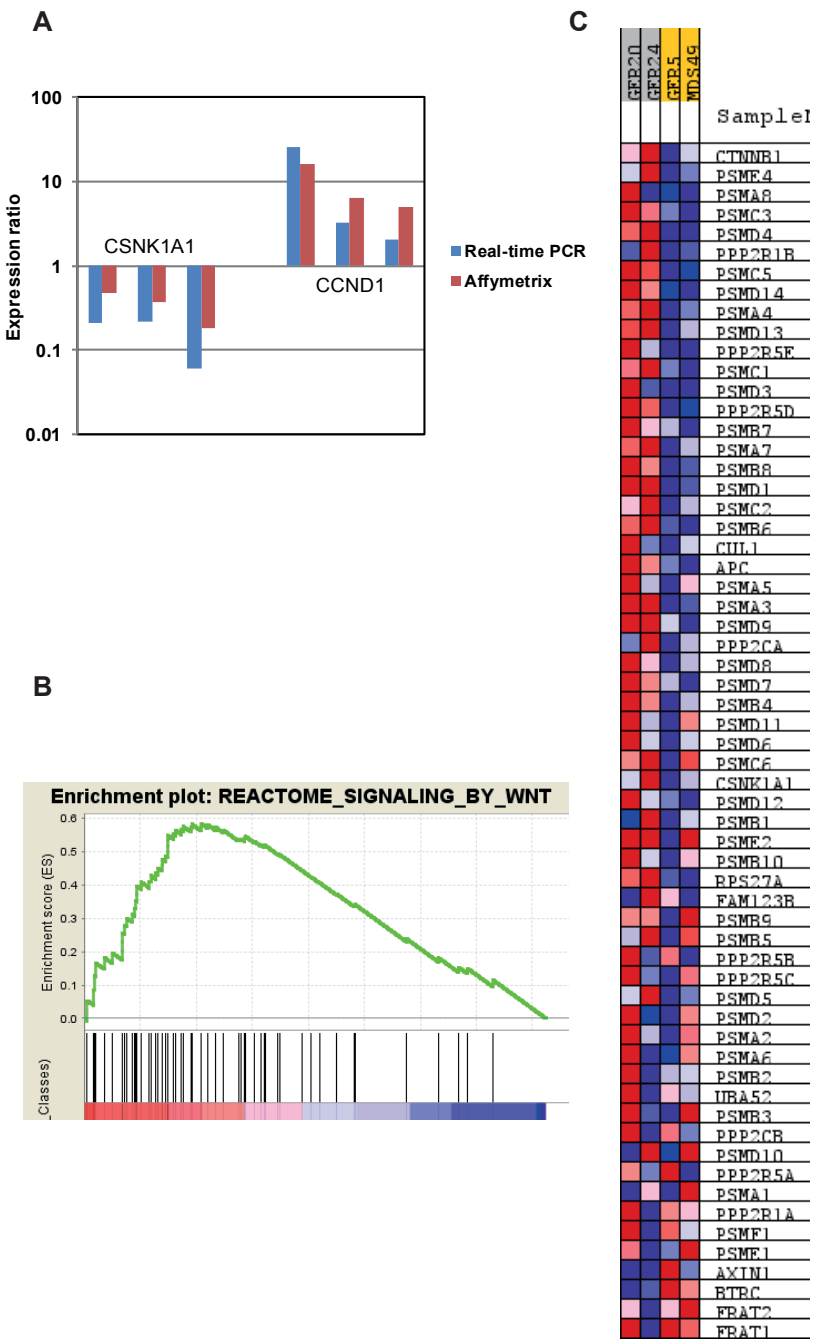


Figure 3.2. Expression of Wnt targets in del(5q) MDS patients with and without *CSNK1A1* mutations. A) Values for fold change in expression of *CSNK1A1* and *CCND1* in three patients with del(5q) MDS measured by Real-Time PCR (RT-qPCR) (blue bars) (ΔΔCT method, values normalised to expression in three healthy controls) and Affymetrix (red bars). n=3. **B)** Enrichment plot of the 'REACTOME_SIGNALING_BY_WNT' gene set from analysis using GSEA. n=2. C) GSEA-generated heat map showing the expression levels (red=high, blue=low) of the genes in the 'REACTOME_SIGNALING_BY_WNT' gene set in two del(5q) MDS cases with *CSNK1A1* mutations and two del(5q) MDS cases wild-type for *CSNK1A1*.

Table 3.3. Cell cycle related gene sets identified by GSEA. GSEA was performed on two del(5q) MDS patients with *CSNK1A1* mutations and two del(5q) MDS cases wild-type for *CSNK1A1*. All gene sets reported contain *CCND1*. For each gene set the enrichment score (ES) is reported as well as the normalised enrichment score (NES) and the false discovery rate (FDR) q-value.

GENE SET NAME	SIZE	ES	NES	FDR q-value
REACTOME DNA REPLICATION	166	0.6682	2.8936	<0.00 00929
REACTOME CELL CYCLE MITOTIC	271	0.6053	2.8569	<0.00 00929
REACTOME DNA STRAND ELONGATION	27	0.7936	2.6173	<0.00 00929
REACTOME S PHASE	95	0.6453	2.6094	<0.00 00929
REACTOME CELL CYCLE	345	0.5434	2.6011	<0.00 00929
REACTOME CELL CYCLE CHECKPOINTS	100	0.6418	2.5112	<0.00 00929
REACTOME SYNTHESIS OF DNA	80	0.6726	2.5154	<0.00 00929
REACTOME G1 S TRANSITION	94	0.6073	2.5083	<0.00 00929
REACTOME MITOTIC G1/ S PHASES	115	0.5450	2.3151	<0.00 00929
REACTOME SIGNALING BY WNT	61	0.5829	2.1649	0.001 8
REACTOME CYCLIN E ASSOCIATED EVENTS DURING G1/S TRANSITION	60	0.5204	1.9373	0.001 8

3.5. Discussion

The overall frequency of *CSNK1A1* mutations in the patient cohort was 4.8% (5/104 of cases) (Table 3.1) and this is consistent with reports subsequent and parallel to this study and confirms that *CSNK1A1* mutations are rare events in del(5q) MDS (Heuser *et al.*, 2015, Meggendorfer *et al.*, 2017, Schneider *et al.*, 2014b). Studies using whole-exome sequencing have identified *CSNK1A1* mutations in around 7% of patients, confirming the

low frequency of *CSNK1A1* mutations in MDS with del(5q) (Heuser *et al.*, 2015, Schneider *et al.*, 2014b). A higher frequency of *CSNK1A1* mutations in MDS with del(5q) (18%) was found in a study by Smith *et al.* (2015), however the patient cohort investigated included a significantly lower number of cases (39) than our cohort (104). Consistent with a number of studies of *CSNK1A1* mutations in del(5q) MDS, substitutions affecting amino acid residues E98 and D140 of CK1α seem to be particularly recurrent (Heuser *et al.*, 2015, Meggendorfer *et al.*, 2017, Smith *et al.*, 2015, Schneider *et al.*, 2014b). The *CSNK1A1* mutations identified at amino acid E98 in this study (Table 3.1) were also reported both in MDS with del(5q) (Heuser *et al.*, 2015, Hosono *et al.*, 2017, Schneider *et al.*, 2014b, Smith *et al.*, 2015) and oesophageal adenocarcinoma (Dulak *et al.*, 2013). Moreover, a previously identified *CSNK1A1* mutation at amino acid 140 (D140A) (Graubert *et al.*, 2012, Heuser *et al.*, 2015, Smith *et al.*, 2015) was found in one case of MDS with isolated del(5q) in this study. Mutations affecting amino acids E98 and D140 were likely to be damaging to CK1α protein function (Heuser *et al.*, 2015) as suggested by the Polyphen-2 analysis and by the fact that they were both located in the critical region for CK1α kinase function (Figure 3.1B). However, it is worth noting that in this and other studies (Heuser *et al.*, 2015, Smith *et al.*, 2015), screening of *CSNK1A1* mutations was focussed on exons 3 and 4 of the gene and the identification of hotspots for *CSNK1A1* mutations was therefore limited. A mutation at codon 134 (H134L), which had not been previously reported, was also identified in our study. The H134L mutation was not present in the COSMIC database (Catalogue of somatic mutations in cancer, http://cancer.sanger.ac.uk/cosmic) but PholyPhen-2 analysis suggested that it could have a damaging effect on the function of CK1α (PholyPhen-2 score 1, probably damaging). Unfortunately, live cells derived from the patients analysed were not available and therefore further investigation of the effect of the H134L mutation on the del(5q) MDS phenotype was not possible. If more time and

resources had been available, the function of the H134L mutation in del(5q) MDS pathogenesis could have been investigated by retroviral expression of the mutant cDNA or by the introduction of the mutation using CRISPR-Cas9 in *CSNK1A1* haploinsufficient haematopoietic cells. The effect of *CSNK1A1* mutations on haematopoietic stem cell differentiation, proliferation, p53 activation, cell cycle and apoptosis could have been studied. Mutations of *CSNK1A1* were found to be associated with non-synonymous substitutions in other genes, such as *SF3B1* and *ASXL1* (Heuser *et al.*, 2015, Meggendorfer *et al.*, 2017). In this study, no specific association of *CSNK1A1* mutations with other myeloid gene mutations was observed. However, the number of cases analysed was clearly small, and the study of a larger cohort of del(5q) MDS cases with CSNK1A1 mutations is required to determine whether a robust association with other gene mutations exists.

In this study, gene expression profiling and RT-qPCR data showed that haploinsufficiency of *CSNK1A1* was associated with increased expression of the Wnt/β-catenin transcriptional target *CCND1* in the HSC of MDS patients with del(5q) (Figure 3.2A). Consistent with this observation, *Csnk1a1*$^{+/-}$ hematopoietic cells transplanted into wild-type mice showed increased expression of cyclin D1, accompanied by β-catenin accumulation (Schneider *et al.*, 2014b). It was also found that the Wnt/β-catenin pathway was significantly up-regulated in del(5q) MDS patients with mutation of *CSNK1A1* (Figure 3.2B, 3.2C), suggesting that *CSNK1A1* mutations may lead to an increase in the expression of genes involved in Wnt signalling. Future work would include investigating the expression of genes involved in Wnt signalling in a larger cohort of del(5q) MDS patients with and without *CSNK1A1* mutations. Expression of mutant *CSNK1A1* (E98V mutation) in *Csnk1a1*$^{-/-}$ hematopoietic stem cells was found to cause an increase in levels of nuclear β-catenin and significant HSC cell cycle progression, providing a possible mechanism for selective clonal advantage of del(5q) MDS cells (Schneider *et al.*, 2014b). In a recent study

by Bhagat *et al.* (2017), it was found that stroma cells from MDS patients exhibit a lower expression of the Wnt pathway antagonist FRZB, which leads to activation of the Wnt/ β-catenin pathway in the HSCs surrounding the stroma. β-catenin activation resulted in a lethal MDS or leukaemia phenotype in a mouse model of MDS (Bhagat *et al.*, 2017). Moreover, a gene expression signature correlated with activation of the Wnt pathway was found in patients with advanced MDS and was correlated with adverse prognosis (Bhagat *et al.*, 2017). Interestingly, cell cycle related gene sets were also found to be significantly upregulated in del(5q) MDS patients with *CSNK1A1* mutations (Table 3.1), consistent with the role of cyclin D1 in cell cycle progression (Alao, 2007). Moreover, expression of mutant *CSNK1A1* rescues haematopoiesis in *Csnk1a*$^{-/-}$ mice, which otherwise die from bone marrow failure (Schneider *et al.*, 2014b). Finally, studies on the clonal architecture of MDS with del(5q) showed that *CSNK1A1* mutations occur after the del(5q), suggesting that *CSNK1A1* mutations play an important role in the mechanism of expansion of the del(5q) clone (Hosono *et al.*, 2017).

In the patient cohort analysed, one patient with E98K mutation of *CSNK1A1* did not respond to treatment with lenalidomide. Unfortunately, treatment information for the other four *CSNK1A1* mutated patients identified was not available. Nonetheless, this observation is consistent with the findings of another study on MDS del(5q) patients with the same *CSNK1A1* mutation (E98K), suggesting that *CSNK1A1* mutations are correlated with a decreased response to lenalidomide (Smith *et al.*, 2015). In subsequent studies it was found that lenalidomide modulates the CRL4-CRBN complex and induces ubiquitination of CK1α by CRL4-CRBN, resulting in CK1α degradation (Fink *et al.*, 2014, Kronke *et al.*, 2015, Hollenbach *et al.*, 2014). Gain of function mutations of *CSNK1A1* may play a role in the mechanism of resistance to lenalidomide observed in del(5q) MDS patients with *CSNK1A1* mutations (Smith *et al.*, 2015). However, further characterisation of *CSNK1A1*

mutations del(5q) MDS and larger studies on patient cohorts are needed to establish whether *CSNK1A1* mutational status is definitely linked to lenalidomide resistance.

In conclusion, the presence of *CSNK1A1* mutations was confirmed in a small proportion of patients with del(5q) MDS and a substitution which had not been previously reported was identified (H134L). Our study was conducted on HSC of patients with del(5q) MDS and it was shown that in these cells *CSNK1A1* mutations cause upregulation of β-catenin target genes. Our data support a central role for *CSNK1A1* in the pathogenesis of del(5q) MDS.

4. Chapter Four - An investigation into the mode of action of the translation enhancer L-leucine in the MDS with del(5q)

4.1. Introduction

Haploinsufficiency of the ribosomal protein gene *RPS14* underlies the erythroid defect found in the 5q- syndrome and MDS with del(5q) (Ebert *et al.*, 2008). *RPS14* haploinsufficiency leads to a block in processing of pre-ribosomal RNA, aberrant ribosome biogenesis and deregulation of multiple genes involved in ribosome biogenesis and protein synthesis (Ebert *et al.*, 2008, Pellagatti *et al.*, 2008). As a consequence of the rapid cell division and massive translational demand for globin synthesis, erythroid progenitor cells are particularly sensitive to this defect in ribosome biogenesis (Dutt *et al.*, 2011, Yip *et al.*, 2012, Narla *et al.*, 2011). Treatment with the amino acid L-leucine has been shown to increase erythroid differentiation, cell proliferation and mRNA translation efficiency in animal and cellular models of the 5q- syndrome and MDS with del(5q) (Yip *et al.*, 2012, Payne *et al.*, 2012, Narla *et al.*, 2014, Zhang *et al.*, 2014, Ear *et al.*, 2016). Our group has shown that L-leucine treatment of cultured human erythroblasts derived from CD34+ bone marrow cells from two del(5q) MDS patients results in an increase in the cell proliferation, erythroid differentiation and mRNA translation (Yip *et al.*, 2012). L-leucine is a substrate for protein synthesis as well as a regulator of protein metabolism (Boultwood *et al.*, 2013). L-leucine activates the mTOR (mammalian target of rapamycin) signalling pathway which controls cell growth and mRNA translation (Payne *et al.*, 2012, Narla *et al.*, 2014, Boultwood *et al.*, 2013, Yip *et al.*, 2013). A reduction in intracellular leucine levels is sufficient to suppress the mTOR signalling protein mTOR complex 1 (mTORC1) signalling (Dodd and Tee, 2012). It has been hypothesised that the promotion of translation via activation of mTORC1 by L-leucine underlies the enhanced erythroid progenitor cell

growth and differentiation observed in animal and cellular models of 5q- syndrome and DBA treated with this amino acid (Payne *et al.*, 2012, Narla *et al.*, 2014, Boultwood *et al.*, 2013, Yip *et al.*, 2013). Moreover, an increase in haemoglobin levels might underlie the mechanisms of action of L-leucine-mediated enhancement of erythropoiesis in the 5q-syndrome (Payne *et al.*, 2012, Zhang *et al.*, 2014). The mTOR pathway specifically up-regulates the translation of mRNA transcripts with established 5' terminal oligopyrimidine tracts (5'TOP motifs), many of which code for ribosomal proteins and translation factors (Meyuhas and Kahan, 2015, Fonseca *et al.*, 2014, Thoreen *et al.*, 2012). The mechanism of regulation of 5'TOP-mRNAs translation by mTORC1 remains unclear but several studies have shown that RNA-binding proteins such as the T-cell intracellular antigen-1 (TIA-1), TIA-1-related (TIAR) (Damgaard and Lykke-Andersen, 2011) and the lupus autoantigen (La)-related protein 1 (LARP1) (Tcherkezian *et al.*, 2014, Gentilella *et al.*, 2017, Hong *et al.*, 2017) are regulated by mTORC1 and might mediate translation initiation of 5'TOP mRNAs (Fonseca *et al.*, 2014). In a recent study it was found that the RNA-binding protein LARP1 binds the 40S ribosome subunit and stabilises 5'TOP mRNAs (Gentilella *et al.*, 2017).

The aim of this study was to investigate the mechanism of L-leucine-mediated enhancement of erythropoiesis in the MDS with del(5q). Firstly, I studied the effect of L-leucine treatment on erythroid differentiation of cultured erythroblasts from an expanded cohort of patients with del(5q) MDS. Secondly, I applied the polysome profiling technique to investigate the hypothesis that L-leucine increases translation of mTORC1-regulated 5'TOP mRNAs in cultured erythroblasts from patients with MDS with del(5q).

4.2. Results

4.2.1. L-leucine increases erythroid differentiation

Bone marrow mononuclear cells from eight MDS patients with del(5q) and eight healthy controls were differentiated along the erythroid lineage and treated with either L-leucine or the inactive isomer D-leucine (as an experimental control). The eight MDS patients with del(5q) analysed included five cases with 5q- syndrome and three cases with del(5q) and either excess blasts or other karyotypic abnormalities. (Materials and Methods, Table 2.2). Erythroid differentiation was measured by the expression of the erythroid cell surface markers CD36, CD71 and CD235a. In normal haematopoiesis, CD36 and CD71 are expressed on cells committed to the erythroid lineage and expression of these two markers is gradually downregulated during the terminal maturation stage of enucleated red blood cells (RBCs) (Fajtova *et al.*, 2013). CD235a is instead an erythroid specific marker expressed in late stages of erythropoiesis (Fajtova *et al.*, 2013).Consistent with previous studies from our group (Yip *et al.*, 2012), showing that erythroblasts from 5q- syndrome patients exhibit impaired erythroid differentiation, it was found that the percentage of progenitor cells committed to the erythroid lineage (CD36+/CD71+) was significantly lower in cultured erythroblasts from del(5q) MDS patients than in erythroblasts from healthy controls (Figure 4.1A, 4.1B and 4.1C, $p<0.001$). Indeed, the percentage of CD36-/CD71- erythroblasts from patients with del(5q) MDS was significantly higher than the percentage of CD36-/CD71- erythroblasts from healthy controls (Figure 4.1D, $p< 0.001$). Moreover, erythroblasts from healthy controls had a lower percentage of CD36-/CD235a- and CD71-/CD235a- cells compared to patient-derived erythroblasts (Figure 4.2A, 2B, 2C and 2D). Expression of CD36 and CD235a cell surface markers in erythroblasts from healthy controls trended towards higher values than in patient erythroblasts (non-significant) (Figure 4.2A, 2B, 2E and 2F).

L-leucine treatment of erythroblasts from patients with del(5q) MDS was found to significantly decrease the percentage of CD36-/CD71- erythroid progenitor cells (Figure 4.1D, $p<0.01$) and to increase the number of early erythroid progenitors (CD36+/CD71+) (albeit not significantly) (Figure 4.1C). L-leucine treatment also significantly increased the number of intermediate erythroid cells (CD36+/CD235a+ and CD71+/CD235a+) (Figure 4.2A and 2B, $p< 0.01$) and significantly decreased the number of CD36-/CD235a- (Figure 4.2A, $p<0.01$) and CD71-/CD235a- (Figure 4.2B, $p<0.05$) in erythroblasts from patients with del(5q) MDS. L-leucine treatment had an effect on the expression of erythroid markers in erythroblasts from a few of the healthy controls analysed (Figure 4.1A, 2E and 2F) but, overall (n=8), no significant changes in expression of CD36, CD71 or CD235a were found in L-leucine treated erythroblasts from healthy controls compared to healthy erythroblasts treated with D-leucine (Figure 4.2A, 2B, 2E and 2F). This observation suggests a patient specific effect of L-leucine on erythroid differentiation. These data show that L-leucine increases erythroid differentiation of cultured erythroblasts from del(5q) MDS patients.

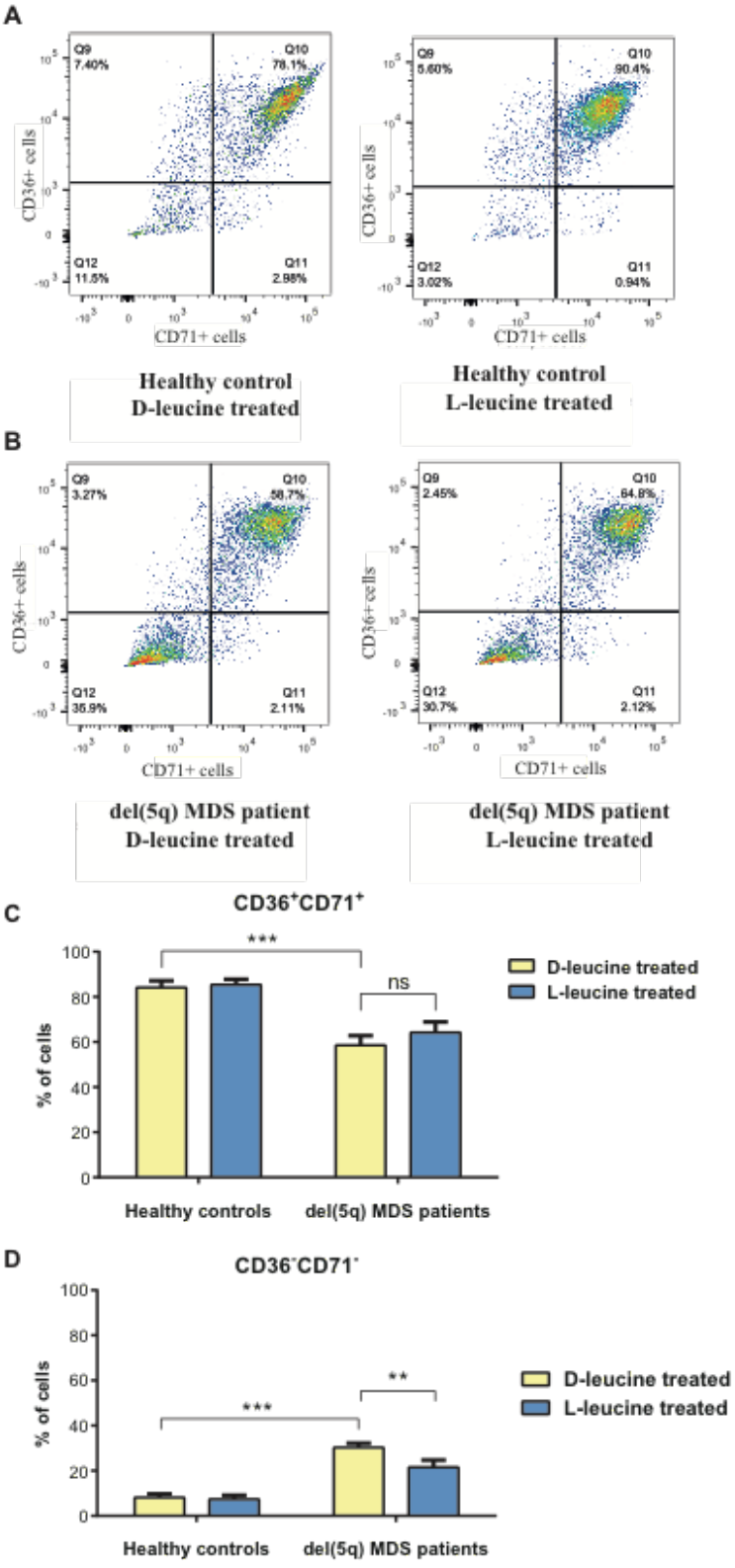
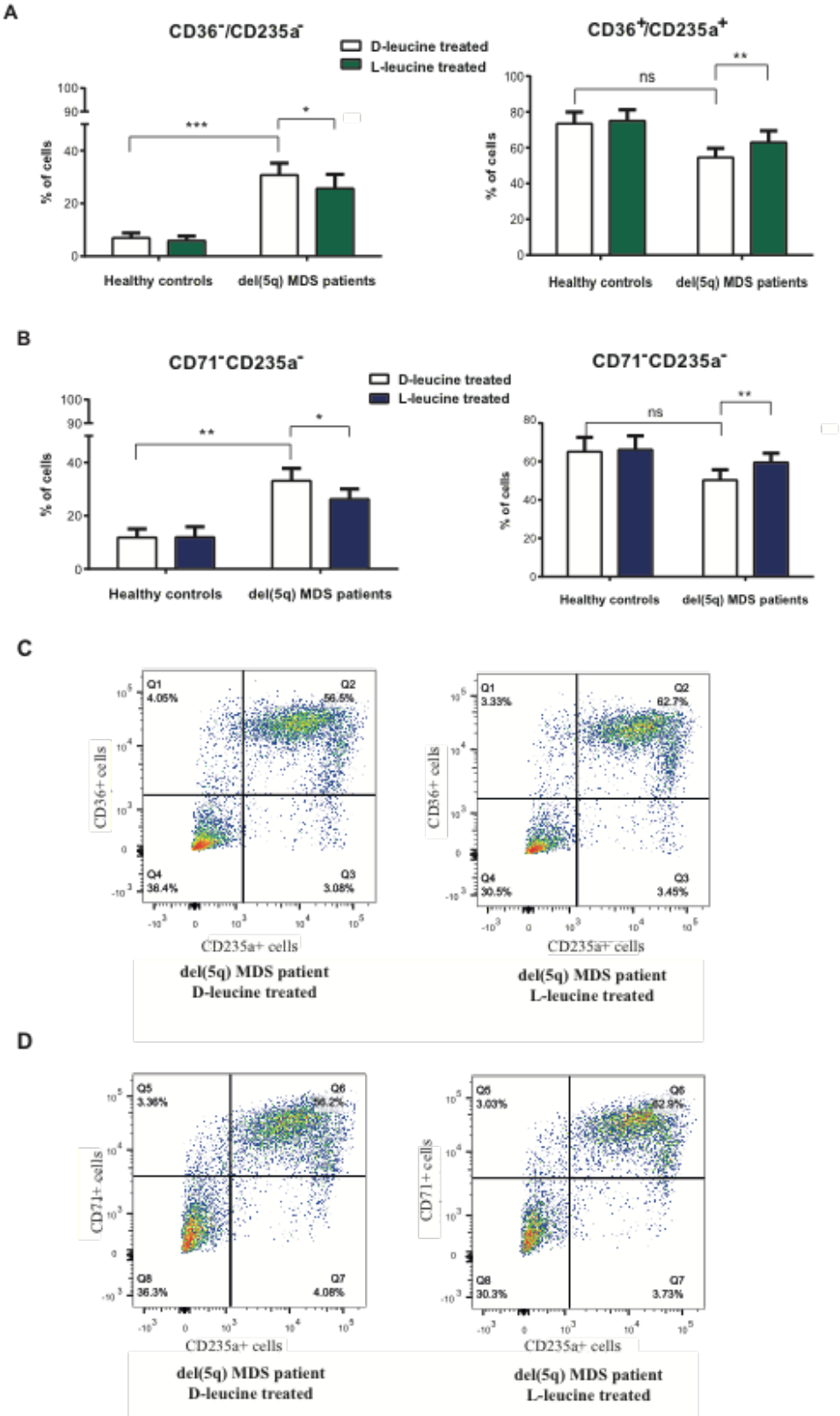


Figure 4.1. Expression of erythroid markers CD36 and CD71 in erythroblasts from patients with MDS with del(5q) and healthy controls treated with D-leucine or L-leucine. A&B) Representative scatter plots showing results of flow cytometry analysis of expression of CD36 and CD71 cell surface markers in erythroblasts from healthy controls (A) and del(5q) MDS patients (B)

treated with either D-leucine (left) or L-leucine (right). **C&D)** Bar graphs showing the percentage of CD36, CD71 double-positive (C) and double-negative (D) cells in healthy controls and del(5q) MDS patients, treated with either D-leucine (yellow) or L-leucine (blue). n=8. ** = $p<0.01$, *** = $p<0.001$; ns = not significant. Error bars represent the SEM of eight biological replicates.


A
CD36-/CD235a-
D-leucine treated
L-leucine treated
% of cells
Healthy controls
del(5q) MDS patients
CD36+/CD235a+
ns
B
CD71-CD235a-
C
CD36+ cells
CD235a+ cells
Q1 4.05%
Q2 56.5%
Q4 38.4%
Q3 3.08%
del(5q) MDS patient
D-leucine treated
Q1 3.33%
Q2 62.7%
Q4 30.5%
Q3 3.45%
del(5q) MDS patient
L-leucine treated
D
CD71+ cells
Q5 3.36%
Q6 56.2%
Q8 36.3%
Q7 4.08%
Q5 3.03%
Q6 62.9%
Q8 30.3%
Q7 3.73%

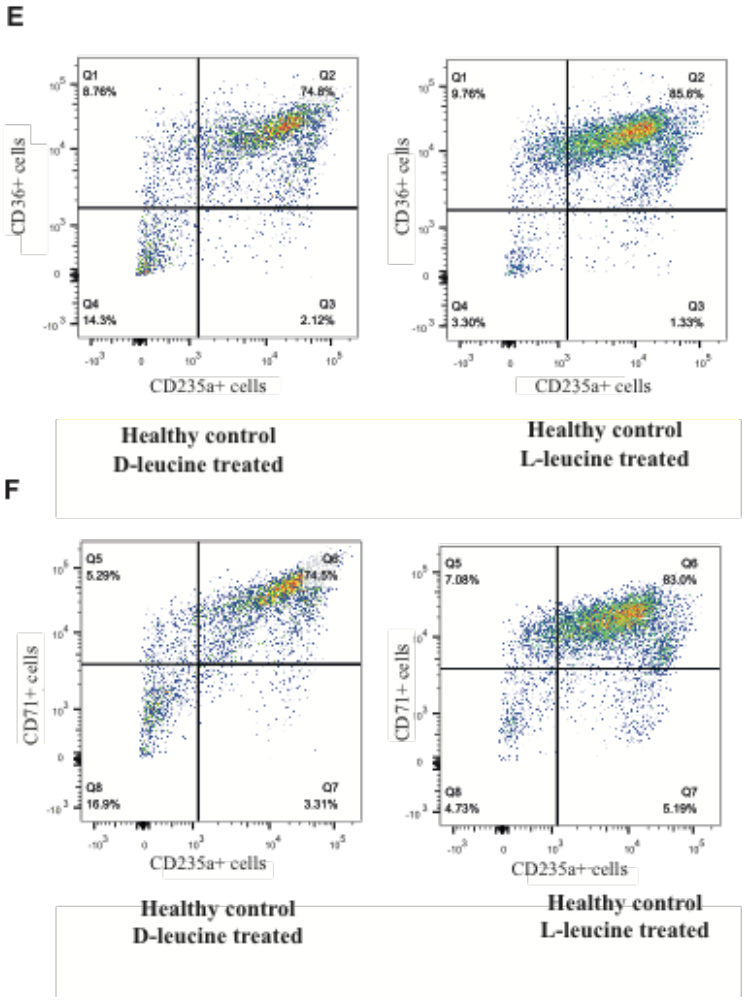


Figure 4.2. Expression of erythroid markers CD36, CD71 and CD235a in erythroblasts from patients with MDS with del(5q) and healthy controls treated with D-leucine or L-leucine. A) Bar graphs showing the percentage of CD36, CD235a double-negative (left) and double-positive (right) cells in healthy controls and del(5q) MDS patients, treated with either D-leucine (white) or L-leucine (green). n=8. * = $p<0.05$; ** = $p<0.01$, *** = $p<0.001$; ns = not significant. Error bars represent the SEM of eight biological replicates. **B)** Bar graphs showing the percentage of CD71, CD235a double-negative (left) and double-positive (right) cells in healthy controls and del(5q) MDS patients, treated with either D-leucine (white) or L-leucine (blue). n=8. * = $p<0.05$; ** = $p<0.01$, *** = $p<0.001$; ns = not significant. Error bars represent the SEM of eight biological replicates. **C) D) E) F)** Representative plots showing results of flow cytometry analysis of expression of CD36, CD71 and CD235a cell surface markers in erythroblasts from del(5q) MDS patients (C and D) and healthy controls (E and F) treated with either D-leucine (left) or L-leucine (right).

Consistent with the finding that erythroblasts from del(5q) MDS patients show impaired erythroid differentiation, both alpha and beta globin levels were found to be significantly decreased in patient erythroblasts than in erythroblasts from healthy controls (Figure 4.3A and 4.3B). Alpha and beta globin levels were significantly higher in patient erythroblasts treated with L-leucine than in D-leucine treated erythroblasts from del(5q) MDS patients (Figure 4.3A and 4.3B). L-leucine treatment had a patient-specific effect and did not alter globin levels in erythroblasts from healthy controls (Figure 4.3A and 4.3B).

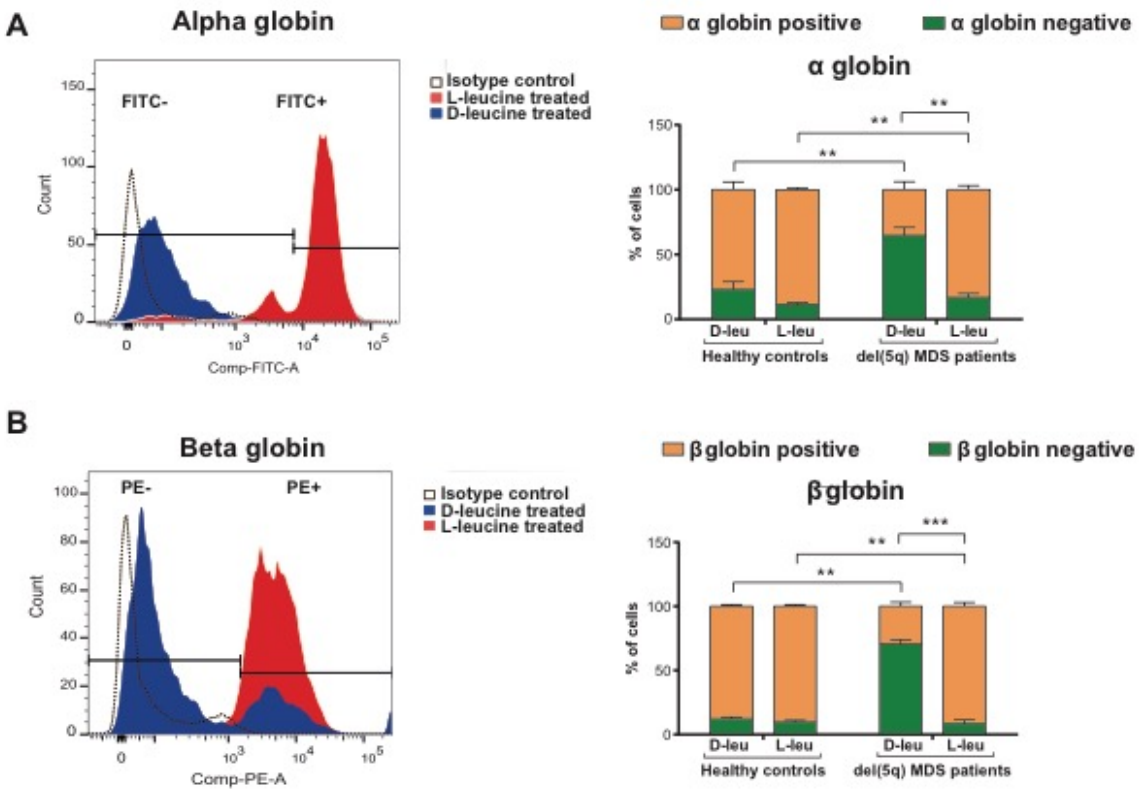


Figure 4.3. Reduced globin levels in del(5q) MDS patient erythroblasts are increased following L-leucine treatment. A) Left panel: Representative histogram showing alpha globin-positive and negative cells in D-leucine and L-leucine treated patient. Alpha globin was stained with a FITC conjugated antibody. Isotype control is also included. Right panel: Bar graph showing the percentage of alpha globin-positive (yellow) and –negative (green) cells in healthy control and del(5q) MDS patients, treated with either D-leucine or L-leucine measured by intracellular staining and flow cytometry. n=3. ** = $p<0.01$. Error bars represent the SEM of three biological replicates. **B)** Left panel: Representative histogram showing beta globin-positive and negative cells in D-leucine and L-leucine treated patient (left). Beta globin was stained with a PE conjugated antibody. Isotype control is also included. Right panel: Bar graph showing the percentage of beta globin-positive (yellow) and –negative (green) cells in healthy control and del(5q) MDS patients, treated with either D-leucine or L-leucine measured by intracellular staining and flow cytometry. n=3. ** = $p<0.01$; *** = $p<0.001$. Error bars represent the SEM of three biological replicates.

4.2.2. L-leucine increases phosphorylation of mTOR targets

A previously published study by our group showed that L-leucine activates the mTOR pathway in *RPS14*-deficient erythroblasts and in erythroblasts derived from CD34+ cells of two patients with 5q- syndrome (Yip *et al.*, 2013). Here, the cohort of erythroblasts derived from del(5q) MDS patients treated with L-leucine was expanded and phosphorylation of the mTOR targets 4EBP1 and S6K1 and the S6K1 target RPS6 was measured in cultured erythroblasts from three del(5q) MDS patients. I found that phosphorylation of 4EBP1, S6K1 and RPS6 was lower in patient erythroblasts than in erythroblasts from healthy controls (Figure 4.4). Moreover, L-leucine treatment increased phosphorylation of all three mTOR targets measured in cultured erythroblasts from del(5q) MDS patients, while having no significant effect on phosphorylation of 4EBP1, S6K1 and RPS6 in erythroblasts from healthy controls (Figure 4.4).

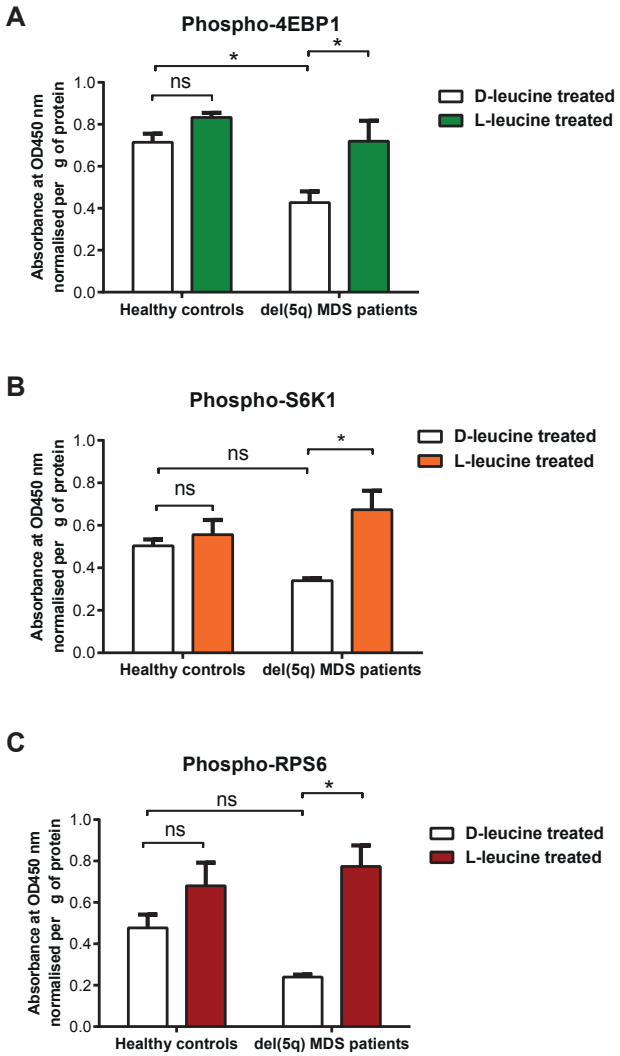


Figure 4.4. Phosphorylation of mTOR targets in patient-derived erythroblasts and erythroblasts from healthy controls treated with D-leucine or L-leucine. A, B & C) Normalised levels of phosphorylated 4EBP1 (A), S6K1 (B) and RPS6 (C), in healthy controls and del(5q) MDS patients, that were treated with either D-leucine (white) or L-leucine (coloured) measured by sandwich ELISA. n=3. * = $p<0.05$; ns = not significant. Error bars represent the SEM of three biological replicates.

4.2.3. Optimisation of polysome profiling technique in cultured erythroblasts from patients with del(5q) MDS

The mTOR pathway is involved in regulation of both mRNA translation and cell proliferation (Gingras *et al.*, 2004). Since L-leucine has been shown to increase proliferation and mRNA translation in *RPS14*-deficient erythroblasts and erythroblasts derived from CD34+ cells of patients with 5q- syndrome (Yip *et al.*, 2012, Yip *et al.*, 2013), I investigated the hypothesis that L-leucine promotes translation via activation of the mTOR signalling pathway in cultured erythroblasts derived from patients with del(5q) MDS. The L-leucine-mediated increase in mRNA translation could constitute the mechanism underlying the enhancement of cell growth and erythroid differentiation of progenitor cells from patients with del(5q) MDS treated with L-leucine.

In order to study the effect of L-leucine treatment on mRNA translation, I attempted to optimise the ribosome profiling technique for use in del(5q) MDS patient-derived erythroblasts. Ribosome profiling involves nuclease footprinting of polysomal RNA followed by purification of ribosome-protected mRNA fragments (RPF) using a sucrose gradient (Ingolia *et al.*, 2012). Total RNA is also isolated from each sample, in order to normalise the number of footprints produced from one transcript to the transcript abundance (Ingolia *et al.*, 2012). However, a minimum input of around 10^7 cells is required for sucrose gradient separation of RPFs and our erythroblast culture system did not yield a sufficient number of cells for the use of sucrose gradient techniques. The total number of patient-derived cells obtained with our culture system was limited by the fact that del(5q) MDS patient-derived erythroblasts show an arrest in erythroid differentiation and an increase in apoptosis compared to erythroblasts from healthy controls. Moreover, the possibility of scaling up the cellular model was limited by the number of bone marrow mononuclear cells available from each patient with del(5q) MDS. Therefore, I tested the

TruSeq Ribo Profile Library Prep Kit (Illumina) which allows purification of RPFs using a gel filtration column followed by isolation of RPFs by polyacrylamide gel electrophoresis (PAGE) (Figure 4.5A). In this protocol isolation of RPFs by chromatography allows significant reduction of the input number of cells compared to RPFs purification by sucrose gradient. The addition of universal sequence adapters to the cDNA footprints for next-generation sequencing (NGS) is achieved using multiple PAGE purification and nucleic acid precipitation steps. The TruSeq Ribo Profile Library Prep Kit was tested on 0.2 $x10^6$ and $0.5x10^6$ cultured erythroblasts from healthy controls. However, I failed to obtain a sufficient DNA yield for NGS, since a significant amount of RNA or cDNA was lost during PAGE purification of RPFs (Figure 4.5B).

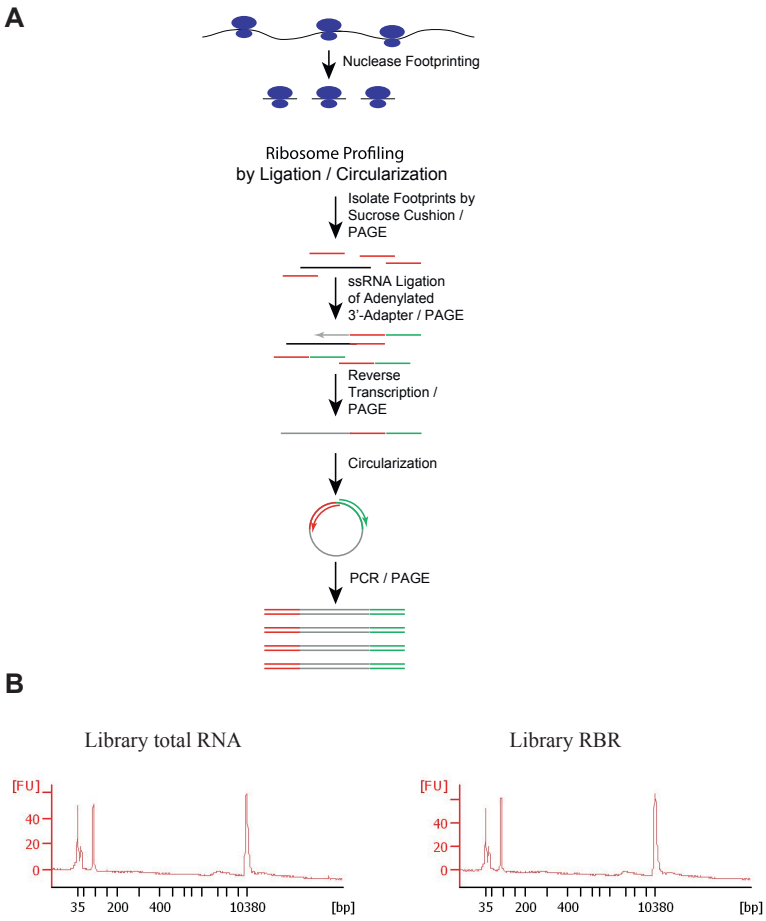
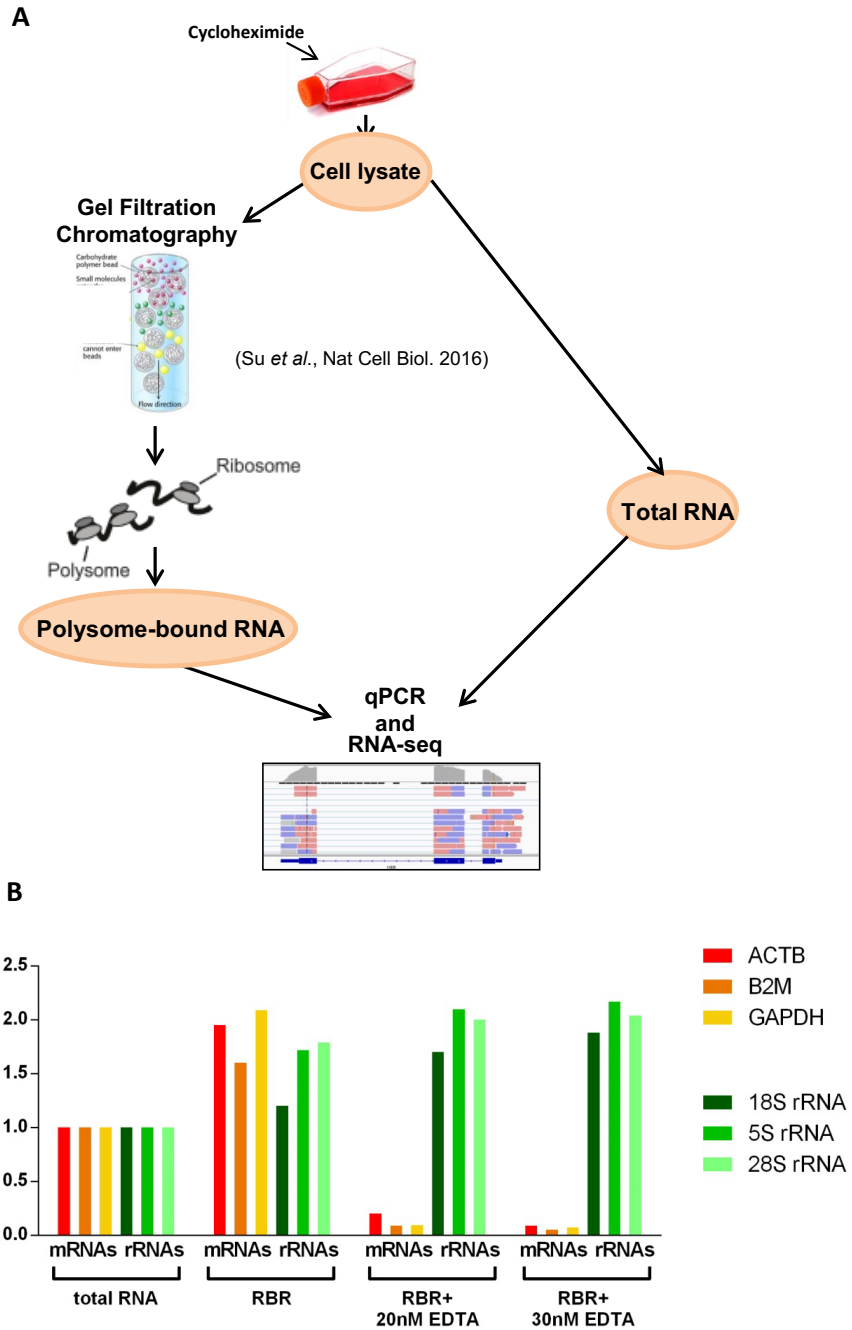
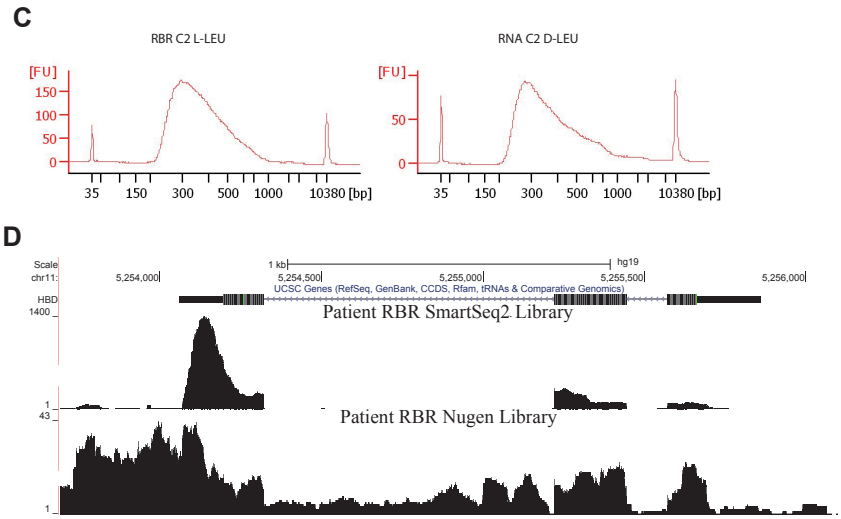


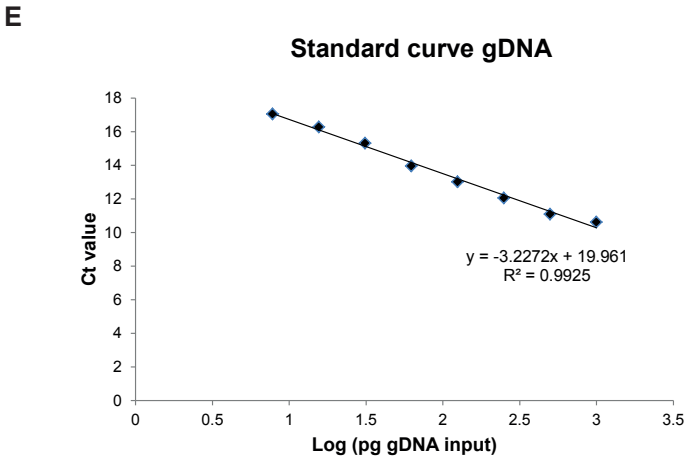
Figure 4.5. The ribosome profiling technique using the TruSeq Ribo Profile Library Prep Kit (Illumina) failed to produce sufficient DNA yield for NGS. A) Schematic showing the experimental workflow of ribosome profiling using the TruSeq Ribo Profile Library Prep Kit (Illumina) (modified from Hornstein *et al.*, 2016). **B)** Plots from the analysis of DNA libraries obtained using ribosome profiling. Libraries were analysed using the Agilent High Sensitivity DNA Kit on an Agilent 2100 Bioanalyser System (Agilent Technologies). Left plot: DNA library obtained from preparation of total RNA. Right plot: library obtained from processing of ribosome-bound RNA (RBR). Both libraries were obtained from erythroblasts from healthy controls using the TruSeq Ribo Profile Library Prep kit (Illumina).

Next, instead of digesting the mRNA with nucleases in order to generate RPFs, I tried using the sephacryl gel filtration columns recommended in the TruSeq Ribo Profile Library Prep Kit to purify undigested ribosome-bound mRNA (RBR) (Figure 4.6A). Analysis of RBR is defined as polysome profiling (Chasse *et al.*, 2017) and a protocol for isolation of RBR by gel filtration had been previously developed (Su *et al.*, 2016). Since this protocol did not require any PAGE purification step, it could be optimised for use on a low number of cells. RBR was isolated using gel filtration, purified by column chromatography and analysed by either RT-qPCR or RNA sequencing (RNA-seq) (Materials and Methods section 2.17) (Figure 4.6A). For RNA-seq analysis of RBR, universal sequence adapters for NGS were added using a low-input RNA library preparation kit. For each sample, the total mRNA fraction (total RNA) was also purified by column chromatography and analysed by RT-qPCR or RNA-seq (Figure 4.6A). Polysome profiling allows measurement of translation efficiency of transcripts as the ratio of abundance of actively translated RBR and total RNA (Chasse *et al.*, 2017). Ribosome profiling is instead based on deep sequencing of RPFs and determines the position of the ribosome on the mRNA transcript, allowing discoveries of new coding transcripts as well as measurement of translation rates (Ingolia *et al.*, 2012). For the purpose of the present study the use of the polysome profiling technique was suitable for the investigation of the effects of L-leucine on translation efficiency of mRNA targets in erythroblasts derived from patients with del(5q) MDS. I tested the validity of the gel filtration method of isolation of RBR by treating cell lysates with various concentrations of the magnesium-chelating agent ethylenediaminetetraacetic acid (EDTA). Since EDTA causes dissociation of the ribosome subunits, EDTA treatment of cell lysates would lead to the release of mRNA from ribosome complexes and its exclusion from the eluate of the gel filtration column. Therefore, mRNA yield should be absent or dramatically decreased in column eluates from EDTA-treated cell lysates. I

determined the presence of mRNA in the samples by measuring the expression of the housekeeping genes *ACTB*, *B2M* and *GAPDH*. Expression of these three genes was decreased in the sample extracted from EDTA-treated cell lysates purified by gel filtration compared to RNA extracted from column-purified untreated lysates, demonstrating that the RBR isolated by gel filtration was in fact ribosome-bound (Figure 4.6B). Expression of ribosomal RNAs (rRNA) 18S, 5S and 28S was instead detected in all RNA samples, suggesting that ribosomal subunits and rRNA were isolated by gel filtration in EDTA-treated lysates as well as in untreated samples. The optimised polysome profiling protocol described above was tested on cultured erythroblasts from healthy controls. RBR and total RNA fractions were isolated for each sample and adapters were added for sequencing using the Ovation Human Blood RNA-Seq Library System (Nugen). Suitable libraries for NGS were obtained from a minimum of $0.2x10^6$ cells (Figure 4.6C).


A
Cycloheximide
Cell lysate
Gel Filtration
Chromatography
(Su et al., Nat Cell Biol. 2016)
Ribosome
Polysome
Polysome-bound RNA
Total RNA
qPCR
and
RNA-seq
B
mRNA expression
(relative to total RNA sample)
2.5
2.0
1.5
1.0
0.5
0.0
mRNAs rRNAs
total RNA
RBR
RBR+
20nM EDTA
RBR+
30nM EDTA
ACTB
B2M
GAPDH
18S rRNA
5S rRNA
28S rRNA





Sample	% gDNA contamination
RNA untreated	17.3256
RNA DNAse I Invitrogen	0.0263
Library Nugen DNAse Control	17.4267
Library Nugen DNAse Patient	25.3753

Figure 4.6. Optimisation of Polysome Profiling for the analysis of mRNA translation in erythroblasts from del(5q) MDS patients. A) Schematic of the experimental workflow of polysome profiling using gel filtration chromatography for purification of ribosome-bound RNA. **B)** Bar graph showing the level of beta actin (ACTB), B2M and GAPDH mRNA (mRNAs) and 18S, 5S, and 28S rRNA (rRNAs) in total RNA, untreated Ribosome-bound RNA (RBR) fraction and RBR treated either with 20nM EDTA or 30nM EDTA. Total RNA and RBR fractions were extracted from erythroblasts from healthy controls. RNA levels were measured by RT-qPCR and are shown relative to the absolute amounts of RNA found in the total RNA fraction from the same sample. n=1. **C)** Plots from the analysis of DNA libraries obtained using polysome profiling. Libraries were analysed using the Agilent High Sensitivity DNA Kit on an Agilent 2100 Bioanalyser System (Agilent Technologies). Left plot: DNA library obtained from preparation of ribosome-bound RNA from erythroblasts from a healthy control treated with L-leucine (RBR C2

L-LEU). Right plot: library obtained from processing of total RNA from erythroblasts from a healthy control treated with D-leucine (RNA C2 D-LEU). Both libraries were obtained using the Ovation Human Blood RNA-Seq Library System (Nugen) **D)** Tracks from the UCSC Genome Browser showing RNA sequencing reads mapping to the HBD (hemoglobin subunit delta) gene in a sample prepared with the SmartSeq2 Library Prep kit (SmartSeq2 library) (top) and in a sample prepared with the Ovation Human Blood RNA-Seq Library System (Nugen library) (bottom). Both samples analysed were derived from ribosome-bound RNA from del(5q) MDS patient erythroblasts (Patient RBR). The bottom track shows mapping of sequencing reads to gene introns, suggesting the presence of DNA contamination in the Nugen Library. **E)** Detection of gDNA by RT-qPCR. Top graph: RT-qPCR standard curve for relative quantification of gDNA. Bottom table: percentage of gDNA contamination in each sample. Samples analysed were: untreated total RNA from erythroblasts from a healty control (RNA untreated), total RNA from healthy erythroblasts treated with Invitrogen DNAse I (RNA DNAseI Invitrogen), DNA library from total RNA from erythroblasts from a healthy control prepared with the Ovation Human Blood RNA-Seq Library System (Nugen) (Library Nugen DNAse Control) and DNA library from total RNA from patient erythroblasts prepared with the Ovation Human Blood RNA-Seq Library System (Library Nugen DNAse Patient). All samples but the RNA DNAseI Invitrogen show DNA contamination. n=1.

The removal of contaminating DNA by treatment with DNase I is essential prior to library preparation for RNA sequencing. Moreover, due to the sensitivity of RNA sequencing, even trace DNA contaminants can compromise the quality of the final library. The Ovation Human Blood RNA-Seq Library System includes an integrated DNAse I treatment step coupled with the reverse transcription (RT) reaction. The DNAse reaction buffer is compatible with the RT reaction mix, eliminating the need for an intermediate RNA purification step and decreasing the loss of material. However, DNA contamination was found in libraries for total RNA and RBR fractions from erythroblasts derived from healthy controls and prepared with the Ovation Human Blood RNA-Seq Library System (Nugen), as shown by the high percentage of intergenic and intronic reads (Figure 4.6D, bottom track). RT-qPCR analysis confirmed that genomic DNA (gDNA) was present in libraries from samples treated with the DNAse I integrated in the Ovation Human Blood RNA-Seq

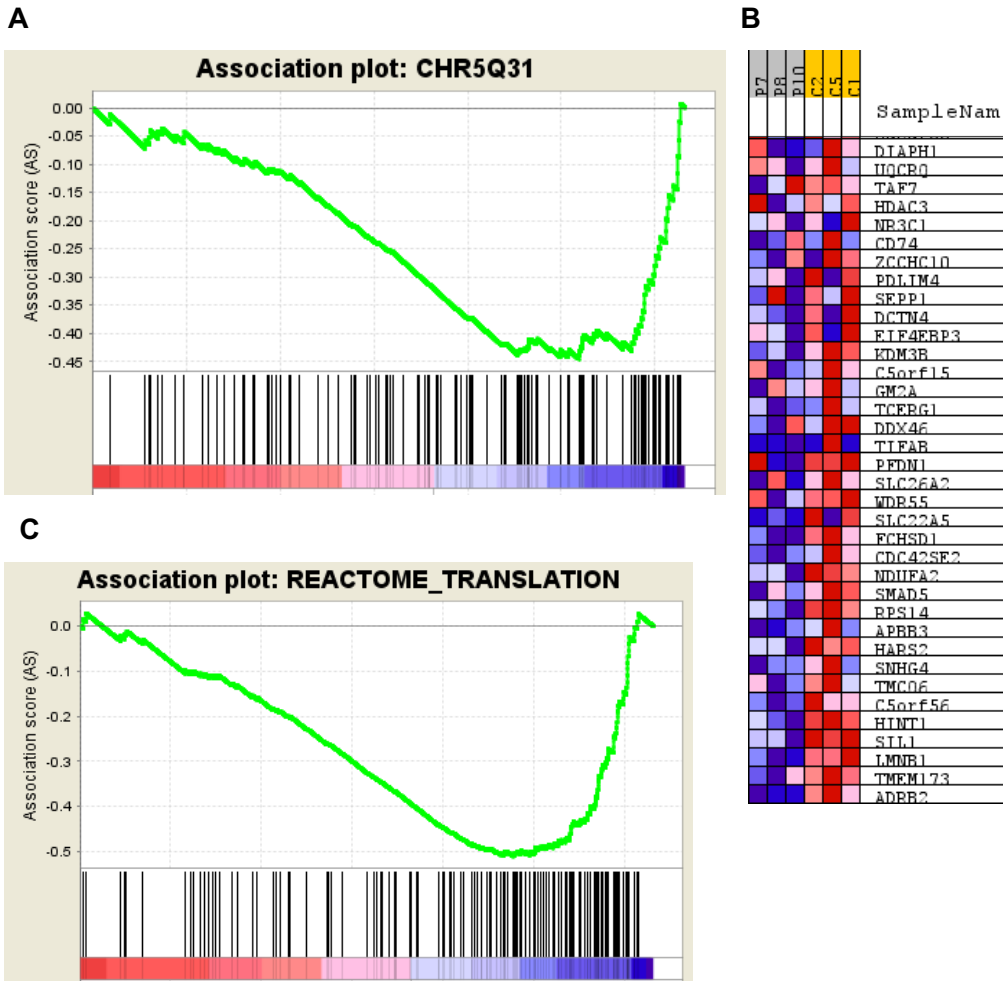
Library System (Figure 4.6E). A different DNAse I treatment (Invitrogen) was therefore tested and it proved effective in removing gDNA from purified RNA samples (Figure 4.6E). However, a drawback of the Invitrogen DNAse I protocol is that it requires an RNA purification step using magnetic beads (Materials and Methods section 2.8), which results in partial loss of RNA input. Therefore, the low input SMART-seq2 library preparation protocol (Picelli *et al.*, 2014) was utilised on column purified RNA samples after DNA-treatment. The SMART-seq2 protocol allows preparation of libraries for NGS from RNA input as low as 10pg and is utilised is RNA sequencing of single cells (Picelli *et al.*, 2014). In conclusion, polysome profiling was optimised for use on a minimum of 150,000 erythroblasts derived from del(5q) MDS patients and healthy controls.

In order to investigate the effects of L-leucine treatment on mRNA translation in MDS with del(5q), RBR from erythroblasts derived from three patients with del(5q) MDS and from three healthy controls treated with either L-leucine or D-leucine was sequenced together with total RNA for each sample. Changes in mRNA translation efficiency in MDS with del(5q) were also studied in patient-derived erythroblasts treated with the inactive L-leucine isomer D-leucine compared to D-leucine treated erythroblasts from healthy controls. The total RNA fraction was used to normalise the amount of RBR to the abundance of each mRNA transcript. Moreover, sequencing of total RNA was utilised to investigate whether changes in protein expression in disease and treatment could be explained by transcript levels changes or whether they were due to translational regulation mechanisms.

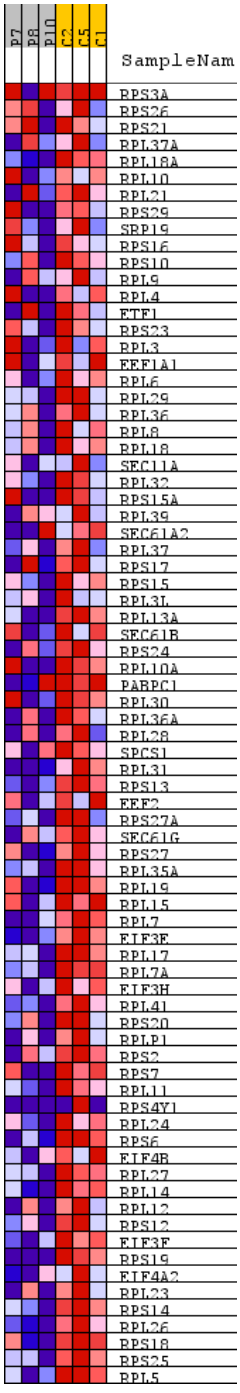
4.2.4. Expression of genes in the CDR of 5q- syndrome is decreased in cultured erythroblasts from del(5q) MDS patients

In order to study gene expression changes in MDS with del(5q), total RNA fractions of patient-derived erythroblasts and erythroblasts from healthy controls treated with D-leucine were analysed. The software package Kallisto (Bray *et al.*, 2016) was used to estimate transcript abundance as transcripts per million (TPM). TPM values for 18361 genes were analysed using Gene Set Association Analysis for RNA-Seq with Sample Permutation (GSAASeqSP). GSAASeqSP identifies pathways and gene sets significantly associated with a phenotype using RNA-seq data and provides a q-value for each gene set identified (q-values lower than 0.25 are considered significant) (Xiong *et al.*, 2014). GSAASeqSP analysis identified the gene set corresponding to cytogenetic band chr5q31 to be significantly downregulated in patient erythroblasts treated with D-leucine compared to D-leucine treated erythroblasts from healthy controls (Figure 4.7A and 4.7B, q-value=0.068). MDS with del(5q) and 5q- syndrome are both characterised by deletions of chromosome 5q (Boultwood *et al.*, 2002). The commonly deleted region (CDR) of the 5q-syndrome and MDS with del(5q) comprises an interval flanked by the DNA marker D5S413 and the *GLRA1* gene, which is located at 5q32-q33 according to the Ensembl database (Boultwood *et al.*, 2002). However, the positional gene sets contained in the Molecular Signatures Database (MSigDB version 6, http://www.broadinstitute.org/gsea/msigdb/) and used in GSAASeqSP analysis were generated by merging the relevant annotations from HUGO (October 2006) and UniGene (build 197) databases to derive a single cytogenetic band location for every gene symbol. Therefore, I found that the several genes included in the CDR of 5q- syndrome, such as *RPS14*, *CD74*, *SLC26A2* are included in the chr5q31 gene set and downregulated in patient-derived erythroblasts compared to those of healthy controls (Figure 4.7B). Moreover, the

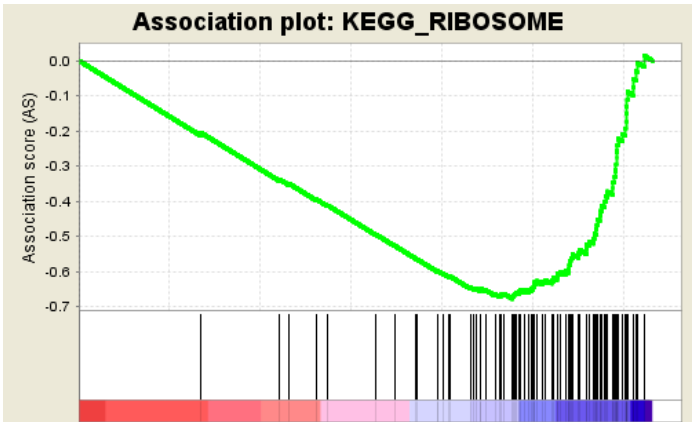
karyotype of the three del(5q) MDS patients analysed in this experiment shows that the 5q deletion in all patients comprises the cytogenetic bands 5q32-q33 as well as the 5q31 band (Table 4.1).



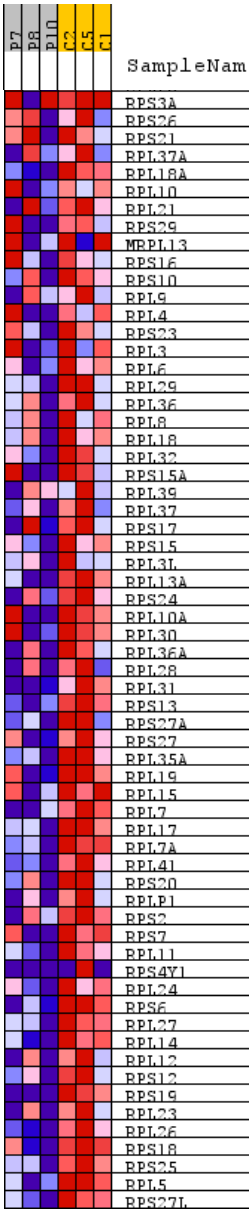
D


P7
P8
P10
C2
C5
C1
SampleNam
RPS3A
RPS26
RPS21
RPL37A
RPL18A
RPL10
RPL21
RPS29
SRP19
RPS16
RPS10
RPL9
RPL4
ETF1
RPS23
RPL3
EEF1A1
RPL6
RPL29
RPL36
RPL8
RPL18
SEC11A
RPL32
RPS15A
RPL39
SEC61A2
RPL37
RPS17
RPS15
RPL3L
RPL13A
SEC61B
RPS24
RPL10A
PABPC1
RPL30
RPL36A
RPL28
SPCS1
RPL31
RPS13
EEF2
RPS27A
SEC61G
RPS27
RPL35A
RPL19
RPL15
RPL7
EIF3E
RPL17
RPL7A
EIF3H
RPL41
RPS20
RPLP1
RPS2
RPS7
RPL11
RPS4Y1
RPL24
RPS6
EIF4B
RPL27
RPL14
RPL12
RPS12
EIF3F
RPS19
EIF4A2
RPL23
RPS14
RPL26
RPS18
RPS25
RPL5

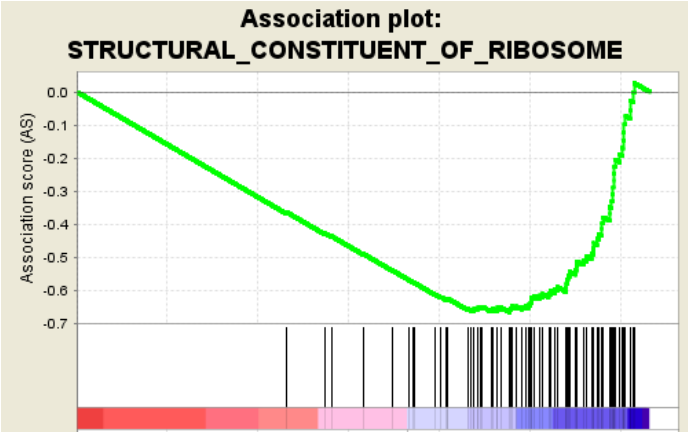
E


Association plot: KEGG_RIBOSOME
Association score (AS)
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7

F


P7
P8
P10
C3
C5
C1
SampleNam
RPS3A
RPS26
RPS21
RPL37A
RPL18A
RPL10
RPL21
RPS29
MRPL13
RPS16
RPS10
RPL9
RPL4
RPS23
RPL3
RPL6
RPL29
RPL36
RPL8
RPL18
RPL32
RPS15A
RPL39
RPL37
RPS17
RPS15
RPL3L
RPL13A
RPS24
RPL10A
RPL30
RPL36A
RPL28
RPL31
RPS13
RPS27A
RPS27
RPL35A
RPL19
RPL15
RPL7
RPL17
RPL7A
RPL41
RPS20
RPLP1
RPS2
RPS7
RPL11
RPS4Y1
RPL24
RPS6
RPL27
RPL14
RPL12
RPS12
RPS19
RPL23
RPL26
RPS18
RPS25
RPL5
RPS27L

G


Association plot:
STRUCTURAL_CONSTITUENT_OF_RIBOSOME
Association score (AS)
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7

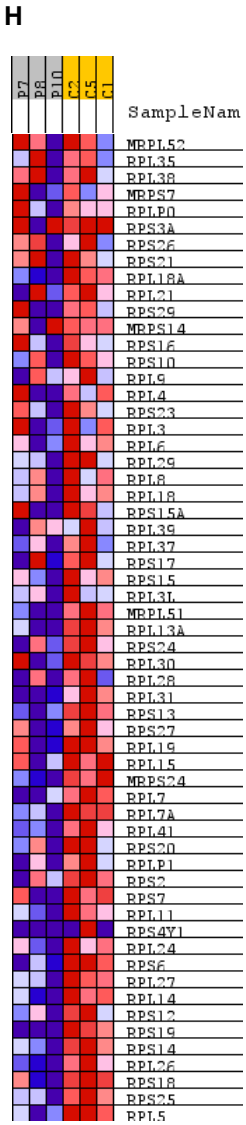


Figure 4.7. GSAASeqSP analysis of gene expression in patients with MDS with del(5q). A), C), E), G) Enrichment plots generated by GSAASeqSP analysis of RNA-seq data in D-leucine treated erythroblasts from del(5q) MDS patients compared to D-leucine treated erythroblasts from healthy controls. n=3. Four of the gene sets identified are shown: the "CHR5Q31" (A) (q value = 0.068), the "REACTOME_TRANSLATION" gene set (C) (q value < 0.04), the "KEGG_RIBOSOME" gene set (E) (q value < 0.04) and the "STRUCTURAL_CONSTITUENT_OF_RIBOSOME" gene set (G) (q value < 0.04). **B), D), F), H)** Heat maps generated by GSAASeqSP analysis of RNA-seq data in D-leucine treated erythroblasts from del(5q) MDS patients (grey, left three columns) compared to D-leucine treated erythroblasts from healthy controls (yellow, right three columns). n=3. The heat maps show the relative expression of genes in the "CHR5Q31" gene set (B), in the "REACTOME_TRANSLATION" gene set (D), in the "KEGG_RIBOSOME" gene set (F) and in the "STRUCTURAL_CONSTITUENT_OF_RIBOSOME gene set (H)". Red = high expression; blue = low expression.

Table 4.1. Karyotypes of del(5q) MDS patients analysed by polysome profiling.

Patient	Karyotype
MDS1	46,XX,del(5)(q13q33) [19] / 46,XX [1]
MDS2	46,XX,del(5)(q22q35) [19] / 46,XX [1]
MDS3	46,XX,del(5)(q13q33) [20]

4.2.5. MDS with del(5q) is associated with decreased expression of ribosomal genes

GSAASeqSP analysis of RNA-seq data in D-leucine treated erythroblasts from del(5q) MDS patients compared to D-leucine treated erythroblasts from healthy controls identified a number of ribosomal and translation-related gene sets significantly downregulated in patient erythroblasts (Table 4.2). In particular, the "REACTOME_TRANSLATION" gene set, which includes RPL (large ribosomal protein subunit) genes, RPS (small ribosomal protein subunit) genes, EIF (eukaryotic translation initiation factor) genes and EEF (eukaryotic translation elongation factor) genes, had a q score < 0.04 (Figure 4.7C and 4.7D). Consistent with the haploinsufficiency of *RPS14* and defective ribosome biogenesis observed in patients with the 5q- syndrome, gene sets "KEGG_RIBOSOME" and "STRUCTURAL CONSTITUENT OF RIBOSOME" were also significantly downregulated (q scores < 0.04) in D-leucine treated erythroblasts from del(5q) MDS patients compared to D-leucine treated erythroblasts derived from healthy controls (Figure 4.7E, F, G and H).

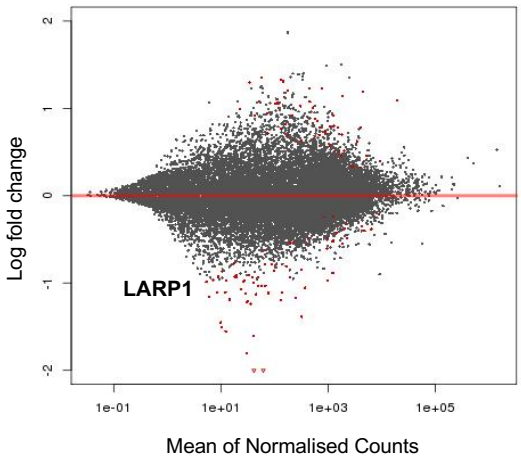
Table 4.2. Ribosomal and translation-related gene sets identified by GSAASeqSP analysis. GSAASeqSP was performed on RNA-seq data from three D-leucine treated samples from del(5q) MDS patients and three D-leucine treated samples from healthy controls. For each gene set the association score (AS) is reported as well as the normalised association score (NAS) and the false discovery rate (FDR) q-value.

NAME	AS	NAS	FDR q-value
REACTOME PEPTIDE CHAIN ELONGATION	-0.69	-2.73	<0.04
REACTOME 3 UTR MEDIATED TRANSLATIONAL REGULATION	-0.58	-2.39	<0.04
STRUCTURAL CONSTITUENT OF RIBOSOME	-0.67	-2.63	<0.04
KEGG RIBOSOME	-0.68	-2.65	<0.04
REACTOME TRANSLATION	-0.51	-2.20	<0.04
RIBOSOME	-0.54	-1.84	0.04
REACTOME FORMATION OF THE TERNARY COMPLEX AND SUBSEQUENTLY THE 43S COMPLEX	-0.52	-1.83	0.04
REACTOME ACTIVATION OF THE mRNA UPON BINDING OF THE CAP BINDING COMPLEX AND EIFS AND SUBSEQUENT BINDING TO 43S	-0.50	-1.82	0.05

Differential expression analysis was also performed using the software DESeq2 (Love *et al.*, 2014) on the read counts per gene calculated using featureCounts (Liao *et al.*, 2014). 140 genes were found to be differentially expressed in D-leucine treated erythroblasts from del(5q) MDS patients compared to D-leucine treated erythroblasts from healthy controls (normalised p value<0.05) (Figure 4.8A). Expression of *LARP1*, which encodes a RNA binding protein involved in regulation of 5'TOP mRNAs stability, was also downregulated in patient erythroblasts compared to erythroblasts from healthy controls (normalised p value= 0.048 and log FC= -0.549) (figure 4.8A). The expression of several ribosomal genes was lower in D-leucine treated patient erythroblasts compared erythroblasts from healthy controls treated with D-leucine (logFC<0) (Table 4.3). However, changes in expression of ribosomal genes were not statistically significant (Table 4.3). Hierarchical clustering

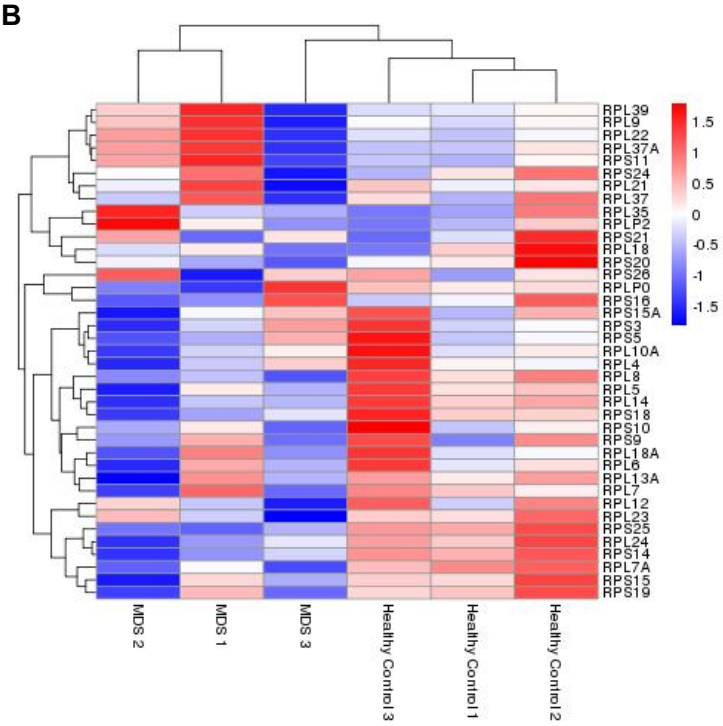
performed using normalised read counts from DESeq2 for several ribosomal genes grouped the del(5q) MDS patients together (Figure 4.8B).

A



Gene	Log2 Fold Change	Normalised p value
SLC1A4	0.574994	0.004754
GATA2	0.624556	0.005432
CD34	1.308431	0.001867
AKR1C1	0.967134	0.000222
MPO	1.095268	0.010258
LMO4	0.917574	0.021165

B



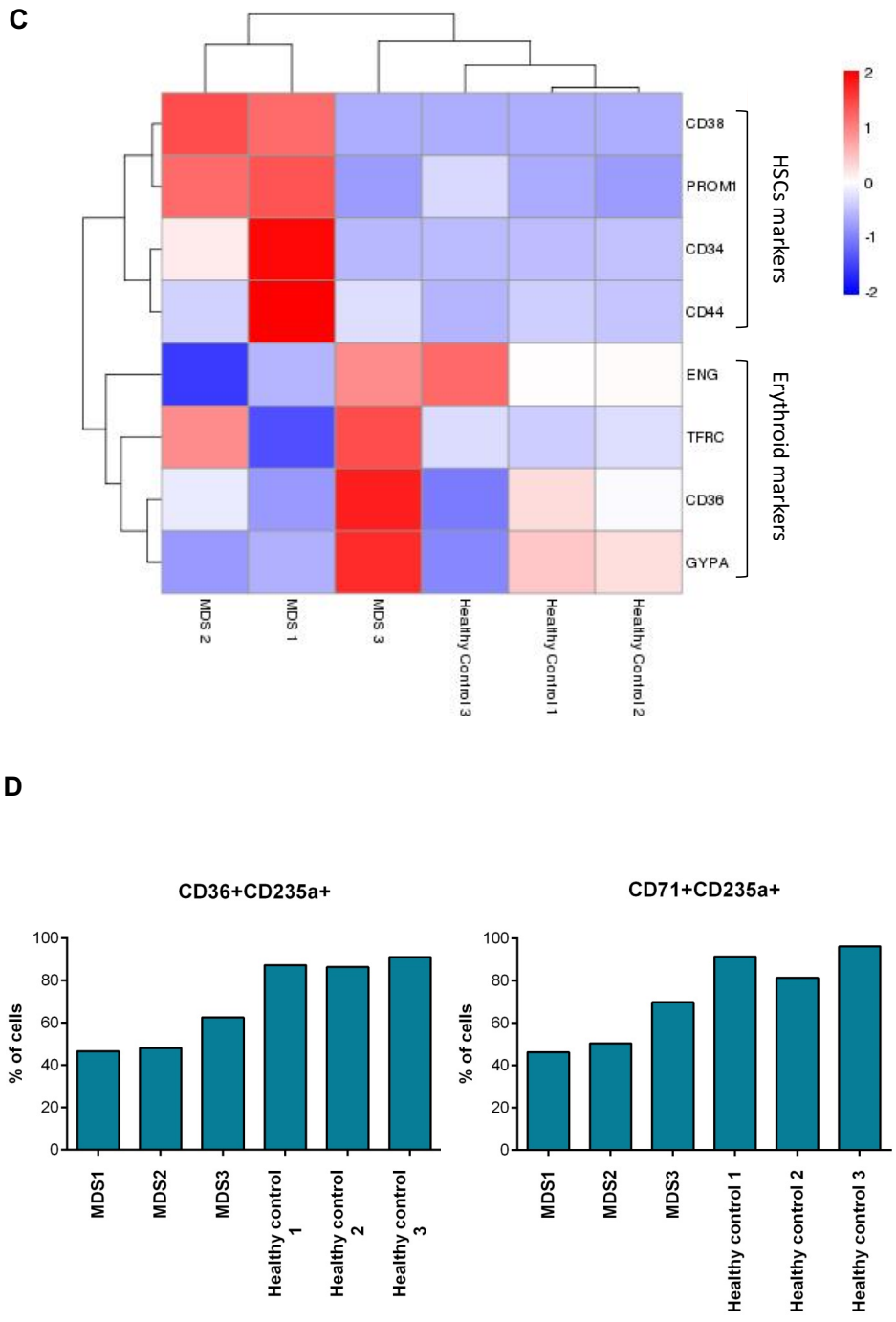


Figure 4.8. Analysis of gene expression in del(5q) MDS patients using DESeq2. A) Top: MA plot showing the log fold-change (FC) calculated using DESeq2 against the mean of the normalised RNA-seq read counts for all genes in MDS del(5q) patient erythroblasts compared to healthy control treated with D-leucine. n=3. Grey dots= non-significant genes; red dots= genes with p value <0.05. LARP1 (log FC=-0.549) is one of the genes significantly downregulated in patient erythroblasts (p value=0.048). Bottom: Table showing Log FC and normalised p values for genes upregulated in haematopoietic stem cells and early erythroid cells (Novershtern *et al.*, 2011). The log FC and p values were calculated using DESeq2 in MDS del(5q) patient erythroblasts compared

to healthy control treated with D-leucine. n=3. **B)** Heat-map of normalised read counts from DESeq2 showing the relative expression of ribosomal genes in MDS del(5q) patients (MDS1, 2 and 3; left three columns) and healthy controls (right three columns), treated with D-leucine. n=3. Red = high expression; blue = low expression. Dendrograms show clustering of similar samples. **C)** Heat-map of normalised read counts from DESeq2 showing the relative expression of haematopoietic stem cell (HSCs markers, top four) and erythroid cells markers (bottom four) in MDS del(5q) patients (left three columns) and healthy controls (right three columns) treated with D-leucine. n=3. Red = high expression; blue = low expression. Dendrograms represent clustering of similar samples. **D)** Bar graphs showing the percentages of CD36 and CD235a double-positive (left) and CD71, CD235a double-positive (right) cells measured by flow cytometry, in samples from MDS del(5q) patients and healthy controls treated with D-leucine included in the polysome profiling experiments.

Table 4.3. Expression of ribosomal genes in del(5q) MDS patients. Log_2 of the fold change (FC) in expression of each gene and normalised p values were calculated using DESeq2. Gene expression in three L-leucine treated del(5q) MDS patients was compared to expression in three healthy controls treated with D-leucine.

Gene	**log_2FC**	**normalised p value**
RPL10A	-0.02029	0.954351
RPL12	-0.02878	0.952466
RPL13A	-0.02523	0.933422
RPL14	-0.02273	0.938854
RPL18	-0.05566	0.866049
RPL18A	-0.06721	0.796514
RPL21	-0.18169	0.466225
RPL22	-0.27412	0.457396
RPL23	-0.0993	0.827508
RPL24	-0.02279	0.952466
RPL35	-0.2157	0.42009
RPL37	-0.11723	0.712423

RPL37A	-0.20401	0.459424
RPL39	-0.26141	0.480188
RPL4	-0.0121	0.974799
RPL5	-0.11247	0.426523
RPL6	-0.0958	0.668426
RPL7	-0.0425	0.896963
RPL7A	-0.05702	0.854586
RPL8	-0.02953	0.927714
RPL9	-0.13192	0.654509
RPLP0	-0.05621	0.884218
RPLP2	-0.24151	0.381875
RPS10	-0.17178	0.497777
RPS11	-0.31792	0.21195
RPS14	-0.04896	0.876605
RPS15	-0.10768	0.715889
RPS15A	-0.06503	0.858034
RPS16	-0.11956	0.690686
RPS18	-0.07332	0.809917
RPS19	-0.00925	0.978173
RPS20	-0.04899	0.890235
RPS21	-0.20436	0.423511
RPS24	-0.18111	0.566984
RPS25	-0.01645	0.966666
RPS26	-0.04322	0.92547
RPS3	-0.08435	0.755857
RPS5	-0.05161	0.927675
RPS9	-0.17007	0.421232

4.2.6. Erythroblasts from patients with del(5q) MDS show increased expression of HSC genes and decreased expression of markers of erythroid differentiation

Flow cytometry analysis showed that erythroblasts from del(5q) MDS patients exhibit impaired erythroid differentiation compared to erythroblasts from healthy controls (Figure 4.2). Consistent with this observation, patient erythroblasts showed upregulation of genes, such as *GATA2* and *CD34* (logFC 0.6 and 1.3 respectively), which have been shown to be upregulated in HSCs and early erythroid cells (Novershtern *et al.*, 2011) (DESeq2 analysis, figure 4.8A). Hierarchical clustering performed using normalised read counts from DESeq2 for eight cell surface markers grouped the patient samples together (Figure 4.8C). *CD34*, *CD44*, *CD38* and *PROM1* (coding for CD133) have been found to be highly expressed in HSCs while high expression of *CD36*, *CD71*, *CD235a* and *ENG* (coding for CD105) is found in erythroid cells (Novershtern *et al.*, 2011). Consistent with impaired erythroid differentiation observed in MDS with del(5q), the expression of *CD34*, *CD44*, *CD38* and *PROM1* was higher in erythroblasts from del(5q) MDS patients than in erythroblasts from healthy controls (Figure 4.8C). Erythroid markers *CD36*, *CD71*, *CD235a* and *ENG* showed instead lower expression in patient erythroblasts compared to healthy controls (Figure 4.8C). However, the expression of these cell surface markers was variable in del(5q) MDS patients and the MDS3 patient clustered away from healthy controls but not in the same group as MDS1 and MDS2 (Figure 4.8C). MDS3 showed a higher expression of late erythroid markers compared to the other two del(5q) MDS patients (Figure 4.8C), consistent with flow cytometry analysis of the erythroid markers CD36, C71 and CD235a (Figure 4.8D). The MDS3 sample showed the highest percentage of CD36+CD235a+ and CD71+CD235a+ erythroblasts amongst the three patient samples analysed (Figure 4.8D).

4.2.7. MDS with del(5q) is characterised by decreased translation efficiency of ribosomal and erythroid genes

Read counts per gene were calculated using featureCounts and normalised using DEseq2 for RBR and total RNA from erythroblasts derived from patients with del(5q) MDS and from healthy controls treated with D-leucine. For each sample, the translation efficiency (TE) of each gene was calculated as the ratio between the read count per gene in the RBR fraction and the read count for the same gene in the total RNA fraction (Materials and Methods section 2.30.2). The translation efficiency of 9868 transcripts was measured and the $\log_2$ fold change in translation efficiency of each mRNA was calculated in patient erythroblasts treated with D-leucine compared to D-leucine-treated erythroblasts from healthy controls. A z-score cut-off of ±1.5 was utilised to identify transcripts differentially translated in erythroblasts from del(5q) MDS patients compared to erythroblasts from healthy controls (Thoreen *et al.*, 2012) (Figure 4.9A). The Compute Overlaps tool of the molecular signatures database (MSigDB, Broad Institute) was used to evaluate the overlap of mRNAs significantly suppressed in del(5q) MDS patient-derived erythroblasts with collections of genes in the MSigDB. This analysis identified a significant overlap of gene ontology (GO) gene sets related to the ribosome and ribosomal subunits (Figure 4.9B). The decrease in expression of alpha and beta globin in del(5q) MDS patient erythroblasts (Figure 4.3) could, therefore, be a consequence of the decrease in translation efficiency of mRNAs encoding haemoglobin transcripts (Figure 4.9B). Interestingly, the translation efficiency of several genes was increased in erythroblasts from del(5q) MDS patients compared to erythroblasts from healthy controls (Figure 4.9A). Transcripts with increased TE in del(5q) MDS patient-derived erythroblasts were found to overlap significantly with only one list in the MSigDB, and interestingly this was the list of upregulated genes in

$CD34^+$ hematopoietic progenitor cells after knockdown of *RPS14* (MSigDB compute overlap analysis, p value=0.035) (Ebert *et al.*, 2008).

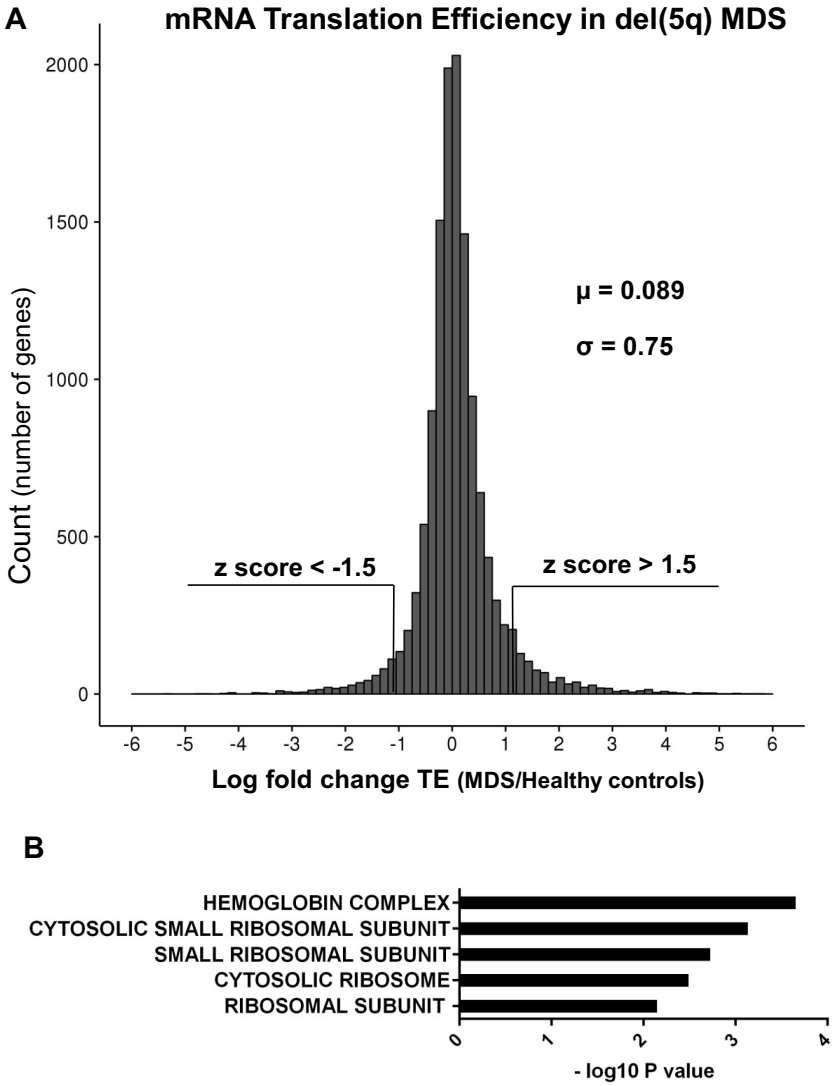
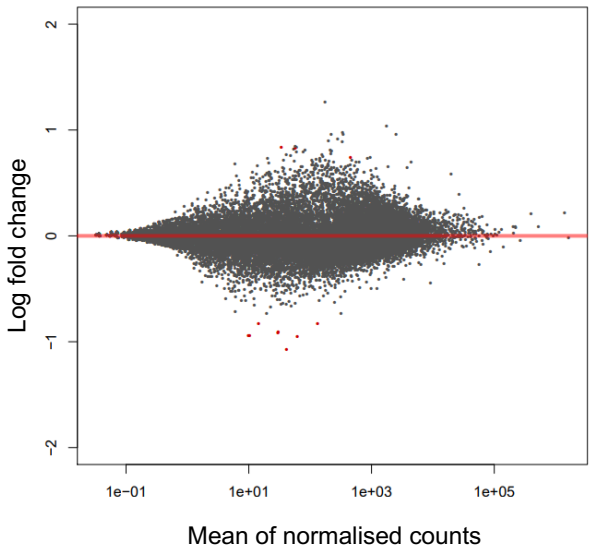


Figure 4.9. Analysis of mRNA translation efficiency in MDS del(5q) patients. A) Histogram showing the log2 fold-change in TE of all genes in MDS del(5q) patient erythroblasts treated with D-leucine compared to D-leucine treated healthy controls. n=3. The y- axis represents the number of genes (count) for a given log2 fold-change value. Log2 FC between -6 and 6 are shown. Extreme log2FC values were not shown for visualisation purposes. Transcripts with a Z score <-1.5 (corresponding to a LogFC of -1.03) or > 1.5 (corresponding to a LogFC of 1.21) were considered differentially translated. μ is the mean and σ the standard deviation. **B)** GO analysis of transcripts showing a significant decrease (z score < -1.5) in TE in erythroblasts derived from del(5q) MDS patients.

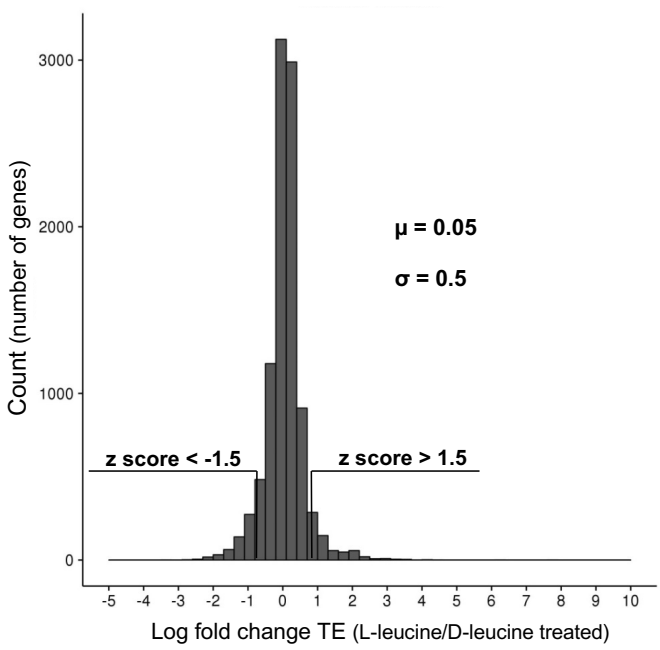
4.2.8. L-leucine treatment affects mRNA translation but not mRNA transcription in cultured erythroblasts from del(5q) MDS patients

DESeq2 identified only 12 significantly differentially expressed genes between patient-derived erythroblasts treated with L-leucine or with D-leucine (Figure 4.10A). The $\log_2$ fold change in mRNA translation efficiency was calculated for every mRNA in erythroblasts from del(5q) MDS patients treated with L-leucine compared to patient erythroblasts treated with D-leucine (Figure 4.10C) (Materials and Methods section 2.30.2). L-leucine was found to influence mRNA translation in patient erythroblasts, while having no significant impact on mRNA transcription (Figure 4.10D).

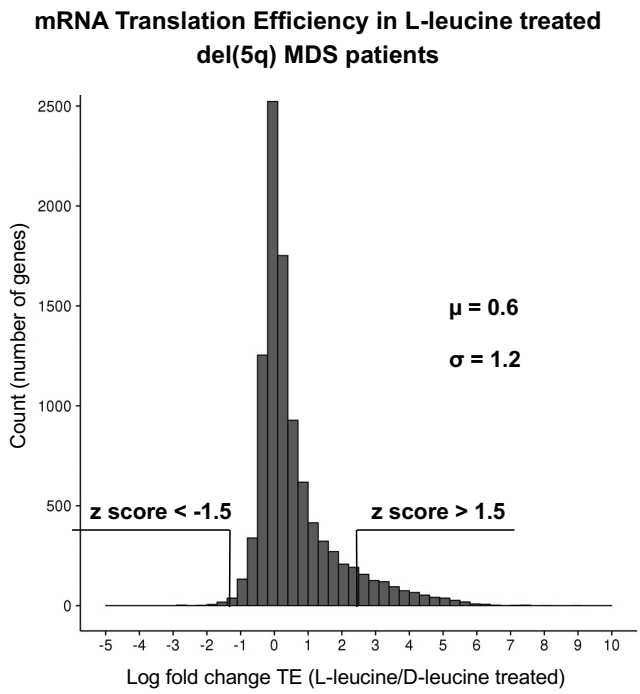
A


Log fold change
2
1
0
−1
−2
1e−01
1e+01
1e+03
1e+05
Mean of normalised counts

B


mRNA Translation Efficiency in L-leucine-treated controls
Count (number of genes)
3000
2000
1000
0
μ = 0.05
σ = 0.5
z score < -1.5
z score > 1.5
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
Log fold change TE (L-leucine/D-leucine treated)

C



D

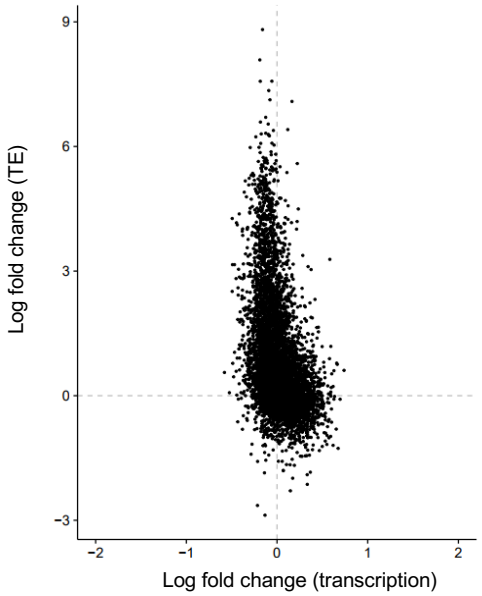


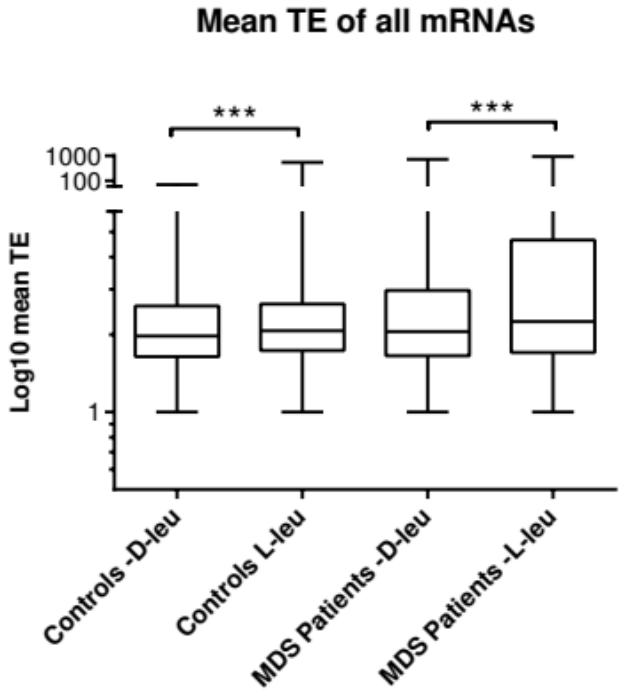
Figure 4.10. L-leucine treatment affects mRNA translation but not transcription. A) MA plot showing the log fold-change (FC) calculated using DESeq2 against the mean of the normalised RNA-seq read counts for all genes in L-leucine-treated patient erythroblasts compared to patient erythroblasts treated with D-leucine. n=3. Grey dots= non-significant genes; red dots= genes with p value <0.05. **B & C)** Histograms showing the log2 fold-change in TE of all genes in L-leucine against D-leucine-treated healthy controls (B) and del(5q) MDS patient samples (C). The y- axis represents the number of genes (count) for a given log2 fold-change value. Log2 FC between -10

and 10 are shown. Extreme log2FC values were not shown for visualisation purposes. Transcripts with a Z score <-1.5 (corresponding to a LogFC of -0.77 for Controls and -1.2 for Patients) or > 1.5 (corresponding to a LogFC of 0.89 for Controls and 2.39 for Patients) were considered differentially translated. n=3. μ is the mean and σ the standard deviation. **D)** Scatter plot showing log fold-change in gene expression (transcription) (calculated using DESeq2) against the log-fold change in TE for all genes in L-leucine-treated against D-leucine-treated patient samples. n=3.

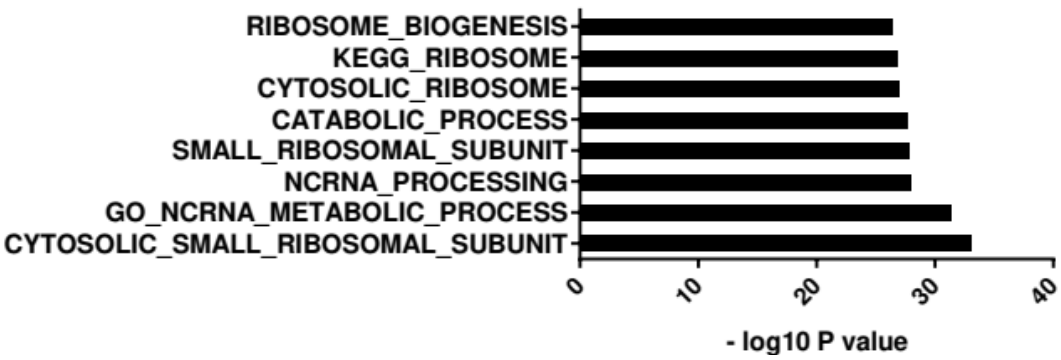
4.2.9. L-leucine treatment increases mRNA translation in cultured erythroblasts from del(5q) MDS patients

The $\log_2$ fold change in translation efficiency of each mRNA was calculated as previously described in erythroblasts from healthy controls (Figure 4.10B) or del(5q) MDS patients (Figure 4.10C) treated with L-leucine compared to D-leucine-treated erythroblasts. A significant increase in translation efficiency was observed for 440 transcripts in L-leucine-treated erythroblasts from healthy controls (Figure 4.10B) and for 910 transcripts in patient-derived erythroblasts treated with L-leucine (Figures 4.10C). Only 48 mRNAs showed a significant decrease in translation efficiency in patient-derived erythroblasts treated with L-leucine, suggesting that L-leucine leads to an overall increase in translation efficiency (TE) in patient-derived erythroblasts. In fact, mean TE across patient samples for all mRNAs was found to be increased by L-leucine treatment (Figure 4.11.A). L-leucine treatment also increased mean TE in erythroblasts from healthy controls (Figure 4.11.A), despite the fact that a similar number of transcripts were found to be significantly less or more efficiently translated (576 and 440 transcripts, respectively) in association with L-leucine treatment in erythroblasts derived from healthy controls (Figure 4.10B).

A


Mean TE of all mRNAs
***
***
1000
100
1
Log10 mean TE
Controls -D-leu
Controls L-leu
MDS Patients -D-leu
MDS Patients -L-leu

B


RIBOSOME_BIOGENESIS
KEGG_RIBOSOME
CYTOSOLIC_RIBOSOME
CATABOLIC_PROCESS
SMALL_RIBOSOMAL_SUBUNIT
NCRNA_PROCESSING
GO_NCRNA_METABOLIC_PROCESS
CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT
0
10
20
30
40
- log10 P value

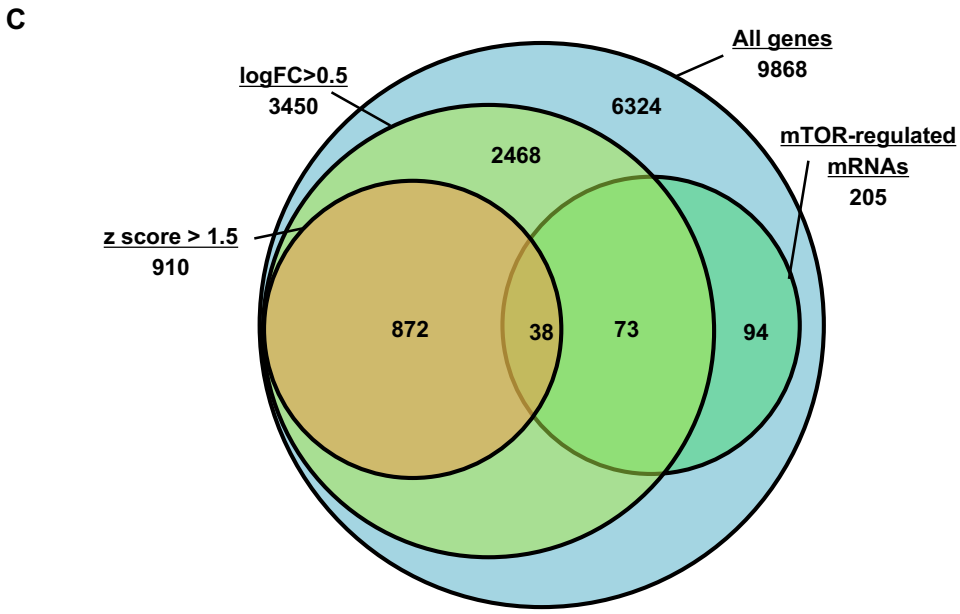


Figure 4.11. L-leucine increases mRNA translation. A) Box-and-whisker plot showing the median, inter-quartile range and range of TE of all mRNAs measured in D-leucine and L-leucine-treated healthy controls and patients. n=3. Mann-Whitney U test, *** = p<0.001. **B)** GO analysis of transcripts showing a significant increase (z score> 1.5) in TE following L-leucine treatment in MDS del(5q) patient samples. **C)** Venn diagram showing all genes measured (9868) and the overlap between transcripts with increased TE (either logFC>0.5 or z score > 1.5) and mTOR-regulated mRNAs (205). Data was analysed using a Chi-square test with Yates correction, p<0.0001.

4.2.10. L-leucine treatment of cultured erythroblasts from del(5q) MDS patients increases translation of 5'TOP mRNAs

L-leucine treatment increases mRNA translation (Figure 4.11A) as well as activation of mTORC1 (Figure 4.4). The mTOR pathway is involved in regulation of mRNA translation and mTOR target genes have been identified in a number of studies using ribosome profiling on mouse and human cells treated with mTORC1 inhibitors (Hsieh *et al.*, 2012, Thoreen *et al.*, 2012). Therefore, I investigated whether translation of mRNAs with significantly increased TE following L-leucine treatment was regulated by mTORC1. The mTOR pathway preferentially regulates translation of 5'TOP mRNAs, which encode ribosomal and translation-related proteins (Meyuhas and Kahan, Fonseca *et al.*, 2014, Thoreen *et al.*, 2012). The list of transcripts which showed a significant increase in TE following L-leucine treatment of erythroblasts from del(5q) MDS patients had a significant overlap with several ribosomal-related GO gene sets in the MSigDB (figure 4.11B). Moreover, of the 910 mRNAs which showed significantly increased TE following L-leucine treatment of patient erythroblasts, 38 were also significantly repressed in cells treated with mTORC1 inhibitors (Hsieh *et al.*, 2012, Larsson *et al.*, 2012, Thoreen *et al.*, 2012) (two-tailed Chi-square test with Yates' correction, $p < 0.0001$) (Figure 4.11C). 73 other transcripts repressed upon inhibition of mTORC1 (Hsieh *et al.*, 2012, Larsson *et al.*, 2012, Thoreen *et al.*, 2012) showed increased TE following L-leucine treatment ($logFC > 0.5$) but did not reach statistical significance (Figure 4.11C). The mRNA binding protein LARP1 has been found to stabilize 5'TOP mRNAs in a complex with the 40S ribosome subunit (Gentilella *et al.*, 2017). A recent study found that LARP1 is phosphorylated by mTORC1 and acts as a molecular switch for translation initiation of mTOR-regulated mRNAs (Hong *et al.*, 2017). Interestingly, the translation efficiency of *LARP1* was also significantly increased in del(5q) MDS patient erythroblasts treated with

L-leucine compared to D-leucine treated patient erythroblasts (log FC= 4.015 and z score=2.84). These data support the hypothesis that L-leucine activates mTORC1 and increases the translation efficiency of genes regulated by the mTOR pathway.

Since the mTOR pathway has been found to preferentially regulate translation of 5'TOP and 5'TOP-like mRNAs (Hsieh *et al.*, 2012, Thoreen *et al.*, 2012), I investigated whether the translation efficiency of 83 known 5'TOP and TOP-like mRNAs (Thoreen *et al.*, 2012) was significantly increased in L-leucine treated erythroblasts from del(5q) MDS patients compared to patient-derived erythroblasts treated with D-leucine. I found that L-leucine treatment of erythroblasts derived from both patients and healthy controls significantly increased the translation efficiency of 5'TOP and 5'TOP-like mRNAs (Figure 4.12A and 4.12B, $p<2.4x10^{-7}$ and $p<7.6x10^{-6}$ respectively). 5'TOP mRNAs which showed a significant increase in TE following treatment with L-leucine include ribosomal proteins, such as *RPS15*, *RPS27A* and *RPS14*, and other translation-related genes, such as *EEF2* and *EIF3E* (Table 4.4). No significant changes in transcription of these 5'TOP and TOP-like mRNAs were instead identified in L-leucine treated 5q- patient erythroblasts compared to patient erythroblasts treated with D-leucine (Figure 4.12C). These data suggest that the mTOR pathway is involved in mediating the effects observed following L-leucine treatment.

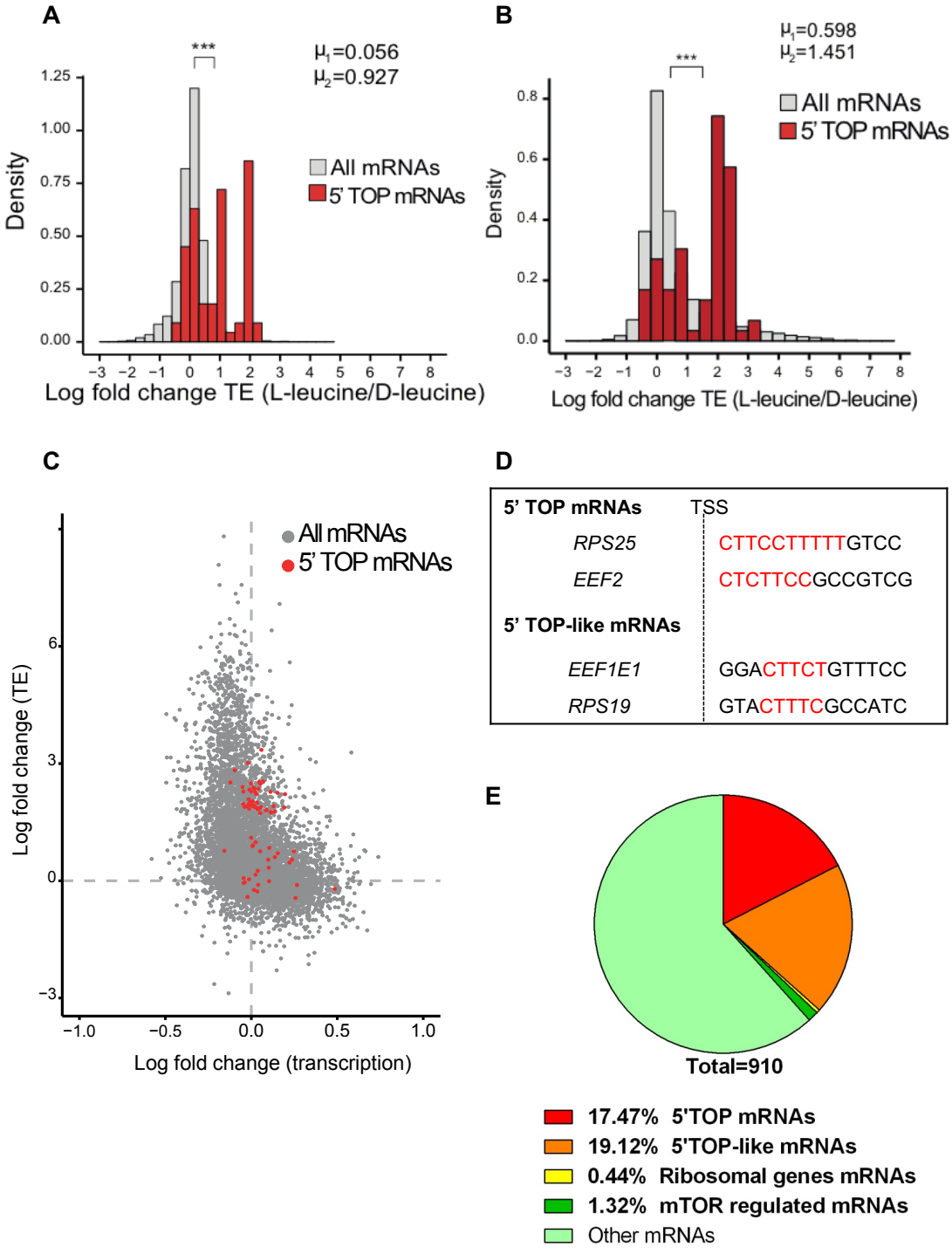


Figure 4.12. L-leucine treatment increases translation of 5' TOP and 5' TOP-like mRNAs. A & B) Histograms showing the log2 fold-change in TE of 83 known 5'TOP and 5'TOP-like mRNAs (red) and the TE of all mRNAs measured (grey) in D-leucine- over L-leucine-treated healthy controls (A) and patient erythroblasts (B). n=3. Mann-Whitney U test was performed. μ1 is the mean of the logFC in TE of all mRNAs while μ2 is the mean logFC in TE of 5'TOP mRNAs. **C)**

Scatter plot showing the log fold-change in transcription (x-axis) and the log fold change in TE (y-axis) of mRNAs following L-leucine treatment in patient samples. n=3. Red dots = 5'TOP and 5'TOP-like mRNAs; Grey dots = all mRNAs. **D)** Table showing four of the mRNAs which had a significant increase in TE (z score>1.5) following L-leucine treatment in patient samples 5'TOP and 5'TOP-like motifs are highlighted in red. TSS = translational start site. **E)** Pie chart showing the categorisation of the 910 mRNAs showing a significant increase in TE (z score>1.5) following L-leucine treatment in patients erythroblasts.

Table 4.4. Top 20 differentially translated known 5'TOP mRNAs in L-leucine treated erythroblasts from del(5q) MDS patients. For each gene, the log fold change (FC) of the translation efficiency (TE) in L-leucine treated compared to D-leucine treated patient erythroblasts is shown. Z scores were calculated using a custom R script (Appendix C) and z scores > 1.5 or <-1.5 were considered significant.

Genes	LogFC of TE in Patients	z score Patients
RPS15	3.55	2.46
RPS27A	3.48	2.40
RPS25	3.47	2.39
RPS20	3.43	2.35
RPL12	3.35	2.29
PABPC4	3.01	2.01
RPS24	2.97	1.98
RPS3	2.95	1.96
EEF2	2.83	1.86
RPS18	2.76	1.80
RPS26	2.75	1.79
RPS5	2.69	1.74
RPS21	2.64	1.70
RPS9	2.54	1.62
EIF3E	2.53	1.61
RPS14	2.52	1.60
EEF1E1	2.52	1.60
RPS19	2.49	1.57
RPS16	2.48	1.57
TPT1	2.46	1.55

Since mTORC1 has been found to preferentially regulate translation mRNAs harbouring a 5'TOP or 5'TOP-like motif (Hsieh *et al.*, 2012, Thoreen *et al.*, 2012), I further investigated the presence of these motifs on the transcripts which showed a significant increase in TE following L-leucine treatment in patient erythroblasts (Figure 4.12D). The aim of this analysis was to investigate the hypothesis that mTORC1 mediates the increase in translation of these mRNAs, resulting from treatment with L-leucine. 5'TOP mRNAs are transcripts harbouring an uninterrupted stretch of 4 to 14 pyrimidines immediately adjacent to the transcription start site (TSS) while 5'TOP-like motifs consist of a stretch of at least five pyrimidines within four nucleotides of the TSS (Thoreen *et al.*, 2012). 36.59% of transcripts which showed increased TE following L-leucine treatment of erythroblasts from del(5q) MDS patients harboured a 5'TOP (17.47%) or 5'TOP-like (19.12%) motif, as defined by Thoreen *et al.* (2012) (Figure 4.12E). A few other mRNAs with significantly increased TE following L-leucine treatment of patient-derived erythroblasts encode translation-related proteins but lack a 5'TOP or TOP-like motif (0.44%) (Figure 4.12E). Finally, 1.32% of transcripts upregulated by L-leucine treatment did not contain a 5'TOP or TOP-like motif but are known to be regulated by the mTOR pathway and were significantly repressed in cells treated with mTORC1 inhibitors (Hsieh *et al.*, 2012, Larsson *et al.*, 2012, Thoreen *et al.*, 2012). These data suggest that mTORC1 is a mediator in the increase in mRNA translation efficiency, following L-leucine treatment of erythroblasts from del(5q) MDS patients.

4.2.11. Translation of 5'TOP mRNAs is decreased in MDS with del(5q) and increased by L-leucine treatment

Polysome profiling analysis showed that the translation efficiency of 5'TOP and 5'TOP like mRNAs was increased by L-leucine treatment of del(5q) MDS patient erythroblasts (Figure 4.12B and Figure 4.13B). Moreover, translation efficiency of 83 known 5'TOP and

TOP-like mRNAs (Thoreen *et al.*, 2012) was lower in erythroblasts from del(5q) MDS patients than in erythroblasts from healthy controls (Figure 4.13A). RT-qPCR analysis of three 5'TOP mRNAs also confirmed that translation efficiency of these transcripts trended towards a decrease in patient erythroblasts compared to erythroblasts from healthy controls (Figure 4.13C). Moreover, translation efficiency of 5'TOP mRNAs was increased in L-leucine treated erythroblasts from two del(5q) MDS patients compared to D-leucine treated patient erythroblasts (Figure 4.13C). Taken together, these data show that in erythroblasts from MDS with del(5q) and characterised by a downregulation of translation of 5'TOP mRNAs, L-leucine treatment increases activation of the mTOR pathway, which leads to upregulation of mTORC1-regulated 5'TOP mRNAs (Figure 4.13B).

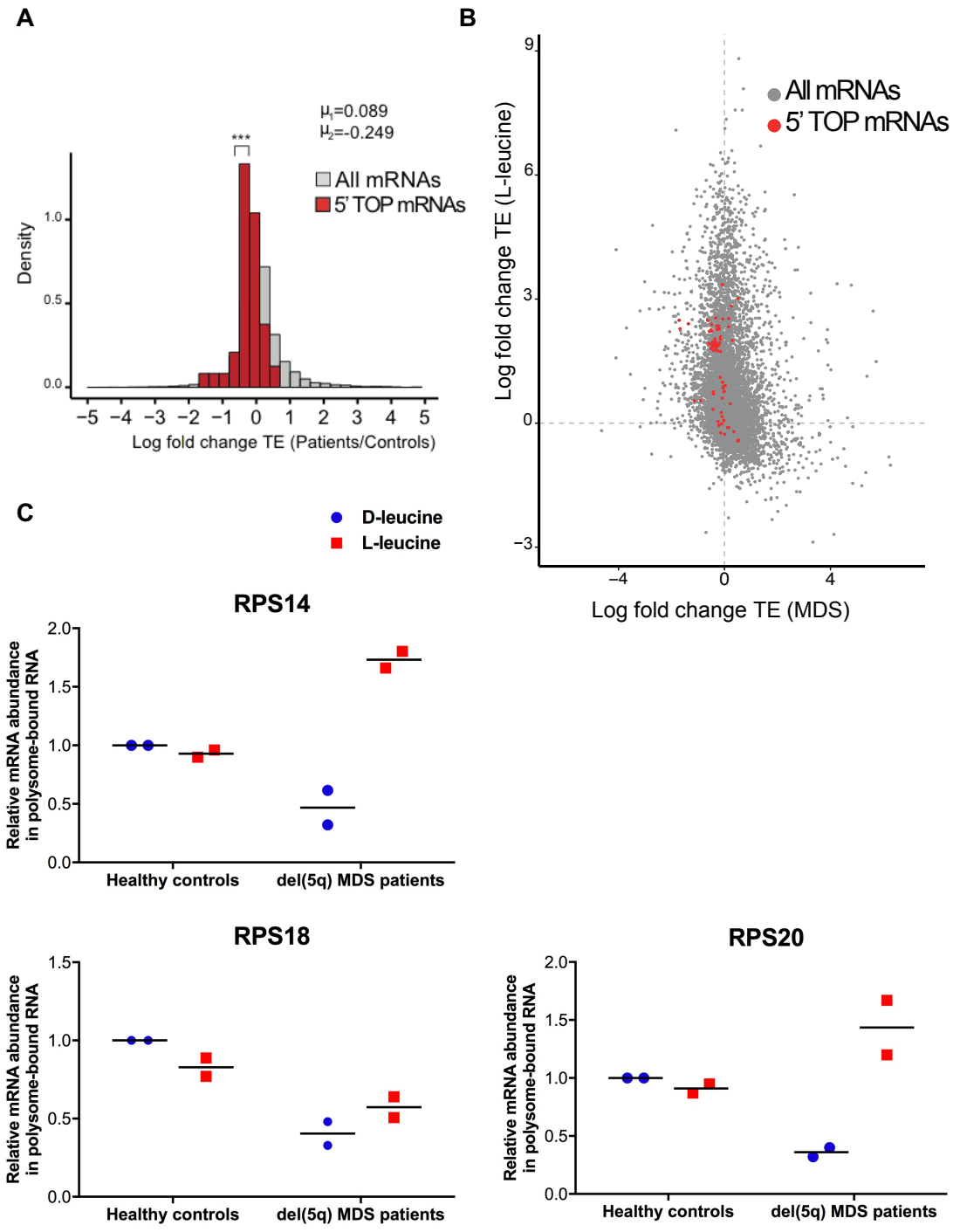


Figure 4.13. Translation of 5'TOP and 5'TOP-like mRNAs is decreased in MDS del(5q) and increased by L-leucine treatment. A) Histogram showing the log2 fold-change in TE of 83 known 5'TOP and 5'TOP-like mRNAs (red) and of all mRNAs measured (grey) in del(5q) MDS patient erythroblasts compared to healthy controls. n=3. Mann-Whitney U test was performed. μ_1 is the mean of the logFC in TE of all mRNAs while μ_2 is the mean logFC in TE of 5'TOP and 5'TOP-like mRNAs. **B)** Scatter plot showing on the y axis the log2 fold-change in TE in L-leucine treated patients compared to D-leucine treated patients. On the x axis the log2 fold change in TE in del(5q) MDS patient erythroblasts compared to healthy controls is shown. Red dots represent 5'TOP and

TOP-like mRNAs while grey dots represent all mRNAs. n=3. **C)** Dot plots showing the relative amount of RPS14, RPS20 and RPS18 mRNA in the polysome-bound fraction of RNA from two healthy controls and two MDS del(5q) patient erythroblasts, treated with either D-leucine (blue) or L-leucine (red) measured by RT-qPCR. n=2. Horizontal black lines represent the mean of the two biological replicates.

4.3. Discussion

In this study I investigated the mechanism of L-leucine-mediated enhancement of erythropoiesis in MDS with del(5q). Following a previously published study from our group (Yip *et al.*, 2012), I expanded the cohort of erythroblasts derived from del(5q) MDS patients and investigated the effects of L-leucine treatment on eight patient samples. Moreover, I optimised the polysome profiling technique for use on a small number of cells and applied it for the analysis of mRNA translation in erythroblasts from del(5q) MDS patients for the first time. However, despite the low input of cells required, the polysome profiling study was limited by the availability of samples from del(5q) MDS patients and of resources for NGS experiments. Further investigation of the effects of L-leucine on mRNA transcription and translation should be performed in a bigger cohort of patients.

In line with a previous study from our group (Yip *et al.*, 2012), erythroblasts from del(5q) MDS patients showed a decrease in erythroid differentiation compared to erythroblasts from healthy controls (Figure 4.1B, 4.1C, 4.2A and 4.2B). Consistent with impaired erythroid differentiation, erythroblasts from del(5q) MDS patients had a higher expression of genes upregulated in HSCs compared to erythroblasts from healthy controls, which showed instead a higher expression of erythroid genes (Figure 4.8C). For example, increased expression of the HSC marker CD44 had been previously reported in MDS patients (Laranjeira *et al.*, 2015). Levels of alpha and beta globin were found to be significantly lower in erythroblasts from del(5q) MDS patients (Figure 4.3C), in line with the decreased percentage of mature erythroid cells in patient-derived samples (Figure 4.2A and 4.2B). This observation is consistent with previous studies in zebrafish models of del(5q) MDS, which show that Rps14-deficient embryos exhibit severe anaemia (Payne *et al.*, 2012). I found that translation efficiency, but not transcription, of haemoglobin genes was significantly decreased in patient erythroblasts (Figure 4.9B), suggesting that

decreased globin levels and anaemia could result from decreased translation of haemoglobin genes. In line with a previous study gene expression profiling study of $CD34^+$ cells from 5q- syndrome patients (Pellagatti *et al.*, 2008), expression of ribosomal and translation-related genes was downregulated in erythroblasts from del(5q) MDS patients compared to erythroblasts from healthy controls (Figure 4.8B). Moreover, translation efficiency of ribosomal genes was also significantly decreased in erythroblasts from del(5q) MDS patients (Figure 4.9B). Decreased translation of ribosomal and translation-related genes might result from haploinsufficiency of the ribosomal gene *RPS14* which characterises MDS with del(5q) and leads to aberrant ribosome biogenesis (Pellagatti *et al.*, 2008, Ebert *et al.*, 2008). Consistent with a previous study (Yip *et al.*, 2012), L-leucine treatment of erythroblasts from del(5q) MDS patients increased expression of markers of erythroid differentiation (Figure 4.1B, 4.1C, 4.2A and 4.2B). Despite the fact that a few samples had a slightly higher percentage of erythroid cells compared to the average of all patient samples (Figure 4.8C), L-leucine treatment resulted in a significant increase in erythroid differentiation in all eight samples analysed (Figure 4.1B, 4.1C, 4.2A and 4.2B). Erythroblasts from del(5q) MDS patients treated with L-leucine showed increased levels of alpha and beta globin (Figure 4.3C), in line with the increase in erythroid differentiation following L-leucine treatment. Previous studies showed that L-leucine treatment increased haemoglobinisation and improved the anaemia in Rps14-deficient zebrafish embryos (Ear *et al.*, 2016, Payne *et al.*, 2012). The mTOR pathway regulates mRNA translation, cell growth and proliferation (Fonseca *et al.*, 2014) and a study from our group showed that L-leucine treatment of two patients with MDS with del(5q) resulted in activation of the mTOR pathway (Yip *et al.*, 2013). In line with the study by Yip *et al.* (2013), phosphorylation of mTOR targets 4EBP1, S6K1 and RPS6 was decreased in patient erythroblasts compared to healthy controls and L-leucine treatment increased

phosphorylation of mTOR targets in my expanded cohort of patient-derived erythroblasts (Figure 4.4A, 4.4B and 4.4C). The number of mTORC1 targets analysed was limited but included the main targets 4EBP1, which regulates initiation of mRNA translation, including translation of 5'TOP mRNAs, S6K1, involved in mRNA translation elongation and RPS6, involved in ribosome biogenesis (Fonseca *et al.*, 2014). It would be interesting to analyse mTOR-regulated components of the translation initiation complex eIF4F, such as eIF4G (Fonseca *et al.*, 2014). Consistent with L-leucine activation of the mTOR pathway, I showed that L-leucine treatment of patient erythroblasts increased mRNA translation (Figure 4.11A), while not influencing gene expression (Figure 4.10D). This result is consistent with a study on a zebrafish model of DBA where L-leucine treatment was found to alter protein production but not mRNA transcription (Zhang *et al.*, 2014). Treatment of patient-derived erythroblasts with L-leucine results in increased translation efficiency of mTOR-regulated 5'TOP and 5'TOP-like mRNAs (Figure 4.12), the majority of which codes for ribosomal proteins and translation-related proteins (Meyuhas and Kahan, Thoreen *et al.*, 2012). In order to investigate further whether the mTOR pathway mediates the increase in translation efficiency observed in L-leucine-treated patient erythroblasts, I analysed the mRNAs which showed increased translation efficiency following L-leucine treatment for the presence of 5'TOP and 5'TOP-like motifs (Figure 4.12D). 37% of transcripts sensitive to L-leucine treatment in patient erythroblasts, harboured a 5'TOP or 5'TOP-like motif (Figure 4.12E), suggesting that the mTOR pathway is likely to mediate the initiation of translation of these mRNAs. This analysis used the definition of 5'TOP and 5'TOP-like mRNAs utilised in the study by Thoreen *et al.* (2012). However, in this study the authors identified several other mRNAs sensitive to mTOR inhibition, which harboured pyrimidine sequences interrupted by a single purine, indicating that the spectrum of mTOR-regulated mRNAs might not be limited to 5'TOP

and 5'TOP-like transcripts following the current definition (Thoreen *et al.*, 2012). The analysis of mRNAs regulated by L-leucine could be expanded to other types of transcripts since conflicting evidence has been collected regarding the identity of mRNAs sensitive to mTOR inhibition. Studies using ribosome profiling showed that mRNAs regulated by the mTOR pathway almost exclusively harbour either 5'TOP or 5'TOP-like motifs (Hsieh *et al.*, 2012, Thoreen *et al.*, 2012). In contrast with these findings, studies using polysome profiling (Larsson *et al.*, 2012) and nanoCAGE technique (Gandin *et al.*, 2016) identified other non-5'TOP mRNAs regulated by mTOR activity. Therefore, non-5'TOP mRNAs which showed an increase in translation efficiency in L-leucine-treated patient erythroblasts may be regulated by the mTOR pathway. Future work would include the use of an inhibitor of the mTOR pathway, such as rapamycin, to investigate whether L-leucine-mediated increase of mRNA translation is dependent on mTOR activity. Erythroblasts from del(5q) MDS exhibit decreased transcription and translation of ribosomal genes (Figure 4.8B and 4.9B) and decreased translation of 5'TOP and 5'TOP-like mRNAs (Figure 4.13A). L-leucine treatment of patient-derived erythroblasts results in increased translation of 5'TOP mRNAs and ribosomal genes, leading to an overall increase in mRNA translation (Figure 4.11A). My data suggests that L-leucine activation of the mTOR pathway is likely to mediate the increase in mRNA translation in L-leucine-treated patient erythroblasts, possibly via an increase in ribosome biogenesis. L-leucine treatment resulted in increased translation of 5'TOP and 5'TOP-like mRNAs (Figure 4.12, which code for ribosomal proteins, and in activation of mTORC1 (Figure 4.4A, 4.4B and 4.4C), which is involved in enhancement of ribosome biogenesis through RPS6 phosphorylation and induction of polymerase (Pol) I and Pol III-dependent transcription of rRNAs (Chauvin *et al.*, 2014, Iadevaia *et al.*, 2014). Moreover, mTORC1 promotes cell growth via regulation of cellular processes, such as nucleotide and protein synthesis, and dysregulation of mTORC1

underlies diseases characterised by uncontrolled proliferation such as cancer (Ben-Sahra and Manning, 2017). In erythroblasts from del(5q) MDS patients, L-leucine treatment promotes ribosome biogenesis which results in increased mRNA translation and in cell proliferation and differentiation of patient erythroblasts (Figure 4.14).

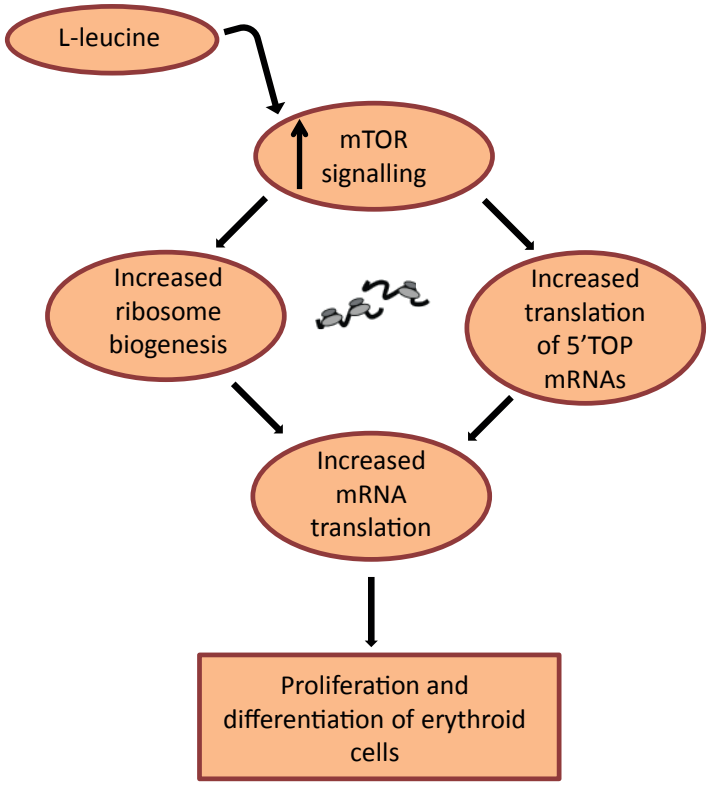


Figure 4.14. Model of the mechanism of L-leucine mediated enhancement of erythroid differentiation in MDS with del(5q).

Pathophysiology of MDS with del(5q) and 5q- syndrome is also characterised by activation of p53 (Barlow *et al.*, 2010, Pellagatti *et al.*, 2010b). Interestingly activation of p53 has been shown to inhibit mTOR activity through activation of AMPK in MEFs (Feng, 2010, Feng and Levine, 2010, Feng *et al.*, 2005). The mRNA translation defect which characterises MDS with del(5q) may, therefore, be exacerbated by p53 activation.

The exact mechanism of regulation of translation of 5'TOP mRNAs by the mTOR pathway is still unclear. Several RNA-binding proteins, such as TIA-1, TIAR and LARP1 have been found to associated with 5'TOP mRNAs (Meyuhas and Kahan, 2015). In a recent study by Gentilella *et al.* (2017) LARP1 expression was found to be downregulated in CD34+ cells of patients with 5q- syndrome. Consistent with this study, I observed a decrease in expression of LARP1 in erythroblasts from del(5q) MDS patients compared to healthy controls (Figure 4.8A). Interestingly, translation efficiency of LARP1 was increased following L-leucine treatment of patient erythroblasts, in line with the previously reported role of LARP1 in stabilisation of 5'TOP mRNAs (Gentilella *et al.*, 2017). The role of LARP1 in mediating L-leucine effect on translation of 5'TOP mRNAs needs to be further investigated using immunoprecipitation techniques in cellular models of del(5q) MDS.

5. Chapter Five – Correction of a *TP53* mutation using the CRISPR/Cas9 system

5.1. Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) (CRISPR/Cas9) system is a tool for genome engineering in various species, and can be used to correct gene mutations in eukaryotic cells (Ran *et al.*, 2013, Mali *et al.*, 2013, Schwank *et al.*, 2013, Wang *et al.*, 2013, Xie *et al.*, 2014). CRISPR/Cas9 could be employed for generation of disease models as well as for correction of causative mutations in patient-derived $CD34^+$ cells or human induced-pluripotent stem cells (hiPSCs), offering an exciting possibility for development of gene therapy for haematological disorders (Osborn *et al.*, 2016, Pellagatti *et al.*, 2015, Zhang and McCarty, 2016). *TP53* mutations are recurrent in patients with MDS with del(5q) (found in 18% of patients) (Jädersten *et al.*, 2011) and they are correlated with poor prognosis and increased risk of AML development (Boultwood *et al.*, 2010a, Sebaa *et al.*, 2012). Moreover, evidence suggests that *TP53* mutations in del(5q) MDS confer resistance to treatment with the thalidomide analogue lenalidomide (Jädersten *et al.*, 2009). Therefore, editing of the *TP53* locus could be employed in developing cellular models for the study of del(5q) MDS phenotype. In order to demonstrate the feasibility of correction of genetic mutations in cancer cells using the genome editing tool CRISPR/Cas9 (Ran *et al.*, 2013), I aimed at correcting a *TP53* mutation in a leukaemic cell line. Immortilised cancer cell lines constitute a good model for genome editing experiments since they are easy to manipulate and transfect. In order to find a suitable cell line for this study harbouring a nonsense mutation in *TP53*, I searched the catalogue of somatic mutations in cancer (COSMIC) (Forbes *et al.*, 2017) and identified the chronic myeloid leukaemia (CML) cell line K562.

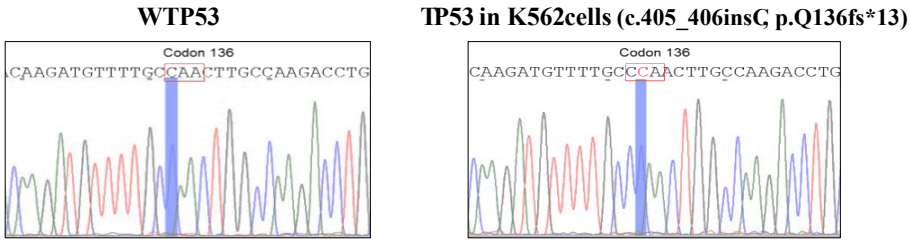
K562 cells were the most suitable model for this study since the COSMIC search did not identify any myeloid leukaemia cell line harbouring nonsense mutations in *TP53*.

5.2. Results

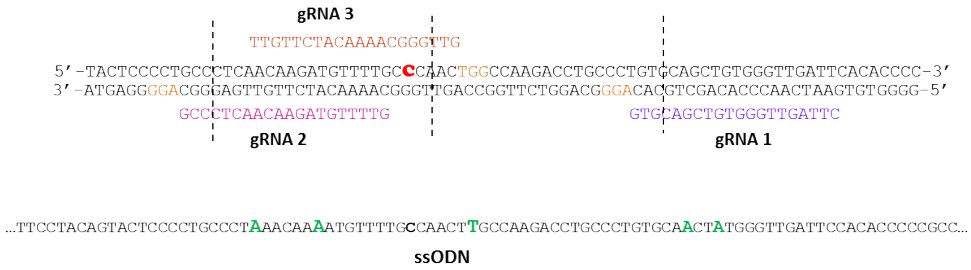
5.2.1. Validation of *TP53* mutations in K562 cells

K562 cells harbour a homozygous cytosine insertion in exon 5 of the *TP53* gene (c.405_406insC, p.Q136fs*13), which creates a premature stop codon and results in lack of p53 protein expression (Law *et al.*, 1993). Sanger sequencing of exonic regions of *TP53* in K562 cells was carried out. The presence of the *TP53* mutation c.405_406insC in codon 136 of exon 5 in K562 cells was confirmed (Figure 5.1A) and no additional mutations in the *TP53* gene were observed.

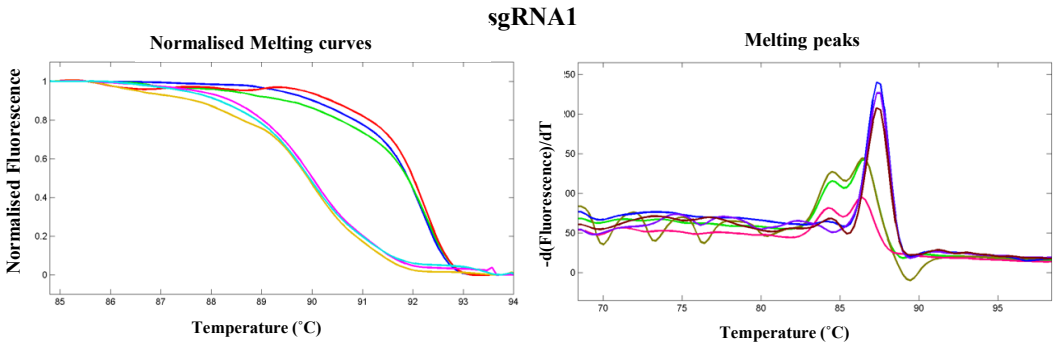
A


WTP53
TP53 in K562cells (c.405_406insC, p.Q136fs*13)
Codon 136
CAAGATGTTTTGCCAACTTGCCAAGACCTG
Codon 136
CAAGATGTTTTGCCCAACTTGCCAAGACCTG

B


gRNA 3
TTGTTCTACAAAACGGGTTG
5′-TACTCCCCTGCCCTCAACAAGATGTTTTGCCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCACACCCC-3′
3′-ATGAGGGGACGGGAGTTGTTCTACAAAACGGGTTGACCGGTTCTGGACGGGACACGTCGACACCCAACTAAGTGTGGGG-5′
GCCCTCAACAAGATGTTTTG
GTGCAGCTGTGGGTTGATTC
gRNA 2
gRNA 1
...TTCCTACAGTACTCCCCTGCCCTAAACAAAATGTTTTGCCAACTTGCCAAGACCTGCCCTGTGCAACTATGGGTTGATTCCACACCCCCGCC...
ssODN

C


sgRNA1
Normalised Melting curves
Normalised Fluorescence
Temperature (°C)
Melting peaks
-d(Fluorescence)/dT
Temperature (°C)

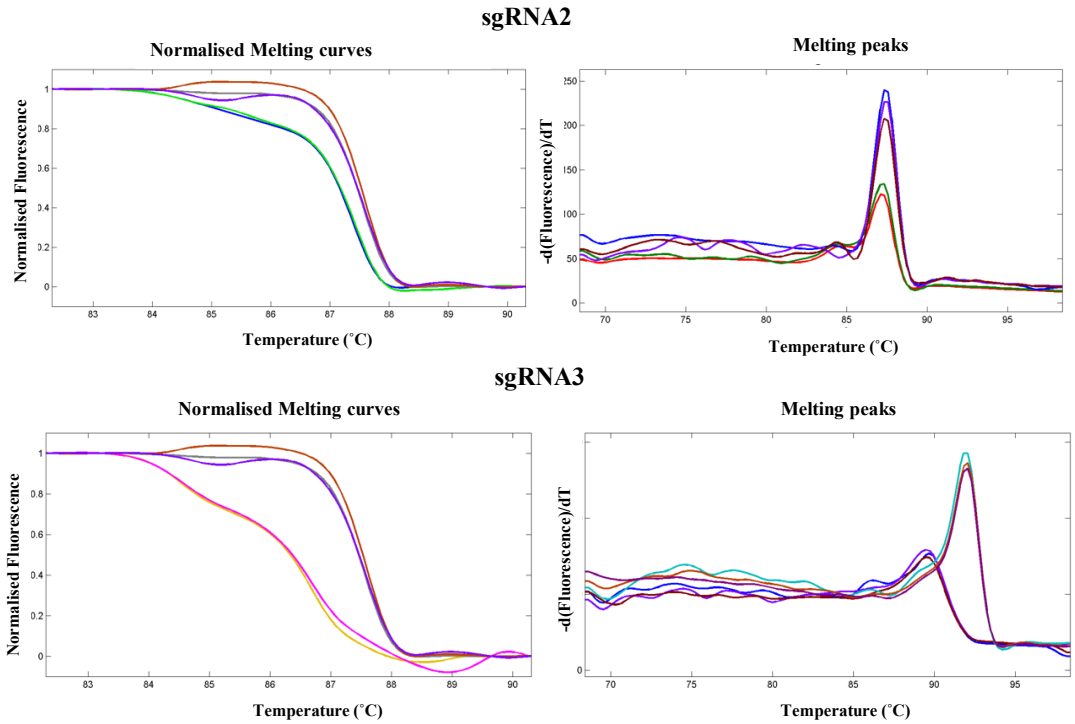


Figure 5.1. Validation of *TP53* mutations in K562 cells using Sanger sequencing and design of the CRISPR/Cas9 system in K562 cells. A). Sanger sequencing traces from NB4 cells (left) with wild type (non-mutant) *TP53* and K562 cells (right) harbouring a homozygous *TP53* point mutation (c.405_406insC). The single nucleotide insertion is highlighted in red and the corresponding peak in the sequencing trace is indicated in blue**. B)** Design of sgRNAs and 139 nt ssODN repair template used for the CRISPR/Cas9-mediated *TP53* mutation correction in K562 cells. The alignment of three designed sgRNAs to the genomic region containing the *TP53* mutation (indicated in red), the PAM sequence of each sgRNA (in orange) and the sequence of the ssODN are shown. The five silent nucleotide changes introduced in the sequence are highlighted in green and the homology arms flanking the site of alteration are indicated in purple. **C)** Plots from HRMA analysis of DNA from untrasfected K562 cells and K562 cells transfected with the Cas9 and each of the sgRNAs designed. On the left: normalised melting curves. On the right: Normalised melting peaks obtained by calculating the derivative of the normalised melting curve for each sample.

5.2.2. CRISPR design

Three different sgRNAs were designed to target the Cas9 nuclease to the genomic region where the *TP53* mutation is located in K562 cells. I transfected K562 cells with the pX458 vector expressing either of the sgRNAs and a ssODN as a repair template for CRISPR/Cas9-mediated HDR to achieve correction of this *TP53* mutation (Figure 5.1B). The ssODN had 70 bp homology arms on either side of position 405 and it also harboured five silent nucleotide changes in the sequence corresponding to either the PAM or the seed sequence of the sgRNAs. These changes were inserted in order to avoid Cas9-mediated re-cleavage of the ssODN or the *TP53* corrected site (Ran *et al.*, 2013). High-Resolution Melting Analysis (HRMA) is a method utilised for the identification of SNPs and InDels which has been employed for testing sgRNA targeting efficiency (Samarut *et al.*, 2016). HRMA is based on the analysis of differences in melting curves of fluorescently-labelled PCR products and can identify even single nucleotide changes which arise as a consequence of Cas9 cutting of a target locus. HRMA (Methods section 2.24) showed that all three designed sgRNAs were efficient at targeting Cas9 to the *TP53* locus (Figure 5.1C).

5.2.3. Evidence of HDR in K562 cells transfected with the CRISPR/Cas9 system

The *TP53* mutation in K562 cells creates a unique BglI restriction site, which is disrupted when the mutation is not present (Figure 5.2A). I exploited the presence of this restriction site to investigate editing of the *TP53* locus my CRISPR/Cas9 system design (Materials and Methods section 2.25). The pX458 vector expresses Cas9 and a sgRNA as well as Green Fluorescent Protein (GFP), which allows selection of transfected cells. Genomic

DNA from GFP positive transfected K562 cells was not completely digested by the BglI enzyme, suggesting that editing of the *TP53* locus was successful (Figure 5.2B).

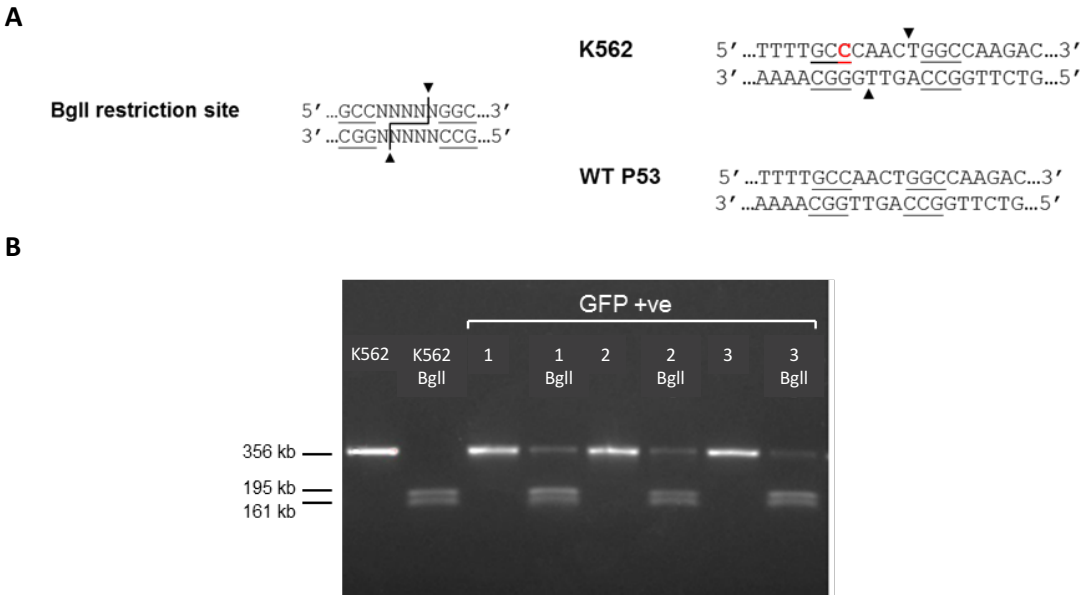


Figure 5.2. Restriction digest assay. A) The restriction site of the enzyme BglI is illustrated. On the right is the BglI restriction site in exon 5 of *T53* in K562 cells (top), where the enzyme digests the template, and in wild type *TP53* (bottom), where the restriction site is disrupted. The C insertion mutation in K562 cells is highlighted in red. **B)** BglI restriction pattern of DNA from untransfected K562 cells and GFP positive transfected cells. Each sample (transfected with sgRNA 1, 2 or 3) was run without treatment with BglI enzyme (e.g. 1) and after treatment with BglI (e.g. 1 BglI). Treated and untreated PCR products for each sample were separated by agarose gel electrophoresis.

Moreover, a ssODN containing the sequence of the promoter for bacteriophage T7 RNA polymerase (T7 promoter) was utilised to transfect K562 cells in order to investigate HDR-mediated modification of the *TP53* locus (Material and Methods section 2.19). A ssODN template was designed, which contained the wild type sequence of *TP53* at codon 136 (no mutation c.405_406insC) as well as the T7 promoter sequence, which was used as a PCR primer binding site for unique amplification of the HDR-modified DNA (Figure 5.3A). PCR amplification of genomic DNA from cells transfected with the CRISPR/Cas9 system and the ssODN containing the T7 sequence showed the presence of HDR-modified template and integration of the T7 primer binding site (Figure 5.3B). It was also found that HDR-mediated modification of the *TP53* locus was detectable starting at 18 hours after transfection (Figure 5.3C).

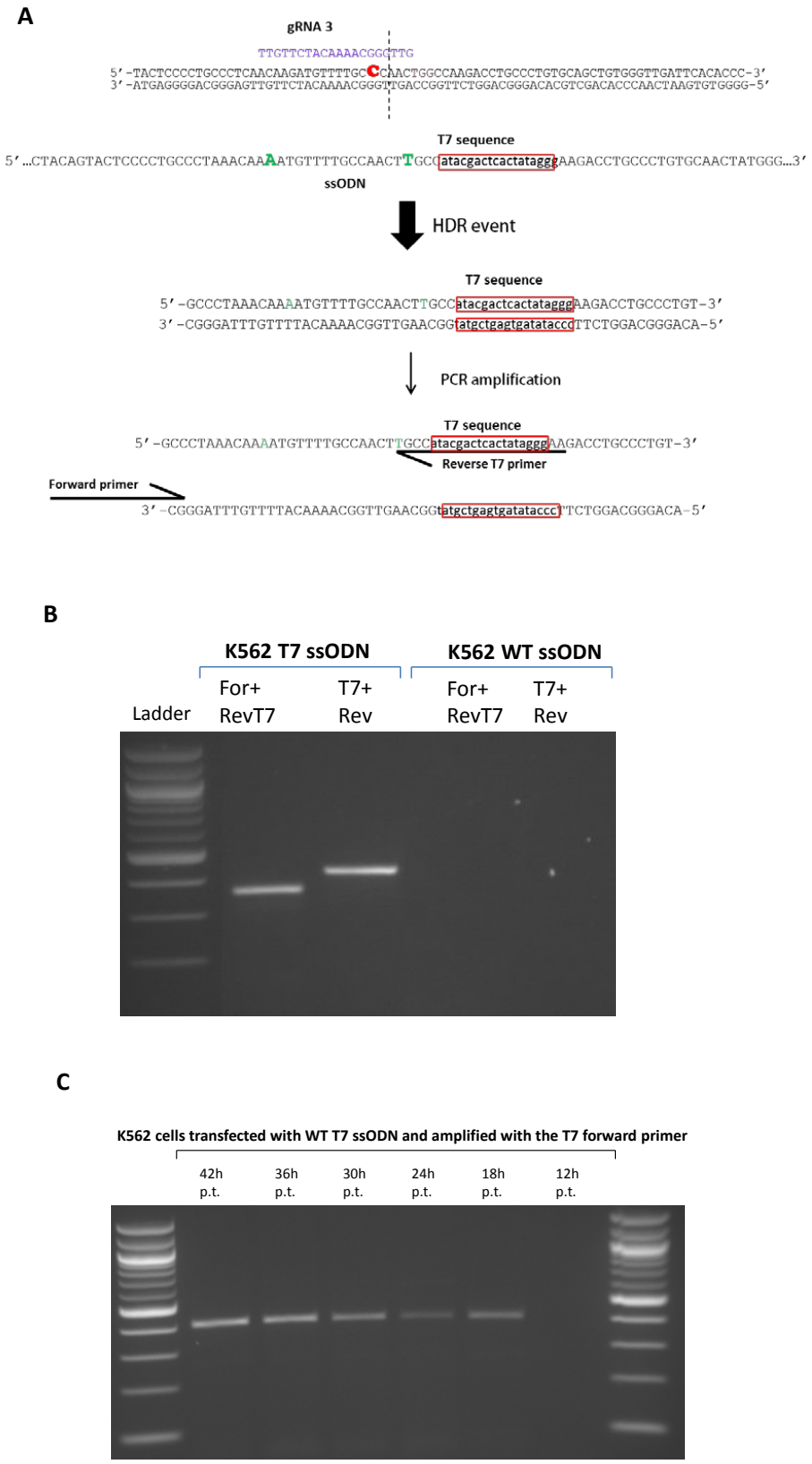


Figure 5.3. Transfection with ssODN containing the T7 primer binding site. A) Design of the T7 assay for the amplification of HDR-modified DNA. K562 were transfected with a plasmid expressing Cas9 and sgRNA3, and a ssODN containing a T7 primer binding site (T7 sequence in

red box). Following a HDR event, the ssODN would be integrated into the gDNA and HDR-modified DNA could be amplified using a primer binding to the T7 sequence (Reverse T7 primer). **B)** DNA from K562 transfected with the ssODN containing the T7 sequence (K562 T7 ssODN) and DNA from K562 cells transfected with the wild type ssODN (K562 WT ssODN) PCR amplified with either a reverse primer binding the T7 sequence (For+RevT7) or a forward primer binding the T7 sequence (T7+Rev). The PCR products were separated by agarose gel electrophoresis. No band is detected from DNA of K562 cells transfected with the wild type ssODN, showing that HDR-modified templated is uniquely amplified by PCR in the assay. **C)** DNA from K562 cells transfected with pX458 vector expressing sgRNA 3 and the ssODN containing the T7 site. DNA was collected every 6 hours and PCR amplified using a forward primer binding to the T7 sequence. The PCR products were separated by agarose gel electrophoresis. p.t.=post transfection.

5.2.4. *TP53* mutation correction and detection of additional mutations

K562 cells were transfected with the pX458 vector expressing either of the sgRNAs and a wild type ssODN. In order to isolate *TP53* mutation corrected clones, single GFP positive transfected cells were FACS sorted into eight 96-well plates per sample (transfected with either of the sgRNAs) and GFP negative transfected cells were sorted into one plate per sample. The majority of GFP positive colonies in each plate did not survive in culture (the average percentage of live cells across all samples was 23.09%) while around 60% of GFP negative colonies were viable 12 days after transfection (Table 5.1). A total of 293 colonies were genotyped by Sanger sequencing and three colonies with homozygous correction of *TP53* mutation were identified (Table 5.2). However, additional changes affecting the *TP53* genomic sequence were identified in all three mutation corrected colonies (Table 5.2).

Table 5.1. Percentage of viable colonies per 96-well plate at day 12 post transfection.

	GFP positive								**GFP negative**
Sample	Plate 1	Plate 2	Plate 3	Plate 4	Plate 5	Plate 6	Plate 7	Plate 8	Plate 1
gRNA 1	13.5	25	10.4	15.6	24.7	15.6	29.1	18.7	60.4
gRNA 2	22.9	13.5	36.5	27	29.1	20.8	20.8	21.8	54.16
gRNA 3	26	19.7	22.5	22.5	30.2	31.2	27	30.2	66.6

Table 5.2. Detection of c.405_406insC *TP53* mutation correction in colonies grown from single cells transfected with the CRISPR/Cas9 system. All mutation corrected colonies identified also harboured additional mutations affecting exon 5 of *TP53*. None of the colonies survived in culture except for Colony 1 (in bold).

	Number of colonies sequenced	c.405_406insC mutation corrected colonies		Additional mutation
sgRNA 3	132	0		
sgRNA 2	76	1	**Colony 1**	**c.388C>A, p.L130I**
sgRNA 1	85	2	Colony 2	20 nt deletion
			Colony 3	c.425C>T, p.P142L

In order to increase the survival of FACS sorted single cells and maximise the possibility of identifying *TP53* mutation corrected colonies without other sequence alterations, I performed a second CRISPR/Cas9 transfection of K562 cells using a different culture media (Materials and Methods 2.2). Single GFP positive transfected cells were FACS sorted into nine 96-well plates per sample (transfected with either of the sgRNAs) and GFP negative transfected cells were sorted into one plate per sample. Approximately 50% of GFP positive transfected cells survived in culture (Table 5.3). Four *TP53* mutation corrected colonies (GFP positive) were identified in this screening, however these colonies also harboured additional mutations in the *TP53* gene (Table 5.4). Overall it was found that in both transfection experiments GFP negative transfected K562 cells survived better than GFP positive cells (Table 5.1 and Table 5.3) and that all mutation corrected colonies identified also harboured additional mutations affecting exon 5 of *TP53*.

Table 5.3. Percentage of viable colonies per 96-well plate at day 14 post transfection in a second experiment.

	GFP positive									GFP negative
Sample	Plate 1	Plate 2	Plate 3	Plate 4	Plate 5	Plate 6	Plate 7	Plate 8	Plate 9	Plate 1
gRNA 1	41.6	51	50	61.4	48.9	47.9	42.7	41.9	54.1	66
gRNA 2	38.5	41.6	35.4	50	47.9	45.8	46.8	45.8	48.9	81.2
gRNA 3	51	55	50	54.1	54.1	47.9	44.7	42.7	54.1	75

Table 5.4. Detection of c.405_406insC *TP53* mutation correction in colonies grown from single cells transfected with the CRISPR/Cas9 system in the second experiment. A higher number of colonies was genotyped, however the four mutation corrected colonies identified (Colony 4 to 7) also harboured additional mutations in exon 5 of *TP53*, close to the site of gene correction.

	Number of colonies sequenced	c.405_406insC mutation corrected colonies		Additional mutation
sgRNA 3	179	0		
sgRNA 2	168	4	Colony 4	c.392_393insA, p.N131fs*17
			Colony 5	c.412_413insG, p.A138fs*10
			Colony 6	11nt deletion
			Colony 7	c.380C>T, p.S127F
sgRNA 1	168	0		

Previous reports showed that CRISPR/Cas9 system can potentially generate undesired off-target effects at sites with homologous sequences to the sgRNA (Mali *et al.*, 2013, Hsu *et al.*, 2013, Schwank *et al.*, 2013). Therefore, I investigated whether sgRNA 2 had off-target effects which could be associated with the additional *TP53* mutations found in *TP53* mutation corrected colonies. sgRNA 2 was predicted to have 402 possible off-target sites, of which 11 were located within exons (CRISPR Design Tool, http://crispr.mit.edu/) but none of the off-target sites identified was located in proximity to *TP53*. Moreover, the top

five off-targets sites predicted by the CRISPR Design Tool for sgRNA 2 were analysed by Sanger sequencing and no sequence alterations were detected (Materials and Methods section 2.26).

5.2.5. Restoration of *TP53* protein expression in a corrected clone with an additional mutation

None of the *TP53* mutation corrected colonies except for Colony 1 survived in culture. Colony 1 showed homozygous correction of *TP53* mutation and the presence of three out of five silent nucleotide changes introduced in the ssODN template; however, a substitution at position 388 of *TP53* was introduced in proximity of the Cas9 cut site and resulted in a leucine to isoleucine amino acid change at position 130 (Figure 5.4A).

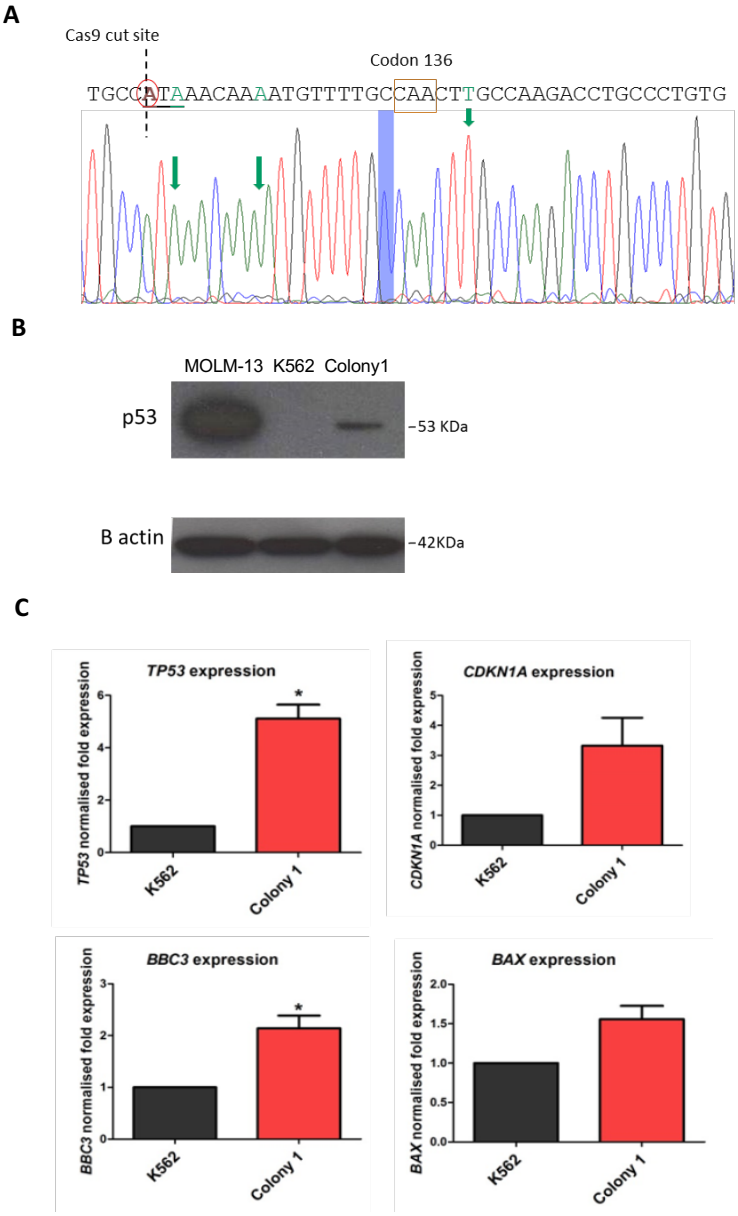


Figure 5.4. Analysis of Colony 1 with correction of the *TP53* mutation. A) Sanger sequencing trace of isolated Colony 1. Correction of C insertion mutation at codon 136 of *TP53* is indicated by the orange box and the additional mutation (C>A) at codon 130 is highlighted by the red circle on the left. This additional mutation (c.388C>A, p.L130I) is located adjacent to Cas9 cut site, illustrated by the dashed line. Green arrows indicate the presence of three (out of five) silent nucleotide changes from the ssODN sequence. **B)** Evaluation of p53 protein expression by Western blotting. From left to right: p53 protein expression in the MOLM-13 cell line (acute myeloid leukaemia cell line with wild type *TP53*), in untransfected K562 cells and in mutation corrected Colony 1. β-actin was used as loading control. **C)** Bar graphs showing expression levels of *TP53*, *CDKN1A*, *BBC3* and *BAX* genes using RT-qPCR. The values in Colony 1 cells are relative to the untransfected K562 cells (expression=1). Error bars represent the SEM of three technical replicates. A paired t-test was performed. * = $p<0.05$.

Analysis of the p.L130I *TP53* mutation observed in Colony 1 cells using the software PolyPhen-2 (Adzhubei *et al.*, 2010) predicted that the amino acid substitution could be damaging to the protein function (score 0.998). CRISPR/Cas9-mediated correction of *TP53* mutation at codon 136 in Colony 1 restored expression of p53 protein (Figure 5.4B) and increased mRNA levels (Figure 5.4C). Expression of the *TP53* target genes *CDKN1A*, *BBC3* and *BAX* was also increased in Colony 1 cells compared to untransfected K562 cells but the increase in expression of *CDKN1A* and *BAX* did not reach significance (Figure 5.4C). The fact that Colony 1 expressed p53 but harboured an additional *TP53* mutation (c.388C>A) provides a possible explanation for the survival of Colony 1 in culture.

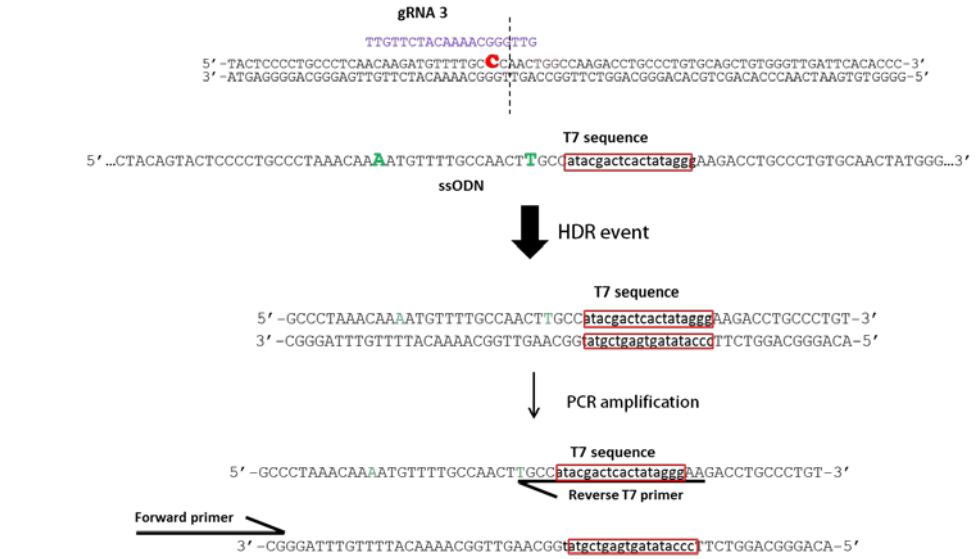
5.2.6. Investigation of HDR-mediated modification of the *TP53* locus in K562 cells

We hypothesised that additional mutations in corrected K562 cells arose as a consequence of the biological function of *TP53*. *TP53* is a tumour suppressor gene but *TP53* mutations have been found to have an oncogenic role (Soussi and Wiman, 2015). Therefore, the effects of *TP53* mutation correction may be context dependent. K562 cells harbour a number of mutations and cytogenetic abnormalities, including the Philadelphia chromosome (Naumann *et al.*, 2001). Restoration of p53 expression and tight regulation of the cell cycle in the genetic context of K562 cells might lead to apoptosis of the corrected cells expressing wild type *TP53*. Additional compensatory mutations in *TP53* mutation corrected K562 cells might provide a proliferative advantage, leading to isolation of only single cell colonies with *TP53* mutation correction which harbour additional genetic changes. In order to investigate this hypothesis, I studied the occurrence of additional genetic changes within the *TP53* locus in K562 cells expressing wild type *TP53* as opposed to K562 cells CRISPR- modified with a ssODN which contained a mutated sequence of *TP53*. I designed a ssODN containing the wild type sequence for *TP53* and the T7 promoter

sequence as described above (wild type T7 ssODN) (Figure 5.5A) and a ssODN containing the T7 primer binding site and an additional insertion of two nucleotides (mutant T7 ssODN) (Figure 5.5B). In the wild type T7 ssODN the T7 sequence was inserted in frame while the two additional nucleotides included in the mutant T7 ssODN caused a frameshift and no restoration of *TP53* expression in CRISPR-modified K562 cells. K562 cells were transfected with the pX458 vector expressing sgRNA 3 and either the wild type T7 ssODN or the mutant T7 ssODN. sgRNA 3 was chosen over the other two sgRNAs designed because the PAM sequence on the ssODN had been silently mutated, decreasing the likelihood of Cas9 re-cutting of the locus after HDR (Ran *et al.*, 2013). Genomic DNA was extracted from transfected GFP positive cells and PCR amplified using specific primers binding the T7 sequence on each ssODN (Methods section 2.5, Table 2.4).

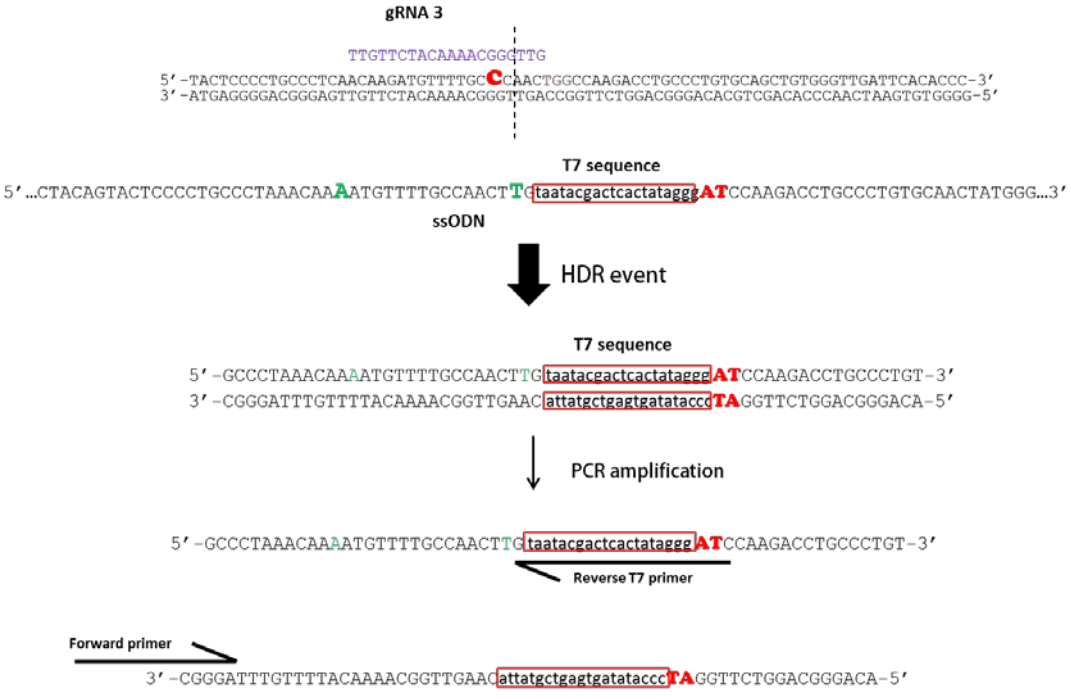
A

WT T7 ssODN



B

Mutant T7 ssODN



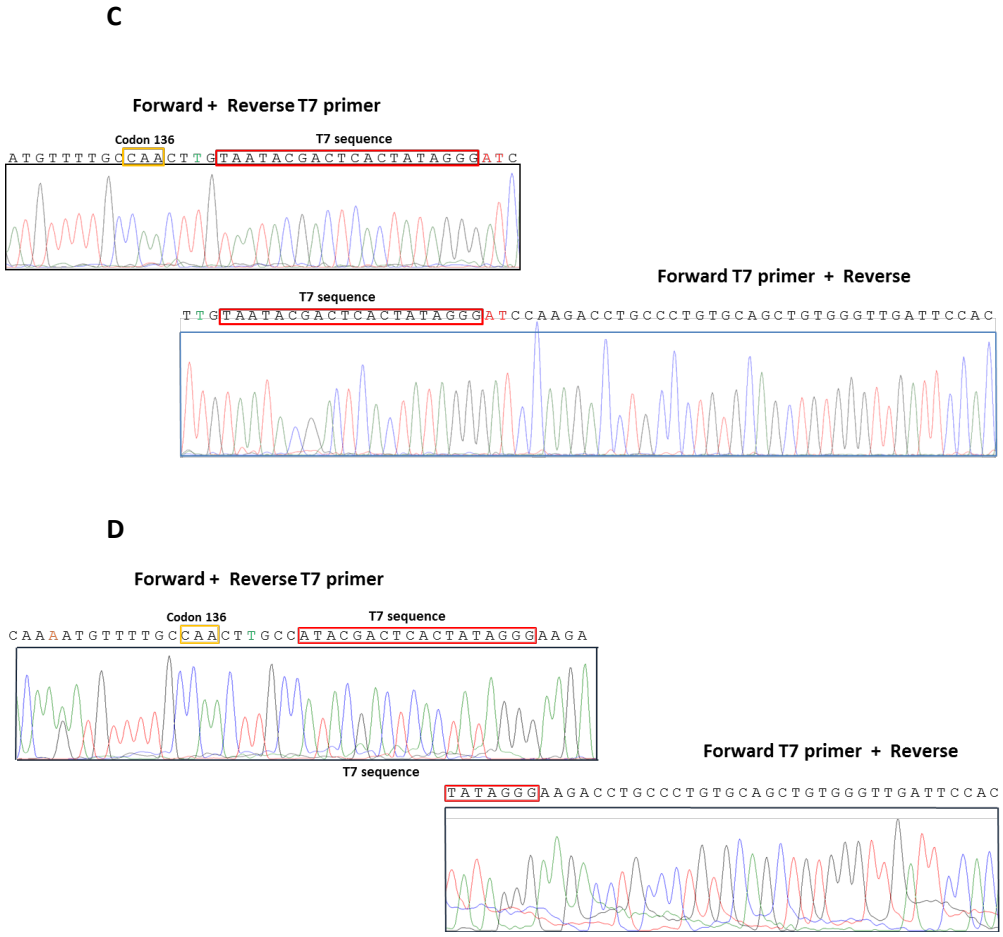
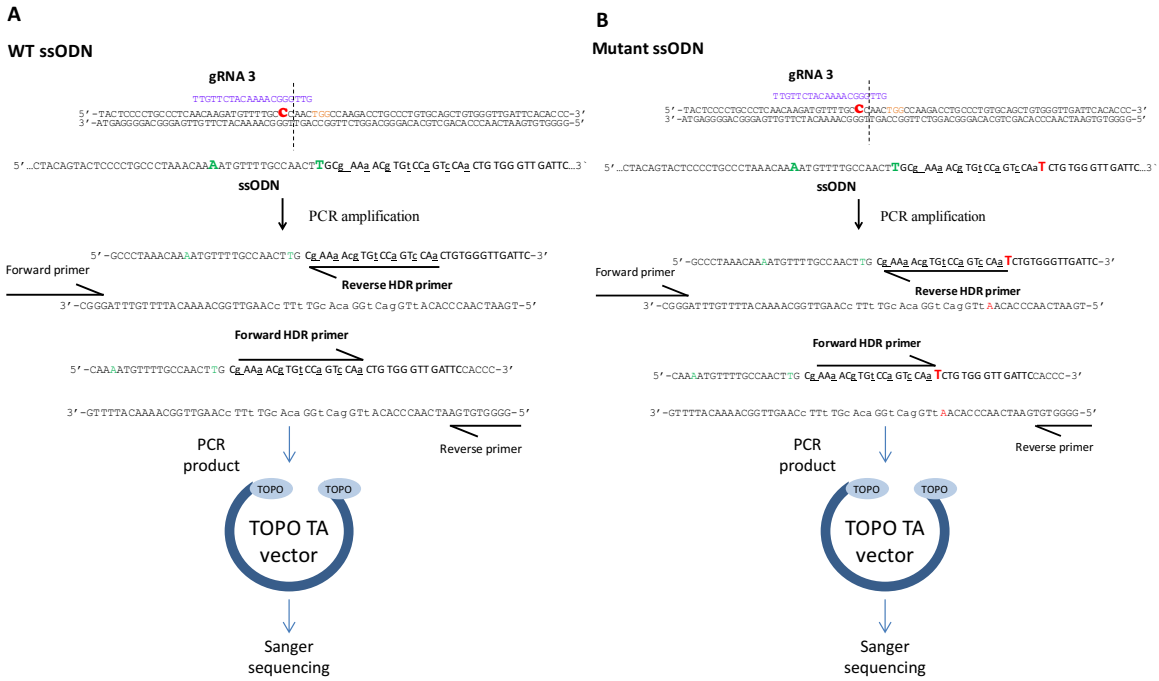


Figure 5.5. HDR-mediated modification of the *TP53* locus using a ssODN containing the T7 sequence. A&B) Design of T7 assay to investigate the effect of HDR- mediated modification of *TP53* using a wild type (WT T7 ssODN) (A) or a mutant (Mutant T7 ssODN) (B) ssODN containing the T7 sequence (red box). The insertion of two nucleotides from the Mutant T7 ssODN is shown in red (B). K562 were transfected with a plasmid expressing Cas9 and sgRNA3, and either of the ssODNs. Following a HDR event, the ssODN would be integrated into the gDNA and HDR-modified DNA could be amplified using a primer binding to the T7 sequence (Reverse T7 primer). **C&D)** Sanger sequencing traces from DNA from K562 cells transfected with the mutant T7 ssODN (C) and the wild type T7 ssODN (D) amplified with primers containing the T7 sequence. Codon 136 is highlighted a yellow box, the T7 sequence in shown in a red box and the insertion of two nucleotides from the mutant T7 ssODN is shown in red (C).

Sanger sequencing of DNA from K562 cells transfected with the mutant T7 ssODN showed scarless HDR-mediated modification of the *TP53* locus and no additional mutations were detected in the amplicon sequenced (Figure 5.5C). Surprisingly, Sanger sequencing of genomic DNA from K562 cells transfected with wild type T7 ssODN also revealed the absence of additional mutations in the HDR-modified *TP53* amplicon (Figure 5.5D).

The absence of detectable additional mutations in the DNA of K562 cells transfected with the wild type T7 ssODN is in contrast with the results of Sanger sequencing of K562 transfected single cell colonies (Table 5.2 and Table 5.4). We hypothesised that the reason why the additional *TP53* mutations were not detectable in this assay was that the in frame insertion resulting from the T7 sequence (coding for five amino acids) was deleterious for p53 protein function. The absence of expression of wild type p53 in K562 cells modified with the wild type T7 ssODN might have prevented the occurrence of additional genetic changes, such as the ones observed in *TP53* mutation corrected single cell colonies. In order to investigate this hypothesis, I designed two additional ssODN templates containing either the wild type sequenced of *TP53* (Figure 5.6A) or a missense mutation (c.432insT, Figure 5.6B), and barcodes of silent mutations, which created unique primer-binding sites for PCR amplification of HDR-modified genomic DNA. We investigated whether restoration of *TP53* expression in K562 cells increases apoptosis, by flow-cytometry apoptosis analysis of K562 cells transfected with the wild type ssODN compared to cells transfected with the mutant ssODN. The number of apoptotic cells in the sample transfected with the wild type ssODN did not differ from the number of apoptotic cells in the sample transfected with mutant ssODN, at six different time points after transfection (Figure 5.7A and Figure 5.7B). Moreover, cell proliferation was measured for 10 days after transfection and no differences were found in the proliferation of K562 cells transfected with the wild type or mutant ssODN (Figure 5.7C). DNA from K562 cells transfected with either wild

type ssODN or mutant ssODN was amplified with primers binding the barcodes of silent mutations and PCR products were cloned into TOPO-TA vectors (Thermo Fisher Scientific) (Figure 5.6A and Figure 5.6B). Sanger sequencing of TOPO-TA vectors allowed analysis of each HDR-modified DNA template separately, as opposed to sequencing of the bulk of CRISPR-transfected cells (Figure 5.6A, B and C). As expected, Sanger sequencing of DNA extracted from TOPO-TA clones showed that 11 sequences from DNA of K562 cells transfected with wild type ssODN harboured additional genetic changes in the *TP53* amplicon and 14 sequences from DNA of K562 cells transfected with mutant ssODN did not harbour any additional mutation in *TP53* (Figure 5.6C). However, I found 9 sequences from DNA of K562 cells transfected with the wild type ssODN which did not harbour any additional genetic changes in the *TP53* amplicon (Figure 5.6C).



C

Sample	HDR sequences with mutations	HDR sequences without mutations
HDR silent WT	11	9
HDR silent Mutated	6	14

Figure 5.6. HDR-mediated modification of *TP53* using a ssODN containing a barcode of silent mutations. A&B) Experimental design. K562 cells were transfected with sgRNA3 and either a wild type (WT ssODN) (A**)** or a mutant (Mutant ssODN) (B) ssODN. Both ssODNs contained a barcode of silent mutations (mutated nucleotides shown in lower case and underlined). The T insertion mutation in the Mutant ssODN (B) is shown in red. DNA was extracted and amplified with specific primers binding the barcodes of silent mutations (Reverse HDR primer and Forward HDR primer). PCR products were cloned into TOPO TA vectors and sequenced. **C)** Table showing the number of TOPO TA vectors containing the sequence of the wild type ssODN (HDR silent WT) and the mutant ssODN (HDR silent Mutated) with or without additional mutations in *TP53*.

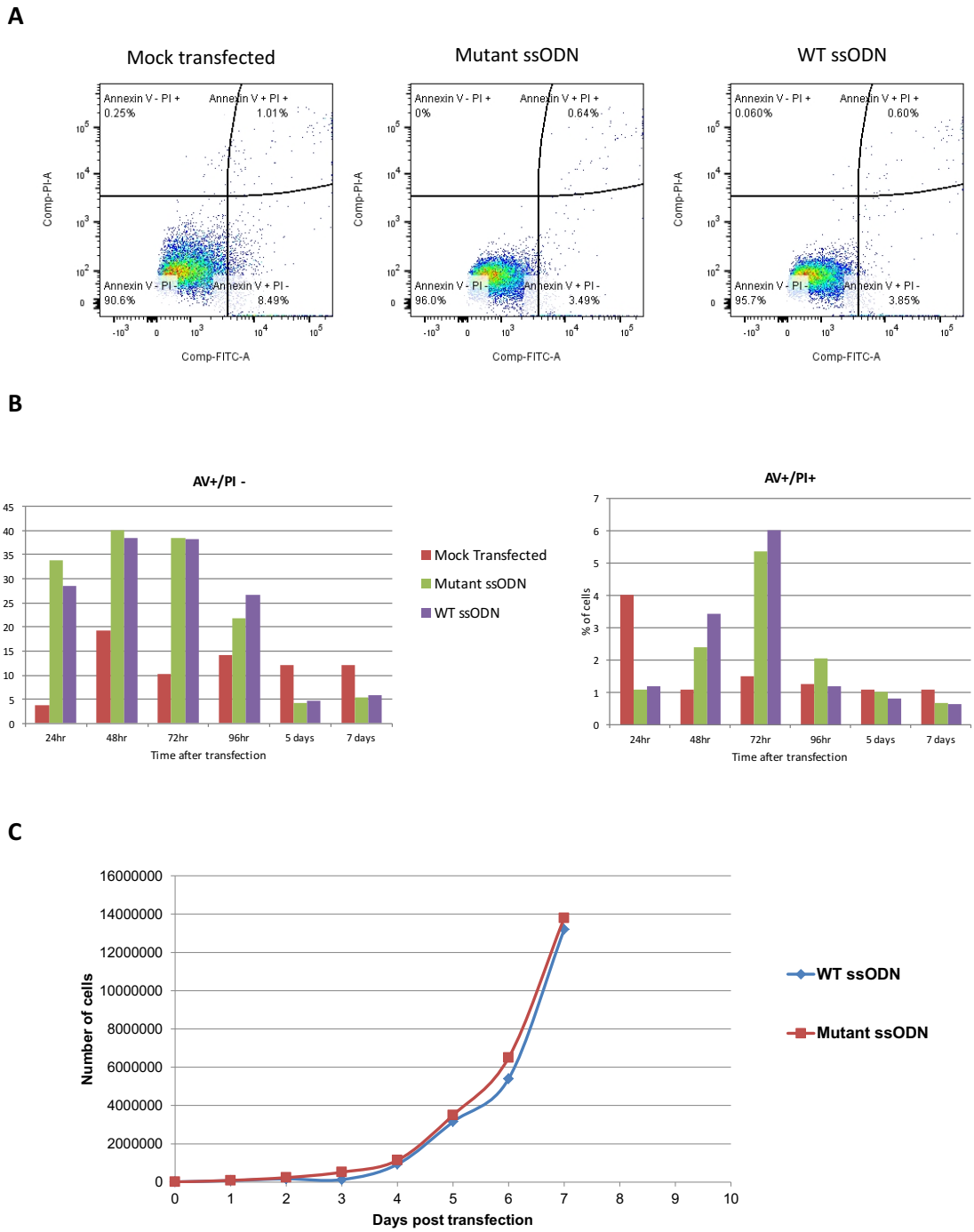


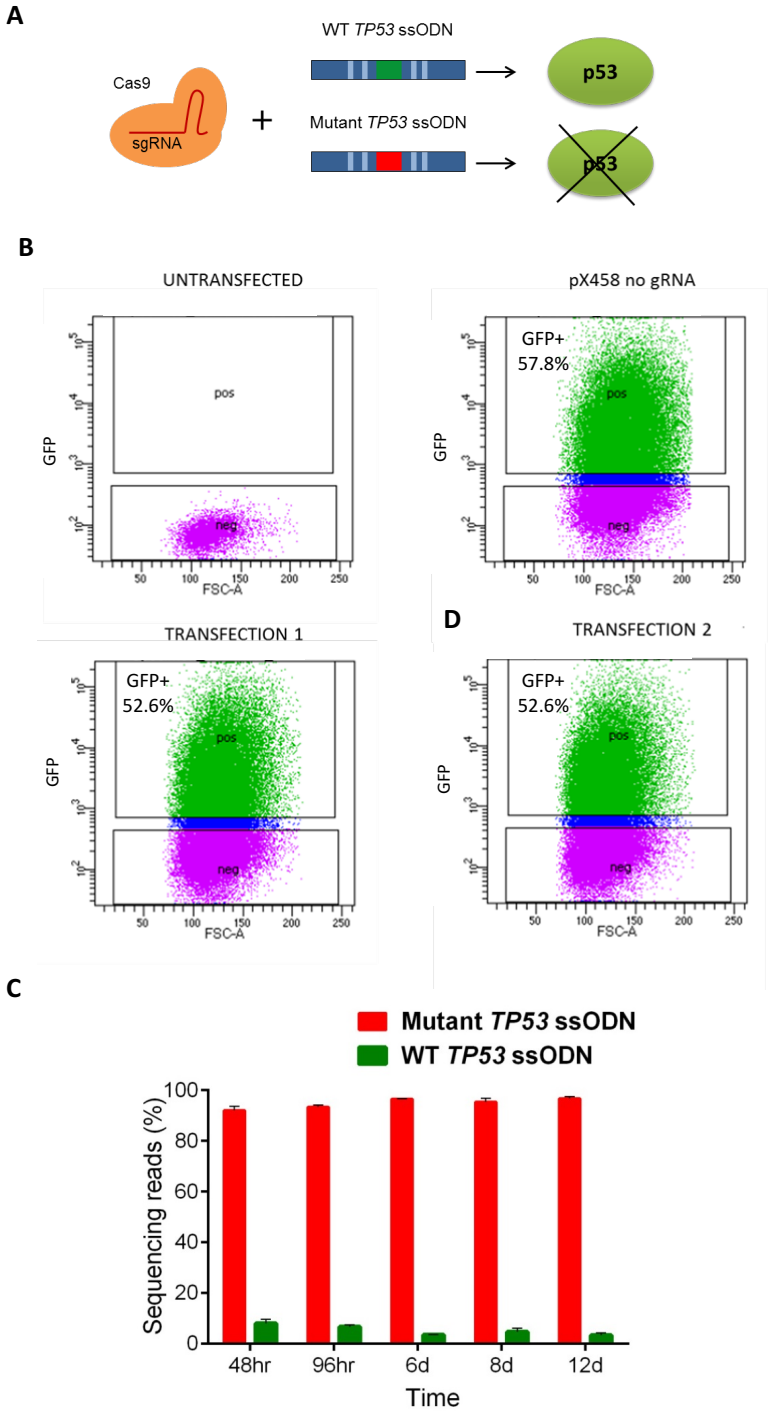
Figure 5.7. Analysis of cellular apoptosis in K562 cells transfected with wild type or mutant ssODNs. A) Scatter plots showing the percentage of cells in the different stages of apoptosis, as measured by Annexin V (AV) and Propidium Iodide (PI) staining and flow cytometry. AV positive and PI negative cells were considered early apoptotic and AV positive PI positive cells are late apoptotic. The analysis was performed 7 days after transfection with either a mutant ssODN (Mutant ssODN) or a wild type ssODN (WT ssODN). Electrophorated K562 cells not transfected with DNA were also analysed (Mock transfected). **B)** Bar graphs showing the percentage of early apoptotic (AV+/ PI -) and late apoptotic (AV+/PI+) K562 cells 7 days after transfection measured by flow cytometry. K562 cells were either mock transfected, transfected with the mutant ssODN

(Mutant ssODN) or transfected with the wild type ssODN (WT ssODN). n=1. **C)** Proliferation curve of transfected K562 cells generated using Trypan Blue staining of cells for 10 days after transfection. K562 cells were transfected with either the wild type ssODN (WT ssODN) or the mutant ssODN (Mutant ssODN). n=1.

5.2.7. Investigation of HDR-mediated modification of *TP53* and *ASXL1* using CRISPR-introduced barcodes

Next Generation Sequencing (NGS) technology was used to investigate the effects of HDR-mediated modification of the *TP53* locus in K562 cells. K562 cells were transfected with pX458 plasmid, expressing GFP, Cas9 and sgRNA3 (Materials and Methods section 2.21, Table 2.9), and a mix of ssODNs containing either the wild type sequence of *TP53* (WT *TP53* ssODN) or a missense mutation (Mutant *TP53* ssODN) (Figure 5.8A). A barcode of silent point mutations was also inserted to allow identification of the integrated ssODNs by targeted DNA sequencing (Figure 5.8A). GFP positive transfected cells (52.6%, Figure 5.8B) were sorted and DNA was extracted at five different time points after transfection Figure 5.8C). The percentage of HDR events detected was 1%. It was found that the percentage of K562 cells with integration of the mutant *TP53* ssODN was significantly higher than the percentage of cells with integration of the WT *TP53* ssODN (Figure 5.8C). K562 cells were also transfected with a ssODN which contained an "N" base: ssODNs containing either a G, an A, a T or a C at position 422 were synthesised and mixed in equal amounts (Figure 5.8D). The position 422 was chosen since HDR-mediated modification of *TP53* with the ssODN harbouring an A base was predicted to lead to lack of p53 expression (via formation of a premature stop codon). The presence of a G at position 422 caused instead a missense mutation in *TP53* in HDR-modified cells, the presence of a C caused a synonymous mutation and T was the wild type nucleotide at position 422 (Figure 5.8D). DNA from transfected cells was sequenced at five time points after transfection and it was found that the percentage of K562 cells with integration of ssODNs which restored wild type p53 expression (containing a T or a C at position 422) was markedly lower than the percentage of cells where the integration of the ssODN lead to a lack of wild type p53 expression (N=A or G) (Figure 5.8E). Sequencing of the mix of

ssODNs transfected confirmed that the ssODNs (wild type and mutant (Figure 5.8A), and "N" (Figure 5.8D)) were mixed in equal amounts (around 33% each) and thus did not generate a bias in the detection of HDR events (Figure 5.8F).



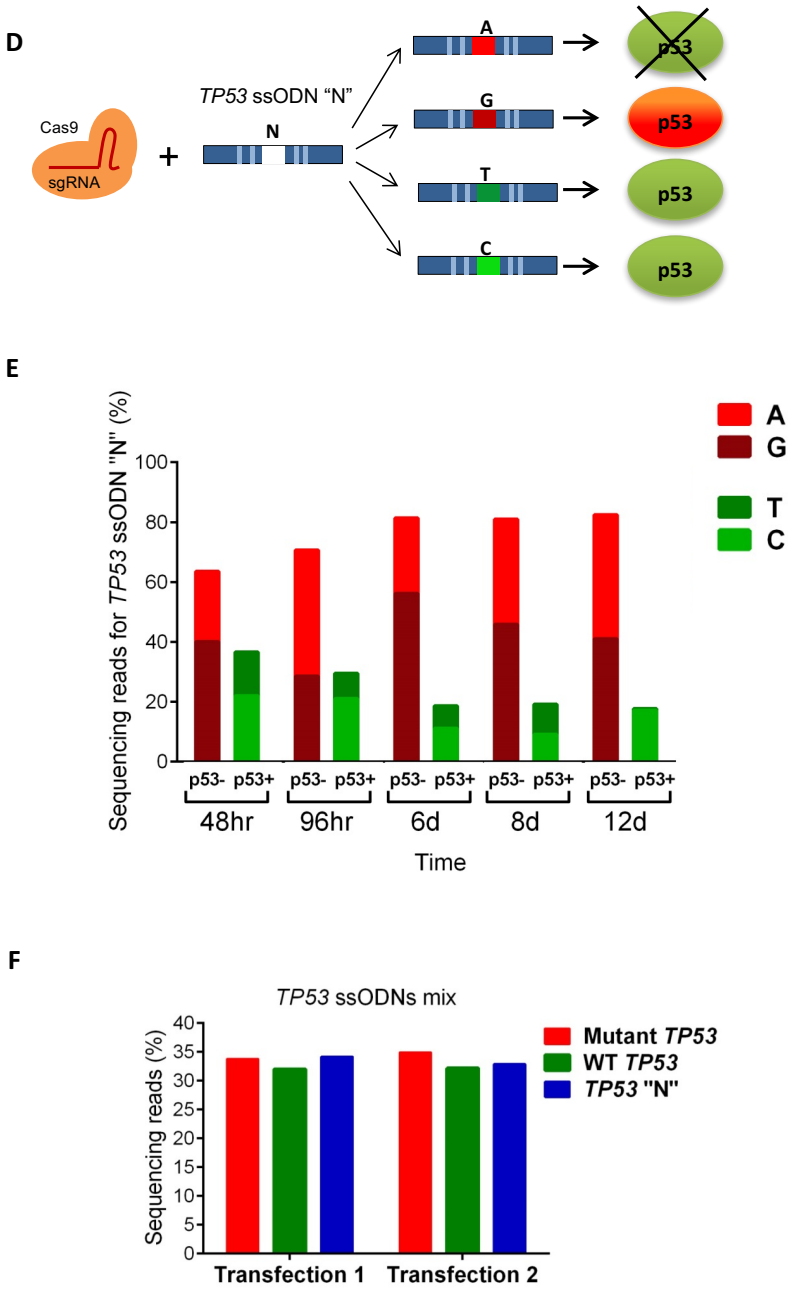


Figure 5.8. Transfection of K562 cells with a mix of two ssODNs for HDR-mediated editing. **A)** K562 cells were transfected with a plasmid expressing Cas9 and sgRNA 3, and a mix of a wild type (WT *TP53* ssODN) and a mutant (Mutant *TP53* ssODN) ssODN both harbouring barcodes of silent mutations for the identification of HDR events. The sequence of the WT ssODN was predicted to restore *TP53* expression while the mutant sequence was predicted to generate a stop codon and no *TP53* expression. The transfection was repeated twice (n=2). **B)** Scatter plots showing the percentage of GFP positive K562 cells 48 hours after transfection. Samples analysed were: K562 cells untransfected (UNTRANSFECTED), cells transfected with the pX458 plasmid expressing Cas9 but no sgRNA (pX458 no gRNA) and cells transfected with pX458 expressing sgRNA3, and a mix of the wild type and mutant ssODNs (TRANSFECTION 1 AND

TRANSFECTION 2). **C)** Bar graph showing the percentage of sequencing reads from transfected K562 cells, containing the sequence of the wild type ssODN (WT *TP53* ssODN, green) and the mutant ssODN (Mutant *TP53* ssODN, red) at different timepoints after transfection. Error bars are Standard Error of the Mean (SEM) between two technical replicates (two transfections were performed). **D)** K562 cells were also transfected with a plasmid expressing Cas9 and sgRNA 3, and a ssODN containing an "N" nucleotide (*TP53* ssODN "N"). Depending on the nt present, integration of the ssODN was predicted to cause a nonsense mutation (N=A) (light red), a missense mutation (N=G) (dark red) or no changes to the protein coding sequence (N=T or C) (dark and light green). **E)** Bar graph showing the percentage of sequencing reads from transfected K562 cells, containing the sequence of the "N" ssODN corresponding to each N nucleotide (A, T, C or G). **F)** Bar graph showing the percentage of sequencing reads in the mix of ssODNs used in each of the two transfections, containing the sequence of the mutant (Mutant *TP53*, red), wild type (WT *TP53*, green) or "N" ssODN (*TP53* "N", blue).

In a study published by our group, a homozygous nonsense substitution in *ASXL1* (Figure 5.9A) was successfully corrected using CRISPR/Cas9 in the human chronic myeloid leukaemia cell line KBM5 (Valletta *et al.*, 2015). Correction of the *ASXL1* mutation in KBM5 cells restored protein expression and function (Valletta *et al.*, 2015). Therefore, I investigated whether HDR-modification of *ASXL1* in KBM5 cells using either a mutant or a wild type ssODN template (Figure 5.9B) resulted in a similar effect to the one observed for *TP53*. GFP positive transfected KBM5 cells were FACS-sorted (Figure 5.9C) and it was found that the percentage of KBM5 cells with integration of wild type *ASXL1* ssODN was not significantly different from the percentage of cells with integration of the mutant ssODN (Figure 5.9D). Integration of the wild type ssODN was slightly higher than integration of the mutant ssODN (Figure 5.9D) and sequencing of the mix of ssODNs transfected confirmed that the ssODNs were mixed in equal amounts (Figure 5.9E).

A

***ASXL1* mutation in KBM5 cells**

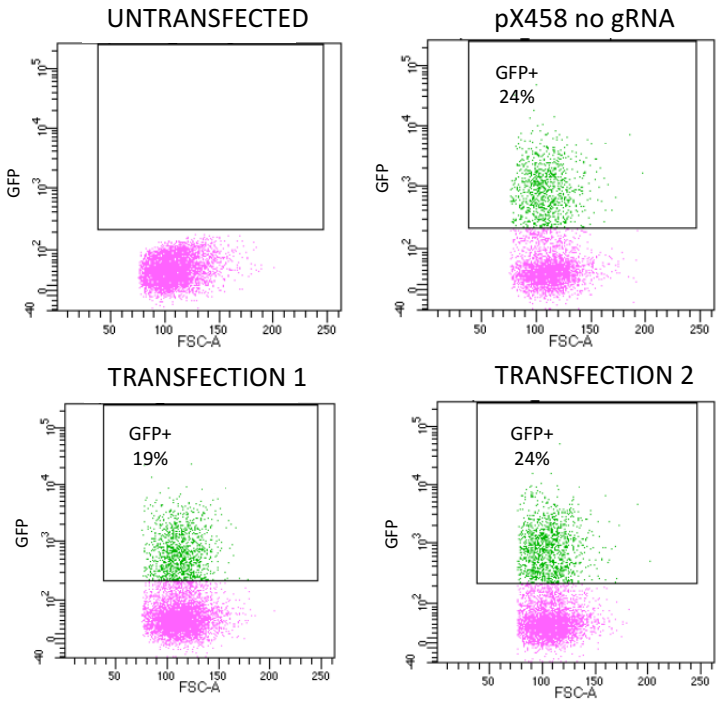
5′-GGGGAGCATACCCAGGCC**T**GAACTGCCATGTCCAGAGCTAG-3′
3′-CCCCTCGTATGGGTCCGGACTTGACGGTACAGGTCTCGATC-5′

c.2128 G>T

B



C



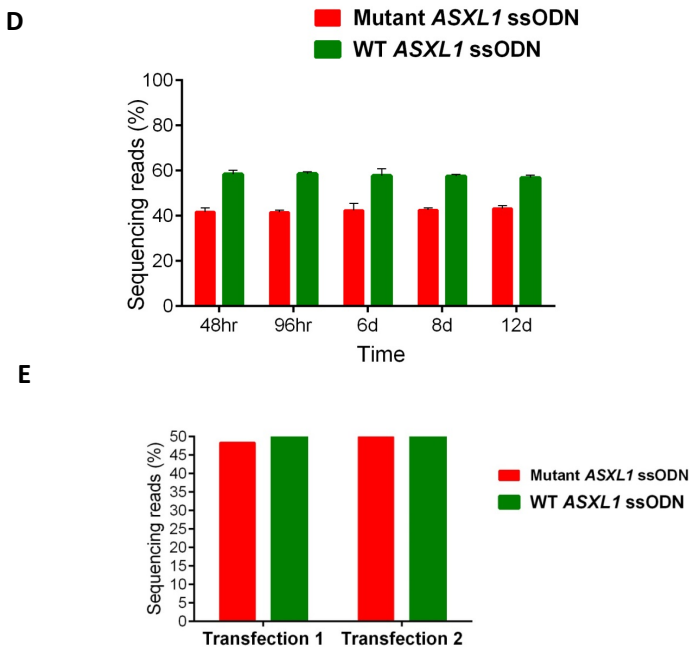


Figure 5.9. Transfection of KBM5 cells with a mix of two ssODNs for HDR-mediated editing of *ASXL1*. A) Mutation in exon 12 of *ASXL1* in KBM5 cells (in red), which leads to a lack of protein expression. **B)** KBM5 cells were transfected with the pX458 plasmid expressing Cas9 and a sgRNA, and a mix of a wild type (WT *ASXL1* ssODN) and a mutant (Mutant *ASXL1* ssODN) ssODN, both harbouring barcodes of silent mutations for the identification of HDR events. The sequence of the WT ssODN was predicted to restore *ASXL1* expression while the mutant sequence was predicted to generate a stop codon and no *ASXL1* expression. The transfection was repeated twice (n=2). **C)** Scatter plots showing the percentage of GFP positive KBM5 cells 48 hours after transfection. Samples analysed were: KBM5 cells untransfected (UNTRANSFECTED), cells transfected with the pX458 plasmid expressing Cas9 but no sgRNA (pX458 no gRNA) and cells transfected with pX458 expressing a sgRNA, and a mix of the wild type and mutant ssODNs (TRANSFECTION 1 AND TRANSFECTION 2). **D)** Bar graph showing the percentage of sequencing reads from transfected KBM5 cells, containing the sequence of the wild type ssODN (WT *ASXL1* ssODN, green) and the mutant ssODN (Mutant *ASXL1* ssODN, red) at different timepoints after transfection. Error bars are Standard Error of the Mean (SEM) between two technical replicates (two transfections were performed). **E)** Bar graph showing the percentage of sequencing reads in the mix of ssODNs used in each of the two transfections, containing the sequence of the mutant (Mutant *ASXL1* ssODN, red) or wild type (WT *ASXL1* ssODN, green) ssODN.

Unlike the percentage of K562 cells with integration of the wild type *TP53* ssODN (Figure 5.10A), the percentage of KBM5 cells with integration of wild type *ASXL1* ssODN did not decrease over time (Figure 5.10B). Lastly, I investigated whether HDR-modified *TP53* and *ASXL1* sequences also harboured additional mutations not contained in the ssODNs used as repair templates. It was found that in both loci the majority of sequences (around 80%) did not harbour additional genetic changes in the edited amplicon (Figure 5.10C and Figure 5.10D). However, since around 20% of HDR-modified sequences analysed did contain additional mutations, I investigated the effects of these changes on p53 or ASXL1 protein function. Analysis using software SIFT and PolyPhen did not identify a significantly higher number of deleterious mutations in HDR-modified sequences of *TP53* (Figure 5.10C) compared to *ASXL1* (Figure 10D). However, it should be noted that the majority of additional mutations found in K562 cells transfected with the wild type ssODN 12 days after transfection were deleterious to p53 function (Figure 5.10C).

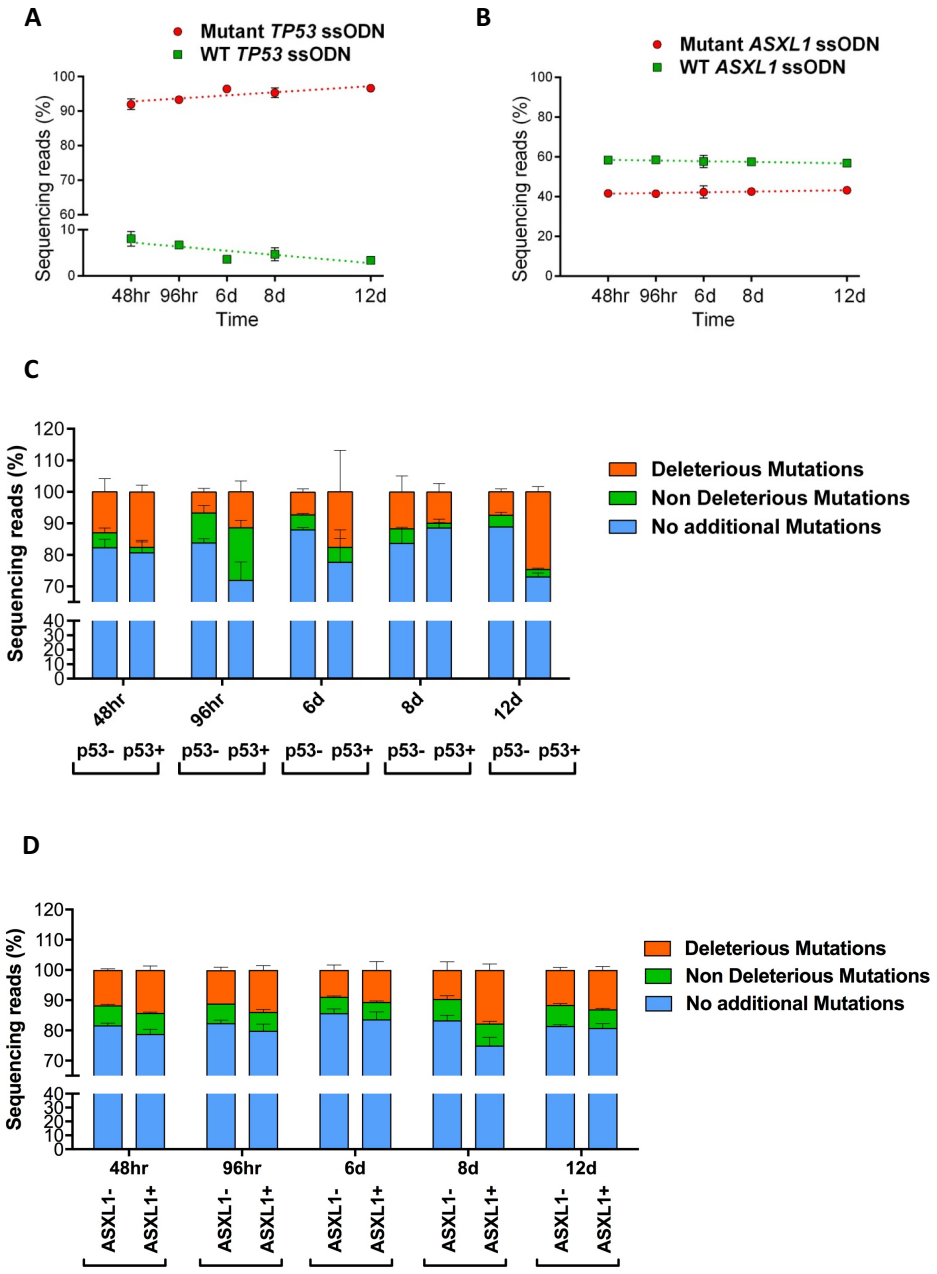


Figure 5.10. Analysis of the percentage of mutation-corrected cells over time. A&B) Dot plot showing the percentage of sequencing reads from transfected K562 (A) and KBM5 (B) cells, containing the sequence of the wild type ssODN (green) (WT *TP53* ssODN or WT *ASXL1* ssODN) and the mutant ssODN (red) (Mutant *TP53* ssODN or Mutant *ASXL1* ssODN) at different timepoints after transfection. Error bars are SEM between two technical replicates (two transfections were performed). Linear regression curves (dashed lines) of the percentage of reads over time are plotted. Linear regression tests in A are statistically significant (both for WT and mutant TP53 ssODNs, p value=0.0145). n=2. Linear regression tests in B were not significant. **C&D)** Analysis of deleterious and non-deleterious additional mutations in HDR reads in *TP53* (C) and *ASXL1*(D) in transfected K562 and KBM5 cells respectively. Stacked bar graph showing the

percentage of sequencing reads of the mutant ssODN (p53- or ASXL1-) and the wild type ssODN (p53+ or ASXL1+) containing additional deleterious mutations (orange), non-deleterious mutations (green) or no additional mutations (blue) at different timepoints after transfection. The data is represented as the percentage of the total reads per ssODN per timepoint. Error bars are SEM between two technical replicates (two transfections were performed).

5.3. Discussion

In this study I attempted to correct a *TP53* mutation in the K562 cell line using the CRISPR/Cas9 system. Correction of *TP53* mutation was achieved in seven colonies expanded from single cells in two different transfection experiments. The occurrence of HDR-mediated modification of the locus was confirmed by the presence of several silent mutations contained in the ssODN template in *TP53* mutation corrected colonies, including Colony 1 (Figure 5.4A). However, additional *TP53* sequence alterations were introduced in all colonies isolated (Table 5.2 and Table 5. 4). These additional changes were not necessarily located in the proximity of the Cas9 cut site and each change was unique to a colony (Table 5.2 and Table 5.4). Sanger sequencing of the *TP53* locus in transfected colonies was limited to an amplicon spanning the Cas9 cut site and therefore the colonies analysed may harbour additional mutations outside this amplicon. The occurrence of additional changes affecting the *TP53* locus in mutation corrected K562 cells could be attributed to several mechanisms. These mutations could arise from NHEJ events happening as a consequence of DSBs caused by Cas9 re-cleavage of the modified *TP53* locus (Kwart *et al.*, 2017). However, it has been shown that the introduction of silent mutations modifying the PAM site on the ssODN is an effective method to avoid Cas9 re-cleavage of the locus, once the ssODN sequence has been integrated into the DNA by HDR (Kwart *et al.*, 2017). Moreover, it was found that mismatches between the target DNA sequence and the seed sequence at the 3' end of the sgRNA abolish target cleavage (Kwart *et al.*, 2017). Changes in the ssODN template used in this study were made for all three sgRNAs sequences, in order to prevent Cas9 re-cleavage of the modified DNA (Figure 5.1B). Therefore the occurrence of subsequence DSBs caused by Cas9 re-cleavage of genomic DNA seems unlikely in this context. When K562 cells were transfected with the CRISPR/Cas9 system and a ssODN containing the wild type sequence of *TP53* (Figure

5.1B), a higher percentage of GFP positive transfected colonies died in culture compared to transfected GFP negative colonies (Table 5.1 and Table 5.3) and only one colony (Colony 1) with correction of *TP53* mutation survived in culture. Moreover, when K562 cells were transfected with a mix of ssODNs (Figure 5.8A), the percentage of cells with integration of the WT *TP53* ssODN was significantly lower than the percentage of cells with integration of the mutant *TP53* ssODN (Figure 5.8C and Figure 5.8E). The percentage of cells with integration of the WT *TP53* ssODN also decreased while the percentage of cells with integration of the mutant *TP53* ssODN increased over time (Figure 5.10A). Taken together, these data support the hypothesis that correction of the *TP53* mutation in K562 cells and resulting restoration of p53 expression leads to apoptosis. This observation is consistent with the role of p53 in regulation of cell cycle and apoptosis (Vousden and Lane, 2007). K562 cells harbour several karyotypic abnormalities, such as the Philadelphia chromosome [t(9;22) (q34.1;q11.2)], and mutations of oncogenes like *RUNX1* (Naumann *et al.*, 2001). The presence of these genetic abnormalities might lead to activation of p53 in *TP53* mutation corrected cells and subsequent induction of cell cycle arrest and apoptosis. CRISPR-mediated editing of *ASXL1* in KBM5 cells revealed that the *ASXL1* mutation does not provide a proliferative advantage in this context. The percentage of cells with integration of wild type *ASXL1* ssODN was not significantly different from the percentage of cells with integration of the mutant ssODN (Figure 5.9D) and the percentage of cells with integration of either ssODNs did not change over time (Figure 5.10B). These results are consistent with the fact that *ASXL1* and *TP53* mutations contribute differently to development and progression of CML. *ASXL1* mutations are found in patients with advanced CML in blast crisis (BC) (Menezes *et al.*, 2013, Togasaki *et al.*, 2017, Boultwood *et al.*, 2010c) but studies using *Asxl1* knockout mice showed that mice with loss of *ASXL1* presented with a mild phenotype with differentiation defects in the myeloid and lymphoid

progenitors and did not progress to a more severe MDS or AML-like disease state (Fisher *et al.*, 2010). Moreover, a recent study using a knock-in mouse model of a *ASXL1* mutation found that *Asxl1* mutant mice do not develop blood disorders, demonstrating that *Asxl1* mutation alone is likely to be insufficient for leukaemogenesis (Hsu *et al.*, 2017). Furthermore, unlike *TP53* alterations, mutations of *ASXL1* are also found in healthy elderly individuals which exhibit clonal haematopoiesis but show normal blood counts, indicating that *ASXL1* mutations are unlikely to be oncogenic causative events (Zink *et al.*, 2017, Genovese *et al.*, 2014). The current model of pathogenesis of MDS, involves the acquisition of initiating mutations, such as *ASXL1* and *TET2* mutations, which lead to asymptomatic clonal expansion of HSCs followed by occurrence of further mutations in the expanded clone, which cause development of the MDS or secondary AML phenotype (Kennedy and Ebert, 2017). *TP53* mutations instead were shown to contribute to rapid progression to CML BC phase in study using a murine model expressing BCR-ABL and with heterozygous knockout of p53 (Honda *et al.*, 2000). In MDS *TP53* mutations are associated with poor prognosis, thrombocytopenia and a complex karyotype (Kennedy and Ebert, 2017).

Analysis of K562 cells transfected with ssODNs containing a barcode of silent mutations for detection of HDR events did not show any difference in apoptosis and proliferation of cells transfected with the wild type ssODN compared to cells transfected with a mutant ssODN (Figure 5.7A, B and C). Given the low rate of HDR (1%) at the *TP53* locus in K562 cells, the effect of *TP53* mutation correction on cell proliferation and apoptosis would be difficult to detect by the analysis of the bulk GFP positive transfected cells.

NGS analysis of additional mutations in the *TP53* edited locus did not identify a consistent increase in the number of deleterious mutations in cells with integration of the WT *TP53* ssODN compared to cells with integration of the mutant *TP53* ssODN at all time points

(Figure 5.10C). However, 12 days after transfection, cells with integration of the WT *TP53* ssODN harboured a higher number of deleterious mutations than cells with integration of the mutant *TP53* ssODN and the majority of additional mutation in WT *TP53*-edited cells were deleterious to the protein function (Figure 5.10C). These results suggest that acquired deleterious compensatory mutations in *TP53* corrected K562 cells might provide a proliferative advantage and result in survival of cells with integration of the WT *TP53* ssODN. A recent study in hiPSCs showed that the cellular toxicity caused by DSBs introduced by Cas9 is dependent on p53 activation (Ihry *et al.*, 2017). *TP53* knock-out (KO) in hiPSCs expressing Cas9 increased the survival of cells transduced with lentiviral vectors expressing sgRNAs and also increased the recovery of CRISPR-edited cells (Ihry *et al.*, 2017). While this evidence suggests that transient knock-down of *TP53* may increase the efficiency of genome editing in hiPSCs, it also implies that *TP53* mutational status should be carefully monitored in CRISPR-based editing of hiPSCs for therapeutic purposes (Ihry *et al.*, 2017). The authors suggest that spontaneously arising *TP53* mutations may be selected for in edited hiPSCs due to the proliferative advantage of *TP53* mutant cells and their increased resistance to Cas9 toxicity (Ihry *et al.*, 2017).

As previously mentioned, the analysis of integration of the ssODN templates in the *TP53* and *ASXL1* locus was limited to NGS of the amplicon spanning the Cas9 cut site and therefore additional deleterious and non-deleterious mutations may have been present in these genes outside of the sequenced amplicon. The analysis of additional mutations using inserted primer binding sites for the amplification of HDR-edited DNA in CRISPR-transfected cells (Figure 5.5A and 5.5B and Figure 5.6A and 5.6B) did not identify any correlation between the occurrence of additional changes and the integration of WT or mutant *TP53* ssODN (Figure 5.5C, Figure 5.5D and Figure 5.6C). This analysis was limited by the use of Sanger sequencing technology and by the fact that PCR amplification was

only possible on either side of the inserted primer binding site (Figure 5.6A and Figure 5.6B) and therefore sequencing of the entire HDR-edited amplicon was not possible.

When performing genome editing experiments aimed at generating disease models or therapeutic applications, the biological function of the gene to be edited should be taken into consideration. The effect of gene editing on cellular phenotypes and cell survival can be analysed to functionally characterise genes and mutations in pooled CRISPR-based assays. A pooled loss-of-function screen has been employed to identify essential genes in Burkitt's lymphoma and CML cell lines, including K562 cells (Wang *et al.*, 2015). Our data shows that restoration of gene expression in the context of a cell line harbouring a homozygous nonsense mutation, such as *TP53* in K562 cells, can also be used to investigate the role of mutations as oncogenic drivers, by analysing the effect of mutation correction on cell survival. The results of our NGS analysis of HDR-modification of *TP53* and *ASXL1* reflect the role of *TP53* as driver oncogene (Honda *et al.*, 2000) while *ASXL1* mutations are non-driver initiating events in CML (Fisher *et al.*, 2010). A similar method based on CRISPR-mediated insertion of barcodes was also used in a study by Guernet *et al.* (2016) to investigate the effect of insertion of a nonsense *TP53* mutation in MCF7 and HCT-116 cell lines. Cells were transfected with ssODNs containing either WT or mutant *TP53* and it was found that cells lacking *TP53* activity showed a selective advantage, as measured by the increased percentage of cells with integration of the mutant *TP53* ssODN upon treatment with the p53 activator Nutlin-3 (Guernet *et al.*, 2016). Poorly characterised mutations in cancer, such as mutations in intergenic regulatory elements, could be studied using this CRISPR-barcoding method in order to define whether they play an oncogenic driver function in cancer.

6. Chapter Six – General Discussion

This D.Phil. project aimed at investigating various aspects of the pathophysiology of MDS with del(5q) and possible therapeutic approaches for this MDS subtype. The 5q deletion is found in approximately 20% of MDS patients (Bernasconi *et al.*, 2005). Patients with de(5q) MDS and 5q- syndrome are affected by refractory hypoplastic anaemia and approximately 10% of 5q- syndrome patients transform to AML (Boultwood *et al.*, 2010a). There is a clinical need for novel treatments for MDS with the del(5q) and investigation into the pathophysiology of this disease could help identifying potential targets for therapeutic intervention.

CSNK1A1 maps to the commonly deleted region (CDR) of the 5q- syndrome and it encodes CK1α, which is involved in regulation of the Wnt/β-catenin pathway. A large cohort of patients with del(5q) MDS was screened for the occurrence of mutation in *CSNK1A1* and it was confirmed that *CSNK1A1* mutations are rare events in 5q- syndrome. It was also found that the Wnt/β-catenin pathway was significantly up-regulated in del(5q) MDS patients with mutation of *CSNK1A1*. However, the analysis was limited to a small number of patients and further investigation should include a larger patient cohort. The most commonly used therapeutic agent for treatment of 5q- syndrome is the thalidomide analogue lenalidomide (List *et al.*, 2006b). Gain of function mutations of *CSNK1A1* might play a role in the mechanism of resistance to lenalidomide observed in del(5q) MDS patients with *CSNK1A1* mutations (Smith *et al.*, 2015). Therefore, further investigation into the cellular phenotypes associated with *CSNK1A1* mutations in 5q- syndrome would help elucidating the mechanisms of lenalidomide resistance in 5q- syndrome. Our data support the notion *CSNK1A1* plays a central role in the pathogenesis of del(5q) MDS.

Evidence from several sources suggests that treatment with the amino acid L-leucine increases erythroid differentiation, cell proliferation and mRNA translation efficiency in

animal and cellular models of the 5q- syndrome and Diamond–Blackfan anaemia (DBA) (Yip *et al.*, 2012, Payne *et al.*, 2012, Narla *et al.*, 2014, Zhang *et al.*, 2014). In this study, I gained useful insights into the mechanism of action of L-leucine in cultured erythroblasts derived from an expanded cohort of del(5q) MDS patients. Treatment with L-leucine was shown to activate the mTOR pathway and increase translation of mTOR-regulated 5'TOP and 5'TOP-like mRNAs. Promotion of ribosome biogenesis, resulting from activation of mTORC1 and increased translation of 5'TOP mRNAs, may be the mechanism behind the increase in mRNA translation in patient erythroblasts treated with L-leucine. Increased mRNA translation may lead to promotion of proliferation and differentiation of L-leucine-treated patient erythroblasts and result in amelioration of the anaemia. However, the role of mTORC1 activity in the L-leucine-mediated increase of mRNA translation will require further investigation using inhibitors of mTORC1. These data support the rationale for use of L-leucine in treatment of patients with DBA, who exhibit defective mRNA translation. Clinical trials for the therapeutic use of L-leucine in DBA are underway in the US and in Russia (NCT02386267 and NCT01362595, www.clinicaltrials.gov) (Pellagatti and Boultwood, 2015b). In a recent study three MDS patients with isolated del(5q) were treated with L-leucine and showed no adverse effects during treatment (Steensma and Ebert, 2013). However, none of the patients showed improvements in peripheral blood cytopenia or transfusion requirements (Steensma and Ebert, 2013). Absorption tests should be utilised to determine the optimal dose for therapy, analysis of basal translation levels could help identify patients more likely to respond to therapy (Pospisilova *et al.*, 2007) and further studies should involve a higher number of patients. Moreover, in the study by Steensma and Ebert (2013) the effects of L-leucine treatment on blood counts were assessed after two to three months of therapy while clinical trials for the use of L-leucine for DBA treatment involve administration of the drug for

nine months before evaluation of the effects of treatment (www.clinicaltrials.gov). Moreover, L-leucine could be used for the treatment of other non-del(5q) MDS subtypes which show downregulated expression of *RPS14* (Czibere *et al.*, 2009) and may be characterised by a decrease in mRNA translation.

The present study also uncovered new aspects of the pathogenesis of del(5q) MDS. Erythroblasts derived from del(5q) MDS patients showed a decrease in translation of mTOR-regulated 5'TOP and 5'TOP-like mRNAs, in line with decreased activation of mTORC1. Moreover, the expression of the RNA-binding protein LARP1, involved in regulation of 5'TOP mRNAs translation was also decreased. These data are consistent with pathogenesis del(5q) MDS being characterised by decreased mTOR signalling and a defect in ribosome biogenesis.

18% of patients with MDS with del(5q) harbour *TP53* mutations (Jädersten *et al.*, 2011), which have been correlated with poor prognosis, resistance to lenalidomide treatment (Jädersten *et al.*, 2009) and increased risk of AML development (Sebaa *et al.*, 2012, Boultwood *et al.*, 2010a). The use of genome editing techiques such as the CRISPR/Cas9 system to correct or insert recurrent mutations in MDS is a potential approach for generating cellular models of the disease (Kotini *et al.*, 2015, Kotini *et al.*, 2017). Allogeneic haematopoietic stem cell transplantation (HSCT) is currently the only curative treatment for MDS and innovations in stem cell therapy are needed to improve patients outcome (Bejar and Steensma, 2014). Despite the current technical challenges, the correction of gene mutations in HSCs from MDS patients could represent a strategy for therapeutic intervention in the future (Valletta *et al.*, 2015). I used CRISPR/Cas9 gene editing technology to correct a *TP53* mutation by HDR in the CML cell line K562. K562 cells were used to conduct proof-of-principle experiments because they are easier to handle

and transfect than $CD34^+$ or hiPSCs. I achieved correction of the *TP53* mutation in K562 cells by HDR but I found that corrected cells also harboured additional genetic changes near the Cas9 cut site. NGS-based investigation of the effects of HDR-mediated modification of the *TP53* locus showed that correction of the *TP53* mutation in K562 cells leads to apoptosis of the corrected cells. These data indicate that, in genome editing experiments aimed at disease modelling or for therapeutic applications, it is important to consider the biological function of edited genes. Moreover, CRISPR/Cas9-mediated genome editing can be exploited as a tool to identify oncogenic driver genes in cancer (Wang *et al.*, 2015). The effects of HDR-mediated modification of *TP53* and *ASXL1* in K562 and KBM5 cells respectively were investigated using an NGS-based CRISPR-barcoding system. Mutations affecting intergenic regulatory elements, such as mutations in the promoter region of *hTERT* in melanoma (Horn *et al.*, 2013), have a causative role in development of cancer (Weinhold *et al.*, 2014). A similar CRISPR-barcoding method to the one I developed in this project could be used to investigate the effects of correction of these intergenic mutations, in order to elucidate their role as oncogenic drivers in cancer.

In conclusion, this study provided new insights into the pathogenesis and therapy of 5q-syndrome and MDS with del(5q) (Figure 6.1). *CSNK1A1* mutations and gene expression profiling (GEP) in del(5q) MDS patients were analysed and it was found that *CSNK1A1* is likely to play an important role in the pathogenesis of MDS with del(5q). A more complete understanding of the pathophysiology of del(5q) MDS will aid the development of new therapeutic approaches and can guide therapeutic strategies. CK1α has been implicated in the mechanism of action of lenalidomide in 5q- syndrome and further work is needed in order to establish whether *CSNK1A1* mutations affect lenalidomide resistance. The mechanism of action of L-leucine in del(5q) MDS was also investigated and it was found that L-leucine increases mRNA translation and differentiation of erythroblasts from del(5q)

MDS patients. This drug is currently in clinical trial and it could be used as a potential therapy to alleviate the anaemia in MDS patients with del(5q). Lastly, I have shown that the genome editing system CRISPR/Cas9 can be utilised to modify frenquently mutated genes in cancer cell lines. Using a similar approach, CRISPR/Cas9 could be also utilised to create models of del(5q) MDS to study the pathogenesis of this disease. The data presented in this thesis contribute to clinically-relevant research into MDS with del(5q).

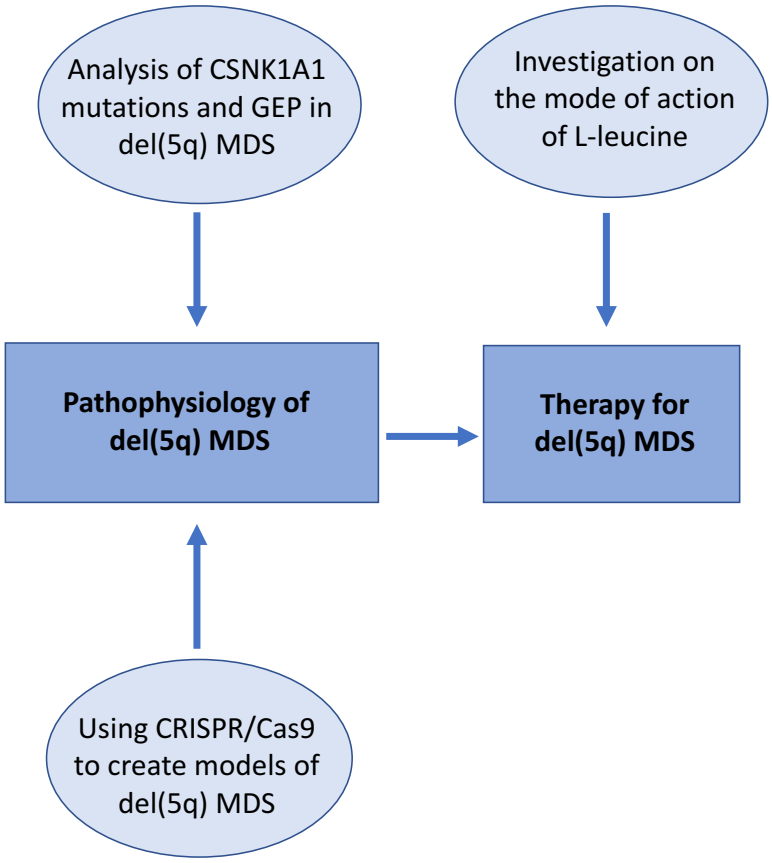


Figure 6.1. Contribution of the work presented in this thesis to del(5q) MDS research.

References

ABDEL-WAHAB, O., ADLI, M., LAFAVE, L. M., GAO, J., HRICIK, T., SHIH, A. H., PANDEY, S., PATEL, J. P., CHUNG, Y. R., KOCHE, R., PERNA, F., ZHAO, X., TAYLOR, J. E., PARK, C. Y., CARROLL, M., MELNICK, A., NIMER, S. D., JAFFE, J. D., AIFANTIS, I., BERNSTEIN, B. E. & LEVINE, R. L. 2012. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell,* 22**,** 180-93.

ABDEL-WAHAB, O., GAO, J., ADLI, M., DEY, A., TRIMARCHI, T., CHUNG, Y. R., KUSCU, C., HRICIK, T., NDIAYE-LOBRY, D., LAFAVE, L. M., KOCHE, R., SHIH, A. H., GURYANOVA, O. A., KIM, E., LI, S., PANDEY, S., SHIN, J. Y., TELIS, L., LIU, J., BHATT, P. K., MONETTE, S., ZHAO, X., MASON, C. E., PARK, C. Y., BERNSTEIN, B. E., AIFANTIS, I. & LEVINE, R. L. 2013. Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *J Exp Med,* 210**,** 2641-59.

ABOU ZAHR, A., KAVI, A. M., MUKHERJEE, S. & ZEIDAN, A. M. 2017. Therapy-related myelodysplastic syndromes, or are they? *Blood Rev,* 31**,** 119-128.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

ALAO, J. P. 2007. The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. *Molecular Cancer,* 6**,** 24.

ARBER, D. A., ORAZI, A., HASSERJIAN, R., THIELE, J., BOROWITZ, M. J., LE BEAU, M. M., BLOOMFIELD, C. D., CAZZOLA, M. & VARDIMAN, J. W. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,* 127**,** 2391-405.

BARLOW, J. L., DRYNAN, L. F., HEWETT, D. R., HOLMES, L. R., LORENZO-ABALDE, S., LANE, A. L., JOLIN, H. E., PANNELL, R., MIDDLETON, A. J., WONG, S. H., WARREN, A. J., WAINSCOAT, J. S., BOULTWOOD, J. & MCKENZIE, A. N. J. 2010. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nat Med,* 16**,** 59-66.

BEJAR, R. 2015. Myelodysplastic Syndromes Diagnosis: What Is the Role of Molecular Testing? *Curr Hematol Malig Rep,* 10**,** 282-91.

BEJAR, R. & STEENSMA, D. P. 2014. Recent developments in myelodysplastic syndromes. *Blood,* 124**,** 2793-803.

BELLO, E., PELLAGATTI, A., SHAW, J., MECUCCI, C., KUSEC, R., KILLICK, S., GIAGOUNIDIS, A., RAYNAUD, S., CALASANZ, M. J., FENAUX, P. & BOULTWOOD, J. 2015. CSNK1A1 mutations and gene expression analysis in myelodysplastic syndromes with del(5q). *Br J Haematol*.

BEN-SAHRA, I. & MANNING, B. D. 2017. mTORC1 signaling and the metabolic control of cell growth. *Curr Opin Cell Biol,* 45**,** 72-82.

BERNASCONI, P., KLERSY, C., BONI, M., CAVIGLIANO, P. M., CALATRONI, S., GIARDINI, I., ROCCA, B., ZAPPATORE, R., CARESANA, M., QUARNA, J., LAZZARINO, M. & BERNASCONI, C. 2005. Incidence and prognostic significance of karyotype abnormalities in de novo primary myelodysplastic syndromes: a study on 331 patients from a single institution. *Leukemia,* 19**,** 1424-31.

BHAGAT, T. D., CHEN, S., BARTENSTEIN, M., BARLOWE, A. T., VON AHRENS, D., CHOUDHARY, G. S., TIVNAN, P., AMIN, E., MARCONDES, A. M., SANDERS, M. A., HOOGENBOEZEM, R. M., KAMBHAMPATI, S., RAMACHANDRA, N., MANTZARIS, I., SUKRITHAN, V., LAURENCE, R., LOPEZ, R., BHAGAT, P., GIRICZ, O., SOHAL, D., WICKREMA, A., YEUNG, C., GRITSMAN, K., APLAN, P., HOCHEDLINGER, K., YU, Y., PRADHAN, K., ZHANG, J., GREALLY, J. M., MUKHERJEE, S., PELLAGATTI, A., BOULTWOOD, J., WILL, B., STEIDL, U., RAAIJMAKERS, M., DEEG, H. J., KHARAS, M. G. & VERMA, A. 2017. Epigenetically Aberrant Stroma in MDS Propagates Disease via Wnt/beta-Catenin Activation. *Cancer Res*.

BIEGING, K. T., MELLO, S. S. & ATTARDI, L. D. 2014. Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer,* 14**,** 359-70.

BOULTWOOD, J., DOLATSHAD, H., VARANASI, S. S., YIP, B. H. & PELLAGATTI, A. 2014. The role of splicing factor mutations in the pathogenesis of the myelodysplastic syndromes. *Adv Biol Regul,* 54**,** 153-61.

BOULTWOOD, J., FIDLER, C., LEWIS, S., KELLY, S., SHERIDAN, H., LITTLEWOOD, T. J., BUCKLE, V. J. & WAINSCOAT, J. S. 1994. Molecular mapping of uncharacteristically small 5q deletions in two patients with the 5q- syndrome: delineation of the critical region on 5q and identification of a 5q-breakpoint. *Genomics,* 19**,** 425-32.

BOULTWOOD, J., FIDLER, C., STRICKSON, A. J., WATKINS, F., GAMA, S., KEARNEY, L., TOSI, S., KASPRZYK, A., CHENG, J. F., JAJU, R. J. & WAINSCOAT, J. S. 2002. Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood,* 99**,** 4638-41.

BOULTWOOD, J., PELLAGATTI, A., CATTAN, H., LAWRIE, C. H., GIAGOUNIDIS, A., MALCOVATI, L., DELLA PORTA, M. G., JADERSTEN, M., KILLICK, S., FIDLER, C., CAZZOLA, M., HELLSTROM-LINDBERG, E. & WAINSCOAT, J. S. 2007. Gene expression profiling of CD34+ cells in patients with the 5q-syndrome. *British Journal of Haematology,* 139**,** 578-89.

BOULTWOOD, J., PELLAGATTI, A., MCKENZIE, A. N. & WAINSCOAT, J. S. 2010a. Advances in the 5q- syndrome. *Blood,* 116**,** 5803-11.

BOULTWOOD, J., PERRY, J., PELLAGATTI, A., FERNANDEZ-MERCADO, M., FERNANDEZ-SANTAMARIA, C., CALASANZ, M. J., LARRAYOZ, M. J., GARCIA-DELGADO, M., GIAGOUNIDIS, A., MALCOVATI, L., DELLA PORTA, M. G., JADERSTEN, M., KILLICK, S., HELLSTROM-LINDBERG, E., CAZZOLA, M. & WAINSCOAT, J. S. 2010b. Frequent mutation of the polycomb-

associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia,* 24**,** 1062-5.

BOULTWOOD, J., PERRY, J., ZAMAN, R., FERNANDEZ-SANTAMARIA, C., LITTLEWOOD, T., KUSEC, R., PELLAGATTI, A., WANG, L., CLARK, R. E. & WAINSCOAT, J. S. 2010c. High-density single nucleotide polymorphism array analysis and ASXL1 gene mutation screening in chronic myeloid leukemia during disease progression. *Leukemia,* 24**,** 1139-45.

BOULTWOOD, J. & WAINSCOAT, J. S. 2007. Gene silencing by DNA methylation in haematological malignancies. *Br J Haematol,* 138**,** 3-11.

BOULTWOOD, J., YIP, B. H., VUPPUSETTY, C., PELLAGATTI, A. & WAINSCOAT, J. S. 2013. Activation of the mTOR pathway by the amino acid (L)-leucine in the 5q- syndrome and other ribosomopathies. *Adv Biol Regul,* 53**,** 8-17.

BRAY, N. L., PIMENTEL, H., MELSTED, P. & PACHTER, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol,* 34**,** 525-7.

CACERES, G., MCGRAW, K., YIP, B. H., PELLAGATTI, A., JOHNSON, J., ZHANG, L., LIU, K., ZHANG, L. M., FULP, W. J., LEE, J.-H., AL ALI, N. H., BASIORKA, A., SMITH, L. J., DAUGHERTY, F. J., LITTLETON, N., WELLS, R. A., SOKOL, L., WEI, S., KOMROKJI, R. S., BOULTWOOD, J. & LIST, A. F. 2013. TP53 suppression promotes erythropoiesis in del(5q) MDS, suggesting a targeted therapeutic strategy in lenalidomide-resistant patients. *Proceedings of the National Academy of Sciences,* 110**,** 16127-16132.

CAZZOLA, M. & INVERNIZZI, R. 2011. Ring sideroblasts and sideroblastic anemias. *Haematologica,* 96**,** 789-92.

CHASSE, H., BOULBEN, S., COSTACHE, V., CORMIER, P. & MORALES, J. 2017. Analysis of translation using polysome profiling. *Nucleic Acids Res,* 45**,** e15.

CHAUVIN, C., KOKA, V., NOUSCHI, A., MIEULET, V., HOAREAU-AVEILLA, C., DREAZEN, A., CAGNARD, N., CARPENTIER, W., KISS, T., MEYUHAS, O. & PENDE, M. 2014. Ribosomal protein S6 kinase activity controls the ribosome biogenesis transcriptional program. *Oncogene,* 33**,** 474-483.

CHEN, D., ZHANG, Z., LI, M., WANG, W., LI, Y., RAYBURN, E. R., HILL, D. L., WANG, H. & ZHANG, R. 2007. Ribosomal protein S7 as a novel modulator of p53-MDM2 interaction: binding to MDM2, stabilization of p53 protein, and activation of p53 function. *Oncogene,* 26**,** 5029-37.

CHEONG, J. K. & VIRSHUP, D. M. 2011. Casein kinase 1: Complexity in the family. *Int J Biochem Cell Biol,* 43**,** 465-9.

CHUNG, J., BAUER, D. E., GHAMARI, A., NIZZI, C. P., DECK, K. M., KINGSLEY, P. D., YIEN, Y. Y., HUSTON, N. C., CHEN, C., SCHULTZ, I. J., DALTON, A. J., WITTIG, J. G., PALIS, J., ORKIN, S. H., LODISH, H. F., EISENSTEIN, R. S., CANTOR, A. B. & PAW, B. H. 2015. The mTORC1/4E-BP pathway coordinates hemoglobin production with L-leucine availability. *Sci Signal,* 8**,** ra34.

CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZESNIAK, M. W., GAFFNEY, D. J., ELO, L. L., ZHANG, X. & MORTAZAVI, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol,* 17**,** 13.

CZIBERE, A., BRUNS, I., JUNGE, B., SINGH, R., KOBBE, G., HAAS, R. & GERMING, U. 2009. Low RPS14 expression is common in myelodysplastic syndromes without 5q- aberration and defines a subgroup of patients with prolonged survival. *Haematologica,* 94**,** 1453-5.

DAI, M. S. & LU, H. 2004. Inhibition of MDM2-mediated p53 ubiquitination and degradation by ribosomal protein L5. *J Biol Chem,* 279**,** 44475-82.

DAI, M. S., ZENG, S. X., JIN, Y., SUN, X. X., DAVID, L. & LU, H. 2004. Ribosomal protein L23 activates p53 by inhibiting MDM2 function in response to ribosomal perturbation but not to translation inhibition. *Mol Cell Biol,* 24**,** 7654-68.

DAMGAARD, C. K. & LYKKE-ANDERSEN, J. 2011. Translational coregulation of 5′TOP mRNAs by TIA-1 and TIAR. *Genes & Development,* 25**,** 2057-2068.

DAVIES, C., YIP, B. H., FERNANDEZ-MERCADO, M., WOLL, P. S., AGIRRE, X., PROSPER, F., JACOBSEN, S. E., WAINSCOAT, J. S., PELLAGATTI, A. & BOULTWOOD, J. 2013. Silencing of ASXL1 impairs the granulomonocytic lineage potential of human CD34(+) progenitor cells. *Br J Haematol,* 160**,** 842-50.

DE LA CRUZ, J., KARBSTEIN, K. & WOOLFORD, J. L., JR. 2015. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu Rev Biochem,* 84**,** 93-129.

DELUCA, D. S., LEVIN, J. Z., SIVACHENKO, A., FENNELL, T., NAZAIRE, M. D., WILLIAMS, C., REICH, M., WINCKLER, W. & GETZ, G. 2012. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics,* 28**,** 1530-2.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DODD, K. M. & TEE, A. R. 2012. Leucine and mTORC1: a complex relationship. *Am J Physiol Endocrinol Metab,* 302**,** E1329-42.

DRAGNEV, K. H., FENG, Q., MA, Y., SHAH, S. J., BLACK, C., MEMOLI, V., NUGENT, W., RIGAS, J. R., KITAREEWAN, S., FREEMANTLE, S. & DMITROVSKY, E. 2007. Uncovering novel targets for cancer chemoprevention. *Recent Results Cancer Res,* 174.

DULAK, A. M., STOJANOV, P., PENG, S., LAWRENCE, M. S., FOX, C., STEWART, C., BANDLA, S., IMAMURA, Y., SCHUMACHER, S. E., SHEFLER, E., MCKENNA, A., CARTER, S. L., CIBULSKIS, K., SIVACHENKO, A., SAKSENA, G., VOET, D., RAMOS, A. H., AUCLAIR, D., THOMPSON, K., SOUGNEZ, C., ONOFRIO, R. C., GUIDUCCI, C., BEROUKHIM, R., ZHOU, Z., LIN, L., LIN, J., REDDY, R., CHANG, A., LANDRENAU, R., PENNATHUR,

A., OGINO, S., LUKETICH, J. D., GOLUB, T. R., GABRIEL, S. B., LANDER, E. S., BEER, D. G., GODFREY, T. E., GETZ, G. & BASS, A. J. 2013. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genetics,* 45**,** 478-86.

DUTT, S., NARLA, A., LIN, K., MULLALLY, A., ABAYASEKARA, N., MEGERDICHIAN, C., WILSON, F. H., CURRIE, T., KHANNA-GUPTA, A., BERLINER, N., KUTOK, J. L. & EBERT, B. L. 2011. *Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells*.

EAR, J., HSUEH, J., NGUYEN, M., ZHANG, Q., SUNG, V., CHOPRA, R., SAKAMOTO, K. M. & LIN, S. 2016. A Zebrafish Model of 5q-Syndrome Using CRISPR/Cas9 Targeting RPS14 Reveals a p53-Independent and p53-Dependent Mechanism of Erythroid Failure. *J Genet Genomics,* 43**,** 307-18.

EBERT, B. L., PRETZ, J., BOSCO, J., CHANG, C. Y., TAMAYO, P., GALILI, N., RAZA, A., ROOT, D. E., ATTAR, E., ELLIS, S. R. & GOLUB, T. R. 2008. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature,* 451**,** 335-339.

ELYADA, E., PRIBLUDA, A., GOLDSTEIN, R. E., MORGENSTERN, Y., BRACHYA, G., COJOCARU, G., SNIR-ALKALAY, I., BURSTAIN, I., HAFFNER-KRAUSZ, R., JUNG, S., WIENER, Z., ALITALO, K., OREN, M., PIKARSKY, E. & BEN-NERIAH, Y. 2011. CKIα ablation highlights a critical role for p53 in invasiveness control. *Nature,* 470.

FAJTOVA, M., KOVARIKOVA, A., SVEC, P., KANKURI, E. & SEDLAK, J. 2013. Immunophenotypic profile of nucleated erythroid progenitors during maturation in regenerating bone marrow. *Leuk Lymphoma,* 54**,** 2523-30.

FARIÑA SARASQUETA, A., FORTE, G. I., CORVER, W. E., DE MIRANDA, N. F., RUANO, D., VAN EIJK, R., OOSTING, J., TOLLENAAR, R. A., VAN WEZEL, T. & MORREAU, H. 2013. Integral analysis of p53 and its value as prognostic factor in sporadic colon cancer. *BMC Cancer,* 13**,** 277.

FENG, Z. 2010. p53 regulation of the IGF-1/AKT/mTOR pathways and the endosomal compartment. *Cold Spring Harb Perspect Biol,* 2**,** a001057.

FENG, Z. & LEVINE, A. J. 2010. The regulation of energy metabolism and the IGF-1/mTOR pathways by the p53 protein. *Trends Cell Biol,* 20**,** 427-34.

FENG, Z., ZHANG, H., LEVINE, A. J. & JIN, S. 2005. The coordinate regulation of the p53 and mTOR pathways in cells. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 8204-8209.

FERNANDEZ-MERCADO, M., BURNS, A., PELLAGATTI, A., GIAGOUNIDIS, A., GERMING, U., AGIRRE, X., PROSPER, F., AUL, C., KILLICK, S., WAINSCOAT, J. S., SCHUH, A. & BOULTWOOD, J. 2013. Targeted re-sequencing analysis of 25 genes commonly mutated in myeloid disorders in del(5q) myelodysplastic syndromes. *Haematologica,* 98**,** 1856-64.

FINK, E. C., KRÖNKE, J., HURST, S. N., UDESHI, N. D., SVINKINA, T., SCHNEIDER, R. K., MCCONKEY, M. E., JÄRÅS, M., BULLINGER, L., CARR, S. A. & EBERT, B. L. 2014. Lenalidomide Induces Ubiquitination and Degradation of CSNK1A1 in MDS with Del(5q). *Blood,* 124**,** 4.

FISHER, C. L., PINEAULT, N., BROOKES, C., HELGASON, C. D., OHTA, H., BODNER, C., HESS, J. L., HUMPHRIES, R. K. & BROCK, H. W. 2010. Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. *Blood,* 115**,** 38-46.

FLANAGAN, S. E., PATCH, A. M. & ELLARD, S. 2010. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers,* 14**,** 533-7.

FONSECA, B. D., SMITH, E. M., YELLE, N., ALAIN, T., BUSHELL, M. & PAUSE, A. 2014. The ever-evolving role of mTOR in translation. *Seminars in Cell & Developmental Biology,* 36**,** 102-112.

FORBES, S. A., BEARE, D., BOUTSELAKIS, H., BAMFORD, S., BINDAL, N., TATE, J., COLE, C. G., WARD, S., DAWSON, E., PONTING, L., STEFANCSIK, R., HARSHA, B., KOK, C. Y., JIA, M., JUBB, H., SONDKA, Z., THOMPSON, S., DE, T. & CAMPBELL, P. J. 2017. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research,* 45**,** D777-D783.

FREED-PASTOR, W. A. & PRIVES, C. 2012. Mutant p53: one name, many proteins. *Genes & Development,* 26**,** 1268-1286.

GANDIN, V., MASVIDAL, L., HULEA, L., GRAVEL, S. P., CARGNELLO, M., MCLAUGHLAN, S., CAI, Y., BALANATHAN, P., MORITA, M., RAJAKUMAR, A., FURIC, L., POLLAK, M., PORCO, J. A., JR., ST-PIERRE, J., PELLETIER, J., LARSSON, O. & TOPISIROVIC, I. 2016. nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res,* 26**,** 636-48.

GANGAT, N., PATNAIK, M. M. & TEFFERI, A. 2016. Myelodysplastic syndromes: Contemporary review and how we treat. *Am J Hematol,* 91**,** 76-89.

GELSI-BOYER, V., TROUPLIN, V., ROQUAIN, J., ADELAIDE, J., CARBUCCIA, N., ESTERNI, B., FINETTI, P., MURATI, A., ARNOULET, C., ZERAZHI, H., FEZOUI, H., TADRIST, Z., NEZRI, M., CHAFFANET, M., MOZZICONACCI, M. J., VEY, N. & BIRNBAUM, D. 2010. ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia. *Br J Haematol,* 151**,** 365-75.

GENOVESE, G., KAHLER, A. K., HANDSAKER, R. E., LINDBERG, J., ROSE, S. A., BAKHOUM, S. F., CHAMBERT, K., MICK, E., NEALE, B. M., FROMER, M., PURCELL, S. M., SVANTESSON, O., LANDEN, M., HOGLUND, M., LEHMANN, S., GABRIEL, S. B., MORAN, J. L., LANDER, E. S., SULLIVAN, P. F., SKLAR, P., GRONBERG, H., HULTMAN, C. M. & MCCARROLL, S. A. 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med,* 371**,** 2477-87.

GENTILELLA, A., MORON-DURAN, F. D., FUENTES, P., ZWEIG-ROCHA, G., RIANO-CANALIAS, F., PELLETIER, J., RUIZ, M., TURON, G., CASTANO, J., TAULER, A., BUENO, C., MENENDEZ, P., KOZMA, S. C. & THOMAS, G. 2017. Autogenous Control of 5'TOP mRNA Stability by 40S Ribosomes. *Mol Cell,* 67**,** 55-70 e4.

GERSTUNG, M., PELLAGATTI, A., MALCOVATI, L., GIAGOUNIDIS, A., PORTA, M. G., JADERSTEN, M., DOLATSHAD, H., VERMA, A., CROSS, N. C., VYAS, P., KILLICK, S., HELLSTROM-LINDBERG, E., CAZZOLA, M., PAPAEMMANUIL, E., CAMPBELL, P. J. & BOULTWOOD, J. 2015. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun,* 6**,** 5901.

GIAGOUNIDIS, A., MUFTI, G., FENAUX, P., GERMING, U., LIST, A. & MACBETH, K. 2014. Lenalidomide as a disease-modifying agent in patients with del(5q) myelodysplastic syndromes: linking mechanism of action to clinical outcomes. *Annals of Hematology,* 93**,** 1-11.

GIAGOUNIDIS, A. A., GERMING, U., HAASE, S., HILDEBRANDT, B., SCHLEGELBERGER, B., SCHOCH, C., WILKENS, L., HEINSCH, M., WILLEMS, H., AIVADO, M. & AUL, C. 2004. Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. *Leukemia,* 18**,** 113-9.

GINGRAS, A. C., RAUGHT, B. & SONENBERG, N. 2004. mTOR signaling to translation. *Curr Top Microbiol Immunol,* 279.

GLENTHOJ, A., ORSKOV, A. D., HANSEN, J. W., HADRUP, S. R., O'CONNELL, C. & GRONBAEK, K. 2016. Immune Mechanisms in Myelodysplastic Syndrome. *Int J Mol Sci,* 17.

GRAUBERT, T. A., SHEN, D., DING, L., OKEYO-OWUOR, T., LUNN, C. L., SHAO, J., KRYSIAK, K., HARRIS, C. C., KOBOLDT, D. C., LARSON, D. E., MCLELLAN, M. D., DOOLING, D. J., ABBOTT, R. M., FULTON, R. S., SCHMIDT, H., KALICKI-VEIZER, J., O'LAUGHLIN, M., GRILLOT, M., BATY, J., HEATH, S., FRATER, J. L., NASIM, T., LINK, D. C., TOMASSON, M. H., WESTERVELT, P., DIPERSIO, J. F., MARDIS, E. R., LEY, T. J., WILSON, R. K. & WALTER, M. J. 2012. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature Genetics,* 44**,** 53-7.

GREENBERG, P. L., TUECHLER, H., SCHANZ, J., SANZ, G., GARCIA-MANERO, G., SOLE, F., BENNETT, J. M., BOWEN, D., FENAUX, P., DREYFUS, F., KANTARJIAN, H., KUENDGEN, A., LEVIS, A., MALCOVATI, L., CAZZOLA, M., CERMAK, J., FONATSCH, C., LE BEAU, M. M., SLOVAK, M. L., KRIEGER, O., LUEBBERT, M., MACIEJEWSKI, J., MAGALHAES, S. M., MIYAZAKI, Y., PFEILSTOCKER, M., SEKERES, M., SPERR, W. R., STAUDER, R., TAURO, S., VALENT, P., VALLESPI, T., VAN DE LOOSDRECHT, A. A., GERMING, U. & HAASE, D. 2012. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood,* 120**,** 2454-65.

GUERNET, A., MUNGAMURI, S. K., CARTIER, D., SACHIDANANDAM, R., JAYAPRAKASH, A., ADRIOUCH, S., VEZAIN, M., CHARBONNIER, F., ROHKIN, G., COUTANT, S., YAO, S., AINANI, H., ALEXANDRE, D., TOURNIER, I., BOYER, O., AARONSON, S. A., ANOUAR, Y. & GRUMOLATO, L. 2016. CRISPR-Barcoding for Intratumor Genetic Heterogeneity Modeling and Functional Analysis of Oncogenic Driver Mutations. *Mol Cell,* 63**,** 526-38.

HAFERLACH, T., KOHLMANN, A., WIECZOREK, L., BASSO, G., KRONNIE, G. T., BÉNÉ, M.-C., VOS, J. D., HERNÁNDEZ, J. M., HOFMANN, W.-K., MILLS, K. I., GILKES, A., CHIARETTI, S., SHURTLEFF, S. A., KIPPS, T. J., RASSENTI, L. Z., YEOH, A. E., PAPENHAUSEN, P. R., LIU, W.-M., WILLIAMS, P. M. & FOÀ, R. 2010. Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Subclassification of Leukemia: Report From the International Microarray Innovations in Leukemia Study Group. *Journal of Clinical Oncology,* 28**,** 2529-2537.

HAFERLACH, T., NAGATA, Y., GROSSMANN, V., OKUNO, Y., BACHER, U., NAGAE, G., SCHNITTGER, S., SANADA, M., KON, A., ALPERMANN, T., YOSHIDA, K., ROLLER, A., NADARAJAH, N., SHIRAISHI, Y., SHIOZAWA, Y., CHIBA, K., TANAKA, H., KOEFFLER, H. P., KLEIN, H. U., DUGAS, M., ABURATANI, H., KOHLMANN, A., MIYANO, S., HAFERLACH, C., KERN, W. & OGAWA, S. 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia,* 28**,** 241-7.

HAY, N. & SONENBERG, N. 2004. Upstream and downstream of mTOR. *Genes Dev,* 18.

HEUSER, M., MEGGENDORFER, M., CRUZ, M. M. A., FABISCH, J., KLESSE, S., KOHLER, L., GOHRING, G., GANSTER, C., SHIRNESHAN, K., GUTERMUTH, A., CERNY-REITERER, S., KRONKE, J., PANAGIOTA, V., HAFERLACH, C., KOENECKE, C., PLATZBECKER, U., THIEDE, C., SCHROEDER, T., KOBBE, G., EHRLICH, S., STAMER, K., DOHNER, K., VALENT, P., SCHLEGELBERGER, B., KROEGER, N., GANSER, A., HAASE, D., HAFERLACH, T. & THOL, F. 2015. Frequency and prognostic impact of casein kinase 1A1 mutations in MDS patients with deletion of chromosome 5q. *Leukemia,* 29**,** 1942-1945.

HOLLENBACH, P., LU, L., GANDHI, A. K., CHOPRA, R. & MACBETH, K. J. 2014. Lenalidomide Promotes Degradation of Casein Kinase 1a (CK1a) through Cereblon: Implications for the Efficacy of Lenalidomide in MDS and AML. *Blood,* 124**,** 3606-3606.

HONDA, H., USHIJIMA, T., WAKAZONO, K., ODA, H., TANAKA, Y., AIZAWA, S., ISHIKAWA, T., YAZAKI, Y. & HIRAI, H. 2000. Acquired loss of p53 induces blastic transformation in p210(bcr/abl)-expressing hematopoietic cells: a transgenic study for blast crisis of human CML. *Blood,* 95**,** 1144-50.

HONG, S., FREEBERG, M. A., HAN, T., KAMATH, A., YAO, Y., FUKUDA, T., SUZUKI, T., KIM, J. K. & INOKI, K. 2017. LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. *Elife,* 6.

HORN, S., FIGL, A., RACHAKONDA, P. S., FISCHER, C., SUCKER, A. & GAST, A. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science,* 339.

HOSONO, N., MAKISHIMA, H., MAHFOUZ, R., PRZYCHODZEN, B., YOSHIDA, K., JEREZ, A., LAFRAMBOISE, T., POLPRASERT, C., CLEMENTE, M. J., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., SANADA, M., CUI, E., VERMA, A. K., MCDEVITT, M. A., LIST, A. F., SAUNTHARARAJAH, Y., SEKERES, M. A., BOULTWOOD, J., OGAWA, S. & MACIEJEWSKI, J. P. 2017. Recurrent genetic defects on chromosome 5q in myeloid neoplasms. *Oncotarget,* 8**,** 6483-6495.

HSIEH, A. C., LIU, Y., EDLIND, M. P., INGOLIA, N. T., JANES, M. R., SHER, A., SHI, E. Y., STUMPF, C. R., CHRISTENSEN, C., BONHAM, M. J., WANG, S., REN, P., MARTIN, M., JESSEN, K., FELDMAN, M. E., WEISSMAN, J. S., SHOKAT, K. M., ROMMEL, C. & RUGGERO, D. 2012. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature,* 485**,** 55-61.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech,* 31**,** 827-832.

HSU, Y. C., CHIU, Y. C., LIN, C. C., KUO, Y. Y., HOU, H. A., TZENG, Y. S., KAO, C. J., CHUANG, P. H., TSENG, M. H., HSIAO, T. H., CHOU, W. C. & TIEN, H. F. 2017. The distinct biological implications of Asxl1 mutation and its roles in leukemogenesis revealed by a knock-in mouse model. *J Hematol Oncol,* 10**,** 139.

HUART, A.-S., MACLAINE, N. J., MEEK, D. W. & HUPP, T. R. 2009. CK1alpha plays a central role in mediating MDM2 control of p53 and E2F-1 protein stability. *J Biol Chem,* 284.

IADEVAIA, V., LIU, R. & PROUD, C. G. 2014. mTORC1 signaling controls multiple steps in ribosome biogenesis. *Semin Cell Dev Biol,* 36**,** 113-20.

IHRY, R. J., WORRINGER, K. A., SALICK, M. R., FRIAS, E., HO, D., THERIAULT, K., KOMMINENI, S., CHEN, J., SONDEY, M., YE, C., RANDHAWA, R., KULKARNI, T., YANG, Z., MCALLISTER, G., RUSS, C., REECE-HOYES, J., FORRESTER, W., HOFFMAN, G. R., DOLMETSCH, R. & KAYKAS, A. 2017. P53 toxicity is a hurdle to CRISPR/CAS9 screening and engineering in human pluripotent stem cells. *bioRxiv*.

INGOLIA, N. T., BRAR, G. A., ROUSKIN, S., MCGEACHY, A. M. & WEISSMAN, J. S. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc,* 7**,** 1534-50.

ISSA, J. P. 2010. Epigenetic changes in the myelodysplastic syndrome. *Hematol Oncol Clin North Am,* 24**,** 317-30.

ITO, E., KONNO, Y., TOKI, T. & TERUI, K. 2010. Molecular pathogenesis in Diamond-Blackfan anemia. *Int J Hematol,* 92**,** 413-8.

ITZYKSON, R. & FENAUX, P. 2014. Epigenetics of myelodysplastic syndromes. *Leukemia,* 28**,** 497-506.

JAAKO, P., DEBNATH, S., OLSSON, K., BRYDER, D., FLYGARE, J. & KARLSSON, S. 2012. Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. *Blood,* 120**,** 2225-8.

JÄDERSTEN, M. 2010. *Pathophysiology and treatment of the myelodysplastic syndrome with isolated 5q deletion*.

JÄDERSTEN, M., SAFT, L., PELLAGATTI, A., GÖHRING, G., WAINSCOAT, J. S., BOULTWOOD, J., PORWIT, A., SCHLEGELBERGER, B. & HELLSTRÖM-LINDBERG, E. 2009. *Clonal heterogeneity in the 5q- syndrome: p53 expressing progenitors prevail during lenalidomide treatment and expand at disease progression*.

JÄDERSTEN, M., SAFT, L., SMITH, A., KULASEKARARAJ, A., POMPLUN, S., GÖHRING, G., HEDLUND, A., HAST, R., SCHLEGELBERGER, B., PORWIT, A., HELLSTRÖM-LINDBERG, E. & MUFTI, G. J. 2011. TP53 Mutations in Low-Risk Myelodysplastic Syndromes With del(5q) Predict Disease Progression. *Journal of Clinical Oncology,* 29**,** 1971-1979.

JAJU, R. J., BOULTWOOD, J., OLIVER, F. J., KOSTRZEWA, M., FIDLER, C., PARKER, N., MCPHERSON, J. D., MORRIS, S. W., MULLER, U., WAINSCOAT, J. S. & KEARNEY, L. 1998. Molecular cytogenetic delineation of the critical deleted region in the 5q- syndrome. *Genes Chromosomes Cancer,* 22**,** 251-6.

JÄRÅS, M., MILLER, P. G., CHU, L. P., PURAM, R. V., FINK, E. C., SCHNEIDER, R. K., AL-SHAHROUR, F., PEÑA, P., BREYFOGLE, L. J., HARTWELL, K. A., MCCONKEY, M. E., COWLEY, G. S., ROOT, D. E., KHARAS, M. G., MULLALLY, A. & EBERT, B. L. 2014. Csnk1a1 inhibition has p53-dependent therapeutic efficacy in acute myeloid leukemia. *J Exp Med,* 211.

KENNEDY, J. A. & EBERT, B. L. 2017. Clinical Implications of Genetic Mutations in Myelodysplastic Syndrome. *J Clin Oncol,* 35**,** 968-974.

KNIPPSCHILD, U., KRÜGER, M., RICHTER, J., XU, P., GARCÍA-REYES, B., PEIFER, C., HALEKOTTE, J., BAKULEV, V. & BISCHOF, J. 2014. The CK1 Family: Contribution to Cellular Stress Response and Its Role in Carcinogenesis. *Frontiers in Oncology,* 4**,** 96.

KOMROKJI, R. S., PADRON, E., EBERT, B. L. & LIST, A. F. 2013. Deletion 5q MDS: Molecular and therapeutic implications. *Best Practice & Research Clinical Haematology,* 26**,** 365-375.

KOTINI, A. G., CHANG, C. J., BOUSSAAD, I., DELROW, J. J., DOLEZAL, E. K., NAGULAPALLY, A. B., PERNA, F., FISHBEIN, G. A., KLIMEK, V. M., HAWKINS, R. D., HUANGFU, D., MURRY, C. E., GRAUBERT, T., NIMER, S. D. & PAPAPETROU, E. P. 2015. Functional analysis of a chromosomal deletion

associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. *Nat Biotechnol,* 33**,** 646-55.

KOTINI, A. G., CHANG, C. J., CHOW, A., YUAN, H., HO, T. C., WANG, T., VORA, S., SOLOVYOV, A., HUSSER, C., OLSZEWSKA, M., TERUYA-FELDSTEIN, J., PERUMAL, D., KLIMEK, V. M., SPYRIDONIDIS, A., RAMPAL, R. K., SILVERMAN, L., REDDY, E. P., PAPAEMMANUIL, E., PAREKH, S., GREENBAUM, B. D., LESLIE, C. S., KHARAS, M. G. & PAPAPETROU, E. P. 2017. Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. *Cell Stem Cell,* 20**,** 315-328 e7.

KRESSLER, D., HURT, E. & BASSLER, J. 2010. Driving ribosome assembly. *Biochim Biophys Acta,* 1803**,** 673-83.

KRONKE, J., FINK, E. C., HOLLENBACH, P. W., MACBETH, K. J., HURST, S. N., UDESHI, N. D., CHAMBERLAIN, P. P., MANI, D. R., MAN, H. W., GANDHI, A. K., SVINKINA, T., SCHNEIDER, R. K., MCCONKEY, M., JARAS, M., GRIFFITHS, E., WETZLER, M., BULLINGER, L., CATHERS, B. E., CARR, S. A., CHOPRA, R. & EBERT, B. L. 2015. Lenalidomide induces ubiquitination and degradation of CK1alpha in del(5q) MDS. *Nature,* 523**,** 183-8.

KWART, D., PAQUET, D., TEO, S. & TESSIER-LAVIGNE, M. 2017. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat Protoc,* 12**,** 329-354.

LARANJEIRA, P., RODRIGUES, R., CARVALHEIRO, T., CONSTANÇO, C., VITÓRIA, H., MATARRAZ, S., TRINDADE, H., ÓRFÃO, A. & PAIVA, A. 2015. Expression of CD44 and CD35 during normal and myelodysplastic erythropoiesis. *Leukemia Research,* 39**,** 361-370.

LARSSON, O., MORITA, M., TOPISIROVIC, I., ALAIN, T., BLOUIN, M. J., POLLAK, M. & SONENBERG, N. 2012. Distinct perturbation of the translatome by the antidiabetic drug metformin. *Proc Natl Acad Sci U S A,* 109**,** 8977-82.

LAW, J. C., RITKE, M. K., YALOWICH, J. C., LEDER, G. H. & FERRELL, R. E. 1993. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk Res,* 17**,** 1045-50.

LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-93.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIN, S.-Y., XIA, W., WANG, J. C., KWONG, K. Y., SPOHN, B., WEN, Y., PESTELL, R. G. & HUNG, M.-C. 2000. β-Catenin, a novel prognostic marker for breast cancer: Its roles in cyclin D1 expression and cancer progression. *Proceedings of the National Academy of Sciences,* 97**,** 4262-4266.

LIST, A. 2008. *Clinician's Manual on Myelodysplastic Syndromes,* London, Current Medicine Group.

LIST, A., DEWALD, G., BENNETT, J., GIAGOUNIDIS, A., RAZA, A., FELDMAN, E., POWELL, B., GREENBERG, P., THOMAS, D., STONE, R., REEDER, C., WRIDE, K., PATIN, J., SCHMIDT, M., ZELDIS, J. & KNIGHT, R. 2006a. Lenalidomide in the Myelodysplastic Syndrome with Chromosome 5q Deletion. *New England Journal of Medicine,* 355**,** 1456-1465.

LIST, A., DEWALD, G., BENNETT, J., GIAGOUNIDIS, A., RAZA, A., FELDMAN, E., POWELL, B., GREENBERG, P., THOMAS, D., STONE, R., REEDER, C., WRIDE, K., PATIN, J., SCHMIDT, M., ZELDIS, J., KNIGHT, R. & MYELODYSPLASTIC SYNDROME-003 STUDY, I. 2006b. Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. *New England Journal of Medicine,* 355**,** 1456-65.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

MALCOVATI, L., PAPAEMMANUIL, E., BOWEN, D. T., BOULTWOOD, J., DELLA PORTA, M. G., PASCUTTO, C., TRAVAGLINO, E., GROVES, M. J., GODFREY, A. L., AMBAGLIO, I., GALLI, A., DA VIA, M. C., CONTE, S., TAURO, S., KEENAN, N., HYSLOP, A., HINTON, J., MUDIE, L. J., WAINSCOAT, J. S., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J., HELLSTROM-LINDBERG, E., CAZZOLA, M., CHRONIC MYELOID DISORDERS WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. & OF THE ASSOCIAZIONE ITALIANA PER LA RICERCA SUL CANCRO GRUPPO ITALIANO MALATTIE, M. 2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood,* 118**,** 6239-46.

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.

MALLO, M., DEL REY, M., IBANEZ, M., CALASANZ, M. J., ARENILLAS, L., LARRAYOZ, M. J., PEDRO, C., JEREZ, A., MACIEJEWSKI, J., COSTA, D., NOMDEDEU, M., DIEZ-CAMPELO, M., LUMBRERAS, E., GONZALEZ-MARTINEZ, T., MARUGAN, I., SUCH, E., CERVERA, J., CIGUDOSA, J. C., ALVAREZ, S., FLORENSA, L., HERNANDEZ, J. M. & SOLE, F. 2013. Response to lenalidomide in myelodysplastic syndromes with del(5q): influence of cytogenetics and mutations. *Br J Haematol,* 162**,** 74-86.

MASVIDAL, L., HULEA, L., FURIC, L., TOPISIROVIC, I. & LARSSON, O. 2017. mTOR-sensitive translation: Cleared fog reveals more trees. *RNA Biol***,** 1-7.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol,* 17**,** 122.

MEGGENDORFER, M., HAFERLACH, C., KERN, W. & HAFERLACH, T. 2017. Molecular analysis of myelodysplastic syndrome with isolated del(5q) reveals a specific spectrum of molecular mutations with prognostic impact: a study on 123 patients and 27 genes. *Haematologica*.

MENEZES, J., SALGADO, R. N., ACQUADRO, F., GOMEZ-LOPEZ, G., CARRALERO, M. C., BARROSO, A., MERCADILLO, F., ESPINOSA-HEVIA, L., TALAVERA-CASANAS, J. G., PISANO, D. G., ALVAREZ, S. & CIGUDOSA, J. C. 2013. ASXL1, TP53 and IKZF3 mutations are present in the chronic phase and blast crisis of chronic myeloid leukemia. *Blood Cancer J,* 3**,** e157.

MEYUHAS, O. & KAHAN, T. 2015. The race to decipher the top secrets of TOP mRNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*.

MILLS, K. I., KOHLMANN, A., WILLIAMS, P. M., WIECZOREK, L., LIU, W.-M., LI, R., WEI, W., BOWEN, D. T., LOEFFLER, H., HERNANDEZ, J. M., HOFMANN, W.-K. & HAFERLACH, T. 2009. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. *Blood,* 114**,** 1063-1072.

MULLER, P. A. & VOUSDEN, K. H. 2013. p53 mutations in cancer. *Nat Cell Biol,* 15**,** 2-8.

NARLA, A. & EBERT, B. L. 2010. Ribosomopathies: human disorders of ribosome dysfunction. *Blood,* 115**,** 3196-205.

NARLA, A., HURST, S. N. & EBERT, B. L. 2011. Ribosome defects in disorders of erythropoiesis. *Int J Hematol,* 93**,** 144-9.

NARLA, A., PAYNE, E. M., ABAYASEKARA, N., HURST, S. N., RAISER, D. M., LOOK, A. T., BERLINER, N., EBERT, B. L. & KHANNA-GUPTA, A. 2014. L-Leucine improves the anaemia in models of Diamond Blackfan anaemia and the 5q- syndrome in a TP53-independent way. *British Journal of Haematology,* 167**,** 524-528.

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**,** 313-22.

NG, P. C. & HENIKOFF, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res,* 31**,** 3812-4.

NOVERSHTERN, N., SUBRAMANIAN, A., LAWTON, L. N., MAK, R. H., HAINING, W. N., MCCONKEY, M. E., HABIB, N., YOSEF, N., CHANG, C. Y., SHAY, T., FRAMPTON, G. M., DRAKE, A. C., LESKOV, I., NILSSON, B., PREFFER, F., DOMBKOWSKI, D., EVANS, J. W., LIEFELD, T., SMUTKO, J. S., CHEN, J., FRIEDMAN, N., YOUNG, R. A., GOLUB, T. R., REGEV, A. & EBERT, B. L. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell,* 144**,** 296-309.

OSBORN, M. J., BELANTO, J. J., TOLAR, J. & VOYTAS, D. F. 2016. Gene editing and its application for hematological diseases. *Int J Hematol,* 104**,** 18-28.

PAPAEMMANUIL, E., CAZZOLA, M., BOULTWOOD, J., MALCOVATI, L., VYAS, P., BOWEN, D., PELLAGATTI, A., WAINSCOAT, J. S., HELLSTROM-LINDBERG, E., GAMBACORTI-PASSERINI, C., GODFREY, A. L., RAPADO, I., CVEJIC, A., RANCE, R., MCGEE, C., ELLIS, P., MUDIE, L. J., STEPHENS, P. J., MCLAREN, S., MASSIE, C. E., TARPEY, P. S., VARELA, I., NIK-ZAINAL, S., DAVIES, H. R., SHLIEN, A., JONES, D., RAINE, K., HINTON, J., BUTLER, A. P., TEAGUE, J. W., BAXTER, E. J., SCORE, J., GALLI, A., DELLA PORTA, M. G., TRAVAGLINO, E., GROVES, M., TAURO, S., MUNSHI, N. C., ANDERSON, K. C., EL-NAGGAR, A., FISCHER, A., MUSTONEN, V., WARREN, A. J., CROSS, N. C., GREEN, A. R., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J. & CHRONIC MYELOID DISORDERS WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *New England Journal of Medicine,* 365**,** 1384-95.

PAPAEMMANUIL, E., GERSTUNG, M., MALCOVATI, L., TAURO, S., GUNDEM, G., VAN LOO, P., YOON, C. J., ELLIS, P., WEDGE, D. C., PELLAGATTI, A., SHLIEN, A., GROVES, M. J., FORBES, S. A., RAINE, K., HINTON, J., MUDIE, L. J., MCLAREN, S., HARDY, C., LATIMER, C., DELLA PORTA, M. G., O'MEARA, S., AMBAGLIO, I., GALLI, A., BUTLER, A. P., WALLDIN, G., TEAGUE, J. W., QUEK, L., STERNBERG, A., GAMBACORTI-PASSERINI, C., CROSS, N. C., GREEN, A. R., BOULTWOOD, J., VYAS, P., HELLSTROM-LINDBERG, E., BOWEN, D., CAZZOLA, M., STRATTON, M. R., CAMPBELL, P. J. & CHRONIC MYELOID DISORDERS WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood,* 122**,** 3616-27; quiz 3699.

PAYNE, E. M., VIRGILIO, M., NARLA, A., SUN, H., LEVINE, M., PAW, B. H., BERLINER, N., LOOK, A. T., EBERT, B. L. & KHANNA-GUPTA, A. 2012. *L-leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway*.

PELLAGATTI, A., BENNER, A., MILLS, K. I., CAZZOLA, M., GIAGOUNIDIS, A., PERRY, J., MALCOVATI, L., DELLA PORTA, M. G., JADERSTEN, M., VERMA, A., MCDONALD, E. J., KILLICK, S., HELLSTROM-LINDBERG, E., BULLINGER, L., WAINSCOAT, J. S. & BOULTWOOD, J. 2013. Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. *J Clin Oncol,* 31**,** 3557-64.

PELLAGATTI, A. & BOULTWOOD, J. 2015a. The molecular pathogenesis of the myelodysplastic syndromes. *Eur J Haematol,* 95**,** 3-15.

PELLAGATTI, A. & BOULTWOOD, J. 2015b. Recent Advances in the 5q- Syndrome. *Mediterr J Hematol Infect Dis,* 7**,** e2015037.

PELLAGATTI, A., CAZZOLA, M., GIAGOUNIDIS, A., PERRY, J., MALCOVATI, L., DELLA PORTA, M. G., JADERSTEN, M., KILLICK, S., VERMA, A.,

NORBURY, C. J., HELLSTROM-LINDBERG, E., WAINSCOAT, J. S. & BOULTWOOD, J. 2010a. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia,* 24**,** 756-64.

PELLAGATTI, A., CAZZOLA, M., GIAGOUNIDIS, A. A., MALCOVATI, L., PORTA, M. G., KILLICK, S., CAMPBELL, L. J., WANG, L., LANGFORD, C. F., FIDLER, C., OSCIER, D., AUL, C., WAINSCOAT, J. S. & BOULTWOOD, J. 2006. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood,* 108**,** 337-45.

PELLAGATTI, A., DOLATSHAD, H., VALLETTA, S. & BOULTWOOD, J. 2015. Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Arch Toxicol,* 89**,** 1023-34.

PELLAGATTI, A., HELLSTRÖM-LINDBERG, E., GIAGOUNIDIS, A., PERRY, J., MALCOVATI, L., DELLA PORTA, M. G., JÄDERSTEN, M., KILLICK, S., FIDLER, C., CAZZOLA, M., WAINSCOAT, J. S. & BOULTWOOD, J. 2008. Haploinsufficiency of RPS14 in 5q− syndrome is associated with deregulation of ribosomal- and translation-related genes. *British Journal of Haematology,* 142**,** 57-64.

PELLAGATTI, A., JADERSTEN, M., FORSBLOM, A. M., CATTAN, H., CHRISTENSSON, B., EMANUELSSON, E. K., MERUP, M., NILSSON, L., SAMUELSSON, J., SANDER, B., WAINSCOAT, J. S., BOULTWOOD, J. & HELLSTROM-LINDBERG, E. 2007. Lenalidomide inhibits the malignant clone and up-regulates the SPARC gene mapping to the commonly deleted region in 5q-syndrome patients. *Proc Natl Acad Sci U S A,* 104**,** 11406-11.

PELLAGATTI, A., MARAFIOTI, T., PATERSON, J. C., BARLOW, J. L., DRYNAN, L. F., GIAGOUNIDIS, A., PILERI, S. A., CAZZOLA, M., MCKENZIE, A. N., WAINSCOAT, J. S. & BOULTWOOD, J. 2010b. Induction of p53 and up-regulation of the p53 pathway in the human 5q- syndrome. *Blood,* 115**,** 2721-3.

PETZOLD, G., FISCHER, E. S. & THOMA, N. H. 2016. Structural basis of lenalidomide-induced CK1alpha degradation by the CRL4(CRBN) ubiquitin ligase. *Nature,* 532**,** 127-30.

PICELLI, S., FARIDANI, O. R., BJORKLUND, A. K., WINBERG, G., SAGASSER, S. & SANDBERG, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc,* 9**,** 171-81.

POSPISILOVA, D., CMEJLOVA, J., HAK, J., ADAM, T. & CMEJLA, R. 2007. Successful treatment of a Diamond-Blackfan anemia patient with amino acid leucine. *Haematologica,* 92**,** e66-7.

RAMLEE, M. K. 2015. High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci Rep,* 26.

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**,** 2281-308.

SAMARUT, É., LISSOUBA, A. & DRAPEAU, P. 2016. A simplified method for identifying early CRISPR-induced indels in zebrafish embryos using High Resolution Melting analysis. *BMC Genomics,* 17**,** 547.

SAXTON, R. A. & SABATINI, D. M. 2017. mTOR Signaling in Growth, Metabolism, and Disease. *Cell,* 168**,** 960-976.

SCHITTEK, B. & SINNBERG, T. 2014. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. *Molecular Cancer,* 13**,** 231.

SCHNEIDER, R., MCCONKEY, M., CHU, L. & EBERT, B. 2014a. Reduced Protein Synthesis and p53 Activation in Late-Stage Erythroblasts Mediate the Erythroid Differentiation Defect in Mice with Ribosomal Protein S14 Haploinsufficiency. *Blood,* 124

.

SCHNEIDER, REBEKKA K., ADEMÀ, V., HECKL, D., JÄRÅS, M., MALLO, M., LORD, ALLEGRA M., CHU, LISA P., MCCONKEY, MARIE E., KRAMANN, R., MULLALLY, A., BEJAR, R., SOLÉ, F. & EBERT, BENJAMIN L. 2014b. Role of Casein Kinase 1A1 in the Biology and Targeted Therapy of del(5q) MDS. *Cancer Cell,* 26**,** 509-520.

SCHWANK, G., KOO, B.-K., SASSELLI, V., DEKKERS, JOHANNA F., HEO, I., DEMIRCAN, T., SASAKI, N., BOYMANS, S., CUPPEN, E., VAN DER ENT, CORNELIS K., NIEUWENHUIS, EDWARD E. S., BEEKMAN, JEFFREY M. & CLEVERS, H. 2013. Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. *Cell Stem Cell,* 13**,** 653-658.

SEBAA, A., ADES, L., BARAN-MARZACK, F., MOZZICONACCI, M.-J., PENTHER, D., DOBBELSTEIN, S., STAMATOULLAS, A., RÉCHER, C., PREBET, T., MOULESSEHOUL, S., FENAUX, P. & ECLACHE, V. 2012. Incidence of 17p deletions and TP53 mutation in myelodysplastic syndrome and acute myeloid leukemia with 5q deletion. *Genes, Chromosomes and Cancer,* 51**,** 1086-1092.

SENGUPTA, S., PETERSON, T. R. & SABATINI, D. M. 2010. Regulation of the mTOR complex 1 pathway by nutrients, growth factors, and stress. *Mol Cell,* 40**,** 310-22.

SHALEM, O., SANJANA, N. E., HARTENIAN, E., SHI, X., SCOTT, D. A., MIKKELSON, T., HECKL, D., EBERT, B. L., ROOT, D. E., DOENCH, J. G. & ZHANG, F. 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science (New York, N.Y.),* 343**,** 84-87.

SINNBERG, T., MENZEL, M., KAESLER, S., BIEDERMANN, T., SAUER, B., NAHNSEN, S., SCHWARZ, M., GARBE, C. & SCHITTEK, B. 2010. Suppression of Casein Kinase 1α in Melanoma Cells Induces a Switch in β-Catenin Signaling to Promote Metastasis. *Cancer Research,* 70**,** 6999-7009.

SMITH, A. E., KULASEKARARAJ, A. G., JIANG, J., MIAN, S., MOHAMEDALI, A., GAKEN, J., IRELAND, R., CZEPULKOWSKI, B., BEST, S. & MUFTI, G. J. 2015. CSNK1A1 mutations and isolated del(5q) abnormality in myelodysplastic syndrome: a retrospective mutational analysis. *Lancet Haematol,* 2**,** e212-21.

SOUSSI, T. & WIMAN, K. G. 2015. TP53: an oncogene in disguise. *Cell Death Differ,* 22**,** 1239-49.

SPERLING, A. S., GIBSON, C. J. & EBERT, B. L. 2017. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer,* 17**,** 5-19.

STARCZYNOWSKI, D. T., KUCHENBAUER, F., ARGIROPOULOS, B., SUNG, S., MORIN, R., MURANYI, A., HIRST, M., HOGGE, D., MARRA, M., WELLS, R. A., BUCKSTEIN, R., LAM, W., HUMPHRIES, R. K. & KARSAN, A. 2010. Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat Med,* 16**,** 49-58.

STEENSMA, D. P. & EBERT, B. L. 2013. Initial experience with L-leucine therapy in myelodysplastic syndromes with associated chromosome 5q deletion. *Blood,* 121**,** 4428.

STOLZ, A. & BASTIANS, H. 2015. Fresh WNT into the regulation of mitosis. *Cell Cycle,* 14**,** 2566-2570.

SU, K. H., CAO, J., TANG, Z., DAI, S., HE, Y., SAMPSON, S. B., BENJAMIN, I. J. & DAI, C. 2016. HSF1 critically attunes proteotoxic stress sensing by mTORC1 to combat stress and promote growth. *Nat Cell Biol,* 18**,** 527-39.

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. *Proceedings of the National Academy of Sciences of the United States of America* 102**,** 15545-50.

TANG, H., HORNSTEIN, E., STOLOVICH, M., LEVY, G., LIVINGSTONE, M., TEMPLETON, D., AVRUCH, J. & MEYUHAS, O. 2001. Amino Acid-Induced Translation of TOP mRNAs Is Fully Dependent on Phosphatidylinositol 3-Kinase-Mediated Signaling, Is Partially Inhibited by Rapamycin, and Is Independent of S6K1 and rpS6 Phosphorylation. *Molecular and Cellular Biology,* 21**,** 8671-8683.

TCHERKEZIAN, J., CARGNELLO, M., ROMEO, Y., HUTTLIN, E. L., LAVOIE, G., GYGI, S. P. & ROUX, P. P. 2014. Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5′TOP mRNA translation. *Genes & Development,* 28**,** 357-371.

TETSU, O. & MCCORMICK, F. 1999. [beta]-Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature,* 398**,** 422-426.

THOREEN, C. C., CHANTRANUPONG, L., KEYS, H. R., WANG, T., GRAY, N. S. & SABATINI, D. M. 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature,* 485**,** 109-13.

TOGASAKI, E., TAKEDA, J., YOSHIDA, K., SHIOZAWA, Y., TAKEUCHI, M., OSHIMA, M., SARAYA, A., IWAMA, A., YOKOTE, K., SAKAIDA, E., HIRASE, C., TAKESHITA, A., IMAI, K., OKUMURA, H., MORISHITA, Y., USUI, N., TAKAHASHI, N., FUJISAWA, S., SHIRAISHI, Y., CHIBA, K., TANAKA, H., KIYOI, H., OHNISHI, K., OHTAKE, S., ASOU, N., KOBAYASHI, Y., MIYAZAKI, Y., MIYANO, S., OGAWA, S., MATSUMURA, I., NAKASEKO, C. & NAOE, T. 2017. Frequent somatic mutations in epigenetic regulators in newly diagnosed chronic myeloid leukemia. *Blood Cancer J,* 7**,** e559.

VALLETTA, S., DOLATSHAD, H., BARTENSTEIN, M., YIP, B. H., BELLO, E., GORDON, S., YU, Y., SHAW, J., ROY, S., SCIFO, L., SCHUH, A., PELLAGATTI, A., FULGA, T. A., VERMA, A. & BOULTWOOD, J. 2015. ASXL1 mutation correction by CRISPR/Cas9 restores gene function in leukemia cells and increases survival in mouse xenografts. *Oncotarget,* 6**,** 44061-71.

VAN DEN BERGHE, H., CASSIMAN, J. J., DAVID, G., FRYNS, J. P., MICHAUX, J. L. & SOKAL, G. 1974. Distinct haematological disorder with deletion of long arm of no. 5 chromosome. *Nature,* 251**,** 437-8.

VOUSDEN, K. H. & LANE, D. P. 2007. p53 in health and disease. *Nat Rev Mol Cell Biol,* 8.

VOUSDEN, K. H. & PRIVES, C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell,* 137.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell,* 153**,** 910-918.

WANG, T., BIRSOY, K., HUGHES, N. W., KRUPCZAK, K. M., POST, Y., WEI, J. J., LANDER, E. S. & SABATINI, D. M. 2015. Identification and characterization of essential genes in the human genome. *Science,* 350**,** 1096-101.

WEI, S., CHEN, X., MCGRAW, K., ZHANG, L., KOMROKJI, R., CLARK, J., CACERES, G., BILLINGSLEY, D., SOKOL, L., LANCET, J., FORTENBERY, N., ZHOU, J., EKSIOGLU, E. A., SALLMAN, D., WANG, H., EPLING-BURNETTE, P. K., DJEU, J., SEKERES, M., MACIEJEWSKI, J. P. & LIST, A. 2013. Lenalidomide promotes p53 degradation by inhibiting MDM2 auto-ubiquitination in myelodysplastic syndrome with chromosome 5q deletion. *Oncogene,* 32**,** 1110-20.

WEINHOLD, N., JACOBSEN, A., SCHULTZ, N., SANDER, C. & LEE, W. 2014. Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet,* 46.

WEINSTAT-SASLOW, D., MERINO, M. J., MANROW, R. E., LAWRENCE, J. A., BLUTH, R. F., WITTENBEL, K. D., SIMPSON, J. F., PAGE, D. L. & STEEG, P.

S. 1995. Overexpression of cyclin D mRNA distinguishes invasive and in situ breast carcinomas from non-malignant lesions. *Nat Med,* 1.

WOLFSON, R. L., CHANTRANUPONG, L., SAXTON, R. A., SHEN, K., SCARIA, S. M., CANTOR, J. R. & SABATINI, D. M. 2016. Sestrin2 is a leucine sensor for the mTORC1 pathway. *Science,* 351**,** 43-8.

WOLL, P. S., KJALLQUIST, U., CHOWDHURY, O., DOOLITTLE, H., WEDGE, D. C., THONGJUEA, S., ERLANDSSON, R., NGARA, M., ANDERSON, K., DENG, Q., MEAD, A. J., STENSON, L., GIUSTACCHINI, A., DUARTE, S., GIANNOULATOU, E., TAYLOR, S., KARIMI, M., SCHARENBERG, C., MORTERA-BLANCO, T., MACAULAY, I. C., CLARK, S. A., DYBEDAL, I., JOSEFSEN, D., FENAUX, P., HOKLAND, P., HOLM, M. S., CAZZOLA, M., MALCOVATI, L., TAURO, S., BOWEN, D., BOULTWOOD, J., PELLAGATTI, A., PIMANDA, J. E., UNNIKRISHNAN, A., VYAS, P., GOHRING, G., SCHLEGELBERGER, B., TOBIASSON, M., KVALHEIM, G., CONSTANTINESCU, S. N., NERLOV, C., NILSSON, L., CAMPBELL, P. J., SANDBERG, R., PAPAEMMANUIL, E., HELLSTROM-LINDBERG, E., LINNARSSON, S. & JACOBSEN, S. E. 2014. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell,* 25**,** 794-808.

WU, S., CHEN, L., BECKER, A., SCHONBRUNN, E. & CHEN, J. 2012. Casein kinase 1α regulates an MDMX intramolecular interaction to stimulate p53 binding. *Mol Cell Biol,* 32.

XIE, F., YE, L., CHANG, J. C., BEYER, A. I., WANG, J., MUENCH, M. O. & KAN, Y. W. 2014. Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Research,* 24**,** 1526-1533.

XIONG, Q., MUKHERJEE, S. & FUREY, T. S. 2014. GSAASeqSP: a toolset for gene set association analysis of RNA-Seq data. *Sci Rep,* 4**,** 6347.

XU, B., GOGOL, M., GAUDENZ, K. & GERTON, J. L. 2016. Improved transcription and translation with L-leucine stimulation of mTORC1 in Roberts syndrome. *BMC Genomics,* 17**,** 25.

XU, B., LEE, K. K., ZHANG, L. & GERTON, J. L. 2013. Stimulation of mTORC1 with L-leucine rescues defects associated with Roberts syndrome. *PLoS Genet,* 9**,** e1003857.

YIP, B. H., PELLAGATTI, A., VUPPUSETTY, C., GIAGOUNIDIS, A., GERMING, U., LAMIKANRA, A. A., ROBERTS, D. J., FERNANDEZ-MERCADO, M., MCDONALD, E. J., KILLICK, S., WAINSCOAT, J. S. & BOULTWOOD, J. 2012. Effects of L-leucine in 5q- syndrome and other RPS14-deficient erythroblasts. *Leukemia,* 26**,** 2154-2158.

YIP, B. H., VUPPUSETTY, C., ATTWOOD, M., GIAGOUNIDIS, A., GERMING, U., LAMIKANRA, A. A., ROBERTS, D. J., MACIEJEWSKI, J. P., VANDENBERGHE, P., MECUCCI, C., WAINSCOAT, J. S., PELLAGATTI, A. & BOULTWOOD, J. 2013. Activation of the mTOR signaling pathway by L-

leucine in 5q- syndrome and other RPS14-deficient erythroblasts. *Leukemia,* 27**,** 1760-3.

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., HAFERLACH, C., KOEFFLER, H. P., SHIH, L. Y., HAFERLACH, T., CHIBA, S., NAKAUCHI, H., MIYANO, S. & OGAWA, S. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature,* 478**,** 64-9.

ZEIDAN, A. M., LINHARES, Y. & GORE, S. D. 2013. Current therapy of myelodysplastic syndromes. *Blood Rev,* 27**,** 243-59.

ZHANG, H. & MCCARTY, N. 2016. CRISPR-Cas9 technology and its application in haematological disorders. *Br J Haematol,* 175**,** 208-225.

ZHANG, Y., EAR, J., YANG, Z., MORIMOTO, K., ZHANG, B. & LIN, S. 2014. Defects of protein production in erythroid cells revealed in a zebrafish Diamond–Blackfan anemia model for mutation in RPS19. *Cell Death & Disease,* 5**,** e1352.

ZHOU, X., HAO, Q., LIAO, J., ZHANG, Q. & LU, H. 2013. Ribosomal protein S14 unties the MDM2-p53 loop upon ribosomal stress. *Oncogene,* 32**,** 388-96.

ZHOU, X., LIAO, J. M., LIAO, W. J. & LU, H. 2012. Scission of the p53-MDM2 Loop by Ribosomal Proteins. *Genes Cancer,* 3**,** 298-310.

ZINK, F., STACEY, S. N., NORDDAHL, G. L., FRIGGE, M. L., MAGNUSSON, O. T., JONSDOTTIR, I., THORGEIRSSON, T. E., SIGURDSSON, A., GUDJONSSON, S. A., GUDMUNDSSON, J., JONASSON, J. G., TRYGGVADOTTIR, L., JONSSON, T., HELGASON, A., GYLFASON, A., SULEM, P., RAFNAR, T., THORSTEINSDOTTIR, U., GUDBJARTSSON, D. F., MASSON, G., KONG, A. & STEFANSSON, K. 2017. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood,* 130**,** 742-752.

Appendix

A: Custom script for analysis of RNA-seq data from polysome profiling experiments

```
#!/bin/sh

#List the names of fastq files

fastqFilesR1=( "./C1_L_RBR_1.fastq.gz" "./C1_D_RBR_1.fastq.gz"
"./C1_L_RNA_1.fastq.gz" "./C1_D_RNA_1.fastq.gz" )

fastqFilesR2=( "./C1_L_RBR_2.fastq.gz" "./C1_D_RBR_2.fastq.gz"
"./C1_L_RNA_2.fastq.gz" "./C1_D_RNA_2.fastq.gz" )

#Load modules needed

module load trim_galore

module load rna-star/2.5.3a

module load samtools

module load subread

module load python

echo "Modules loaded"

#Quality and adapter trimming

for i in "${!fastqFilesR1[@]}"; do

trim_galore -q 20 --paired -a TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -a2
CTGTCTCTTATACACATCTCCGAGCCCACGAGAC ${fastqFilesR1[i]}
${fastqFilesR2[i]};

done

echo "Done trimming"
```

```
#Mapping

sampleNames=( "C1_L_RNA" "C1_D_RNA" "C1_L_RBR" "C1_D_RBR" )

for i in "${sampleNames[@]}"; do

STAR --genomeDir /databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/STAR --sjdbGTFfile /databank/igenomes/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf --outSAMtype BAM Unsorted --sjdbOverhang 74 --readFilesIn $i"_1_val_1.fq.gz" $i"_2_val_2.fq.gz" --readFilesCommand zcat --outFilterScoreMinOverLread 0.44 --outFilterMatchNminOverLread 0.44 --outSAMattributes All --outFileNamePrefix $i"_" --runThreadN 6

echo "Mapped $i"

#Sort bam file and remove unsorted bam

samtools sort $i"_Aligned.out.bam" -o $i"_Aligned.sortedByCoord.out.bam"

until [ -s $i"_Aligned.sortedByCoord.out.bam" ]

do

sleep 2m

done

rm -f $i"_Aligned.out.bam"

echo "Sorted $i"

#RNA-SeQC

if [ -s C1_L_RNA_Aligned.sortedByCoord.out.bam ] && [ -s C1_L_RBR_Aligned.sortedByCoord.out.bam ] && [ -s C1_D_RNA_Aligned.sortedByCoord.out.bam ] && [ -s C1_D_RBR_Aligned.sortedByCoord.out.bam ]

then

python /t1-data/user/erepapi/python/RNASeq_pipeline_revised/rnaseq_rnaseqc.py -o /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/
```

```
C1_stats/ -mapdir /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/
-map star -build hg19 -suffixbam _Aligned.sortedByCoord.out.bam -user ebello

fi

#Counting reads in genes from BAM files - FeatureCounts

featureCounts -T 5 -a
/databank/igenomes/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf -t exon -g
gene_id -o $i"_featured_counted.txt" -p -s 0 $i"_Aligned.sortedByCoord.out.bam" 2>
$i"_fc_exon.txt"

done

echo "Made FeatureCounts files"

#Counting reads in genes and generating TPMs from BAM files - Kallisto

module load kallisto

kallisto quant -i /t1-
data/databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/Kallisto/refMrna.idx -o
kallisto_C1_L_RBR -b 100 /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/
C1_L_RBR_1.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/
C1_L_RBR_2.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analy
sis/C1_L_RBR_1.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analy
sis/C1_L_RBR_2.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analy
sis/C1_L_RBR_1.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analy
sis/C1_L_RBR_2.fastq.gz /t1-
data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analy
sis/C1_L_RBR_1.fastq.gz /t1-
```

data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analysis/C1_L_RBR_2.fastq.gz

kallisto quant -i /t1-data/databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/Kallisto/refMrna.idx -o kallisto_C1_D_RBR -b 100 /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_220317/C1_D_RBR_analysis/C1_D_RBR_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_220317/C1_D_RBR_analysis/C1_D_RBR_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_D_RBR_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_D_RBR_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_4/C1_analysis/C1_D_RBR_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_4/C1_analysis/C1_D_RBR_2.fastq.gz

kallisto quant -i /t1-data/databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/Kallisto/refMrna.idx -o kallisto_C1_L_RNA -b 100 /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_L_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_L_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analysis/C1_L_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analysis/C1_L_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analysis/C1_L_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analysis/C1_L_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analy

sis/C1_L_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analysis/C1_L_RNA_2.fastq.gz

kallisto quant -i /t1-data/databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/Kallisto/refMrna.idx -o kallisto_C1_D_RNA -b 100 /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_D_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_040517/C1_analysis/C1_D_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analysis/C1_D_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_1/C1_analysis/C1_D_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analysis/C1_D_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_2/C1_analysis/C1_D_RNA_2.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analysis/C1_D_RNA_1.fastq.gz /t1-data/user/ebello/polysome_profiling/polysome_profiling_5q_Smart_230517_3/C1_analysis/C1_D_RNA_2.fastq.gz

echo "All done"

B: R script for analysis of differential gene expression using DESeq2

#Packages required

library("calibrate")

library("RColorBrewer")

library("pheatmap")

library("DESeq2")

```
library("ggplot2")

library("dplyr")

library("tidyr")

library("tibble")

library("viridis")

#Set working directory

setwd("/t1-data/user/ebello/polysome_profiling/")

#DESeq2 analysis

#import table with all counts for all samples

countdata_all <- read.table("featurecounts_raw_counts.txt")

colData_D_RNA <- data.frame(

row.names = colnames(countdata_D_RNA),

disease = as.factor (c(rep("Healthy Control", Healthy), rep("5q- syndrome",Patient))),

sample = as.factor(samplesList),

treatment = as.factor(treatList))

colData_D_RNA_ $disease <- relevel (colData_D_RNA $disease, ref="Healthy
Control")

dds_D_RNA <- DESeqDataSetFromMatrix(countData = countdata_D_RNA,
colData=colData_D_RNA, design=~disease + treatment)

dds_D_RNA_res  <-  DESeq(dds_D_RNA)

res_ treatment <- results(dds_D_RNA_res, contrast = c(0,0,0.5,0,1))

res_ disease <- results(dds_D_RNA_res, contrast = c(0,0,1,0,0.5))

plotMA(res_ treatment, alpha=0.05, ylim=c(-2,+2))
```

```
plotMA(res_ disease, alpha=0.05, ylim=c(-2,+2))
```

C: R script to calculate translation efficiency from read counts

```
#make table of FC of TE from TPMs for l-leucine

setwd("/t1-data/user/ebello/polysome_profiling/")

df <- read.table("deseq_normCounts_nop11.txt")

genes <- row.names(df)

df %>%

tbl_df() %>%

mutate(genes = genes) -> df

df %>%

mutate(Sum = df %>%

rowwise %>%

dplyr::select(contains("R")) %>%

rowSums()  ) %>%

filter(Sum > 0) ->

df.filtered

filter <- rowSums(df.filtered>1)>=3

df.filtered1 <- df.filtered[filter,]

df.filtered1 %>%

transmute (    genes,

C1_D_TE = C1_D_RBR / C1_D_RNA,

C1_L_TE = C1_L_RBR / C1_L_RNA,
```

```
C2_D_TE = C2_D_RBR / C2_D_RNA,

C2_L_TE = C2_L_RBR / C2_L_RNA,

C4_D_TE = C4_D_RBR / C4_D_RNA,

C4_L_TE = C4_L_RBR / C4_L_RNA,

P7_D_TE = P7_D_RBR / P7_D_RNA,

P7_L_TE = P7_L_RBR / P7_L_RNA,

P8_D_TE = P8_D_RBR / P8_D_RNA,

P8_L_TE = P8_L_RBR / P8_L_RNA,

P10_D_TE = P10_D_RB / P10_D_RN,

P10_L_TE = P10_L_RB / P10_L_RN) -> df.te

df.te[df.te == Inf] <- NA

df.te.filtered <- na.omit(df.te)

df.te.filtered %>%

transmute(genes,

C1_D_TE_1 = 1+C1_D_TE,

C1_L_TE_1 = 1+C1_L_TE,

C2_D_TE_1 = 1+C2_D_TE,

C2_L_TE_1 = 1+C2_L_TE,

C4_D_TE_1 = 1+C4_D_TE,

C4_L_TE_1 = 1+C4_L_TE,

P7_D_TE_1 = 1+ P7_D_TE,

P7_L_TE_1 = 1+ P7_L_TE,

P8_D_TE_1 = 1+ P8_D_TE,
```

```
P8_L_TE_1 = 1+ P8_L_TE,

P10_D_TE_1 = 1+ P10_D_TE,

P10_L_TE_1 = 1+ P10_L_TE) ->  df.te.1

df.te.1 %>%

transmute(genes,

C1_FC = (C1_L_TE_1 / C1_D_TE_1),

C2_FC = (C2_L_TE_1 / C2_D_TE_1),

C4_FC = (C4_L_TE_1 / C4_D_TE_1),

P7_FC = (P7_L_TE_1 / P7_D_TE_1),

P8_FC = (P8_L_TE_1 / P8_D_TE_1),

P10_FC = (P10_L_TE_1 / P10_D_TE_1)) -> df.log2

df.log2[df.log2 == Inf] <- NA

df.log2[df.log2 == -Inf] <- NA

df.log2.filtered <- na.omit(df.log2)

FC <- data.matrix(df.log2.filtered[,2:7])

rownames(FC) <- df.log2.filtered$genes

controls <- FC[,1:3]

patients <- FC[,4:6]

mean_controls <- as.matrix(rowMeans(controls))

mean_patients <- as.matrix(rowMeans(patients))

mean_FC <- cbind(mean_patients,mean_controls)

colnames(mean_FC) <- c("Patients", "Controls")

mean_FC <- as.data.frame(mean_FC)
```

```
mean_FC <- tibble::rownames_to_column(mean_FC, var = "genes")

mean_FC %>%

transmute(genes,

Patients_logFC = log2(Patients),

Controls_logFC = log2(Controls)) ->  logFC_leu
```

D: Custom script for analysis of DNA-seq data from CRISPR-edited K562 cells

```
#!/bin/sh

module load bwa

module load samtools

module load sickle

sampleNames=( "12d_A" "12d_B" "12d_C" )

for i in "${sampleNames[@]}"; do

        sickle pe -f $i"_R1.fastq" -r $i"_R2.fastq" -t sanger -o $i"_R1_trim.fastq.gz" -p
        $i"_R2_trim.fastq.gz" -s $i"_R2_singles.fastq.gz" -q 28 -l 50 -n

                bwa mem
                /databank/igenomes/Homo_sapiens/UCSC/hg19/Sequence/BWAIndex/ge
                nome.fa $i"_R1_trim.fastq.gz" $i"_R2_trim.fastq.gz" > $i"_BWA.sam"

        samtools view -bSh $i"_BWA.sam" > $i"_BWA.bam"

        samtools sort $i"_BWA.bam" -o $i"_BWA_sorted.bam"

        samtools index $i"_BWA_sorted.bam" $i"_BWA_sorted.bai"

        samtools view -h $i"_BWA_sorted.bam" "chr17:7578399-7578637" >
        $i"_total.sam"

        samtools flagstat $i"_BWA.sam" > $i"_mapping_stat.txt"
```

```
        samtools view -H $i"_total.sam" > $i"_oligo1_3.sam"
        agrep -i "AAG.TGGC" $i"_total.sam" >> $i"_oligo1_3.sam"
        samtools view -H $i"_total.sam" > $i"_oligo1A.sam"
        agrep -i "TCACGT" $i"_oligo1_3.sam" >> $i"_oligo1A.sam"
        samtools view -H $i"_total.sam" > $i"_oligo1T.sam"
        agrep -i "ACACGT" $i"_oligo1_3.sam" >> $i"_oligo1T.sam"
        samtools view -H $i"_total.sam" > $i"_oligo1G.sam"
        agrep -i "CCACGT" $i"_oligo1_3.sam" >> $i"_oligo1G.sam"
        samtools view -H $i"_total.sam" > $i"_oligo1C.sam"
        agrep -i "GCACGT" $i"_oligo1_3.sam" >> $i"_oligo1C.sam"
        samtools view -H $i"_total.sam" > $i"_oligo2.sam"
agrep -i "AAGT.GGGC" $i"_total.sam" >> $i"_oligo2.sam"
samtools view -H $i"_total.sam" > $i"_oligo3only.sam"
agrep -i "ACAGGT" $i"_oligo1_3.sam" >> $i"_oligo3only.sam"
samtools view -H $i"_total.sam" > $i"_WT.sam"
agrep -i "CAGT.GGGC" $i"_total.sam" >> $i"_WT.sam"
done
echo "DONE"
for file in $(ls *.sam); do bash /t1-home/nuffmed/jkerry/shortcuts/count_paired.sh $file;
done > counts_p53.txt
```