Role of EZH2 in Myelodysplastic Syndromes

Shinde, Sneha

Awarding institution:
King's College London

Role of EZH2 in Myelodysplastic Syndromes

University of London

Sneha Shinde

THESIS PRESENTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

King's College London

Department of Haematological Medicine

2015

Declaration

I hereby declare that I alone composed this thesis and that the work is my own, except where stated otherwise.

Sneha Shinde

March 2015

Acknowledgement

First and foremost, I would like to thank my primary supervisor, Prof. Ghulam J. Mufti for giving me this opportunity. Without your faith in me, this PhD would not have happened. Sincere thanks for your quick feedback during the write up stage of this project in spite of your hectic schedule. Special thanks to my second supervisor and now a friend, Dr. Azim Mohamedali for his critical feedback on my work and for all your help in the crucial last few months of my PhD. Many thanks to Dr. Joop Gaken who was not my supervisor officially but who still helped me immensely in the cloning work of this project. Your kindness cannot be described in words. For me, I am your (adopted!) student.

I would also like to show my gratitude to our small but special group; Terry, Richard, Farooq, Jiang, Shok Ping, Austin and Alex for sharing their reagents in the time of need, for their help, valuable advice and encouraging words. Thank you to Shreyans for providing large amounts of clinical data needed for this project in spite of last minute notifications and to Syed, Lin for proof reading part of my thesis.

I would like to express deepest thanks to my best friends Heba, Nazia, Pilar and Guilia for listening to and supporting me in everything over the past four years. I cannot begin to express my gratitude, appreciation for your friendship. Truly blessed to share my PhD journey with friends like you!

Many thanks to the PhD co-ordinators, Professor Shaun Thomas for being a wonderful mentor, guide and to Dr. Alan Ramsay for the prompt help and kind words. Thanks to everyone in the Department of Haematological Medicine; Thomas, Winston and David for helping with long hours of cell sorting, Aytug, Rick and to the tissue bank (Nigel, Rajani, Bu) for providing the invaluable patient samples.

Last but not the least, I would like to thank my parents for their love, support, encouragement always and especially for the past four years. Without them this dream would not have come true.A big appreciation goes to my husband, friend for his help and advice in the time of need.

Finally, I take this opportunity to acknowledge the financial support of Celgene for this project.

Abstract

Occurrence of mutations in the Polycomb (PcG) gene; *EZH2* (Enhancer of Zeste Homologue 2) represent a new class of molecular lesions associated with instability in the epigenome of patients with Myelodysplastic Syndrome (MDS). Detection of microdeletion at 7q36.1 or 7q Copy Neutral Loss of Heterozygosity [CN (LOH)] led to the identification of *EZH2* mutations. *EZH2* is the catalytic component of Polycomb Repressive Complex 2 (PRC2) that trimethylates lysine 27 of histone 3 (H3K27) resulting in gene silencing and recruitment of the sister complex i.e. Polycomb Repressive Complex 1 (PRC1) to target genes. Discovery of *EZH2* mutations have shed light on the involvement of other PcG and PcG interacting proteins i.e. Jumonji (jmj) family of demethylases and DNA methyltrasferase *3A (DNMT3A)* in MDS. Investigation of Single Nucleotide Polymorphism (SNP) array abnormalities and mutational analysis of these genes have not been ascertained and therefore I examined cytogenetic aberrations affecting twelve Polycomb (PRC1) and seventeen Jumonji genes using high density SNP 6 arrays. SNP6 data analyzed in this study was generated by our group for previous research projects. I visualised this data using CHromosome Analysis Suite (Chas) from Affymetrix and identified five PRC1 genes *(BMI1, PHC1, PHC2, RING1A* and *RING1B)* in 17/91 (19 %) patients with either Copy Number Variations (CNVs) like deletions or amplifications or CN (LOH). Interestingly, the frequency of SNP6 aberrations was high (two times) in Jumonji genes as compared to PRC1. 29/91 patients (31 %) showed either CNVs or CN (LOH) in fifteen (*JMJD3, JMJD4, JMJD1B, JMJD2A, JARID2, JMJD1C, JARID1B, JMJD2C, UTX, JARID1C, JARID1A, JMJD2D, JHJD1A, JARID1D* and *JHJD1B)* Jumonji genes. Mutational analysis of patients with SNP6 aberrations was carried out using Sanger or 454 sequencing but no mutations were detected in either the PRC1 or Jumonji genes. To elucidate changes in gene expression as a result of

amplification or deletion of genomic material, qPCR was performed on 22/29 patients for thirteen Jumonji genes. Gene expression of *JARID1A*, *JARID1C* and *UTX* were modulated concomitant to the CNVs. Deletion of *JARID1A* locus was associated with reduced gene expression (p value <0.0001) in two patients while trisomy of *JARID1C* (n=1) and *UTX* (n=2) were associated with increased expression (p value <0.0001) of both the genes.

Mutational analysis of PRC2 core components (*SUZ12*, *EED*, *EZH1)* and *DNMT3A* was carried out in a cohort of 61 MDS patients previously sequenced by our group for *EZH2* mutations to examine their mutational overlap. 10/61 patients had heterozygous *DNMT3A* mutations (clone size 20-44 %) with two patients showing mutations at the R882 site. Interestingly, these mutations were seen predominantly (n= 6) in patients with monosomy 7/del 7q however only one patient had both *DNMT3A* (R882H) and *EZH2* (V626M) mutations suggesting that there is no specific association between mutations of the two genes. In contrast, PRC2 genes were not mutated in this cohort emphasizing the importance of *EZH2* mutations alone in MDS pathogenesis. Therefore I examined the functional consequences of the commonly occurring *EZH2* (R690C/R690H) and *DNMT3A* (R882H) mutations in myeloid cell lines.

To achieve this, numerous attempts were made to clone *DNMT3A* R882H mutation into p3XFLAG-myc-CMV-26 to allow transfection and in vitro assessment of the mutant in myeloid cells but all attempts to ligate the plasmid failed and therefore work on *DNMT3A* was discontinued.

EZH2 (R690C/R69H) and Flag tagged wild type *EZH2* were constructed in p3XFLAG-myc-CMV-26 vector using a PCR based cloning strategy and transfected into K562 cells. Western blot analysis at 72 hr post transfection, showed elevated levels of both

R690C/R690H mutants and Flag tagged wild type *EZH2* but no alterations in its target H3K27me3 levels. Affymetrix Human Transcriptome 2.0 gene expression profiling was used to identify modulation of gene signature as result of elevated *EZH2* levels and *MLLT10* gene was found to be up regulated in cells transfected with Flag-tagged wild type *EZH2* (2.3 fold) as well as R690C/ R690H (3.6 – 4.6 fold) mutants. In contrast, *PML* (promyelocytic leukaemia) (2.16 fold) and *FANCL* (Fanconi Anaemia, Complementation Group L) (2.18 fold) genes were up regulated exclusively in cells over expressing the Flag tagged wild type *EZH2*. To compare this gene signature to gene expression changes as a result of *EZH2* knock out (KO), shRNA mediated inhibition of *EZH2* was carried out in myeloid cells and 95 % KO of both *EZH2* and H3K27me3 levels were observed at Day 7 post transduction. Microarray gene expression profiling identified *BCL2* (-2.14 fold), *FLT1* (-4.03 fold), *HOXA10* (-2.2 fold), *CD44* (-8.2 fold), *CD83* (-2.1 fold), *TLSP* (-3.24 fold), *IFI16* (-3.11 fold) and *PAG1* (-3.37 fold) inhibition in cells transduced with shRNA against *EZH2* compared to the scrambled and wild type K562 cells. There were no overlapping genes in K562 cells with *EZH2* KO and *EZH2* mutants R690C/R690H. The differences in expression profiling could be due to the difference in H3K27me3 levels modulated by *EZH2*. Comparison of gene signature obtained by *EZH2* KO on patient samples carrying the R690H mutation, showed contrasting results i.e. up regulation of *HOXA10*, *FLT1*, *PAG1B*, *EZH1* and *TLSP* compared to patients with wild type *EZH2* suggesting that *EZH2* R690C/R690H mutants do not mimic the transcriptional changes seen in *EZH2* KO. This strongly suggests the presence of other mechanisms to compensate for the loss of *EZH2* in myeloid cells. However the results obtained here should be examined in additional other myeloid cell lines to validate the findings obtained in K562 cells.

Table of Contents

Declaration 1
Acknowledgement 2
Abstract 3
Table of Contents 6
List of Figures 13
List of Tables 15
List of Abbreviations 17
1 Chapter: Introduction 20
1.1 Myelodysplastic syndrome (MDS) 21
1.1.1 Disease classification and prognostic scoring system 21
1.1.2 Implication of Cytogenetics in MDS 26
1.1.3 Mutational analysis in MDS 27
1.2 Epigenetic aberrations in MDS 30
1.2.1 Post translational histone modifications 31
1.3 Polycomb proteins and polycomb repressive complexes 33
1.3.1 Classification of PcG proteins 33
1.3.2 Mechanism of action of Canonical PcG proteins 36
1.3.3 PcG and DNA methylation 39
1.3.4 Expression patterns of PcG proteins 42
1.4 Subunits of PRC2 44
1.4.1 Embryonic ectoderm enhancement protein 44
1.4.2 Suppressor of Zeste protein 45
1.4.3 Enhancer of Zeste proteins 46
1.5 Subunits of PRC1 47
1.5.1 Posterior sex comb proteins 47
1.5.2 Ring proteins 49
1.5.3 Polyhomeotic proteins 50
1.5.4 Chromobox proteins 52
1.6 Role of PcG proteins in DNA damage 54
1.7 Role in Haematological malignancies 55
1.8 Role in Haematopoietic system 59
1.9 Therapeutic role of PcG proteins 63
1.10 Project Aims 65

2 Chapter: Materials and Methods 67
2.1 Reagents 68
2.1.1 Enzymes and Buffer 68
2.1.2 Plasmids 68
2.1.3 Antibodies 69
2.1.4 PCR, Gel electrophoresis and Sequencing Materials 69
2.1.5 DNA and RNA extraction reagents 70
2.1.6 Cloning reagents 70
2.1.7 Transfection reagents 71
2.1.8 Tissue culture reagents & equipment 71
2.1.9 Cell lines 72
2.1.10 qPCR reagents 72
2.1.11 Western Blot reagents 72
2.1.12 Cell viability, proliferation and cell cycle reagents 72
2.1.13 Buffers and Solutions 73
2.2 siRNA 74
2.3 shRNA used with pSUPER 74
2.4 Patient samples 75
2.5 DNA Techniques 75
2.5.1 DNA Extraction 75
2.5.2 Determination of DNA concentration 75
2.5.3 DNA Amplification 76
2.5.4 DNA quality Testing 76
2.6 Primer Designing 77
2.6.1 Primer Testing and PCR optimisation 77
2.7 Mutational analysis by Sanger Sequencing 78
2.8 Next-generation sequencing of PRC2, *DNMT3A* and Jumonji genes 80
2.8.1 Patient Cohort 80
2.8.2 Gel extraction and purification 81
2.8.3 Bead Purification 82
2.8.4 Library preparation 82
2.8.5 Sequencing 82
2.8.6 Emulsion PCR 83
2.8.7 Vaccum assisted emulsion breaking 84

2.8.8 DNA library Bead Enrichment 85
2.8.9 Preparation of Enrichment Beads 85
2.8.10 Collection of Enriched DNA beads 86
2.8.11 Determine Bead enrichment 86
2.8.12 Sequencing Primer Annealing 86
2.8.13 Preparation of Bead Buffer 2 (BB2) 87
2.8.14 Bead Preparation [Titanium Sequencing Kit LR70 CatNo: 04932315001] 87
2.8.15 Layer Preparation 88
2.8.16 Assembling the BDD 89
2.8.17 Run Requirements 90
2.9 Total RNA isolation 90
2.10 cDNA synthesis 90
2.11 Real Time PCR analysis 91
2.12 Affymetrix Human Transcriptome 2.0 gene expression profiling 92
2.12.1 Reverse transcription to synthesise first-strand cDNA 93
2.12.2 Second-strand cDNA synthesis 94
2.12.3 In Vitro Transcription (IVT) to synthesise cRNA 96
2.12.4 cRNA Purification 96
2.12.5 2nd-Cycle single-strand cDNA synthesis 96
2.12.6 Hydrolyze RNA by RNase H 97
2.12.7 Purification of 2nd- Cycle single-strand cDNA 97
2.12.8 Fragmentation & Labelling single-strand cDNA 98
2.12.9 WT Array Hybrdization 99
2.12.10 Expression microarray data analysis 100
2.13 Protein Blotting 101
2.13.1 Cell lysis 101
2.13.2 Western blot analysis 101
2.14 Cloning 102
2.14.1 Restriction Digests 102
2.14.2 Conversion of extended DNA termini to blunt ends 103
2.14.3 Isolation and purification of DNA from agarose gels 103
2.14.4 Topo cloning 104
2.14.5 Transformation into DH5α-T1 *E.Coli* 104
2.14.6 Ligation 105

2.14.7 Transformation into DH10B cells 105
2.14.8 Miniprep: Alkaline lysis method 105
2.14.9 Miniprep: Wizard Plus SV miniprep 106
2.14.10 Maxiprep: Qiagen Plasmid kit 106
2.15 Cell culture 107
2.15.1 Culturing non-adherent myeloid cells lines 107
2.15.2 Transfection using Lipofectamine LTX reagent 108
2.15.3 RNAiMAX transfection protocol for siRNA experiment 108
2.15.4 Electroporation 108
2.16 Knockdown of *EZH2* 108
2.16.1 Constructing shRNA in pSUPER vector 108
2.17 Lentivirus short-hairpin RNA (LVshRNA) virus particle production & concentration of the lentivirus vector 110
2.18 Mtt cell proliferation assay 111
2.19 Cell Cycle analysis 112
2.19.1 Cell cycle staining 112
2.19.2 Cell cycle analysis by flow cytometry 112
2.20 Annexin V staining 113
3 Chapter: SNP6 and Mutational Analysis of PRC1, PRC2, DNMT3A & JUMONJI genes 114
3.1 Introduction 115
3.2 Involvement of PRC subunits, DNMT3A and the Jumonji genes in MDS/AML 115
3.3 PRC and interacting proteins 115
3.3.1 DNMT3A 115
3.3.2 Jumonji 116
3.4 SNP6 Analysis 117
3.5 Aim 118
3.6 Patient samples 118
3.6.1 Amplification of patient samples 119
3.7 SNP6 and Chas analysis 120
3.7.1 PRC1 122
3.7.2 Sanger sequencing 123
3.7.3 Jumonji genes 124
3.8 PRC2 and DNMT3A 135
3.8.1 Primer testing and PCR optimisation (PRC2) 135

3.8.2 454 parallel sequencing (PRC2) 136
3.8.3 Primer testing and PCR optimisation (DNMT3A) 140
3.8.4 454 parallel sequencing (DNMT3A) 142
3.8.5 Mapping of *DNMT3A* mutations 144
3.8.6 Correlation with EZH2 mutations 145
3.9 Discussion 146
3.9.1 SNP6 and mutational analysis of PRC1 and Jumonji genes 146
3.9.2 Mutational analysis of PRC2 and DNMT3A 150
3.9.3 Mutational overlap 152
3.10 Conclusion 152
4 Chapter: Role of EZH2 mutations (R690C & R690H) in Myeloid Malignancies 154
4.1 Introduction 155
4.2 Aim 155
4.3 Constructing the R690C & R690H mutations in pCMV sport6 vector 156
4.3.1 Generation of fragment containing R690C & R690H mutation 158
4.3.2 Three-way ligation 160
4.4 Cloning R690C & R690H mutants into p3XFLAG-myc-CMV-26 expression vector 163
4.4.1 Linearise the p3XFLAG-myc-CMV-26 expression vector for ligation 163
4.4.2 Isolation of EZH2 cDNA from pCMV sport6 vector 164
4.4.3 Ligation of EZH2 cDNA and p3XFLAG-myc-CMV-26 expression vector 164
4.5 Transfection of R690C/R690H & wild type EZH2 into myeloid cell lines 166
4.5.1 Electroporation& Lipofectamine transfection methods 167
4.6 Cloning EZH2 cDNA into p3XFLAG-myc-CMV-26 containing GFP and puromycin 171
4.6.1 Construction of wild type EZH2 171
4.6.2 Construction of R690C and R690H EZH2 mutants 173
4.7 Transfection of R690C/R690H/wild type EZH2 in p3XFLAG-myc-CMV-26 (GFP + Puromycin) into myeloid cell lines 175
4.8 Impact of EZH2 mutation on H3K27me3 177
4.9 Microarray analysis of *EZH2* mutants and over expressed wild type *EZH2* reveals novel *EZH2* targets 179
4.10 Discussion 184
4.10.1 Cloning R690C/R690H and Flag tagged wild type *EZH2* into p3XFLAG-myc-CMV-26 (neomycin) 185
4.11 Conclusion 190
5 Chapter: Role of EZH2 inhibition in Myeloid malignancies 191

5.1 Introduction 192
5.2 Knock out (KO) of EZH2 in cancer 192
5.3 RNA interference 193
5.4 Aim 195
5.5 Transient knockdown of EZH2 196
5.5.1 Silencing Efficiency of siRNA (Dharmacon) in MOLM13 & K562 cells using RNAiMAX 196
5.5.2 Silencing Efficiency of siRNA (Dharmacon) in MOLM13 & K562 cells using electroporation 197
5.6 Impact of EZH2 KO on H3K27me3 200
5.7 Cell Cycle analysis 201
5.8 Silencing Efficiency of siRNA (Santacruz) in MOLM13 & K562 cells using electroporation 202
5.9 Stable knockdown of EZH2 using shRNA in pSUPER 203
5.10 Stable knockdown of EZH2 using shRNA in lentivirus 206
5.10.1 shRNA (Insight Biotechnology) 206
5.10.2 shRNA (Fisher Scientific) 207
5.11 Effect of shRNA_D on EZH1 expression 209
5.12 Impact of EZH2 KO on H3K27me3 211
5.13 Microarray gene expression analysis of shRNA transduced cells reveals novel *EZH2* targets 212
5.13.1 Validation of *EZH2* targets 215
5.14 Impact of EZH2 KO on cell viability & proliferation 218
5.14.1 Trypan blue staining 218
5.14.2 MTT tetrazolium cell proliferation assay 219
5.14.3 Annexin V staining 220
5.14.4 Cell Cycle analysis 222
5.15 Confirmation of gene signature in patient samples 223
5.16 Discussion 225
5.16.1 Transcription and gene expression profiling 225
5.16.2 Effect of EZH2 on cell cycle, cell viability and cell cycle kinetics 229
5.16.3 Gene signature in patient samples 233
5.17 Conclusion 235
6 Chapter: Role of DNMT3A (R882H) mutation in Myeloid malignancies 236
6.1 Introduction 237
6.2 Aim 237

6.3 Constructing the R88H mutant in pCMV sport6 vector 237
6.3.1 Generating the fragment containing the R882 mutation 239
6.3.2 Ligation of mutation into pCMV-SPORT6 242
6.3.3 Cloning R882H mutation into p3XFLAG-myc-CMV-26 expression vector 244
6.3.4 Ligation of DNMT3A cDNA and p3XFLAG-myc-CMV 249
6.3.5 Ligation of Flag tag into pCMV-SPORT6 249
6.4 Transfection of R882H-mutant into myeloid cell lines 251
6.4.1 Electroporation 251
6.4.2 Lipofectamine LTX reagent 252
6.5 Discussion 254
6.5.1 Literature survey 255
6.6 Conclusion 257
7 Chapter: General Conclusions and Future work 258
7.1 Identification of SNP6 abnormalities and mutations affecting Polycomb genes and their binding partners (*DNMT3A* and Jumonji) 259
7.2 Functional analysis in myeloid malignancies: R690C/R690H mutations, over expression and knock down of EZH2. 262
7.3 Functional analysis of DNMT3A in myeloid malignancies 266
7.4 Future Work 268

List of Figures

Figure Name	Page No
Figure 1.1: Epigenetic regulation of transcription	30
Figure 1.2: Post translational modifications of overhanging N-terminal of histone tails	32
Figure 1.3: Canonical and non canonical PRC1 and PRC2 complexes	35
Figure 1.4: Mechanism of action; PRC1 and PRC2.	38
Figure 1.5: Interaction of PRC2 and DNMT3A.	40
Figure 1.6: Illustrates inhibitory effect of PcG proteins on the INK4A/ARF locus.	44
Figure 1.7: Deletion of BMI-1 domains.	48
Figure 1.8: Polyubiquitination of RING1B	50
Figure 1.9: Propagation of transcriptional repression	51
Figure 1.10: Identification 7q CN (LOH) and EZH2 involvement.	55
Figure 1.11: Dual effect of KO of RING1B on the haematopoietic system	61
Figure 2.1: Assembled set-up for vaccum assisted emulsion breaking and bead recovery	85
Figure 2.2: Assembling the BDD	89
Figure 2.3: Work-flow for Affymetrix Human Transcriptome 2.0 gene expression array	93
Figure 2.4: Components of the array chip. The square in the centre indicates the probe arrays.	100
Figure 2.5: Four shRNAs were designed against different positions along the gene	109
Figure 3.1:Human DNA OK results.	120
Figure 3.2:Visualisation of SNP6 data for the Jumonji and PRC1 genes by Chromosome Analysis Software (ChAS).	121
Figure 3.3: Total number of patients with SNP6 abnormalities.	125
Figure 3.4: Visualisation of PCR product for JMJD4 and JMJD1B on 2 % agarose for 454 sequencing.	129
Figure 3.5: Gene expression analysis of Jumonji genes by qPCR.	132
Figure 3.6: Primer optimisation for EZH1, EED & SUZ12 using different primer and template DNA concentrations.	135
Figure 3.7: Visualisation of the PCR product for EED after second round of PCR	136
Figure 3.8: SUZ12, EED, EZH1 library purification	138
Figure 3.9: Optimisation of DNMT3A primers	139
Figure 3.10: DNMT3A library preparation	140
Figure 3.11: DNMT3A (R882C) mutation in patient (LSL4062) visualised by Vnode.	143
Figure 3.12: Mapping of the mutations to the different domains of the DNMT3A protein.	143
Figure 4.1: Cloning strategy for introducing the R690C and R690H mutation in pCMV-SPORT6	157
Figure 4.2: Digestion of pCMV-SPORT6 vector containing EZH2 cDNA	157
Figure 4.3: Generation of mutational fragment R690C and R690H	159
Figure 4.4: Verification of the correct colony after cloning the PCR fragment in pCRII- TOPO vector	160
Figure 4.5: Confirmation of R690C and R690H mutant in pCMV-SPORT6 vector.	162
Figure 4.6: Cloning Strategy.	163
Figure 4.7: Cloning EZH2 cDNA into p3XFLAG-myc-CMV	164
Figure 4.8: Verification of accurate ligation of EZH2 cDNA into p3XFLAG-myc-CMV plasmid.	165
Figure 4.9: Endogenous expression of EZH2 in myeloid cell lines	166
Figure 4.10: Experimental workflow for transfection by Lipofectamine LTX	167
Figure 4.11: qPCR analysis of EZH2 expression in MOLM13 cells, 48 hr and 21 days after transfection	168
Figure 4.12: qPCR analysis of EZH2 expression in Molm13 cells, 48 and 72 hr after 5-Azacytidine treatment	169
Figure 4.13: Cloning EZH2 cDNA into p3XFLAG-myc-CMV-26 (GFP+puro)	171
Figure 4.14: Confirmation of accurate ligation of wild type EZH2 cDNA into p3XFLAG-myc-CMV plasmid	172

Figure 4.15: Confirmation of the FLAG and EZH2 sequences by Sanger sequencing.	173
Figure 4.16: Ligation of fragment containing R690C or R690H mutant in to wt Flag pX3FLAG-myc-CMV (GFP+puromycin).	174
Figure 4.17: FACS BD Fortessa analysis of GFP+ K562 cells transfected with R690C mutant.	176
Figure 4.18: Western blot analysis at 72 hr post transfection.	177
Figure 5.1:Schematic representation of RNA interference in cells	194
Figure 5.2: *EZH2* expression levels at 48 hr by qPCR	197
Figure 5.3: *EZH2* expression levels by qPCR using 150 nM siRNA (Dharmacon) concentration	198
Figure 5.4: Measurement of *EZH2* expression levels by qPCR using 200 nM concentration of siRNA (Dharmacon)	199
Figure 5.5: Western blot analysis of siRNA mediated KO of ezh2 in Molm13 cells at 24 and 48 hr	200
Figure 5.6: Cell cycle profile of Molm13 at 24 hr post transfection	201
Figure 5.7: *EZH2* expression levels at 24 hr after transfection with siRNA from Santacruz biotechnology by qPCR	202
Figure 5.8:Experimental workflow for transfection of pSUPER containing shRNA against *EZH2* into Molm13 and K562 cells	204
Figure 5.9: *EZH2* expression levels after puromycin selectionin cells transfected with shRNA in pSUPER vector by qPCR	205
Figure 5.10: qPCR analysis of *EZH2* expression after puromycin selectionin cells transfected with shRNA in pSUPER vector	205
Figure 5.11: qPCR analysis of *EZH2* expression levels at day7 post infection by shRNA (Insight Biotechnology)	206
Figure 5.12: qPCR analysis of *EZH2* expression levels at 48 hr and 7 days after transduction with lentivirus containing shRNA (Fisher Scientific) in K562 cells	208
Figure 5.13: qPCR analysis of *EZH2* expression at day 7 after infection with the concentrated (add titre) lentivirus containing shRNA (Fisher Scientific) in K562 cells	209
Figure 5.14:qPCR analysis of *EZH1* expression levels at day 7 after infection with shRNA_D in K562 cells	210
Figure 5.15: Western blot analysis of shRNA mediated KO of ezh2 in K562 cells on day 7 post FACs sorting for GFP positive cells	211
Figure 5.16: Illustrates schematic representation of top enriched pathways (a) (*P*<) and processes (b) for all up and down regulated genes after *EZH2* KO	213
Figure 5.17: qPCR validation of genes modulated due to loss of *EZH2* in K562 cells	216
Figure 5.18: Cell count of K562 cells using Trypan blue staining post shRNA lentiviral infection.	218
Figure 5.19:Cell proliferation assessment by MTT assay on day 7 after transduction in K562 cells	220
Figure 5.20: Annexin V and 7-AAD staining of K562 cells at 48 hr post shRNA infection (top panel) and on day 7 post sorting for GFP positive cells (bottom panel).	221
Figure 5.21: Cell cycle profile in K562 cells on day 7 after sorting for GFP positive cells.	222
Figure 5.22: qPCR analysis of patient samples with (R690H) or without *EZH2* mutation	224
Figure 6.1: Coning strategy for introducing the R882H mutation in pCMV-SPORT6	238
Figure 6.2: Digestion of pCMV-SPORT6 vector containing *DNMT3A* cDNA	239
Figure 6.3: PCR based generation of the mutational (R882H) fragment in *DNMT3A* cDNA	240
Figure 6.4: Second PCR to generate the mutational (R882H) fragment in *DNMT3A* cDNA	241
Figure 6.5 Visualisation of four colonies on 1 % agarose gel aftertreatment with XbaI-NruI restriction enzymes.	242
Figure 6.6: Confirmation of the presence of R882H mutant in pCMV-SPORT6 vector using sanger sequencing.	244
Figure 6.7: Cloning the R882H *DNMT3A* mutation into p3XFLAG-myc-CMV-26 expression vector	245

Figure 6.8: Digestion of pCMV-SPORT6 containing R882H mutant by SalI-Bpu10I restriction enzymes.	246
Figure 6.9: Digestion of colonies obtained after ligation of *DNMT3A*cDNA into p3XFLAG-myc-CMV-26 expression vector by HindIII-NotI restriction enzymes.	249
Figure 6.10: Second attempt to ligate *DNMT3A* into p3XFLAG-myc-CMV-26 plasmid.	247
Figure 6.11: Plasmid DNA from six colonies was digested by ScaI-NotI restriction enzymes followed by partial digest by HindIII	248
Figure 6.12: Digestion of pX3FLAG-myc-CMV-26 plasmid by HindIII-NotI restriction enzymes.	249
Figure 6.13: Cloning strategy for inserting the FLAG tag from pX3FLAG-myc into pCMV-SPORT6.	250
Figure 6.14: (a) Digestion of pX3FLAG-myc-CMV by SnabI-HindIII restriction enzymes(b) SnabI digest followed by HindIII partial digest on pCMV-SPORT6.	250
Figure 6.15: Experimental workflow for transfection by electroporation.	252
Figure 6.16: *DNMT3A* expression in Molm13 (48hr post transfection) and in Kg1(48hr/ Day 9 post transfection) cell lines	253

List of Tables

Table name	Page No
Table 1.1: WHO classification of MDS into its sub types based on changes in bone marrow and peripheral blood	23
Table 1.2: IPSS assignment of score based on bone marrow blasts, karyotype and cytopenia	25
Table 1.3: IPSS-R assignment of score based on bone marrow blasts, karyotype and cytopenia.	25
Table 1.4: Common cytogenetic abnormalities and their frequency of incidence in MDS	27
Table 1.5: Summary of genes commonly mutated in MDS	29
Table 1.6: Sub-units of canonical PRC 1 and PRC2	36
Table 1.7: Expression patterns of the PcG proteins in HSC	43
Table 1.8: Summarization of different CBX proteins.	53
Table 2.1: Preparation of D1 and N1 buffer.	76
Table 2.2: Preparation of Human DNAOK multiplex PCR master mix.	77
Table 2.3: a) PCR reaction mixture b) PCR programme on the thermocycler	78
Table 2.4: ExoSAP-IT clean–up programme.	79
Table 2.5: a) PCR reaction mixture b) PCR programme.	79
Table 2.6: Programme for the second round of PCR for 454 sequencing.	81
Table 2.7: Preparation of a) Live Amplification mix per capture bead solution b) Mock Amplification mix c) 1X Capture bead wash buffer.	83
Table 2.8: PCR conditions for clonal amplification in the emulsions.	84
Table 2.9: Coulter Counter Z1 models settings for bead counting	86
Table 2.10: Components of DBIM buffer	87
Table 2.11: Dilution of beads for layer preparation.	88
Table 2.12: Preparation of master mix for cDNA synthesis	91
Table 2.13: Preparation of Real-time PCR master mix	91
Table 2.14: PCR conditions for qPCR	91

Table 2.15: Reaction mixture for topo cloning	104
Table 3.1: Clinical characteristics of the patients studied (n=60)	119
Table 3.2: SNP6 aberrations [deletion, amplification and CN (LOH)] in PRC1 genes	123
Table 3.3: Gain/ del/ copy neutral loss of heterozygosity the Jumonji genes	127
Table 3.4: Mutational analysis of DNMT3A in ten patients with MDS/tAML using 454 sequencing	143
Table 4.1: Reaction mix for treatment of plasmid DNA with SuperTaq to add dNTP "A" to the fragment.	159
Table 4.2 :Ligation mixture for three fragments (SalI-NotI, SalI-EcoRI & EcoRI-NotI) of pCMV-SPORT6.	161
Table 4.3 :Ligation of EZH2 cDNA (R690C /R690H/ wild type EZH2) into p3XFLAG-myc-CMV-26 expression vector.	165
Table 4.4: Ligation ofwild type EZH2 into p3XFLAG-myc-CMV-26 expression vector	172
Table 4.5: Ligation of EZH2 cDNA (R690C /R690H/ wild type EZH2) into p3XFLAG-myc-CMV-26 expression vector containing GFP+Puromycin.	175
Table 4.6: List of genes modulated by over expression of wild type and mutant EZH2 by micro array gene expression profiling.	182
Table 5.1: List of genes modulated by loss of *EZH2* expression	214
Table 5.2: Characteristics of patients used for confirmation of gene signature obtained by microarray gene expression profiling	223
Table 5.3: Gene expression profiling in different cell types/ cancers after KO of *EZH2*	227
Table 6.1: Reaction mix for treatment of plasmid DNA with SuperTaq to add dNTP "A" to the fragment	241
Table 6.2: Ligation mixture for three fragments (NruI-XbaI, XbaI-NotI&NruI-NotI) of pCMV-SPORT6	243

List of Abbreviations

ASXL1	Additional Sex Combs Like 1
AML	Acute Myeloid Leukaemia
BCL2	B-cell CLL/Lymphoma 2
BSA	Bovine serum albumin
Chas	Chromosome Analysis Suite
CMML	Chronic Myelomonocytic Leukaemia
CN (LOH)	Copy Neutral Loss of Heterozygosity
CNVs	Copy Number Variations
CB	Chromobox proteins
CFUS	Colony Forming Units in Spleen
DLBCL	Diffuse Large B-cell Lymphoma
DMSO	Dimethyl sulfoxide
DNMT3A	DNA methyltransferase 3A
EZH2	Enhancer of Zeste Homologue 2
EED	Embryonic ectoderm development
EVI-1	Ecotropic Viral Integration site 1
FAB	French-American British
FANCL	Fanconi Anaemia,Complementation Group L
FBRS	probable fibrinosin-1
FCS	Fetal Calf serum
FITC	Fluorescein isothiocyanate
FLT1	Fms-Related Tyrosine Kinase 1
FLT3	Fms-Related Tyrosine Kinase 3
HOXA10	Homeobox A10
HOTAIR	HOX Transcript Antisense RNA
HSCs	Haematopoietic Stem cells
HMEC	Human Mammary Epithelial cells
HDFs	Human Diploid Fibroblasts

IFI16	Interferon, Gamma-Inducible Protein 16
IPSS	International Prognostic Scoring System
IPSS-R	International Prognostic Scoring System-R
INK	Inhibitor of cyclin-dependent kinase
IDH1/2	Isocitrate Dehydrogenase ½
Jmj	Jumonji
JHDM	Jumonji Histone Lysine Demethylase
JARID2	Jumonji, ARID domain-containing protein 2
KDM2B	Lysine-specific demethylase 2B
KO	Knock out
L3MBTL2	Lethal (3) malignant brain tumour-like protein
MAPKAP	MAPK-activated protein kinases
MID	Multiplex Identifier
MDS	Myelodysplastic Syndrome
MDS-U	Myelodysplastic syndrome-unclassified
MLLT10	Mixed Lineage Leukaemia T10
MTT	Tetrazoliumsalt[3-(4,5-Dimethylthiazol-2-yl) 2,5diphenyltetrazolium- bromide]
MPC	Magnetic Particle Concentrator
MPN	Myeloproliferative neoplasm
NPM1	Nucleophosmin
NUP98	Nuclear pore complex protein
OS	Overall Survival
PAG1	Phosphoprotein Associated With Glycosphingolipid Microdomains 1
PcG	Polycomb group
PC	Polycomb
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PML	Promyelocytic Leukaemia
PHC	Polyhomeotic
PSC	Posterior Sex combs
PHF1	PHD Finger protein 1

PREs	Polycomb Responsive Elements
PTEN	Phosphatase and TENsin homolog
PI	Propidium Iodide
PBS	Phosphate Buffered Saline
RA	Retionic acid
RA	Refractory Anaemia
RARS	Refractory Anaemia with Ringed Sideroblasts
RAEB	Refractory Anaemia with Excess Blasts
RAEB-T	Refractory Anaemia with Excess Blasts in Transformation
RRM1	Ribonucleotide Reductase
RUNX1	Runt-Related Transcription Factor 1
RYBP1	Ring and YY1-binding protein
RPMI	Roswell Park Memorial Institute Medium
RFS	Relapse Free Survival
SNP	Single Nucleotide Polymorphism
SUZ12	Suppressor of Zeste12
STAT3	Signal Transducer And Activator Of Transcription 3
SET	Su (var), Enhancer of zeste, Trithorax
S-AdoHcy	S-adenosylhomocysteine
S-AdoMet	S-adenosylmethionine
SRSF2	Serine/Arginine-Rich Splicing Factor 2
TET2	Ten-Eleven Translocation-2
tAML	Therapy Related AML
TSLP	Thymic Stromal Lymphopoietin
U2AF	U2 auxilliary Factor
WHO	World Health Organisation
X-Gal	5-bromo-4- chloro-3-indolyl-β-D-Galactoside
5-hmc	5 hydroxymethylcytosine
7-AAD	7-Aminoactinomycin D

1 Chapter: Introduction

1.1 Myelodysplastic syndrome (MDS)

1.1.1 Disease classification and prognostic scoring system

MDS are characterized by abnormal clonal haematopoiesis, aberrant differentiation and increased apoptosis leading to peripheral blood cytopenias despite of a hyperplastic bone marrow (Acquaviva et al., 2010). Percentage of myeloblasts is an important parameter in prognosis, classification of MDS and is defined in terms of several nuclear and cytoplasmic characteristics including a high nuclear/cytoplasmic ratio, visibility of nucleoli and usually fine nuclear chromatin. The nuclear shape of myeoblasts is variable and the cytoplasmic characteristics include variable basophilia, with or without granules or Auer rods and absence of Golgi zone (except in AML linked to t (8; 21) (Mufti et al., 2008). The median age of patients with MDS is 71 years (Meers, 2014) with treatments being tailored for patients because of the clinical heterogeneity of the disorder making accurate classification and prediction of prognosis an essential component of patient care. Classification system by the World Health Organization (WHO) is (Table 1.1) currently in use for stratification of the patients is based on clinical features, peripheral blood/bone marrow findings, and cytogenetic analysis. This system has developed from the French-American British (FAB) system [1982] but differs from it in the following aspects: (1) WHO criterion for diagnosis of acute myeloid leukaemia is 20 % blast percentage (FAB: 30 %). The WHO system does not have Refractory Anaemia with Excess Blasts in Transformation (RAEB-T) category (2) segregation of refractory anaemia (RA) and Refractory Anaemia with Ringed Sideroblasts (RARS) into two categories based on the presence of multilineage (2 or 3 cell lineages) or unilineage (usually affecting the erythroid lineage) dysplasia (3) Dividing Refractory Anaemia with Excess Blasts (RAEB) into RAEBI (5%-9% marrow blasts) and RAEBII (10%-19% marrow blasts) and (4) categorising Chronic Myelomonocytic Leukaemia (CMML) into a new category

which has overlapping features of MDS and Myeloproliferative disorders (MPN) (Bennett et al., 2002).

MDS sub-type	Bone marrow	Peripheral blood	Transformation to AML	Incidence in patients
Refractory anaemia (RA)	Normal myeloid and megakaryocytic lineages, < 5 % blasts	No blasts	rare	20-30 %
Refractory anaemia with ring sideroblasts (RARS)	≥ 15 % of erythroid precursors with ring sideroblasts, erythroid dysplasia only, < 5 % blasts	Anaemia, no blasts	1-2 %	10-12 %
Refractory cytopenia with multilineage dysplasia (RCMD)	Dysplasia in ≥ 10 % of cells in ≥ 2 haematopoietic lineages, ± 15 % ring sideroblasts, < 5% blasts	Cytopenia(s), < 1×10^9/L monocytes	11%	24 %
Refractory anaemia with excess blasts- I (RAEB-I)	Unilineage or multilineage dysplasia, no Auer rods, 5-9 % blasts	Cytopenia(s) ≥ 2-4 % blasts, < 1×10^9/L monocytes	25%	40%
Refractory anaemia with excess blasts-II (RAEB-II)	Unilineage or multilineage dysplasia, Auer rods, ± 10-19 % blasts	Cytopenia(s), 5-19 % blasts, < 1×10^9/L monocytes	33%	
CMML,MDS/MPN, Myelodysplastic syndrome – unclassified (MDS-U)	Unilineage dysplasia or no dysplasia but characteristic MDS cytogenetics, < 5 % blasts, CMML	variable	15-30 %	Variable
MDS associated with isolated del(5q):	Unilineage erythroid dysplasia, isolated del (5q), < 5 % blasts	anaemia, platelet levels normal or increased	Good prognosis	5 %
Therapy-related MDS/AML (t-MDS/tAML)	Blood and bone marrow findings of one of the above diagnostic categories		Not applicable	Depends on history of exposure to cytotoxic drugs/radiation

Table 1.1: WHO classification of MDS into its sub types based on changes in bone marrow and peripheral blood. The table contains additional information on percentage of patients in each category progressing to AML and the incidence rate of each category in MDS. [This information has been collected from multiple research papers (Bennett et al., 2002; Schanz et al., 2012) etc]

In addition to classification, assessing the prognosis of the patients and disease diagnosis is critical and is done by the International Prognostic Scoring System (IPSS) which considers Cytogenetic abnormalities, blasts ratio and cytopenia for this assessment (Greenberg et al., 1997) [Table 1.2]. The IPSS was revised in 2012 to IPSS-R by addition of novel components to refine the old scoring system, however bone marrow cytogenetics, blast percentage, and cytopenias remained the basis of the new system. Some of the changes were classification of the cytogenetic prognostic subgroups into five instead of three categories, addition of a number of less common cytogenetic subsets, splitting the low marrow blast percentage value, and depth of cytopenias. Patient age, performance status, serum ferritin and lactate dehydrogenase were significant additive features for survival but not for transformation to acute myeloid leukaemia. This system comprehensively integrated the numerous known clinical features into a method analyzing MDS patient prognosis more precisely than the initial IPSS (Greenberg et al., 2012) [Table 3].

a)

Score					
Prognostic variable	0	0.5	1	1.5	2
Bone Marrow Blast	<5	5 to 10	--	11 to 20	21 to 30
Karyotype	Good	Intermediate	Poor		
Cytopenia	0 to 1	2 to 3			

b)

Prognosis Score	IPSS Subgroup	Median AML transformation (yrs)	Median Survival (yrs)
0	Low	9.4	5.7
0.5-1.0	Intermediate-1	3.3	3.5
1.5-2.0	Intermediate-2	1.1	1.2
>2.5	High	0.2	0.4

Table 1.2: (a) IPSS assignment of score based on bone marrow blasts, karyotype and cytopenia.(b) IPSS risk groups by sum of scores and there link to transformation to Acute Myeloid Leukaemia (AML)and median survival in patients.

a)

Score							
Prognostic variable	0	0.5	1	1.5	2	3	4
Bone Marrow Blast	<2	--	2 - <5	--	5 to10	>10	--
Karyotype †	Very good		Good		Intermediate	poor	very poor
Haemoglobin (g/dL)	≥10	--	8 -<10	<8	--	--	--
Absolute Neutrophil Count (x109/L)	≥0.8	<0.8	--	--	--	--	--
Platelet count	≥100	50-<100	<50	--	--	--	--

† Very Good Karyotype: del(11q), -Y; Good karyotype: 46XX, 46XY, del(5q), del(20q); poor karyotype: complex (≥ 3 unrelated abnormalities) or chromosome 7 abnormalities; intermediate karyotype: karyotypes not defined as good or poor.

b)

Risk Category	Overall Score	Median AML transformation (yrs)	Median Survival (yrs)
Very Low	≤1.5	15	8.8
Low	2 to 3	10.8	5.3
Intermediate	3.5-4.5	3.2	3
High	5 to 6	1.4	1.6
Very High	>6	0.73	0.8

Table 1.3: (a) IPSS-R assignment of score based on bone marrow blasts, karyotype and cytopenia. (b) IPSS-R risk groups by sum of scores and there link to transformation to Acute Myeloid Leukaemia (AML) and median survival in patients.

1.1.2 Implication of Cytogenetics in MDS

Cytogenetic abnormalities are major determinants in the pathogenesis, diagnosis, prognosis of patients with MDS (as mentioned above) and are increasingly becoming the basis for selection of drugs in individual patients with MDS. Cytogenetic abnormalities are identified in the bone marrow of approximately 50% of patients diagnosed with primary MDS and 80% of patients with therapy-related MDS. These abnormalities can be characterized as being balanced (translocations, inversions & insertions) or unbalanced (gain or loss of chromosomal material). Unbalanced chromosomal abnormalities are particularly common in MDS and the nature of these abnormalities has a profound impact on disease pathogenesis, prognosis and response to treatment (Giagounidis et al., 2006). Amplifications or deletions are called Copy Number Variations (CNVs) and are caused due to gain or loss in genomic material due to error in DNA replication producing either duplications, removal of genetic sequences or unequal crossing-over that occurs during meiosis between misaligned homologous chromosomes. Gain leading to duplication of oncogenes have been associated with development of haematological malignancies, i.e. chr 11q 23 Mixed Lineage Leukaemia (*MLL*).Routine metaphase cytogenetics facilitates detection of chromosomal aberrations and mapping of recurrent lesions to narrow down on the minimally affected regions. However it is time consuming and the yield depends on the sensitivity (proportion of clonal cells in a sample) and resolution (size of the lesion) in the patient samples. In addition, requirement of the technique for dividing cells to obtain chromosomal spread is a limitation since it determines the proportion of abnormal cells in the dividing progenitor pool and may not reflect the actual percentage of abnormal cells. The biggest disadvantage of metaphase cytogenetics is its inability to identify Copy Neutral Loss of Heterozygosity [CN (LOH)] wherein the chromosomal number remains unaltered. These limitations of metaphase cytogenetics are overcome by Single Nucleotide Polymorphism (SNP) arrays which rely on

oligonucelotide based probes which correspond to allelic variants of the selected SNPs. The density of the probes (10K, 50K, 250K, 906K) determine the degree of resolution of the arrays. SNP-A have high resolution, are more sensitive (2-30%) and can therefore detect even low level cryptic chromosomal aberrations. It is not limited by dependence on dividing cells and can identify CN (LOH) (Maciejewski et al., 2009) depending upon the hybridisation of patient DNA to both probe variants (heterozygous) or to a single probe indicating homozygosity. CN (LOH) have been associated with myeloid malignancies i.e. chr 4q Ten-Eleven Translocation-2 (*TET2*) and are associated with duplication of mutations in oncogenic or tumour suppressor genes. A systematic documentation of the abnormalities have been reported (Bejar et al, 2014; Haase et al., 2007) (Mohamedali et al., 2007; Tiu et al., 2011) and the most frequent cytogenetic abnormalities along with their frequency of incidence and prognostic impact are detailed in table 1.4.

Chromosomal abnormalities	Frequency (%)	Prognosis
Trisomy 8	5 to 8	Good
Trisomy 21	2.7	Good
del 5q	15	Good
del 20q	2 to 5	Good
del Y	0.8	Good
del 7/7q-	5 to 10	Intermediate to poor
del 17p	10	Poor
Complex (≥3 abnormalities)	13.4	Poor

Table 1.4: Common cytogenetic abnormalities and their frequency of incidence in MDS and their impact on prognosis (Bejar et al, 2014; Haase et al., 2007)

1.1.3 Mutational analysis in MDS

Mutational analyses of patients with MDS have identified gene mutations recurrently associated with this disease. Mutational status of genes is not currently used in prognostic scoring systems but is likely to be key factor that govern clinical phenotypes and overall survival. High throughput DNA sequencing has revolutionised and vastly contributed to our

understandingof the clonal evolution, clone sizes and allele burden of the individual gene mutations as well as the interplay of different overlapping mutations to generate entire landscape of genetic leisions and their association to MDS(Bejar et al, 2014; Bejar et al., 2011; Haferlach et al., 2014). Some of the common mutations and their clinical significance are summarised in table 1.5.

Mutations	Frequency (%)	Overlap	Exclusive	Prognostic significance	Pathogenesis
ASXL	10 to 15 %	EZH2,RUNX1,STAG2	SF3B1, DNMT3A	Unknown	Unknown
DNMT3A	22%	SF3B1	ASXL1, SRSF2	Poor	Loss of enzymatic activity
EZH2	6%	LUC7L2	SF3B1	Poor	Loss of histone methyltransferase activity
IDH1/IDH2	rare	SRSF2	TET2	Poor	Alters function to convert α-ketoglutarate to 2-hydroxyglutarate
JAK2	5% in RA	SF3B1		Unknown	Constitutive activation of tyrosine
NPM1	4%	EZH2, LUC7L2	SF3B1	Poor	Disruption of nucleolar localization signal leading to cytoplasmic redistribution of the protein
NRAS/KRAS	10%	STAG2	SF3B1	Poor	Activation of serine/threonine kinase
RUNX1	15 to 20 %	SRSF2, ASXL1	SF3B1	Poor	Disrupt DNA and protein binding domains
SF3B1	82 % in RARS	DNMT3A, JAK2	TP53	Good	Part of splicing machinery; loss of protein function
TET2	20%	SRSF2,ZRSF2	IDH1/2	No prognostic value	Suppress the catalytic activity of the enzyme, low 5-hmC levels
TP53	5 to 10 %	NPM1, LUC7L2	SF3B1	Poor	Loss of function of p53 tumour supressor activity

Table 1.5: Summary of genes commonly mutated in MDS. The frequency of mutations are indicated in percentages along with mutational overlap and exclusivity between these genes in patients with MDS (Bains et al., 2011; Bejar et al, 2014; Bejar et al., 2011; Haferlach et al., 2014; Mian et al., 2013).

1.2 Epigenetic aberrations in MDS

Epigenetic regulation of transcription (Fig: 1.1) has attracted particular attention, because of the relatively high frequency of mutations in genes like Isocitrate Dehydrogenase 1 (*IDH1/2),* *TET2,* Additional Sex Combs Like 1*(ASXL1)* and *EZH2* (Table: 1.5) (Itzykson and Fenaux, 2014).

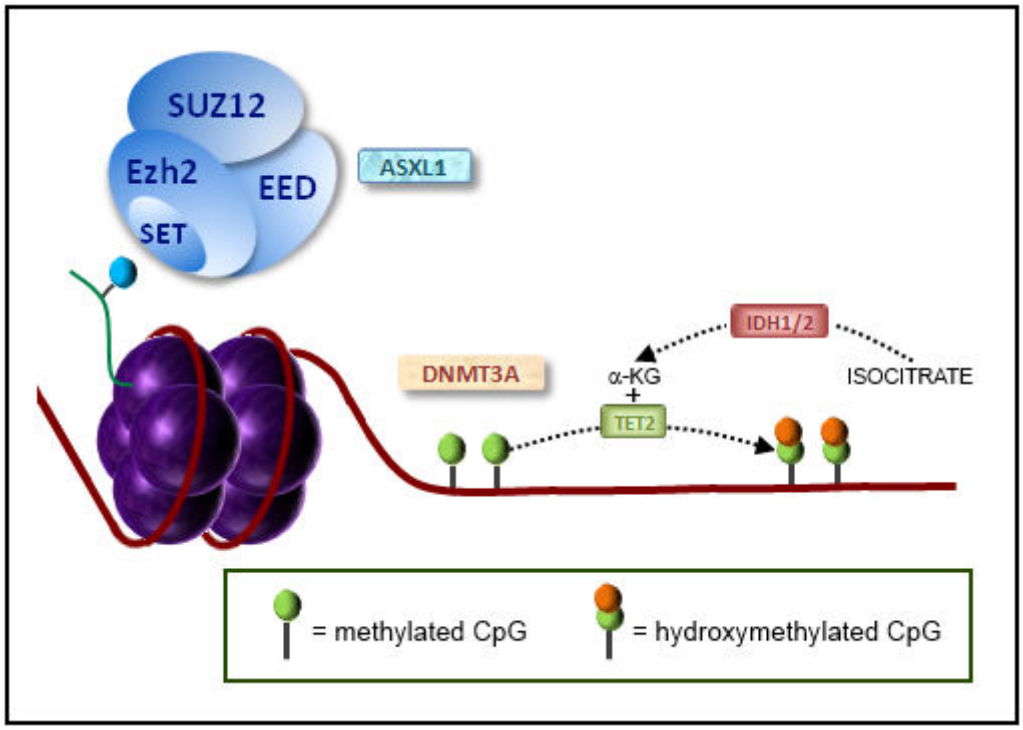


Figure 1.1: Epigenetic regulation of transcription.PRC2 complex (blue) causes trimethylation of H3K27 residues. These marks assist *DNMT3A* localization leading to CpG methylation at target genes. *IDH1/IDH2* are responsible for conversion of isocitrate to α-ketoglutarate which is utilized by *TET2* for hydroxymethylation of methylated CpG resulting in DNA demetylation. *ASXL1* modulates recruitment of PRC2 to H3K27 and mutations in *ASXL1* reduce methylation of H3K27. (Histone octamer is represented in purple while DNA is indicated by a brown string).

Heterozygous mutations of *IDH1/2* result in accumulation of 2-hydroxyglutarate instead of conversion of isocitrate to *α*-ketoglutarate. Elevated 2-hydroxyglutarate levels inhibit the function of *TET2* (Xu et al., 2011) which is responsible for the conversion of 5 methylcytosine to 5 hydroxymethylcytosine (5-hmC) (Fig: 1.1) which is an intermediary in DNA demethylation. This leads to DNA demethylation via specific molecular pathways or

passive demethylation (Itzykson et al, 2014), (Guo et al., 2011).In addition to an inhibitory effect on *TET2,* altered 2-hydroxyglutarate/*α*-ketoglutarate levels impede the demethylation activity of the Jumonji Histone lysine Demethylases *(JHDM)* (Xu et al., 2011).

Mutations in *TET2* suppress the catalytic activity of the enzyme leading to low levels of 5-hmC (Ko et al., 2010).*TET2* mutations are mutually exclusive of *IDH1*/2 mutations and share similar downstream effect (Verena et al, 2010), (Metzeler et al., 2011), (Mohamedali et al., 2009) suggesting development of leukaemia through common mechanisms involving suppression of *TET2* function.

In addition to *TET2* and *IDH1/2*, *ASXL1* has recently emerged as an important epigenetic modulator associated with MDS. The role of *ASXL1* is distinct from that of *TET2* and *IDH1/2*. Mutations of *ASXL1* reduce trimethylation of H3K27me3 and cause over expression of leukaemia promoting genes. It is believed to inhibit the recruitment of polycomb repressive complex 2 (PRC2) to H3K27 (Fig: 1.1) thus reducing trimethylation at this position (Conway O'Brien et al., 2014). This discovery has shed new light on the involvement of histone modifying enzymes in MDS.

1.2.1 Post translational histone modifications

In addition to changes in DNA methylation, epigenetic regulation also occurs at the chromatin level. Nucleosomes are the basic unit of chromatin consisting of a histone octamer (two each of H2A, H2B, H3 and H4) wrapped by 146 bp of double stranded DNA and kept tightly packed by the linker histone H1.These histones have an overhanging N-terminal tail which can be subjected to an array of post-translational modifications like phosphorylation of serine and threonine residues, methylation of arginine and lysine, acetylation of lysine or ubiquitination of lysine.

Histone methylation involves addition of methyl groups to the side chains of arginine (mono- or di-) or lysine (mono-, di-, tri-) creating binding sites either for a transcriptional activator or repressor protein. This association governs the ultimate function of the methyl mark at that location. Lysine methylation marks on H3 can be either repressive (H3K9, H3K27) or might activate transcription (H3K36, H3K79, and H3K4) [Fig 1.2].**This study focuses on lysine methylation at position 27 on histone 3 which is caused by histone methyltransfearse *EZH2* found to be mutated in MDS (detailed later).**

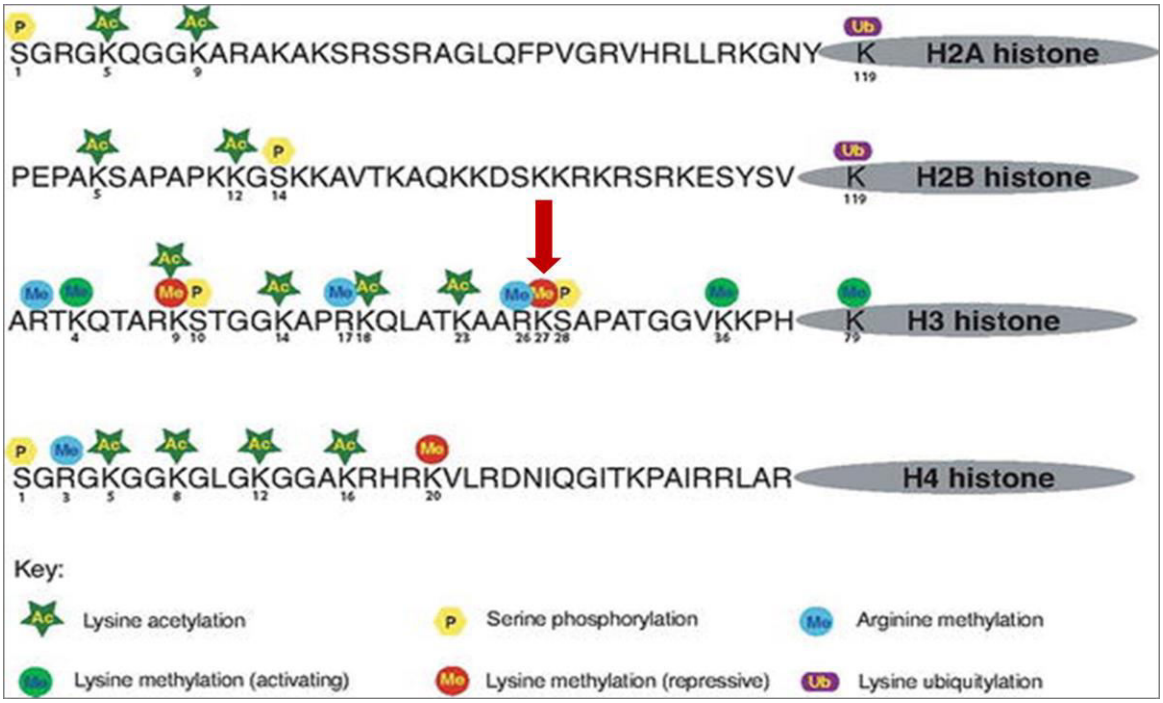


Figure 1.2: Post translational modifications of overhanging N-terminal of histone tails.Serine phosphorylation marks are illustrated in yellow, while lysine acetylation and arginine methylation marks are indicated in green and blue respectively. The focus of my thesis is on H3K27me3 marks (indicated by red arrow) which are caused by *EZH2* histone methyltransferase which is mutated in MDS. Methylation of lysine provides binding sites or acts as sign posts for recruitment of proteins responsible for transcriptional repression (Gagnidze et al., 2013).

1.3 Polycomb proteins and polycomb repressive complexes

PcG proteins were first discovered in Drosophila Melanogaster and governed early embryonic development. They modulate the repression of the homeotic genes needed for embryogenesis via epigenetic mechanisms. This repression is opposed by the trithorax group proteins and the two groups work in tandem to maintain normal homeotic gene expression. In vertebrates, PcG homolog proteins are involved in somatic cell differentiation; especially in adult human haematopoiesis.

1.3.1 Classification of PcG proteins

PcG proteins form multimeric complexes called polycomb repressive complexes (PRC1 and PRC2) based on their interactions with other PcG proteins. The canonical PRC1 complex comprises of 4 core subunits; Polycomb (*PC*) or chromobox proteins (*CB*), Polyhomeotic (*PHC*) proteins, the posterior sex combs (*PSC*) proteins and the RING proteins [Fig: 1.3 (A1)] (Gao et al., 2012). ring proteins especially *RING1B* can bind to other proteins like Ring and YY1-binding protein (*RYBP*) (Gao et al., 2012), lysine-specific demethylase 2B (*KDM2B*) (Farcas et al., 2012; He et al., 2013), lethal (3) malignant brain tumour-like protein (*L3MBTL2*) (Trojer et al., 2011) and probable fibrinosin-1 (*FBRS*) to form non-cannonical PRC1 complexes. *KDM2B* encodes a DNA binding domain that recognizes non-methylated CpG dinucleotides in the DNA sequence. It interacts with a variant PRC1 complex consisting of *RYBP, RING* and *PSC* proteins to direct PRC1 to CpG islands where PRC1 silences polycomb target genes (Farcas et al., 2012) [Fig 1.3 (A3)].

On the other hand, KO of *L3MBTL2* [Fig: 1.3 (A4)] was associated with reduced H2AK119Ub levels but it does not possess an E3 ligase activity and was therefore implicated in recruitment of the *RING* proteins to H2AK119. However, immunoprecipitation studies demonstrated that *L3MBTL2* did not co-localize with H3K27me3 marks and hence

the recruitment of ring proteins by *L3MBTL2* was considered independent of the presence of H3K27me3 marks and its mechanism of PRC1 recruitment needs further clarification (Trojer et al., 2011).Unlike *L3MBTL2, RYBP* and *RYBP-FBRS* [Fig 1.3 (A2,5)] containing non-canonical PRC1 complexes have a stimulatory effect on the E3 ligase activity of the ring proteins and do not participate in PRC1 recruitment (Gao et al., 2012).

PRC2 contains three core subunits i.e. Enhancer of Zeste (*EZH2*), Suppressor of Zeste (*SUZ12*) and Embryonic Ectoderm Development (*EED*) [Fig: 1.3 (B1)] and associates with interchangeable components like PHD Finger protein 1 (*PHF1*) and Jumonji, ARID domain-containing protein 2 (*JARID2*) [Fig: 1.3 (B2,3)].The role of *JARID2* in non-canonical PRC2 complex is controversial wherein two studies suggest that incorporation of *JARID2* reduces PRC2 catalytic activity towards H3K27me3 (Shen et al., 2009), (Peng et al., 2009) while others demonstrate that *JARID2* forms a stable complex with PRC2, promoting its binding to PcG-target genes (Pasini et al., 2010),(Li et al., 2010). Studies on human and mouse embryonic stem cells demonstrate that *JARID2* interacts with PRC2 via *SUZ12* and promotes its recruitment to target genes and simultaneously inhibits the histone methyltransferase activity of *EZH2*, suggesting that it acts as a "molecular rheostat" fine tuning the function of PRC2 (Peng et al., 2009).

Unlike *JARID2*, *PHF1* is involved in Tudor-dependent stabilization and extension of PRC2 and H3K27me3 marks to H3K36me3 marks promoting the intrusion of PRC2 into active chromatin regions (Cai et al., 2013).Loss of *PHF1* decreases H3K27me3 at PcG-target genes (Cao et al., 2008; Walker et al., 2010), by replacing with H3K27me2 (Nekrasov et al., 2007).

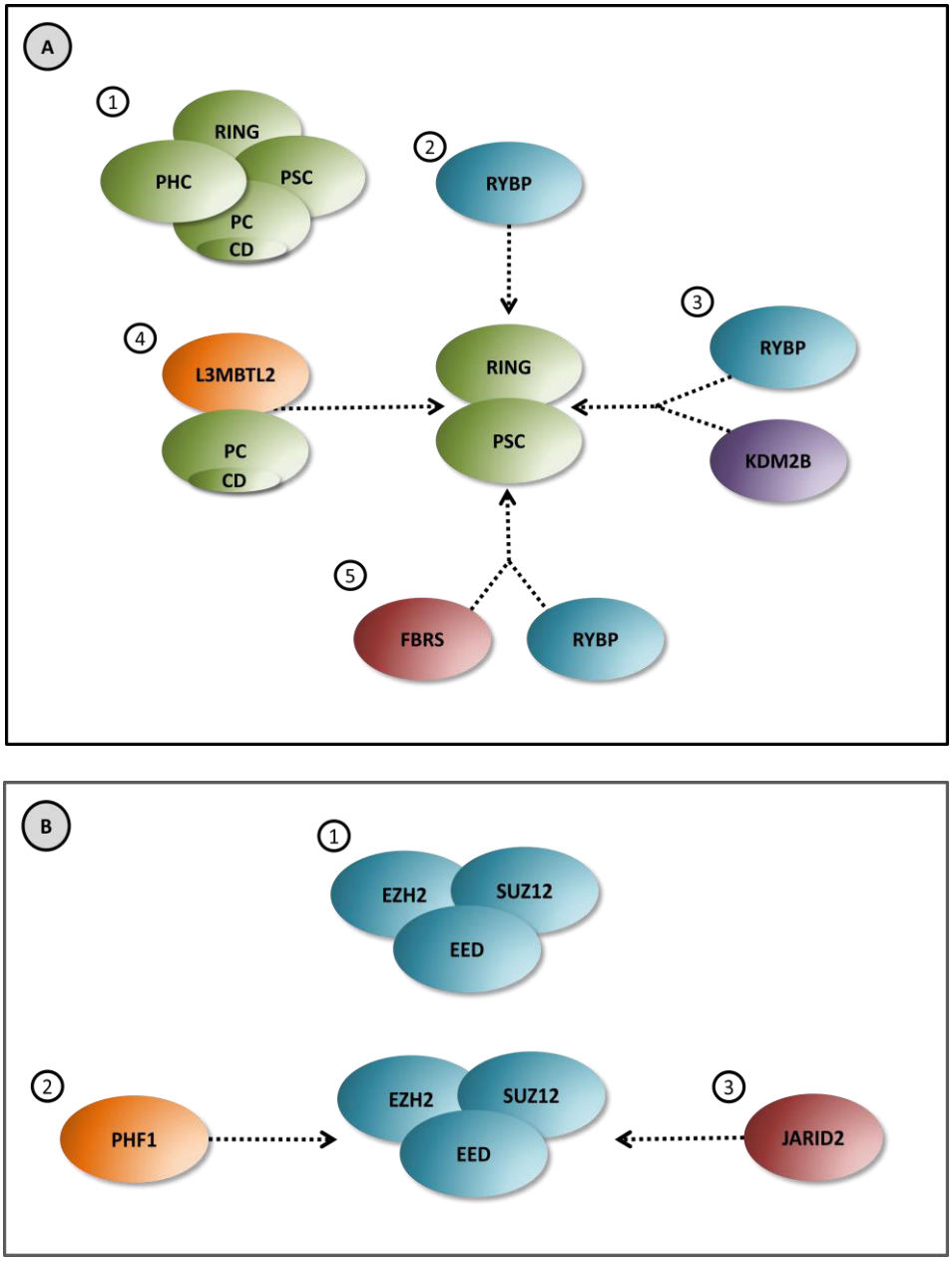


Figure 1.3: Canonical (1) and non canonical (2, 3, 4, 5) PRC1 and PRC2 complexes.(A) The RING and posterior sex comb (PSC) proteins form a back bone on which the non canonical complexes are assembled. Polycomb proteins can interact with only *RYBP* (A2), with *KDM2B* and *RYBP* (A3), with *L3MBTL2* (A4) and *FBRS*, *RYBP* (A5) and to form non canonical complexes with diverse mechanism of action and function (See text for details).

(B) *PHF1* and *JARID2* interact with the core components *(EZH2, SUZ12* and *EED)* of PRC2 to form non canonical PRC2 complexes (See text for details).

The individual members of the canonical complexes along with their nomenclature (gene name) and chromosomal location are illustrated in Table 1.6.The multiplicity of the parallel binding patterns of PcG proteins is perplexing, but it may reflect stages in the process of gene silencing.

a)

PcG	Gene Symbol	Chr location
Chromobox (CB)	**CBX2**/M33/PC	17q25.3
	CBX4/PC2/hPC2	17q25.3
	CBX6	22q13.1
	CBX7	22q13.1
	CBX8/PC3/HPC3	17q25.3
RING (RING)	**RING1A**/RING1/RNF1	6p21.32
	RING1B/RING2/RNF2	1q25.3
Polyhomeotic (PH)	HPH1/**PHC1**/EDR1/RAE28	12p13.31
	HPH2/**PHC2**/EDR2/PH2	1p35.1
	HPH3/**PHC3**/EDR3	3q26.2
Posterior Sex comb (PSC)	PCGF1/**NSPC1**/RNF68	2p13.1
	PCGF2/**MEL18**/RNF110	17q12
	PCGF3/RNF3/RNF3A	4p16.3
	PCGF4/**BMI1**/RNF51	10p12.2
	PCGF5/RNF159	10q23.32
	PCGF6/**MBLR**/RNF134	10q24.33

b)

PcG	Gene Symbol	Chr location
Enhancer of Zeste (EZH1/2)	EZH1	17q21.2
	EZH2	7q36.1
Embronic ectoderm enhancement (EED)	EED	11q14.2
Suppressor of Zeste 12 (SUZ12)	SUZ12	17q11.2

Table 1.6: a) Sub-units of the canonical Polycomb Repressive Complex 1 (names used in the text are indicated in bold)b) Sub-units of the canonical Polycomb Repressive Complex 2along with their chromosomal location.
(Reference: http://eprints.ucl.ac.uk/20210/1/20210.pdf)

1.3.2 Mechanism of action of Canonical PcG proteins

PRC2 contains the histone methyltransferase (HMT); *EZH2* that targets lysine at position 27 of histone 3 and can recruit up to three methyl (CH3) groups at this position causing transcriptional repression. *EZH2* contains the functional "SET [Su (var), Enhancerof zeste, Trithorax] domain" that is critical for its methyltransferase activity and disruption of this domain abolishes the function of PRC2. In addition to mediating extensive trimethylation of H3K27 at PcG target genes, PRC2 is also responsible for widespread dimethylation of H3K27 throughout the genome (Ferrari et al., 2014). In fact H3K27me2 mark accounts for

nearly 60% of all histone H3 in the genome and is probably accompanied by low levels of diffuse H3K27me3 (Peters et al., 2003). The methyltransferase activity of PRC2 depends on inputs from surrounding chromatin wherein the presence of H3K4me3, H3K36me2/3 decreases the catalytic activity of PRC2 (Schmitges et al., 2011), (Yuan et al., 2011) while a basal level of H3K27me2 is a prerequisite for stimulating the timely onset of targeted PcG-mediated repression, thus connecting the two H3K27 methylation states (Peters et al., 2003).Although *EZH2* is the prime histone methyltranferase,the interaction of *SUZ12* and *EED* with *EZH2* is essential for an increased methyltransferase activity of PRC2 (Ketel et al., 2005; Montgomery et al., 2005; Xie et al., 2014).

Apart from its main function of H3K27 trimethylation,PRC2 is also responsible for recruiting PRC1 to the target genes (Fig 1.4).The histone trimethylation marks (H3K27me3) caused by PRC2 are recognized by the *CBX* proteins which in turn directs PRC1 to specific regions of the chromatin (Bernstein et al., 2006).

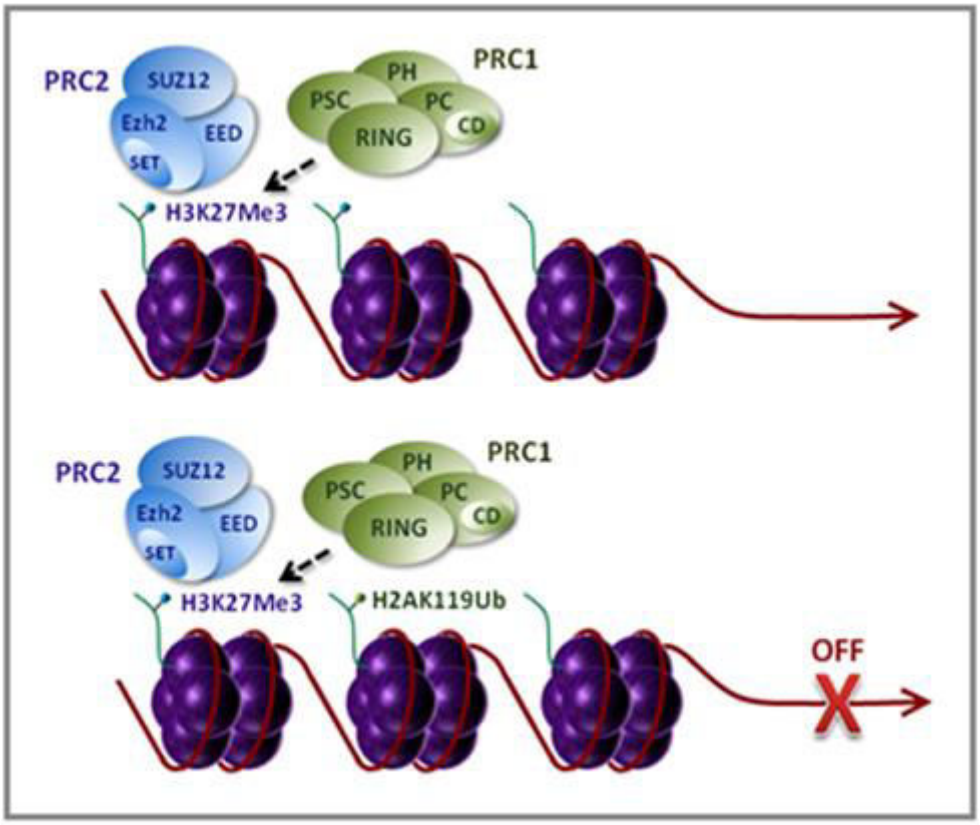


Figure 1.4: Mechanism of action; PRC1 and PRC2.H3K27me3 marks caused by PRC2 (blue complex) assist localization of PRC1 (green complex) to its target genes. PRC1 is responsible for H2AK119Ub resulting in transcriptional repression (indicated by red "X"). Histone octamer is represented in purple and DNA is indicated by brown string.

Once PRC1 is recruited, its subunitcauses ubiquitination of histone H2A at lysine 119 (H2AK119Ub) (Cao et al., 2002).The large size of the ubiquitin groups interfere either directly with the other nucleosome dynamics to assist repression or employ other yet unknown effector proteins to H2AK119Ub which in turn repress transcription of downstream genes (de Napoles et al., 2004). In addition to transcriptional repression, *RING1B*-PRC1 also restrains RNA polymerase II at the promoters of target genes. Inactivation of the RING1 proteins result in global loss of the H2AK119Ub mark and release of the RNA pol II enzyme (Stock et al., 2007).RNA polymerase II is also targeted by *EZH1* (paralogue of *EZH2*) and depletion of *EZH1* is associated with global loss of pol II occupancy of promoters in skeletal muscle cells (Mousavi et al., 2012).

PcG proteins can act as either tumour suppressors or oncogenes depending upon target genes (Federico et al., 2009; Scott et al., 2007), tumour microenvironment (stroma) etc. This 'dual biological function' has been identified for both the canonical and non-canonical PRC1.The varied association of PRC core components with other proteins to form non canonical complexes indicate a heterogeneous route of transcriptional repression.

Apart from variations in binding partners directing the PRC complexes to different targets, the complexes also differ in their target localisations in relation to one another. Wang et al (2010) demonstrated that PRC1 and PRC2 may not always co-localize at the same target genes and additional elements such as the Polycomb Responsive Elements (PREs), might be required to recruit the PRC1 complex. PRE sequences are present in the target genes which can be directly recognized by PRC1 in the absence of PRC2 (Wang et al., 2010).Non coding RNAs like HOX Transcript Antisense RNA (*HOTAIR*) (Chiyomaru et al., 2014) and Kcnq1ot1(Pandey et al., 2008) also aid in directing PRC complexes to their targets. Studies on protein-protein interactions suggest a high occurrence of positively charged amino acids (Arginine/Lysine) in the interacting domains of PcG proteins especially *CBX2* (M33 in mice) to be partly responsible for interacting with and inhibiting the negatively charged activation domains of the transcriptional machinery (Grau et al., 2011).Signalling pathways resulting in phosphorylation of the PcG proteins also impact their interaction with their binding partners and target genes (Kolybaba and Classen, 2014). The effects of phosphorylation on individual PcG sub-units are discussed later. This PRC1 independence suggests that H3K27me3 is not specifically required for transcriptional repression in all instances.

1.3.3 PcG and DNA methylation

DNA methylation involves addition of methyl groups to any of the four DNA bases. In mammals, only the 5^{th} position of the cytosine bases undergo methylation (Li, 2002). This

methylation is catalyzed by DNA methyltransferase (*DNMT*) enzymes. De novo methylation is initiated by *DNMT3A & 3B* while *DNMT1* has a preference for hemimethylated regions and hence involved only in maintenance (Jeltsch, 2006). However, this demarcation in function is not very clear. The H3K27me3 mark caused by PRC2 functions as a sign-posts for the recruitment of DNA methyltransferase (*DNMT*). The SET domain of *EZH2* can bind with *DNMT3A* (Fig 1.5), affecting DNA methylation. *DNMT3A* in turn is responsible for further "deep silencing" of the target regions. Vire et al (2006) showed that loss of *EZH2* causes a reduction in the H3K27me3 levels and a subsequent decrease in the DNA methylation. In the absence of *EZH2* RNA poly II is recruited to the promoters of the target genes resulting in activation (Vire et al., 2006).

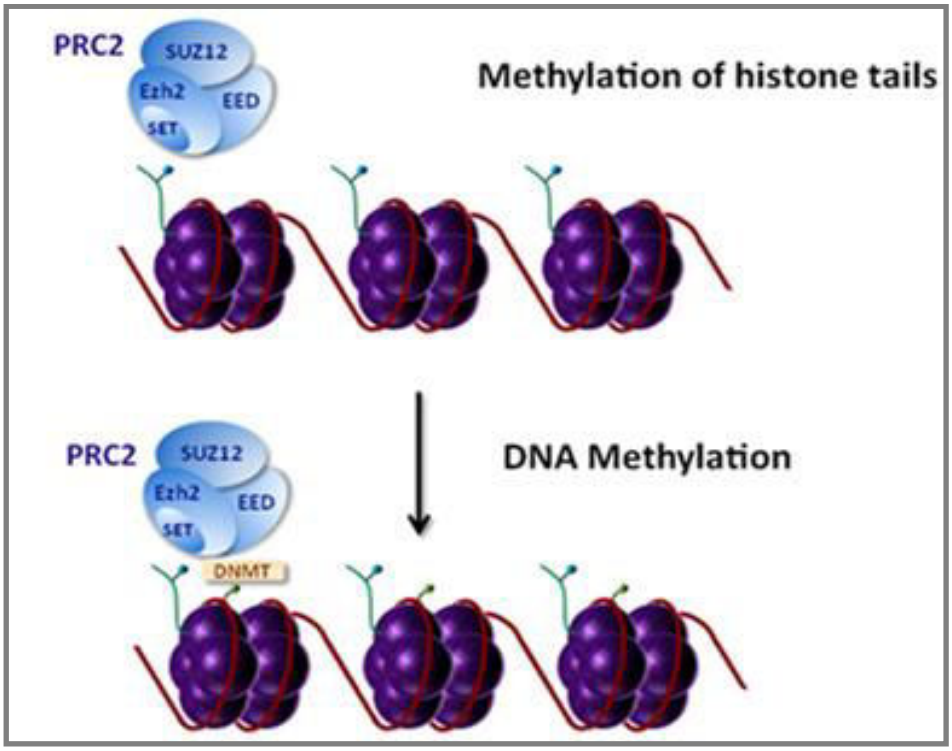


Figure 1.5: Interaction of PRC2 and *DNMT3A*.PRC2 (blue complex) recognises H3K27 sites from the histone octamer (purple) and the SET domain of *EZH2* histone methyltransferase cause H3K27me3 marks that act as sign posts for recruitment of *DNMT3A* (cream). The SET domain of *EZH2* interacts with *DNMT3A* to localise it to the target genes where it leads to DNA methylation and transcriptional repression.

Epigenome profiling indicates that high levels of DNA methylation and H3K27me3 are rarely found at the same loci in mammalian genomes, suggesting the presence of one mark is antagonistic to other (Meissner et al., 2008).The removal of DNA methylation has profound influence on the distribution of the H3K27me3 mark throughout the genome (Brinkman et al., 2012; Kondo et al., 2008; Lindroth et al., 2008; Reddington et al., 2013). And results in accumulation of the PRC2 and H3K27me3 at genomic locations that were previously DNA methylated (Reddington et al., 2013), suggesting that DNA methylation is capable of attenuating PRC2 targeting to chromatin. However, reduction in PRC2 cause only modest changes in DNA methylation (Hagarman et al., 2013), suggesting that H3K27me3 mark does not have a similar reciprocal effect on DNA methylation.

Another factor influencing the H3K27me3 marks is their proximity to existing H3K27me3 which determines whether a gene gets targeted by PRC recruitment following loss of DNA methylation. DNA methylation surrounding PRC2 bound targets prevent the spread of H3K27me3 into adjacent regions (Hagarman et al., 2013; Simon et al, 2013). The correlation of histone and DNA methylation is convoluted in cancer cells compared to normal cells due to elevated expression (hypermethylation of the genome as been observed in some cancers) and the presence of mutations in both *EZH2* and *DNMT3A* genes adding a further layer of complexity to their association (Reddington et al., 2014). Unlike PRC2, there is no direct link between DNA methylation and PRC1 but since the canonical PRC1 depends on PRC2 for its recruitment, reduction in H3K27me3 marks as a consequence of DNA methylation might impact PRC1 recruitment.

In vivo studies on Haematopoietic Stem cells (HSCs) functional potential and DNA methylation in old versus young mice bone marrow has revealed slight high degree of global DNA hypermethylation in old HSCs.DNA methylation did not influence the function but affected ongoing developmental restriction. It concentrated on genomic regions associated

with open chromatin in lymphoid (B cell and T cell) and erythroid lineages, suggesting inhibition of differentiation towards these lineages. A concomitant reduced expression of the PRC2 core components at these loci was observed during aging, suggesting that PRC2 mediated repression diminished with HSC age and allowed DNA methylation machinery to access these sites (Beerman et al., 2013). Thus an enhancer as well as inhibitory effect exists between these two important silencing mechanisms.

1.3.4 Expression patterns of PcG proteins

Different sub-units of PRCs aggregate in the sub nuclear foci as polycomb bodies. Polycomb bodies vary in number and appearance according to the different stages of development and cell types. The number of these bodies is smaller in magnitude compared to the frequency of the target genes suggesting that either numerous genes associate with a single polycomb body or, the formation of polycomb bodies takes place only at a small subset of target genes (Kerppola, 2009). The aggregation of the polycomb bodies was studied by Gonalez et al (2014) using RNAi and high-resolution microscopy wherein they demonstrated that the SUMO protease "velo" was essential for the formation of stable nuclear aggregates of PcG proteins. SUMOylation of *CBX4* helps in targeting it to the relevant genes. A limited degree of SUMOylation prevents the coalescence of PRC1 complexes within the nucleus, whereas excess of SUMO solubilises *CBX4* away from its target chromatin (Gonzalez et al., 2014).

Expression patterns of different PcG proteins have been studied in purified populations of HSCs. *EZH2* and *BMI-1* are highly expressed in HSCs while *CBX2*, *MEL18* and *PHC1* were prominent in the mature bone marrow cells. The expression of these proteins varied based on the stage of maturation and the lineage of the marrow cells (Table: 1.7) (Lessard et al., 1998). Ezhkova et al (2009) examined *EZH2* expression and concluded that it was expressed in proliferating epidermal progenitor cells and controlled their proliferative potential by

suppressing the inhibitor of cyclin-dependent kinase [(*INK4A*)/alternative reading frame (ARF)] (*CDKN2A*) locus (Fig: 1.6). Stage-specific expression of *EZH2* suggests repression of undesirable differentiation programs to orchestrate establishment of other terminal differentiation programs (Ezhkova et al., 2009).

Subpopulations	CBX2	MEL18	BMI-1	PHC1	EZH2
$CD34^+$	Not detected	Not detected	Highest	Lowest	High
$CD34^+ CD45^- CD71^{hi}$	Not detected	Not detected	Reduced	Increased	Not detected
CD34+ CD45+ $CD71^{lo}$	Moderate	High	Reduced	High	Not detected
CD34-	High	Highest	Lowest	Highest	High

Table 1.7: Expression patterns of the PcG proteins in Haematopoietic Stem Cells (HSC). *EZH2* and *BMI-1* show similar trends in their expression patterns where both are high in the HSCs but their expression reduces as the cells differentiate into erythriod and granulocyte macrophage progenitors. While *EZH2* expression recovers in mature $CD34^-$cells, *BMI-1* expression remains low. Similar patterns of expression are also observed of *CBX2* and *MEL18* both of which are highly expressed in mature bone marrow $CD34^-$ cells but not detected in $CD34^+$ HSCs. Unlike *CBX2* and *MEL18*, *PHC1* expressions are high in all erythriod, granulocyte and macrophage progenitor cells but absent in $CD34^+$ cells.
[$CD34^-$= mature bone marrow cells, $CD34^+$= HSC, $CD34^+$ $CD45^-CD71^{hi}$ = enriched in erythriod progenitors, $CD34^+CD45^+CD71^{lo}$= enriched in granulocyte and macrophage progenitors] (Lessard et al., 1998)

Similar studies on maturational stages of B-cells in the germinal centre have also confirmed a cell and maturation dependant PcG protein expression (Raaphorst et al., 2000).

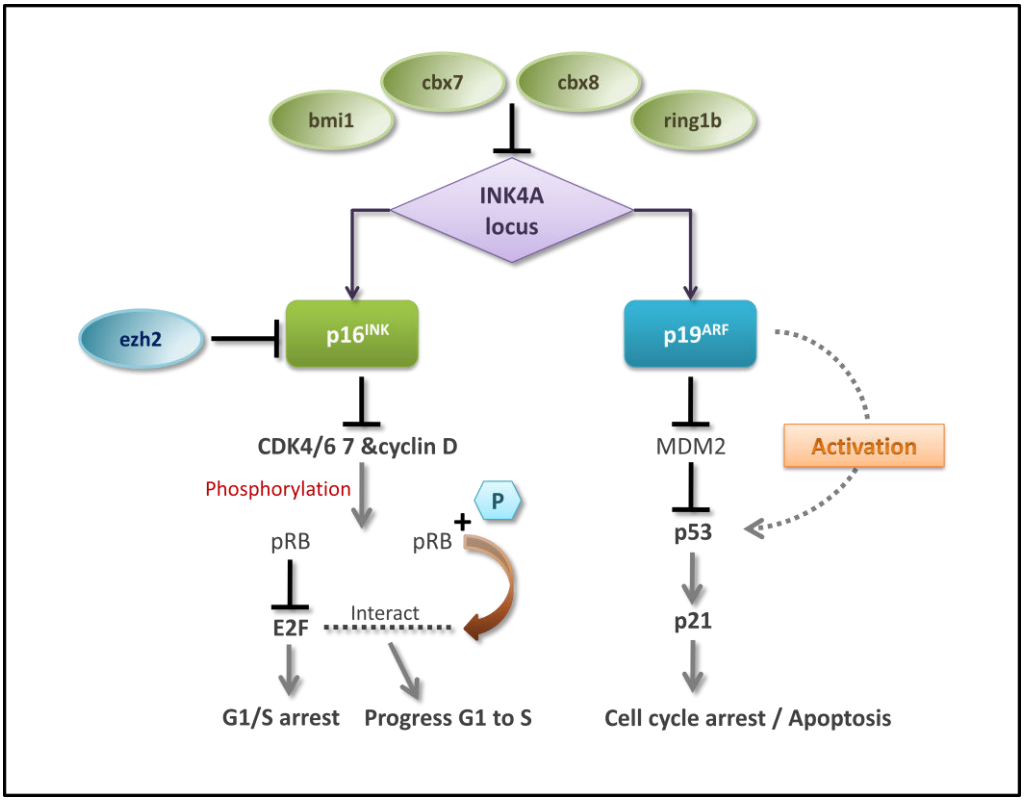


Figure 1.6: Illustrates inhibitory effect of PcG proteins on the *INK4A/ARF* locus. p16INK locus is involved in cell cycle regulation wherein it binds to and inhibits the CDK 4, 6 kinase activity. This prevents the interaction of CDK kinases with cyclin D and downstream phosphorylation of pRB. Lack of phosphorylated pRB stalls the E2F transcription factor and causes S phase cell cycle arrest. Both PRC1 (green) and PRC2 (blue) inhibit p16INK locus allowing formation of CDK 4, 6 - cyclin D complexes and progression of cells through S phase of the cell cycle. PRC1 (green) has an additional inhibitory effect on p19ARF resulting inup regulation of the ubiquitin ligase MDM2 which causes degradation of p53. Reduced levels of p53 are associated with decreased apoptosis and increased cell proliferation. In the absence of PRC1 mediated p53 inhibition, p19ARF downregulates MDM2 and enhances p53 expression inducing cell cycle arrest at both G1/S and G2/M check points. Therefore it can be concluded that both PRC1 and PRC2 mediated inhibition of INK4A/ARF locus enhances cell proliferation and prevents apoptosis.

1.4 Subunits of PRC2

1.4.1 Embryonic ectoderm enhancement protein

EED has four isoforms; *EED1, EED2, EED3* and *EED4*, all differing in the length of the N-terminal. PRC2 contains *EED1* and *EED2* which were believed to have an affinity for H1K26 while *EED3* and *EED4* showed affinity towards H3K27 (Kuzmichev et al., 2004). However; Montgomery et al (2007) showed that there was no binding preference of the *EED* isoforms towards the different histones. The study also confirmed that different states of histone

methylation like mono, di and tri methylation were not correlated to or were caused by any specific *EED* isoform (Montgomery et al., 2007).

EED contains five WD-40 domains that allow interaction with HDACs and *EZH2* (van der Vlag and Otte, 1999). The interaction of *EED* and *EZH2* is crucial for inducing mono, di and tri methylation of specifically H3K27, suggesting that *EED* must be responsible for the histone methyltransferase activity of *EZH2* and for the assembly, stability of PRC2 (Montgomery et al., 2005).

EED also binds to pre-existing H3K27me3 marks causing stimulation of the PRC2 methyltransferase activity for the surrounding unmodified H3K27. The WD 40 repeats of *EED* mediate the recognition and binding to the N-terminal domains of the histone tails. It also has a weak binding affinity for H1K26me3, H3K9me3 and H4K20me3 but requires interaction with other PRC2 subunits for a stable conformation (Margueron et al., 2009).

1.4.2 Suppressor of Zeste protein

SUZ12 has unique structural features; It contains two significant stretches of amino acids; zinc finger binding motif and the VEFS box (Pasini et al., 2004). It also has two promoters, an upstream promoter having a bidirectional function that produces the majority of the *SUZ12* mRNAs while the downstream promoter which works only in the forward direction (Kirmizis et al., 2003). *SUZ12* is crucial for maintaining the integrity of PRC2 by inhibiting the proteolytic degradation of *EZH2*. Lack of *SUZ12* causes complete absence of global di and trimethylated H3K27. However, the monomethylation levels remain unaffected. In the absence of *SUZ12*, *EZH2* levels decrease while *EED* levels remain unchanged (Pasini et al., 2004).*SUZ12* is also essential for the formation of H3K9me3 in differentiated cells. Absence of *SUZ12* causes 5-fold reduction in H3K9me3 levels and induces chromosomal instability (de la Cruz et al., 2007).

SUZ12 can be activated either by its translocation as seen in endometrial cancer or through the β-catenin mediated pathway. WNT/β catenin associate with *SUZ12* promoters in solid tumours (Kirmizis et al., 2003) with non-canonical members of the WNT pathway (*WNT5A* and *WNT11*) being responsible for maximal transcriptional enhancing effect on *SUZ12* (Pizzatti et al., 2010).

1.4.3 Enhancer of Zeste proteins

Two proteins constitute this family namely; *EZH1* and *EZH2*. Margueron et al (2013) demonstrated that *EZH2* has a central role in catalysing H3K27me2/3 and its loss has a global effect on the H3K27 me2/3 status. On the other hand *EZH1* containing PRC2 has a weak (20-fold less) role in methylating and maintaining H3K27me3. *EZH1* is found ubiquitously in all tissues including the G0 resting cells (Son et al., 2013) while *EZH2* is strongly associated with proliferating tissues. There is an overlap between the target genes of *EZH1* and *EZH2*, however, *EZH1* recruitment is independent of *EZH2*. *EZH1* also associates with *EED* and *SUZ12* and hence is a part of a similar complex as *EZH2*. In contrast to *EZH2*, *EZH1* mediated repression is caused by direct alteration of the chromatin which is independent of covalent methylation marks on the histones (H3K27me3). Conservation of amino acids and SET domains of *EZH1* and *EZH2* is 65% and 94%, respectively. The presence of two paralogues in humans as opposed to just one in Drosophila may be explained by the function of *EZH2* to maintain the H3K27me3 mark while that of *EZH1* to be more involved in direct repression of target genes (Margueron et al., 2008).

Analyses of PRC2-*EZH2* substrates indicate a preferential bias towards oligonucleosomes with H1 rather than mononucleosomes, octamers and H3 alone. H1 may enhance the methyltransferase activity towards H3 (Martin, Cao et al., 2006). However, previous studies have shown contrasting results indicating that H1 incorporation into nucleosomes inhibits H3

methylation (Kuzmichev et al., 2004). These discrepancies may be due to the use of varying concentrations of histones or magnesium used in both studies. At low concentrations of magnesium, H1 has a stimulatory effect on H3 methylation while high magnesium concentrations inhibited H3 methylation stimulation by H1 (Martin et al., 2006).

1.5 Subunits of PRC1

1.5.1 Posterior sex comb proteins

MEL18, *BMI-1*, *NSPC1* and *MBLR* belong to the PSC group of proteins. This group is characterized by the presence of an N-terminal RING domain, two α helices that form the HTH domain and a C-terminal domain. *MEL18* has both transcriptional repressor and enhancer activity and is therefore included in the enhancer of trithorax and polycomb (ETP) group (Miyazaki et al., 2005). *BMI-1* and *MEL18* have 93% homology but exhibit a 6 amino acid and a 30 amino acid difference between their N-terminal and the alpha helix respectively (Akasaka et al., 1997; Fujisaki et al., 2003).The RING finger and HTH (Fig: 1.7) domains of *BMI-1* are required for down regulation of *p16INK4A* tumour suppressor gene while the Proline-Serine rich domain (PS) domain of *BMI-1* is involved in its stability. ΔPS mutant of *BMI-1* has an increased half-life and is physiologically more active augmenting its pro-proliferative activity. Specifically, over expression of the ΔPS mutant in Human Mammary Epithelial cells (HMECs) results in AKT upregulation and its targets such as cyclin D1. In primary Human Diploid Fibroblasts (HDFs), which express *p16INK4A*, over expression of the ΔPS mutant results in a decrease of *p16INK4A* expression (Fig: 1.7) accompanied by increased proliferation and decreased senescence (Dimri et al., 2002; Itahana et al., 2003; Yadav et al., 2010).

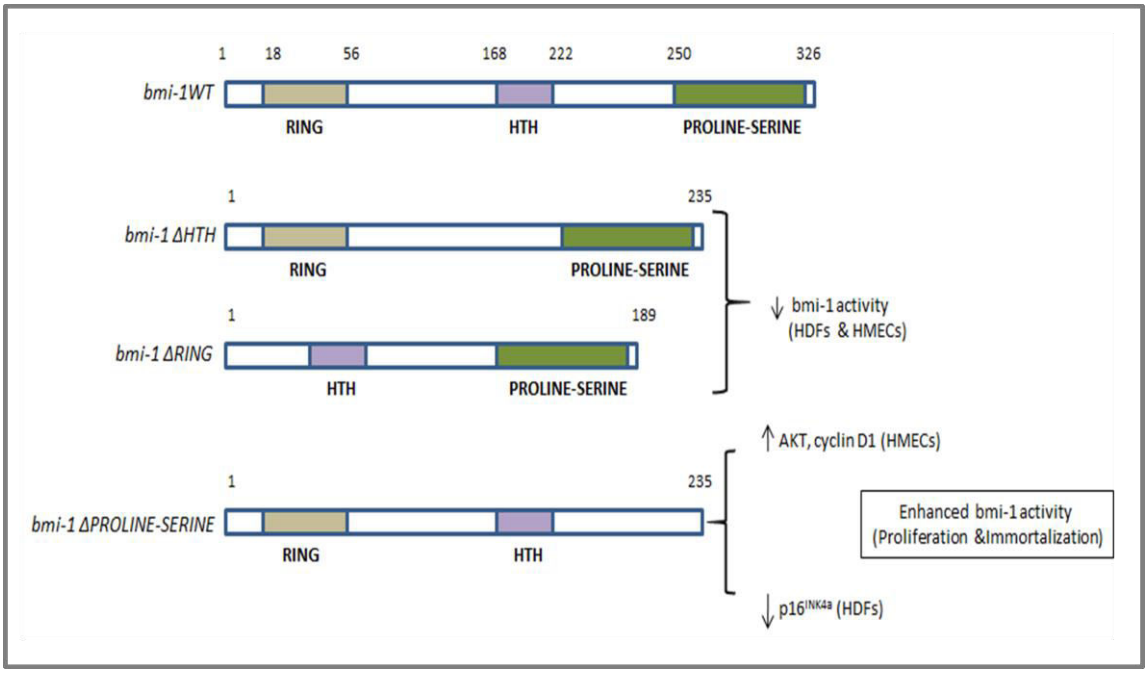


Figure 1.7: Deletion of BMI-1 domains. Loss of *BMI-1* domains result in varying effects on its activity in human diploid fibroblasts (HDFs) and human mammary epithelial cells (HMECs).Lack of HTH or RING domains of *BMI-1* reduced its activity while absence of the proline-serine domain augments *BMI-1* activity causing down regulation of *p16INK4A* and increased AKT levels.

In addition to the effect on the *INK4A/ARF* locus, *BMI-1* and *MEL18* can enhance the E3 ligase activity of ring1b. However, *MEL18* has to be phosphorylated to form the PRC1 complex. Phosphorylation of *MEL18* is needed for a conformational change in PRC1 to enable ring1b to be in close contact with the nucleosome to induce ubiquitylation of H2AK119. While phosphorylation aids *MEL18* to associate with other PRC1 proteins, *BMI-1* phosphorylation caused by MAPK-activated protein kinases (MAPKAP) causes its dissociation from PRC1 (Elderkin et al., 2007).

NSPC1 (Nervous system polycomb1) contains the RING finger domain but lacks the HTH domain (Gong et al., 2005). Depletion of *NSPC1* and *DNMT1* does not modify the H3K27me3 marks by *EZH2*, but has a significant effect in decreasing the levels of H2AK119Ub (Wu et al., 2008). *NSPC1* is highly expressed in the nervous system and exerts

a stronger repressional effect as compared to *BMI-1* and *MEL18* (Gong et al., 2005). *NSPC1* repression is responsible for enhanced cell growth and an increase in the S phase cells that is independent of the p53 pathway (Gong et al., 2006).

The *MBLR* protein consists of three domains; the N-terminal *RING* domain which interacts with *RING1B* and histone H2A, HTH domain and the C-terminus which interact weakly with *PHC2*. Akasaka et al (2002) demonstrated that *MBLR* interacts with *RING1B* causing transcriptional repression. However, it does not aggregate at PcG bodies like other polycomb proteins and may be sequestered by other proteins (Akasaka et al., 2002).

1.5.2 Ring proteins

PcG protein family contains two isoforms of the *RING1* proteins. The *RING1A* and *RING1B* proteins consisting of 406 and 336 amino acids respectively (Vidal, 2009) and have a C-terminal domain which binds to both *RYBP* and chromobox of the *CBX* proteins. The *CBX* and the *RYBP* proteins exhibit differences in both amino acid sequences and structure suggesting that the *RING* proteins are capable of recognizing non-redundant nucleotide sequences. Therefore, the genomic target of *RING1B* and the mechanism of binding (either via histone or chromatin) depends critically on its binding partner (Wang et al., 2010).

The primary function of *RING1B* is to regulate self polyubiquitination and more importantly to cause monoubiquitination of histone H2A (Fig: 1.8) at lysine 119 (H2AK119Ub) causing transcriptional repression as previously described (Stock et al. 2007). Self-ubiquitylation of *RING1B* is mediated by an unknown E3 ligase and must occur concurrently with H2AK119Ub. Prior ubiquitination of *RING1B* reduces the efficiency of the enzyme (Ben-Saadon et al., 2006).

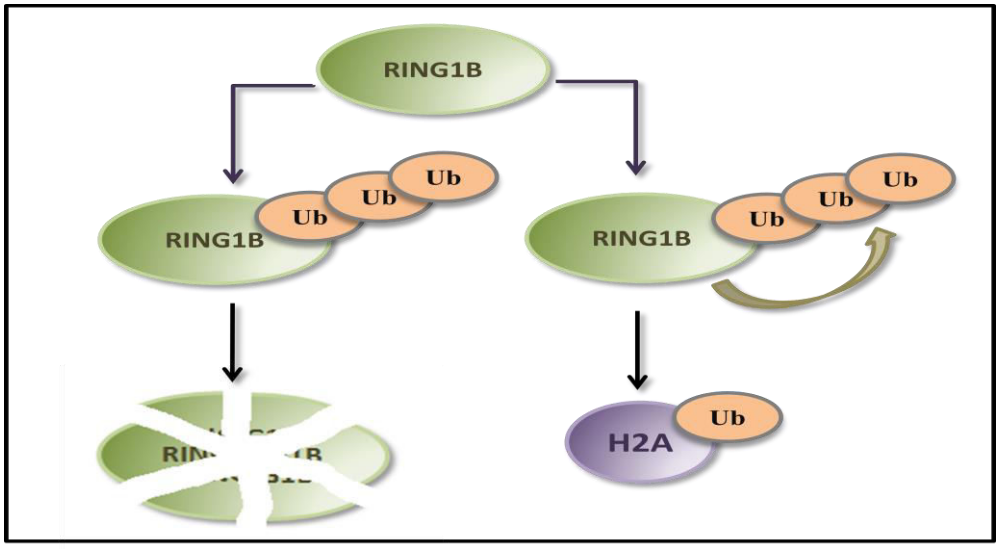


Figure 1.8: Polyubiquitination of *RING1B*. It leads to either degradation of *RING1B* or formation of H2AK119ub repressive marks.

Cao et al (2005) observed that *RING1A* is essential for the optimal functioning of *RING1B* but does not have strong E3 ligase activity. *BMI-1* and *RING1B* associate and stabilizes each other. This stabilization requires intact *RING* domains of both the proteins (Cao et al., 2005). However, *BMI-1* can only enhance the E3 ligase activity of *RING1B* as previously mentioned. Conversely, the association of *RING1A* is independent of *RING1B* E3 ligase activity but is crucial in reducing self-polyubiquitination of *RING1B* thus preventing its degradation (Ben-Saadon et al., 2006). Though *RING1A* reduces self-polyubiquitination of *RING1B*, caspase 3 and 9 can degrade *RING1B* reducing its repressive effects (Wong et al., 2007).

1.5.3 Polyhomeotic proteins

The polyhomeotic family of proteins consists of *PHC1*, *PHC2* and *PHC3*. *PHC1* and *PHC2* form a part of the PRC1 complex and their absence abrogates its repressive effects suggesting that *PHC* is critical in the formation of the PRC1 complex (Isono et al., 2005).

PHC proteins consist of three highly conserved domains; the SAM domain at the C-terminal also known as HDII, the Cys-type Zn coordination domain and the HDI domain of unknown function (Tonkin et al., 2002).*PHC2* and *PHC3* have two N-terminal domains called N-A and N-B which are poorly conserved in *PHC1*. The SAM domains can associate with themselves or with other proteins with or without SAM domains (Fig: 1.9). The complex formed by this domain has a left-handed helical structure, suggesting that chromatin wraps itself around the six-fold screw like structure making it thermodynamically stable (Kim et al., 2002).

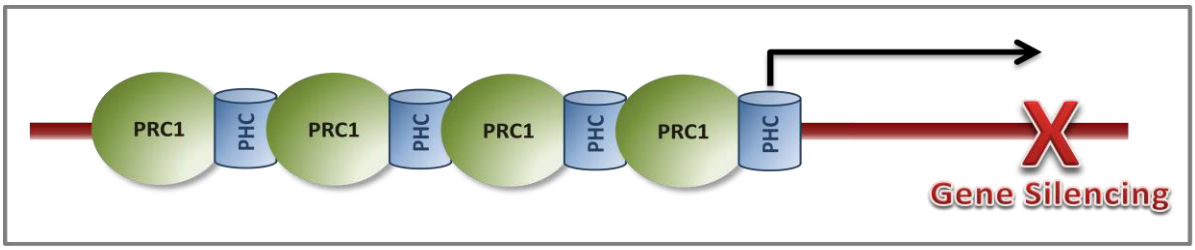


Figure 1.9: Propagation of transcriptional repression. Interaction of polyhomeotic proteins with other PRC1 proteins through its SAM domain facilitates polymerisation of proteins leading to extension of transcriptional repression and gene silencing by PcG proteins to sites outside of the immediate target regions.

Sex Comb on Midleg (*SCM*) is another protein belonging to the PcG proteins, but does not form part of the core PRC1 in all instances. Structural studies of *PHC* SAM domains indicate that *PHC*-Sam and *SCM*-SAM domains bind to each other with high affinity forming a single junction. The regulation of this interaction is unknown. *PHC* is usually localized around the targets but the exact location of *SCM* is unclear. Hence polymerisation can be used to extend the effect of the PcG proteins outside the immediate target regions (Kim et al., 2005).

Martinez et al (2009) illustrated that the *PHC* proteins repress the Notch pathway in drosophila eye. Loss of *PHC* proteins caused reduction in the repression of members of the Notch pathway giving rise to uncontrolled proliferation, loss of cell polarity and failure to exit cell cycle (Martinez et al., 2009). Interestingly, a single nucleotide polymorphism in exon 27 of the *NOTCH* gene has been observed but no mutations are seen in MDS cases (Fu

et al., 2007). Mutational studies in Drosophila embryos have revealed that lack of *PHC* proteins lead to increased number of chromosome bridges not only in embryonic stages but also in subsequent cell cycles, alluding to a role in DNA damage checkpoint (Beck et al., 2010).

1.5.4 Chromobox proteins

The *CBX* family of proteins form a part of the PRC1 complex and are responsible for recognising H3 lysine trimethylation. These proteins have a chromodomain and a chromobox that interacts with histones and other components of the PRC1, respectively(Vincenz and Kerppola, 2008). Different members of this family have different binding targets (Table: 1.8). *CBX2* binds to K4 and K9 on H3 with high affinity as compared to K27 (Vincenz and Kerppola, 2008). Both K9me3 and K27me3 interact with *CBX7* while *CBX4* has a preference for K9me3. *CBX6* and *CBX8* does not bind to any of the known sites on the histones and the targets of these two proteins have yet to be ascertained (Bernstein et al., 2006).

CBX2 has a unique hook like motif (AT-hook) present near its N-terminal chromodomain. Other *CBX* proteins also have a similar motif called AT-Hook-like domain. The AT-hook or AT-hook-like structures trap and stabilize the DNA between the *CBX* protein and the histone (Senthilkumar and Mishra, 2009)**.** In *CBX2*, the phosphorylation of the AT-hook in the chromodomain also governs its target affinity.The AT-hookphosphorylation increases its interaction with H3K27me but also reduces its binding capacity to H3K9me3 (Hatano et al., 2010).

CBX proteins not only show variation in their target preferences but also in their distribution patterns (Table:1.8). *CBX2*-H3 is more localized in chromocentres in contrast to the total population of *CBX2*. On the other hand *CBX6*-H3 as well as the total population of *CBX6* is homogenously distributed in the nucleoplasm. The *CBX4*, *CBX7* and *CBX8* show varied

distribution depending on the cell-type (Vincenz and Kerppola, 2008). Moreover, all of these proteins except *CBX2* are capable of binding to RNA if their secondary structures are intact. *CBX7* has a high affinity to bind to single stranded RNA. Hence these proteins can not only bind to histones but also to the DNA target sequences via non-coding RNAs (Bernstein et al., 2006)**.**

PcG	Binding partners	Distribution	Effect on Ink-Arf locus
CBX2	H3K4 and K9, K27 (low affinity)	chromocentres	Repression
CBX4	H3K9	cell type dependant	Weak repression
CBX6	unknown	Nucleoplasm	Weak repression
CBX7	H3K9 and K27	cell type dependant	Stronger repression than *CBX2*. Causes immortalization of cells. Lack of repression cause cell cycle arrest in human fibroblasts.
CBX8	unknown	cell type dependant	Weak repression compared to *CBX7* but stronger than *CBX2*. Lack of repression maintains cell cycle arrest in human fibroblasts

Table 1.8: Summarization of different *CBX* proteins. Their binding preference to histones, distribution and effect on repression (Yap et al., 2010).

CBX4 possesses a unique function to act as E3 ligase for *DNMT3A*. *DNMT3A* contains a PWWP region and a linker region of amino acids which interacts exclusively with *CBX4*. *CBX4* also recognises H3K9 hence there is a possibility that it may assist in *DNMT3A* localization to the heterochromatin (Li et al., 2007).

1.6 Role of PcG proteins in DNA damage

Recent findings by Chandler et al (2014) question the role of PcG proteins in DNA damage repair, who demonstrated that there was no localization of PRC1 proteins at the sites of DNA damage in fibroblasts (Chandler et al., 2014). However earlier reports implicated *BMI-1* in the maintenance of the redox balance and its deficiency in mice leads to increased levels of reactive oxygen species resulting in DNA damage and double strand repair activation (Liu et al., 2009). Ismail et al (2010, 2012) demonstrated that *CBX4* is an early DNA Damage Response (DDR) protein that mediates SUMO conjugation of *BMI-1* which is required for the accumulation of *BMI-1* at DNA damage sites. *RING1B* is also recruited to sites of DNA double-strand breaks (DSBs) along with *BMI-1* where both contribute to H2AK119Ub in osteosarcoma and kidney cells. In the absence of *BMI-1*, recruitment of several proteins dependent on ubiquitin signalling like *BRAC1* to the DSBs is hampered with a concomitant increase in radiation sensitivity (Ismail et al., 2010; Ismail et al., 2012).Interestingly, other research groups have shown that *BMI-1* mediated ubiquitylation of H2A does not affect the canonical ubiquitin pathway but impacts a alternative pathway involving the recruitment of yet unidentified factors (Ginjala et al., 2011).Unlike *BMI-1* and *RING1B*, *MEL18* is only transiently recruited to DSBs, indicating recruitment and maintenance of different PRC1 complexes via distinct kinetics (Chou et al., 2010). In addition to PRC1, PRC2 has also been implicated in DNA damage repair. Role of *EZH2* in DSB repair has also been demonstrated in breast cancer cells were it modulates repression of the homologous recombination enzyme *RAD51* (Chang et al., 2011) DSBs and of its paralogues (Zeidler et al., 2005) which are crucial for homologous recombination repair. Localization of PcG proteins at DNA DSBs might suggest an additional function of PcG proteins in preventing the transcription of broken genes that would otherwise lead to aggregation of defective mRNAs.

1.7 Role in Haematological malignancies

EZH2 mutations are detected in 8 % of patients with MDS/MPN, 3 % each in MPN, MDS and 2 % in AML (Khan et al., 2013; Wang et al., 2013). These mutations were first discovered using SNP 6.0 arrays on peripheral blood /bone marrow samples and were associated with CN(LOH) 7q and 7q36.1 microdeletion (Fig: 1.10).

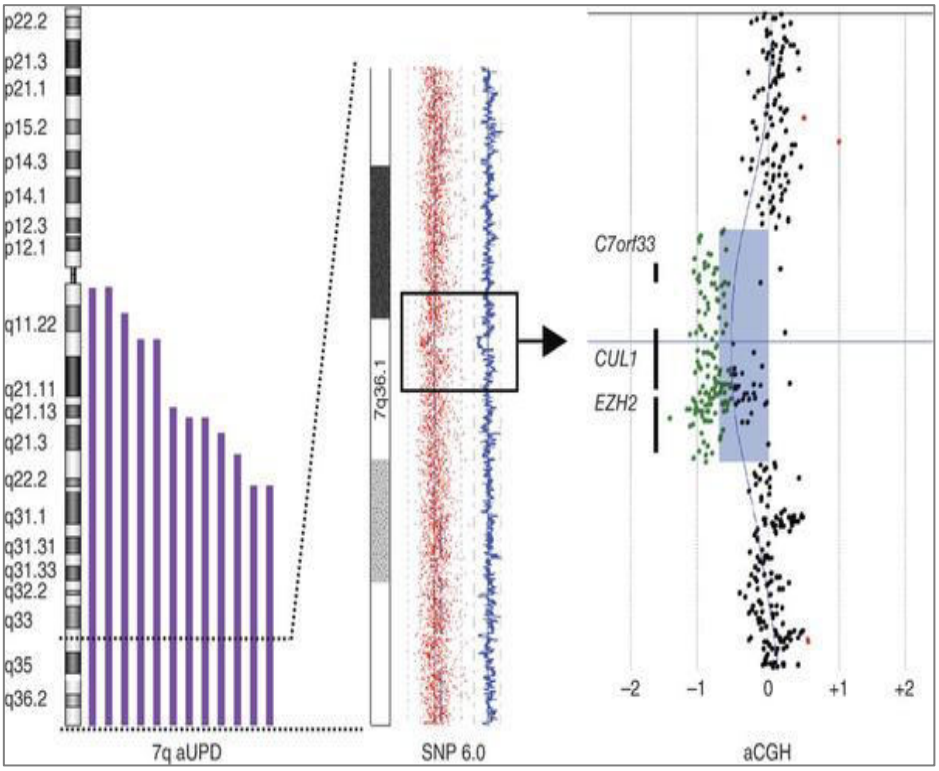


Figure 1.10: Identification 7q CN (LOH) and EZH2 involvement. CN (LOH) was observed in 12 patients (left) with MDS/MPN using SNP6 analysis. (middle) Identification of 7q36.1 microdeletion in one patient and (right) targeted aCGH analysis in the same patient reveal heterozygous microdeletion of *EZH2* (the x axis indicates log copy number ratios) (Ernst et al., 2010).

Truncation mutations were identified throughout the gene while missense mutations were commonly found in the conserved SET domain. A frame-shift mutation of exon 19 was detected with high frequency. Individuals with homozygous mutations have poor survival as compared to those with heterozygous mutations(Ernst et al., 2010; Khan et al., 2013; Xu et al., 2011) with increased risk of transformation to AML (Wang et al., 2013). *EZH2* mutations

are observed in some cases of refractory anaemia indicating their acquisition in the early stages of MDS and have been associated with low blast percentage in the bone marrow (Ernst et al., 2010; Nikoloski et al., 2010; Wang et al., 2013).MDS is an evolutionary process, evolving over variable time frames and driven by stepwise acquisition of somatic cell mutations with sequential, sub-clonal selection. The genotype of patients with MDS shifts over time by obtaining additional mutations, changes in allele burden of existing mutations or complete disappearance of certain mutations post treatment. Sequential analysis of *EZH2* mutations in patient samples to study their clonal evolution and mutational overlap to would help conclude whether *EZH2* mutations are "driver mutations" i.e. playing a major role in initiating and maintaining the disease or "passenger mutations" i.e. assisting the driver mutations in their role. Nikoloski et al (2010) sequenced 126 MDS patients and demonstrated that 8 patients had *EZH2* mutations. In 5 of the 8 (40 %) of patients with *EZH2* mutation an associated *TET2* abnormality was also observed (Nikoloski et al., 2010), suggesting that *EZH2* mutations may not be driver mutations (Makishima et al., 2010). Mutations of *ASXL1*, *TET2*, Runt-Related Transcription Factor 1 (*RUNX1)* and *LUC7L2* (Haferlach et al., 2014) (Table: 1.5) overlap with *EZH2* mutations in $CD34^{+}$ cells obtained from AML cases (Khan et al., 2013). Other studies have shown no correlation between *ASXL1, RUNX1, TET2* and *EZH2* mutations nor do they suggest a poor prognosis of MDS patients related specifically to *EZH2* mutations (Wang et al., 2013). Inactivation of *EZH2* by these mutations has been correlated with absence of the H3K27me3 mark suggesting its role as a tumour suppressor gene (Ernst et al., 2010; Makishima et al., 2010). *EZH2* dysfunction (reduced H3K27me3) can be attributed not just to its mutations (37 %) but also to mis-splicing in patients with U2 auxiliary factor (*U2AF1)*/ Serine/Arginine-Rich Splicing Factor 2 (*SRSF2*) mutations (11 %) and in 24% of cases because of haploinsufficient expression (7q del) (Khan et al., 2013).

Apart from *EZH2*, mutational analysis of other PcG genes in MDS has not been carried out and therefore the aim of my PhD was to examine SNP6 and mutational status of PRC1, PRC2 genes.

In addition to myeloid malignancies, *EZH2* has also been implicated in lymphomagenesis wherein RNA sequencing studies recently identified heterozygous somatic point mutations targeting the *EZH2* SET domain (Hot spot: Y641) in 30% of germinal centre Diffuse Large B-cell Lymphoma (DLBCL) and 10% of Follicular lymphoma (Morin et al., 2010). These mutations altered the enzymatic activity of *EZH2* resulting in a protein that does not recognize unmodified H3K27 and causes increased H3K27 trimethylated state resulting in gain of function (Sneeringer et al., 2010; Yap et al., 2011). Increased expression of mutant *EZH2* in mice induced germinal centre hyperplasia and accelerated lymphomagenesis (Beguelin et al., 2013). High-levels of *EZH2* (70% increase) are associated with increased Overall Survival (OS) of 85.8% compared to low expression (70% reduction), with OS of 44.5% (p=0.005) in DLBCL (Lee et al., 2014). Apart from its role in DLBCL, *EZH2* inhibition by DZNep/shRNA in MLL fusion leukaemia reduces proliferation of leukemic cells via u pregulation of p16 (Ueda et al., 2014) and its deletion in acquired aplastic anaemia decreased the production of destructive Th1 cells in vivo, decreased bone marrow-infiltrating Th1 cells, and rescued mice from bone marrow failure (Tong et al., 2014).

Lack of *EZH2* and *EED* is modulated by the pRB-E2F pathway and causes disturbances in the cell cycle cyclins. Down regulation of pRB is also seen in cases of acute leukaemia (Tonini et al., 2008). Signal Transducer And Activator Of Transcription 3 (STAT3), Oct-3/4 transcription factors modulate *EED* expression by binding to the promoter and enhancing transcription. Suppression of these transcription factors cause a reduction of *EED* resulting in substantial loss of the H3K27me3 mark and incomplete differentiation, indicating that not all the differentiation genes are activated (Ura et al., 2008). STAT3, Oct-3/4 are oncogenic

transcriptional factors exhibiting elevated expression in some AML cases; their association with *EED* may have additional epigenetic consequences in haematological malignancies (Steensma et al., 2006).

In adult life, *EED* regulates the development of myeloid and B cell precursors. Another significant impact of *EED* is in modulating the thymocyte differentiation and inhibiting lymphoma generation following carcinogen exposure. However, *EED* mutants do not cause spontaneous lymphomas. Additional genetic alterations are needed for lymphoma transformation (Richie et al., 2002). Conversely, lack of *SUZ12* impacts granulocyte differentiation. It blocks the differentiation of CML cells and is over expressed in chronic myeloid leukaemic blasts cells (Pasini et al., 2004).

Similar to PRC2, different sub-units of canonical PRC1 have also been reported in haematological malignancies. Flowcytometric analysis of bone marrow CD34+ cells from 50 MDS patients revealed that increasing *BMI-1* levels correlated with disease progression both by WHO and IPSS categorization. *BMI-1* levels were high in cases of refractory anaemia and escalated as the disease progressed to RAEB suggesting that it could be used as a molecular marker for MDS patients. In contrast, *BMI-1* over expression in breast tumours has been associated with a favourable impact on patient survival (Pietersen et al., 2008).

In vivo transplantation studies in mice found that over expression of *CBX7* in HSC and progenitor cells induced leukaemia (Klauke et al., 2013). However, the contribution of *CBX* to leukaemogenesis might vary depending on the tumour-initiating cells or the type of hematologic tumour. For example, *MLL*-induced leukaemogenesis is dependent on *CBX8*, but not on its orthologs (Chubb et al., 2013). *CBX8* and *BMI-1* bind to *MLL* fusion proteins independent of other canonical PRC1 members and induces *MLL–AF9* lymphomagenesis (Smith et al., 2011). *CBX8* is also implicated in leukaemia wherein chromobox on *CBX8* binds to *MLL-AF9* fusion protein to up regulate *HOXA9* expression which is a hallmark of

MLL pathogenesis. The loss of *CBX8* protein prevents leukaemic transformation induced by *MLL-AF9* but did not alter normal haematopoiesis (Tan et al., 2011).*CBX7* on the other hand was identified as a risk locus for developing multiple myeloma (Chubb et al., 2013). Apart from the *CBX* proteins, high levels of *RING1A* were also observed in acute myeloid leukaemia (Ishikawa et al., 2007).

1.8 Role in Haematopoietic system

The role of PcG proteins in haematopoiesis was first suggested by the phenotype of *BMI-1*-/- mice, which displayed reduced numbers of hematopoietic progenitors and increased differentiated cells, eventually leading to haematopoietic failure (van der Lugt et al., 1994). It is associated with HSC self-renewal by regulating symmetrical cell division. Apart from normal haematopoiesis, *BMI-1* is also important for leukemic stem cell self-renewal as evidenced by loss of self-renewal and induction of apoptosis upon *BMI-1* knockdown in primary acute myeloid leukaemia samples (Iwama et al., 2004; Lessard et al., 1999; Park et al., 2003; Rizo et al., 2008; Rizo et al., 2009). Lessard et al (1999) established that *EED* and *BMI-1* have antagonistic effects on the haematopoietic system; *EED* inhibited proliferation whereas, *BMI-1*$^{-/-}$ enhanced proliferation of progenitor cells. *EED* and *BMI-1* mutant mice revealed that *BMI-1* preceded *EED* in the control of primitive $CD34^{+}$ cell proliferation. The interaction of *EED* and *BMI-1* was tissue specific but were expressed in all the hematopoietic cell lines. Mutations in *EED* increased the susceptibility to cancer development; however, additional genetic alterations may be needed for complete transformation of hematopoietic stem cells. Though *BMI-1* and *EED* were co-expressed in HSCs, these proteins did not share common down-stream mediators. The *p16INK4A* and *p19ARF* levels remained unaltered in *EED* mutants compared to *BMI-1* mutants which primarily affect the *INK4A-ARF* locus as previously mentioned (Fig: 1.6). The mediators causing the proliferation inhibition by *EED*

are still unknown (Lessard et al., 1999). Recent studies using ChIP analysis have confirmed the presence of *BMI-1* along with *EZH2*, *SUZ12*, *EED*, *RING1A*, *RING1B* and the transcriptional factor Ecotropic Viral Integration site 1 (*EVI1*) at the promoters of the Phosphatase and TENsin homolog (*PTEN)* gene (Yoshimi et al., 2011).The *PTEN* tumour suppressor gene is located at 10q23.31 and functions as a negative regulator of the phosphatidylinositol-3-OH kinase (PI (3) K)–Akt pathway, which is crucial in cell survival, differentiation, cell proliferation and migration (Zhang et al., 2006). *EVI1* is essential for proliferation of haematopoietic stem cells and *PTEN* has a crucial role in restricting the activation of HSCs, in lineage determination, and prevention of leukaemogenesis. Both *EVI1* and *PTEN* have been implicated in myeloid leukaemogenesis (Yoshimi et al., 2011).*PTEN* mutants in mice demonstrate increased cell populations that match those that become dominant in acute myeloid leukaemia (Zhang et al., 2006).It would be important to comprehend the effect of the PcG proteins on *PTEN* and its relationship to the pathogenesis of MDS.

MEL18 KO in vivo, produces a hypoplastic bone marrow phenotype with increased number of G0 phase cells (Kajiume et al., 2004). Re-expression of *MEL18* normalises the HSC self-renewal capacity (Kajiume et al., 2004; Mihara et al., 2006).In contrast to *MEL18*, lack of *BMI-1* exhibits defective trabecular bone with a reduction in osteoblasts that are replaced by adipocyctes (Oguro et al., 2006).

Unlike *PSC* subunits, the *RING* proteins have a dual function in the haematopoietic system (Fig: 1.11). Role of *RING1B* in mouse models indicates that it maintains equilibrium between the positive and negative proliferation signals.

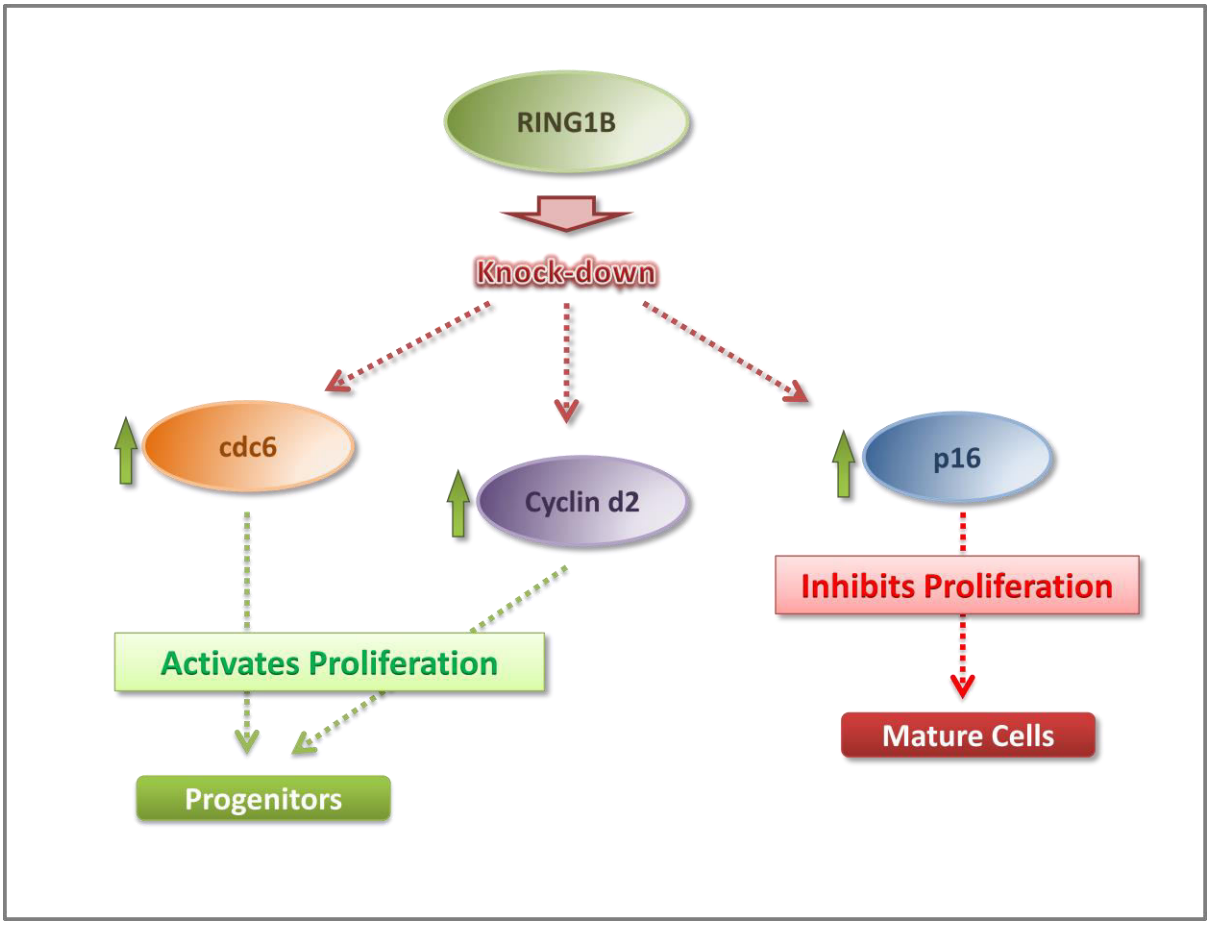


Figure 1.11: Dual effect of KO of *RING1B* on the haematopoietic system. KO of *RING1B* causes up regulation of cyclin D2 and cdc6 which cause enhanced proliferation of the haematopoietic progenitor cells while a concomitant upregulation of *p16* reduces the proliferation of mature myeloid and lymphoid cells.

RING1B knock out leads to up-regulation of cyclin D2 and cdc6 which has a positive effect on proliferation. This is reflected in increased myeloid progenitor cells leading to premature lymphoma development. Moreover, *RING1B* also inhibits the expression of *p16INK4A* which is a proliferation inhibitor during haematopoiesis. In the absence of *RING1B*, there is up-regulation of *p16INK4A* and hence a decrease in the number of mature myeloid and lymphoid precursor cells (Cales et al., 2008). Mutations of *RING1B* have yet to be identified, but with the aberrant haematopoietic phenotype of conditional knockout mice and its crucial H2AK119Ub function, additional studies may be warranted to understand its impact in MDS.

Expression of *PHC1*/mouse *PHC1* (*MPH1*) is significantly high in HSCs during adulthood. Some studies have demonstrated that *PHC1*/*MPH1*$^{-/-}$ mice display arrest of B-cell maturation (Tokimasa et al., 2001).Ohta et al (2002) observed that absence of *PHC1/MPH1* was lethal in mouse embryos. However, reduction in its levels caused no significant differences in the myeloid and erythroid lineages in the mice neonates. FACs analysis showed no difference in the liver tissue but demonstrated a reduction in B-cells in the spleen. Colony forming units in spleen (CFU-S12) were decreased to one-tenth of their normal levels. A reduced activity of PHC1$^{-/-}$HSCs was presumed to be at least in part responsible for the reduced expansion of the (CFU-S12) and the progressive decline of clonogenic progenitors during fetal development. However, detailed analysis revealed poor expansion of colony forming units in only a small population of HSCs indicating an overall conservation of haematopoiesis. The exactmechanisms of development of reduced numbers of aberrant HSCs remains ambiguous. The possible suggestions could be that *PHC1/MPH1* controls repression of genes which are needed for the development of the HSCs and in their absence the HSCs are unable to respond to cytokines essential for their normal regulation (Ohta et al., 2002).

Interestingly, the *CBX* proteins function via an autoregulatory feedback loop wherein *CBX* proteins are PcG targets in stem cells; *CBX7* represses its homolog proteins in HSCs and, in a negative feedback loop while *CBX2*, *CBX4*, and *CBX8* repress *CBX7* in differentiated cells (Klauke et al., 2013). *CBX* proteins dictate the balance between self-renewal and differentiation in HSC. *CBX7* is mainly involved in self-renewal in HSCs which is replaced by its partners *CBX2*, *CBX4*, and *CBX8* during lineage commitment (Klauke et al., 2013). Among the *CBX* paralogues, *CBX2* knockdown resulted in the most prominent phenotype, reflected as loss of HSCs and progenitors, decreased cell proliferation, and an increase in apoptotic events (van den Boom et al., 2013).

Function of *EZH2* in HSCs is to stabilize the chromatin structure and maintain long-term self renewal potential of HSCs by switching off pro-differentiation genes (Ezhkova et al., 2009; Kamminga et al., 2006) and promoting a shift toward proliferation by increasing expression of genes, which facilitate progression through the cell cycle (Bracken et al., 2003).In vivo studies based on deletion of *EZH2* in transformed granulocyte macrophage progenitors indicate compromised growth and increased risk of progression to AML (Tanaka et al., 2012).

1.9 Therapeutic role of PcG proteins

DZNep inhibits *EZH2* by impairing the S-adenosylhomocysteine (S-AdoHcy) hydrolase activity and the synthesis of S-adenosylmethionine (S-AdoMet). DZNep-mediated inhibition of S-AdoMet results in global inactivation of diverse methyltransferases (Chiang, 1998).Fujiwara et al (2014) studied DZNep and its association with erythroid differentiation. They examined differentiation of K562 cells by quantitative RT-PCR after treatment with DZNep which reduced *EZH2* levels inducing erythroid-related genes like globins. The same effects were replicated in primary erythroblasts derived from cord blood CD34-positive cells. However they concluded that these effects were only partially due to *EZH2* inhibition and that additional factor might also be involved (Fujiwara et al., 2014).This might be due to lack of specificity of DZNep towards *EZH2*.Apart from DZNep, small molecule inhibitors like EI1 and GSK126 have also been developed that bind to the S-adenosylmethionine (SAM) pocket of the *EZH2* SET domain and cause reduction of histone H3K27me3 mark. EI1 inhibition *EZH2* in memory B cells lead to G1 growth arrest and apoptosis (Qi et al., 2012). The EI1 molecule like the GSK126 inhibitor targets the SET domain of *EZH2* for its inhibitory activity. Other research groups used a different approach by focusing on the *EZH2-EED* interaction, disrupting it by using a stabilized alpha helix of *EZH2* (SAH-EZH2)

peptides and therefore reducing the protein levels of *EZH2* and H3K27me3 in *MLL-AF9* cells causing cell cycle arrest and monocyte/macrophage differentiation (Kim et al., 2013).

On the other hand Retionic acid (RA) has been shown to cause disruption of the *SUZ12*-RARγ multi-protein complexes removing *SUZ12* from the target genes in embryonic stem cells. The translocation of nuclear to cytoplasmic *SUZ12* is rapid with no net change of total *SUZ12* protein levels. This activates transcription of genes associated with cellular differentiation (Amat et al, 2011). The role of *EZH1* in HSCs is to repopulate the HSC pool thus protecting them from senescence. *EZH1* ablation induces significant loss of adult HSCs, with concomitant impairment of their self-renewal capacity (Hidalgo et al., 2012).

Apart from PRC2, PRC1 is also a promising target for anti cancer therapies. Zhang et al (2014) proved that *RING1B* and *BMI-1* associate with Ribonucleotide Reductase (RRM1) which is predictive of gemcitabine's therapeutic effect. Both the proteins cause polyubiquitination at position 224 and 548 of RRM1 leading to its proteosome-mediated degradation. Loss of RRM1 reduces resistance to gemcitabine treatment (Zhang et al., 2014).

The crucial role of the canonical polycomb proteins in haematological malignancies and their promising therapeutic potential make them important targets in MDS.

1.10 Project Aims

Loss of function mutations of the PcG gene *EZH2* have been identified in MDS and are associated with SNP6 abnormalities i.e. CN (LOH) 7q and microdeletion 7q36.1. *EZH2* is the catalytic component of PRC2 that causes H3K27me3 marks resulting in gene silencing. However identification of changes in gene signature as a result of these mutations has not been examined in myeloid cells. Apart from *EZH2* other PcG family genes i.e. PRC1 and PRC2 have not been investigated for the presence of mutations and may contribute to the development of myeloid malignancies. Jumonji genes are histone demethylases which have opposing function to *EZH2* and some members of this family of proteins form non-canonical complexes with PRC2. The effect of *EZH2* mutations on the histone demethylases and the degree of their mutational overlap if present have not been ascertained till date.

In addition to focusing on histone modifying enzymes, my PhD also examines DNA methyltransferase; *DNMT3A* because of its interaction with the SET domain of *EZH2* and the presence of *DNMT3A* mutations previously reported in MDS. However, these reports have not examined the mutational overlap between *EZH2* and *DNMT3A*. Based on the literature survey available on these genes the following project aims were examined.

- To identify aberrations in Polycomb (PRC1, PRC2) and Jumonji genes using high density SNP array and to determine the prevalence, clonal size of mutations affecting these genes in a cohort of MDS cases pre-selected for SNP6 aberrations using 454 sequencing. To correlate the mutational status of PRC and Jumonji genes to their SNP abnormalities.
- Study modulation of gene expression of Jumonji genes as a consequence of SNP6 abnormalities.

- To determine the prevelance and clonal size of *DNMT3A* gene mutations in patients with MDS using 454sequencing to identify mutational overlap between *DNMT3A* and *EZH2*.
- To examine the functional consequences of SET domain mutant of *EZH2* and KO of *EZH2* in myeloid cells lines using microarray gene expression profiling and verifying this gene signature in patient samples.
- To examine the functional consequences of *DNMT3A* mutation (R882H) in myeloid cell lines.

2 Chapter: Materials and Methods

2.1 Reagents

2.1.1 Enzymes and Buffer

Restriction enzymes: BglI,Bpu10I, BsiWI, BstEII EcoRI HindIII NruI, NotI PstI SalI, ScaI, SfiI, SnaBI XbaI, XhoI NEB buffer 1/2/3/4	New England Biolabs
T4 DNA polymerase (3,000 units/ml)	New England Biolabs
SuperTaq (5U/ µl)	Cambio
dNTPs (10mM)	New England Biolabs
10X PCR buffer	Sigma-Aldrich
T4 DNA polymerase buffer (10X)	New England Biolabs
T4 DNA Ligase (400,000 units/ml)	New England Biolabs
BSA (100X)	New England Biolabs
Alkaline Phosphatase (CIP)	New England Biolabs

2.1.2 Plasmids

Expression plasmid	Plasmid description	Manufacturer
IRATp970F1210D	cDNA EZH2 cloned in pCMV-SPORT6	Source Bioscience (Genome cube)
IRATp970A0473D	cDNA DNMT3A cloned in pCMV-SPORT6	Source Bioscience (Genome cube)
pSUPER.retro.puro	EZH2-ShRNA cloned in pSUPER.retro.puro	Dr. Joop Gaken
p3xFLAG-Myc-CMV-26	Expression plasmid for FLAG tagged wild type and mutant EZH2 and DNMT3A	Dr. Joop Gaken
pBABEPuro	Plasmid contained GFP+ Puromycin	Dr. Joop Gaken
pMDG,pRRE,pREV	Generation of lentivirus	Dr. David Darling

2.1.3 Antibodies

Primary	Dilution for western blotting	Stock conc	Manufacturer
EZH2 (AC22)	1:2000	---	Abcam ab110646
H3 chip-grade antibody	1:2000	100 µg	Abcam ab46765
H3K27me3 chip-grade antibody	1:1000	100 µg	Abcam ab6002
H4	1:1000	200 µg	Upstate 06-598
H3K9me3	1:2000	100 µg	Abcam ab8898
H3K4me3	1:2000	100 µg	Abcam ab8580
γ-Tubulin	1: 2000	200 µg	Santa Cruz Biotechnology sc-7396
β-Actin	1:300	---	Abcam ab8226
Secondary	**Dilution**	**Stock**	**Manufacturer**
HRP-conjugated anti-rabbit antibody	1:1000, 1:2000, 1:5000	1 mg/ml	Sigma
HRP-conjugated anti mouse antibody	1:2000	1 mg/ml	Sigma
HRP-conjugated anti-Rabbit,anti-mouse antibody	1:2000	1 mg/ml	Sigma A9169, A9044

2.1.4 PCR, Gel electrophoresis and Sequencing Materials

BigDye Terminator v3.1 Cycle sequencing kit	Applied biosystems
BigDye XTerminator purification kit	Applied biosystems
Ethidium Bromide	Sigma

Gel cleaning kit QiAEX II, gel extraction kit	Qiagen
GoTaq Colourless Master Mix	Promega
Agarose, molecular biology grade/electrophoresis grade	Melford
O'Gene Ruler ladder mix (0.1µg/µl)	Fermentas
6x Orange DNA loading Dye Fermentas	Fermentas
Tris Acetate-EDTA buffer (TAE buffer) 50X	Sigma
Filtered and unfiltered micropipette tips	Starlabs

2.1.5 DNA and RNA extraction reagents

TRIzol	Invitrogen
RNase free water	Qiagen
Phase lock Gel Heavy 2 ml	5 PRIME
RNeasy Micro Kit	Qiagen
Chloroform	Sigma
2-Propanol (Isopropanol)	Sigma
Ethanol	Sigma
DNeasy blood and tissue kit	Qiagen

2.1.6 Cloning reagents

Nuclease free water	Ambion
Trptone enzymatic digest	Sigma -Aldrich
Bacto Yeast extract	Becton Dickinson
BactoAgar	Becton Dickinson
Carbenicillin solution 100mg/ml in 50% Ethanol (10ml)	Bioline
Maxi prep kit Gen Elute plasmid maxi Prep kit	Sigma
Wizard® Plus SV Minipreps DNA Purification System	Promega
Topo TA cloning kit dual PROMOTER pcri-Topovector	Invitrogen
Pure Yield™ plasmid Mini prep system	Promega
One Shot® Mach1™-T1R Chemically Competent E.coli	Invitrogen
MegaX DH10B Electrocomp™ Cells	Invitrogen
5-bromo-4- chloro-3-indolyl-β-D-Galactoside (X-Gal)	Melford

2.1.7 Transfection reagents

Amaxa cell line neuclofector kit V	Lonza
Lipofectamine® LTX with Plus	Invitrogen
Lipofectamine RNAiMAX Reagent	Invitrogen
Igenio Electroporation solution	Mirus
OptiMEM Reduced Serum Medium	Life technologies
Polybrene	Sigma
G418 (neomycin)	Sigma
Puromycin	Sigma

2.1.8 Tissue culture reagents & equipment

Roswell Park Memorial Institute medium (RPMI-1640)	Sigma Aldrich
Fetal Calf serum (FCS) Penicillin/Streptomycin/Ciprofloxicin	 Invitrogen
15ml & 50ml Falcon tubes Vented tissue culture flasks 5ml,10ml & 20ml Pipettes 6,12 & 24 well plates	 VWR international ltd
1.5 ml microcentrifuge 'Eppendorf tubes'	Starlabs
Phosphate Buffered Saline (PBS)	PPA
L-Glutamine-Penicillin – Streptomycin solution	Sigma
Trypan Blue 0.4% , 0.85% NaCl	Lonza
Dimethyl sulfoxide (DMSO) (freezing cells)	Sigma
Neubauer Improved Haemocytometer	VWR international

2.1.9 Cell lines

Cell line	Source	Provider
K562	Chronic myelogenous leukaemia	Dr. David Darling
MOLM13	Acute monocytic leukaemia	DSMZ
KG1	Acute myelogenous leukaemia	Dr.Terry Gaymes

2.1.10 qPCR reagents

Universal Library probes	Roche Diagnostics
qPCR primers	Integrated DNA technologies
FastStart Universal Probe Master 12.5 ml	Roche Diagnostics
SuperScript® VILO™ cDNA Synthesis Kit	Life Technologies
MicroAmp™ Fast Optical 96 well reactions Plate with Barcode (0.1ml)	Life Technologies

2.1.11 Western Blot reagents

NuPage Bis-Tris gel	Life technologies
Novex Sharp Pre-stained Protein standard	Life technologies
Bovine serum albumin (BSA)	Fisher Scientific
ECL Prime Western Blotting Detection Reagent	GE Healthcare
Non fat dried milk (Marvel)	Sainsbury
2-Mercaptoethanol	Sigma
Methanol	Sigma
Nupage® Tris –Acetate SDS Running buffer (20X)	Invitrogen
Nupage® LDS Sample buffer (4X)	Invitrogen
Nupage® Transfer buffer (20X)	Invitrogen
Nitrocellulose Membrane	GE Healthcare
TWEEN 20	Sigma

2.1.12 Cell viability, proliferation and cell cycle reagents

Tetrazoliumsalt[3-(4,5-Dimethylthiazol-2-yl) 2,5diphenyltetrazolium- bromide] (MTT)	Sigma
Propodium Iodide (IP)	Sigma
Fluorescein isothiocyanate (FITC)	Sigma
Pacific Blue Annexin V	BioLegend

Cell- based Assay Annexin V Binding buffer	Cayman
7-Aminoactinomycin D (7-ADD)	BioLegend

2.1.13 Buffers and Solutions

1X PBS	10 tablets were dissolved in 1 litre of dH_20
X-GAL	20 mg of X-Gal was dissolved in 1 ml of dimethylformamide (DMF) to make a final conc of 20 mg/ml.Solution was stored at -20°C.
1X TAE	Concentrated buffer powder (10X) was dissolved in 1L of dH2O to produce 1x TAE with 40mM Tris-acetate and 1 mM EDTA. Solution was stored at room temperature.
1X sample buffer	20 % (v/v) NuPAGE LDS Samples Buffer (4 x), 10 % (v/v) β-mercaptoethanol were made up in dH20 and stored at - 20°C.
NuPage Transfer buffer	50 ml 20X NuPage transfer buffer, 20 % Methanol (v/v), dH_20 upto 1 litre
1x MES SDS Running buffer	Buffer made by 1: 20 dilution of 20X Running buffer in dH20. The 1X Solution composed of 50 mM MES, 50mM Tris Base, 0.1 % SDS and 1mM EDTA with pH7.
MTT solubilisation buffer	The solution was prepared by mixing 10 % (v/v) Triton X-100 with 90 % (v/v) Isopropanol and 2 drops of HCL 37% (12M). Stored at room temperature.
Luria-Bertoni (LB) broth	1 % (w/v) bactotryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl, 2ml NaOH
LB-Agar	LB broth with 1.5 % (w/v) agar. Addition of ampicillin to a final concentration of 100 µg/ml after autoclaving
Puromycin	1 mg/ml working solution was made in dH20 from10 mg/ml stock solution.
DNA loading dye	30 %(v/v) glycerol and 0.25 % (w/v) Orange R.
MTT	5 mg/ml was made by dissolving MTT powder in PBS. Solution was filtered and aliquots were frozen -20°C in dark.
RNaseA	10 mg/ml dH2O stored at -20°C
FITC	1µg/ml PBS. Stored at 4°C in the dark
Blocking solution	5 % (w/v) Marvel milk powder was dissolved in 1X PBS buffer

2.2 siRNA

ON-TARGETplusSMARTpool Human *EZH2* (Cat No: L-004218-00) was ordered from Dharmacon to target *EZH2* in MOLM13.Non-specific or scrambled siRNA of the same type (smartpool, Dharmacon) was used as control.

ENX-1 siRNA (Cat No: sc-35312) and control siRNA-A (Cat No: sc-37007) against *EZH2* was obtained from Santa Cruz Biotechnology to validate results generated by ON-TARGETplusSMART pool siRNA against *EZH2*.

2.3 shRNA used with pSUPER

EZH2shRNA1_F	5' - GAT CCC CAC CAG TTT GTT GGC GGA AGC GTG TCA AGA GCA CGC TTC CGC CAA CAA ACT GGT TTT TT - 3'	Forward primer for EZH2 shRNA at position 33
EZH2shRNA1_R	5' - TCG AAA AAA ACC AGT TTG TTG GCG GAA GCG TGC TCT TGA CAC GCT TCC GCC AAC AAA CTG GTG GG - 3'	Reverse primer for EZH2 shRNA at position 33
EZH2shRNA2_F	5' - GAT CCC CGA CTT CTG TGA GCT CAT TGC GCG TCA AGA GCG CGC AAT GAG CTC ACA GAA GTC TTT TT - 3'	Forward primer for EZH2 shRNA at position 213
EZH2shRNA2_R	5' - TCG AAA AAA GAC TTC TGT GAG CTC ATT GCG CGC TCT TGA CGC GCA ATG AGC TCA CAG AAG TCG GG - 3'	Reverse primer for EZH2 shRNA at position 213
EZH2shRNA3_F	5' - GAT CCC CCA GGC GCA CTT CCT CCT GAA TGT TCA AGA GAC ATT CAG GAG GAA GTG CGC CTG TTT TT - 3'	Forward primer for EZH2 shRNA at position 758
EZH2shRNA3_R	5' - TCG AAA AAA CAG GCG CAC TTC CTC CTG AAT GTC TCT TGA ACA TTC AGG AGG AAG TGC GCC TGG GG - 3'	Reverse primer for EZH2 shRNA at position 758
EZH2shRNA4_F	5' - GAT CCC CGC GGA AGA ACA CAG AAA CAG CTC TCA AGA GGA GCT GTT TCT GTG TTC TTC CGC TTT TT - 3'	Forward primer for EZH2 shRNA at position 935
EZH2shRNA4_R	5' - TCG AAA AAA GCG GAA GAA CAC AGA AAC AGC TCC TCT TGA GAG CTG TTT CTG TGT TCT TCC	Reverse primer for EZH2 shRNA at position 935

	GCG GG - 3'	

2.4 Patient samples

Bone marrow samples from 70 patients with MDS were collected from King's College tissue bank under project number PR039 and ethical consent of the patients was obtained in accordance with King's College guidelines for human sample collection.

2.5 DNA Techniques

2.5.1 DNA Extraction

DNA was extracted from patient samples using the QIAamp DNA purification kit from Qiagen. The bone marrow cells were disrupted using 200 µl of the cell lysis buffer (AL buffer) and pulse-vortexing, after resuspending them in 200 µl PBS and 20 µl Protease. Incubation at 56 °C for an hour was followed by addition of 200 µl 96-100% (v/v) ethanol. The mixture was applied to a QIAamp Mini spin column and washed with 500 µl of AW1 and AW2 buffer before eluting the DNA in 50 µl EB buffer.

2.5.2 Determination of DNA concentration

DNA concentration was determined by using Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer. The Nanodrop has an absorbance spectrum from 220-750 nm. The absorbance of the DNA solution (nucleic acid concentration) is measured at 260 nm. Typically, 1µl of DNA was placed on the nanodrop pedestal. Light absorbance by the sample is measured against that of the blank (EB/dH_20 buffer) to eliminate background absorbance. The ratio of absorbance at 260/280 nm is a good representation of DNA purity. A ratio of less than 1.8 indicates protein contamination while more than 1.95 suggests RNA contamination. DNA samples (n=61) with ratio of approximately 1.8 were used for sequencing.

2.5.3 DNA Amplification

The REPLI-g Midi Kits from Qiagen were utilised for amplification of samples from 70 patients. The samples were suspended in TE at a concentration of 100 ng. 5 µl of this suspension was added to 5 µl of the D1 buffer (containing KOH & EDTA to denature the DNA) was prepared according to Table 2.1, vortexed and incubated for 3 min at room temperature (RT).

Components	Buffer D1	Buffer N1
Reconstituted DLB buffer (500 µl H_2O + DLB buffer)	9 µl	-
Stop solution	-	12 µl
Nuclease Free water	32 µl	68 µl
Total Volume	41 µl	80 µl

Table 2.1: Preparation of D1 and N1 buffer.

Neutralization was done by adding 10 µl of buffer N1 (Table 2.1) and brief vortexing. Next, 30 µl master mix containing the REPLI-g DNA Polymerase (1 µl), and REPLI-g Midi Reaction Buffer (29 µl) and 20 µl of the denatured DNA was prepared.An isothermal amplification reaction was then carried out at 30° C overnight and the enzyme was later inactivated by increasing the temperature to 65° C for 3 min. A negative control without the template DNA was used to ensure absence of contamination. Manufacturer claims a DNA yield of 40 µg. The amplified product was stored at -20° C.

2.5.4 DNA quality Testing

Human DNAOK multiplex Polymerase chain reaction (PCR) kit from Web Scientific was utilised for assessing the quality of the amplified DNA samples. It consists of five primer pairs which amplify the DNA from five different regions as well as a positive control. If the DNA was partially degraded one or more PCR fragments were absent. Completely

disintegrated DNA sample produced only a positive PCR control product which was visualized on agarose gel. The components of the PCR and cycling profile are described in Table 2.2. The PCR product was loaded on 2 % TAE agarose gel and run alongside a 100 bp ladder for 25 min.

Components	Volume	
Human DNAOK mix	7.5 µl	Denaturation at 95°C for 5 min 30 Cycles: Step 1: 95°C for 30 secs Step 2: 63°C for 30 secs Step 3: 72° C for 45 secs
MegaMix˜ Gold	12.5 µl	
Sample DNA	5 µl	

Table 2.2: Preparation of Human DNAOK multiplex PCR master mix.

2.6 Primer Designing

Exon sequences for *EZH1, BMI1, SUZ12, RING1B*, *EED, DNMT3A*, *RING1A, PHC1*, *PHC2* and fifteen Jumonji genes were obtained from the Ensembl (http://www.ensembl.org/index.html). Primers for these genes were designed using default settings (Tm: 59° C) on Primer3 software (http://primer3.ut.ee/) and a product size of <350bp was achieved. A flanking region of atleast 50 bp was allotted on either side of the target sequence to enable accurate sequencing. Universal tag primers (refer appendix DVD) were incorporated into the primers. All the primers were obtained from Sigma or Integrated DNA technology.

2.6.1 Primer Testing and PCR optimisation

Dnase, Rnase free water was added to the lyophilized primers to make a stock solution of 100 µM.All the primers were tested on 10 ng of genomic control DNA (Affymetrix, UK) using PCR and analysed on 1/ 2 % TAE agarose gel alongside O'Gene Ruler DNA ladder (Thermoscientific,UK).Electrophoresis was carried out at 120V for 20 min in 1X TAE buffer.

a)

Components	Volume
2X Promega Go Taq (Colourless)	10 µl
Primer (5 µM)	2 µl
Sample DNA (10 ng/µl)	1 µl
H_20	7 µl
Total	**20 µl**

b)

Cycles	Temp	Time	Comment
1	95°C	2 mins	Enzyme Activation
35	95°C	30 secs	Denaturation
	60°C	30 secs	Primer annealing
	72°C	30 secs	Elongation
1	72°C	5 mins	Fill missing nucleotides

Table 2.3: a) PCR reaction mixture b) PCR programme on the thermocycler.

2.7 Mutational analysis by Sanger Sequencing

Amplicons were amplified using the primers and PCR programme detailed in table 2.3. The PCR products were cleaned using ExoSAP-IT from SLS (2 µl ExoSAP-IT + 6 µl of PCR product) (Table: 2.4).

Time	Temp	Comment
15 min	37 °C	Activation of (ExoSAP-IT) : Exonuclease I + Shrimp Alkaline Phosphatase and rSAP to remove phosphorylated ends of DNA+ single stranded DNA and dNTPs respectively.
15 min	80 °C	Inactivation of ExoSAP-IT

Table 2.4: ExoSAP-IT clean–up programme.

The purified product was used in another round of PCR to incorporate the dye into the DNA sequence.

a)

Reagents	Volume
5X sequencing buffer	2 µl
Big Dye	0.5 µl
5 µM Universal Primer (aUSF or aUSR)	2 µl
First round PCR product	3 µl
H_2O	2.5 µl
Total	10 µl

b)

Cycle No	Temp	Time	Cycles	Comment
1	96°C	1 mi	X 1	
2	96 °C	10 secs	X 15	Denatruration
	55 °C	5 secs		Primer annealng
	60 °C	1.15 mins		Elongation
3	96 °C	10 secs	X 5	Denatruration
	55 °C	5 secs		Primer annealng
	60 °C	1.30 secs		Elongation
4	96 °C	10 secs	X 5	Denatruration
	55 °C	5 secs		Primer annealng
	60 °C	2 mins		Elongation

Table 2.5: a) PCR reaction mixture b) PCR programme.

Purification of the post-sequencing reaction (10 µl) was carried out using 10 µl of the BigDyeXTerminator and 45 µl of the SAM solution from Applied Biosystems. The mixture

was vortexed on a plate mixer for 30 min followed by centrifugation for 2 min at 1000 xg. Sanger sequencing (Geneservice, Cambridge, United Kingdom) was carried out by ABI 3130X genetic analyzer and the sequences were analyzed for mutations using SeqScape Software (v2.5) and Sequencing Analysis software by Applied Biosystems (Foster City, CA).

2.8 Next-generation sequencing of PRC2, *DNMT3A* and Jumonji genes

2.8.1 Patient Cohort

Amplified DNA samples from the same cohort of patients were utilized for next-generation sequencing of the PRC2 genes; *EZH1, EED*, *SUZ12*, DNA methyltransferase 3A (*DNMT3A*) and fifteen Jumonji genes (*JMJD3, JMJD4, JMJD1B, JMJD2A, JARID2, JMJD1C, JARID1B, JMJD2C, UTX, JARID1C, JARID1A, JMJD2D, JHJD1A, JARID1D* and *JHJD1B).* PCR was performed on 10 ng of amplified DNA as described in table 2.3.Subsequently, a second PCR was performed with 1or 1.3 µl of the amplified product from the first PCR and primers (1 µl) targeted to the aUS primers (refer appendix). These primers contained a 5' tag that identified forward and reverse strands for sequencing and also included a 5-base pair Multiplex Identifier (MIDs) sequence to identify individual patients. The following PCR programme was used.

Cycles	Temperature	Time	Comment
1	94 °C	5 min	Enzyme activation
8	94 °C	35 sec	Denaturaion
	62 °C	40 sec	Primer annealing
	72 °C	45 sec	Elongation
1	72 °C	15 min	Inserting the missing nucleotides

Table 2.6: Programme for the second round of PCR for 454 sequencing.

Post-pcr amplified exons were quantified using picogreen dye (Invitrogen) on Rotor-Gene 6000 Multiplexing System from Corbett Research. The readings were verified with band intensity on 2 % agarose gel. All the exons per patient were equalized in concentration based on the picogreen measurement and pooled. Subsequently, a second a picogreen measurement was performed to equalize the concentration of all the patient samples.

2.8.2 Gel extraction and purification

The mixture was loaded on a 2 % agarose gel in TAE buffer and electrophoresis was carried out at 120V for 45 min. The band was purified using QIAquick Gel Extraction kitfrom Qiagen.Gel slabs were submerged in QG buffer, heated to 50 °C for 15 min and vortexed every 2 min during the incubation.10 µl of (3M) sodium acetate was added to balance the pH of the solution and 100 µl/ slab of 100 % isopropanol was added to increase the yield of DNA fragments <500 bp. 700 µl of the solution was applied to columns containing Silica membrane and centrifugation was carried out at 300 xg for 1 min. The column with the DNA adsorbed onto it was subsequently washed with 500 µl of QG buffer and 750 µl of PE buffer with ethanol to remove all the traces of salt, impurities etc. The DNA was then eluted using 50 µl of Elution buffer in new eppendorf tubes.

2.8.3 Bead Purification

AgencourtAMPure XP magnetic beads (Beckman Coulter (USA)) were used to re-purify the amplicons obtained from gel purification. 60 µl of the gel purified filtrate was mixed with 6 µl of Qiagen 10X buffer to adjust the pH. Magnetic beads (1.8 times) were mixed for every 1 µl of the mixture and incubated at room temperature for 15 min. The mixture was placed on magnet particle concentrator (MPC) [Magnarack, Invitrogen] and the supernatant was discarded. The beads were washed two times with 200 µl 70% (v/v) ethanol and eluted by incubating at RT for 5 min in 60 µl of 1X TE. The mixture was placed on the MPC for 5 min and the supernatant containing the purified amplicons was collected.

2.8.4 Library preparation

A third picogreen measurement and concomitant gel electrophoresis was carried out to confirm the final DNA concentration. The concentration was converted to 1×10^7molecules/µl by adding appropriate amounts of molecular grade water.

2.8.5 Sequencing

GSFLX Titanium LV emPCR Kit (Lib-A) [Catalogue no: 05619114001] from Roche Applied Sciences was used to perform the emulsion PCR on the purified amplicons.Live & mock amplification mix and the capture bead wash buffer were prepared as indicated in Table 2.7.

a)

Reagent	Volume
Molecular Garde water	1200 µl
emPCR additive	1500 µl
5X amplification mix	780 µl
Amplification primer A or B	230 µl
emPCREzzyme Mix	200 µl
Ppiase	5 µl
Total	**3915 µl**

b)

Reagent	Volume
5X Mock amplification mix	2000 µl
Molecular grade water	8000 µl
Total	**10,000 µl**

c)

Reagent	Volume
10XTW Capture Bead wash buffer	1000 µl
Molecular grade water	9000 µl
Total	**10,000 µl**

Table 2.7: Preparation of a) Live Amplification mix per capture bead solution b) Mock Amplification mix c) 1X Capture bead wash buffer.

DNA Capture Beads A (with forward primer) & B (with reverse primer) were vortexed and centrifuged for 1min at 400 xg. The supernatant was discarded and the beads were washed twice with 1 ml of 1X Capture Bead Wash Buffer TW. After the second wash, 50-60 µl of the supernatant was left behind to avoid the beads from drying out.

2.8.6 Emulsion PCR

10 µl of the library mixture at 1×10^7 molecules/ µl (purified before) were added to capture bead A & B each. Next, 3.915 ml of the Live amplification mix (A or B) was added to the respective capture bead (A or B) +DNA library solution. Concurrently, two cups of emulsion oil were installed on the TissueLyzer and shaken for two min at 28 Hz. 5 ml of Mock amplification was added to each cup. Micelle formation was facilitated by shaking the cups at 28 Hz for 5 min on the TissueLyzer.The Live amplification mix A + DNA library + capture beads A were added to one emulsion cup and similarly the solution for capture beads B was added to the other cup. The cups were mixed for 5 min at 12 Hz and the solution was then

dispensed (100 µl/ well) into 96 well plates using multistep dispenser pipette. PCR was carried out over night (Table 2.8).

Cycles	Temperature	Time	Comment
1	94 °C	4 min	Enzyme activation
50	94 °C	30 sec	Denaturation
	68 °C	30 sec	Primer annealing
	58 °C	4 min 30 sec	Elongation
	10 °C	∞	

Table 2.8: PCR conditions for clonal amplification in the emulsions.

2.8.7 Vaccum assisted emulsion breaking

Post-PCR emulsions were broken using the GS FLX Titanium emPCR breaking kitand breaking reagents [Catalogue number: 05233658001]. The breaking kit (shown in the figure) was attached to the vaccum source in the hood and the solution from the PCR plates was collected in tubes. Plates were washed three times with 150 µl of 100 % 2-propanol (v/v). Capture bead A and B solutions were mixed at this stage and the tubes were centrifuged for 10 min at 1000 xg. The supernatant was discarded and the pellet was washed with 40 ml [EF] Enhancing fluid, 100 % (v/v) 2-propanol (two times), 100 % (v/v) ethanol and again with EF by centrifugation at 1000 xg for 10 min. The pellet was transferred to eppendorfs and washed two times with 1 ml EF by centrifugation for 2 min at 1000 xg.

Figure 2.1: Assembled set-up for vaccum assisted emulsion breaking and bead recovery ***(Reference: Roche emPCR method protocol).***

2.8.8 DNA library Bead Enrichment

After the final wash, the supernatant was discarded and 1ml fresh Melt solution (19.5 ml of molecular grade water + 500 µl of 5M NaOH) was added followed by vigorous vortexing and incubation for 2 min at RT. The supernatant was discarded after centrifuging the eppendorfs at 1000 xg for 2 min and the denaturation step using the Melt solution was repeated again. The Melt solution was washed using 1 ml/eppendorf of Annealing buffer XT (two times) and centrifuging for 2 min at 1000 xg. The pellet was resuspended in 45 µl of annealing buffer. 12.5 µl of Enrichment primer A + 12.5 µl of enrichment primer B [from the GSFLX Titanium MV emPCR Kit] were added to each eppendorf and placed on a heating dry-block at 65° C for 7-8 min to allow primer annealing.The reaction was stopped by placing the tubes on ice for 2 min. The beads were washed (two times) using 800 µl EF and centrifuging at 1000 xg for 2 min. The bead pellet was then resuspended in 800 µl EF (*).

2.8.9 Preparation of Enrichment Beads

The eppendorfs containing enrichment beads were vortexed and placed on magnetic particle concentrator (MPC). The supernatant was discarded and the beads were washed three times

with 1 ml EF. After the final wash 80 µl/eppendorf of enrichment beads were added to the pellet suspended in enhancing fluid earlier (*). The eppendorfs were placed on the rotator for 15 min, pooled in 15 ml glass tubes and left on the MPC for 5 min and the supernatant was discarded. These washes were repeated 6-10 times with 3 ml EF till a clear supernatant was obtained.

2.8.10 Collection of Enriched DNA beads

Enriched beads were resuspended in 1.5 ml Melt solution (prepared earlier), vortexed and placed on MPC for 5 min. The supernatant was collected and the denaturation step was repeated once to ensure maximum yield. The beads were washed using 1 ml /eppendorf of Annealing buffer XT (two times) and centrifugation for 2 min at 1000 xg. The pellet was resuspended in 1 ml EF.

2.8.11 Determine Bead enrichment

3 µl aliquot of the beads was counted using Beckham Coulter Z1 particle counter.

Parameter	Dual Threshold
Upper Size	25 µm
Lower Size	15 µm
Count Mode	between
Aperture 100 µm Upper Size	100 µm

Table 2.9: Coulter Counter Z1 models settings for bead counting

$$\% \text{ Bead Enrichment} = \frac{\text{Number of enriched beads} \times 10^6 \text{ beads/cup} \times 100}{35}$$

2.8.12 Sequencing Primer Annealing

EF was discarded and the pellet was re-suspended in 200 µl annealing buffer. 25 µl of sequencing primer A [from the GSFLX Titanium LVemPCR Kit] + 25 µl of sequencing primer B was added to the eppendorf and placed on heating dry-block at 65° C for 7-8 min. The reaction was halted by placing the tubes on ice for 2 min. The library was washed using

1 ml /eppendorf of Annealing buffer XT (two times) and centrifugation for 2 min at 1000 xg and then stored in fridge overnight (ON) in 1 ml annealing buffer at 2°C to 8°C.

2.8.13 Preparation of Bead Buffer 2 (BB2)

Bead wash buffer (200 ml) from the Roche kit was placed on ice for 10 min. 1.2 ml of the Supplement CB and 34 µl of Apyrase was addedto make the Bead Buffer 2 (BB2).

BB2 (µl)	Polymerase Cofactor (µl)	DNA Polymerase (µl)	Total (µl)
1570	150	300	2020

Table 2.10: Components of DBIM buffer.

2.8.14 Bead Preparation [Titanium Sequencing Kit LR70 CatNo: 04932315001]

1) **Packing beads** (stabilize and maintain all the immobilized components of the system within the wells) were washed two times by vortexing, adding 1ml BB2 and centrifugation at 1000 xg for 5 min. Discard the supernatant after each wash. 550 µl BB2 was later added and the beads were left on ice.
2) **PPiase** (scavenge inorganic pyrophosphate (PPi) to reduce interference during each nucleotide flow) and **Enzyme** beads (sulfurylase and luciferase enzymes) were vortexed for 30 sec each and placed on MPC for 2 min. The supernatant was discarded. The PPiase beads were washed three times with 1 ml BB2 and left in 500 µl of BB2. Similarly, the enzyme beads were washed and left in 1 ml BB2. Both the bead solutions were left on ice.
3) **DNA + Control beads** (serve as an internal control for the sequencing reaction)**:** The libraries were thawed the next day and measured two times using the bead counter as described previously. 1 ml of new annealing buffer was added to the libraries. The appropriate amount of DNA library beads (2 million beads per region) were separated and

mixed with (20 µl / region) Amplicon control bead solution from Roche. The final volume should be approximately 50 µl or less. 950 µl/ region of DBIM was added to the mixture of DNA library + control beads and incubated at RT on the rotator for 15 min.

2.8.15 Layer Preparation

Layers 1, 3 and 4 were prepared according to the table in falcon tubes and placed on ice.

Bead Layer	BB2 (µl)	Enzyme Beads (µl)	PPiase Beads (µl)	Total (µl)
Layer 1- Enzyme beads	3250	550	----	3800
Layer 3- Enzyme beads	2500	1300	----	3800
Layer 4- PPiase beads	3340	-----	460	3800

Table 2.11: Dilution of beads for layer preparation.

Layer 2 was prepared by combining 265 µl of packing beads with 1000 µl of the Library+ control beads and 435 µl /region of BB2. The mixture was placed on the rotator for 5 min at RT.

2.8.16 Assembling the BDD

The Roche Pico Titre plate kit (Catalogue number: 05029082001) was washed and assembled as per manufacturer's instructions as indicated in the figure below by placing the PTP device onto the BDD base and aligning the notched corners.

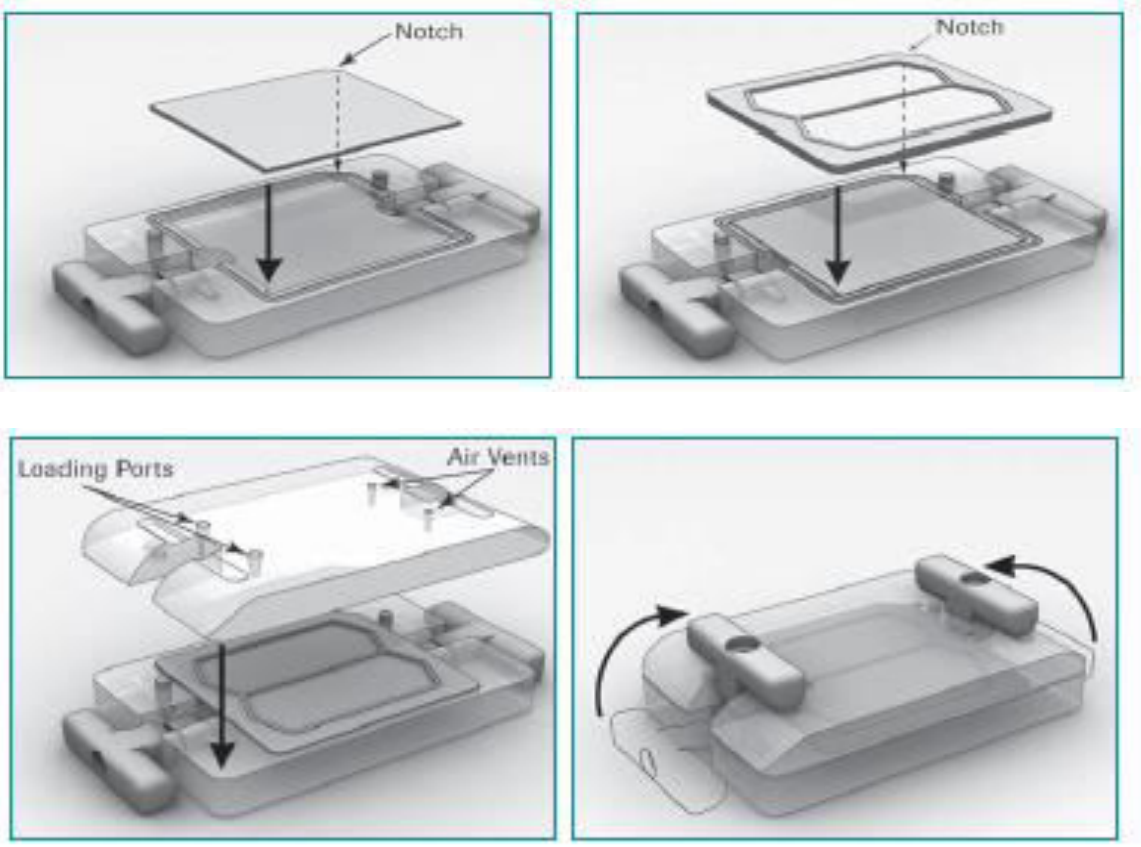


Figure 2.2: Assembling the BDD*(Reference: Roche emPCR method protocol).*

1.86 ml of each layer was loaded onto the PTP through the Loading ports. The plate was centrifuged for 10 min at 1620 xg. Previous layers were aspirated after centrifuging each time before loading the subsequent layer. After the last layer was aspirated out, the plate was removed from the gasket and loaded onto the 454 sequencer.

2.8.17 Run Requirements

One maintenance wash and two pre-washes were carried out before the run. Similarly, the PTP cartridge and camera face plate were cleaned and the sequencing kit reagents LR70 required for each run were loaded on the machine as per the manufacturer's instructions.

2.9 Total RNA isolation

Cell pellets in the range of 10×10^5-1×10^6 cells were resuspended in 1 ml of Trizol (Invitrogen) and vortexed vigorously. These pellets were either stored at -80 °C or the samples were processed immediately by incubating at RT for 5 min. The homogenised samples were mixed with 200 µl of chloroform and then centrifuged at 12,000 xg for 15 min at 4°C. The separated upper aqueous phase was transferred into a fresh tube. 500 µl of 2-Propanol was added to the aqueous phase and the mixture was incubated for 10 min at RT. The samples were centrifuged at 12,000 xg for 10 min at 4°C. The supernatant was discarded and the RNA pellet was washed with 1000 µl of 75% (v/v) ethanol at 7500 xg for 5 min. The supernatant was discarded and the RNA pellet was air-dried for 10-15 min. The pellet was dissolved in 30 µl of RNase free water and the concentration was determined using Thermo Scientific Nanodrop sphectrophotometer.

Alternatively, RNeasy Micro Kit - RNeasyMinElute Spin Columns were used for purification of concentrated total RNA when dealing with small numbers of cells as per manufacturer's instructions.

2.10 cDNA synthesis

100 ng-1µg of total RNA was reverse transcribed with the SuperScript VILO™ cDNA synthesis kit (Catalogue number: 11754050) according to manufacturer's protocol.

Reagent	Volume
10X SuperScript® Enzyme Mix	2 µl
5X VILO™ Reaction Mix	4 µl
RNA (up to 1µg)	x µl
Water	to 20 µl

Table 2.12: Preparation of master mix for cDNA synthesis

The above mix was incubated at 25°C for 10 min and then at 42°C for 120 min. The reaction was terminated for 5 min at 85°C. The cDNa samples were stored at -20°C. Depending upon the starting concentration of the total RNA, the cDNA samples were either diluted (1:25, 1:50 or 1:100) or used undiluted.

2.11 Real Time PCR analysis

Real time PCR was performed using FastStart Universal Probe Master mix and probes from Universal Probe Library (Roche Diagnostics). Each sample was run in triplicate to control for real-time PCR (qPCR) variations.

Reagent	Volume
FastStart Universal Probe Master mix	10 µl
cDNA	3-5 µl
Primer foward& reverse (20 µM)	1 µl
Probe	0.25 µl
dH_2O	Upto 20 µl

Table 2.13: Preparation of Real-time PCR master mix

The thermo cycling conditions on the StepOnePlus Real-Time PCR (Applied Biosystems):

Temperature	Time	No of cycles
50°C 95 °C	2 min 10 min	1 cycle
95 °C 60°C	15 sec 1 min	40 cycles

Table 2.14: PCR conditions for qPCR

For any given cDNA relative target abundance (fold change) was normalized against the house-keeping genes *GAPDH*, *TUBULIN* and *B2M* according to the formula:

$2^{-\Delta\Delta Ct}$ where $\Delta\Delta Ct = \Delta Ct_{\text{target}} - \Delta Ct_{\text{calibrator/wild type}}$.

$\Delta Ct_{target} = Ct_{target} - Ct_{GAPDH\ or\ B2M\ or\ TUBULIN}$,

$\Delta Ct_{\text{calibrator/wild type}} = C_{\text{calibrator/wild type}} - Ct_{\text{GAPDH or B2M or TUBULIN}}$

where Ct is the threshold cycle.

Analysis in patient samples (section 3.7.3.3 & 5.15)

After calculation of $\Delta Ct_{target} = Ct_{target} - Ct_{GAPDH\ or\ B2M\ or\ TUBULIN}$, one patient sample was used as calibrator to obtain the relative fold change for the remaining patient samples and plotted as a scatter plot.

Analysis in gene signature obtained in KO experiments (section 5.13.1)

$2^{-\Delta\Delta Ct}$ where $\Delta\Delta Ct = \Delta Ct_{\text{target}} - \Delta Ct_{\text{calibrator/wild type}}$.

$\Delta Ct_{target} = Ct_{target} - Ct_{GAPDH\ or\ B2M\ or\ TUBULIN}$,

$\Delta Ct_{\text{calibrator/wild type}} = C_{\text{calibrator/wild type}} - Ct_{\text{GAPDH or B2M or TUBULIN}}$

where Ct is the threshold cycle.

The calibrator sample is expressed as zero and the remaining samples are expressed relative to calibrator expression (positive or negative in case of KO)

2.12 Affymetrix Human Transcriptome 2.0 gene expression profiling

Total RNA was extracted from K562 cells on day 7 after transduction for shRNA against *EZH2* in lentivirus and at 72 hr post-transfection for the *EZH2* mutations (R690C & R690H).The quantity and purity of the starting RNA was assessed by the ND-8000 (Nanodrop) spectrophotometer, and the integrity confirmed by Agilent 2100 Bioanalyzer. The protocol recommended between 50 – 500 ng of starting RNA. For all the experiments, 100 - 150 ng of RNA was utilized. The workflow of the experiment is indicated in (Fig: 2.3)

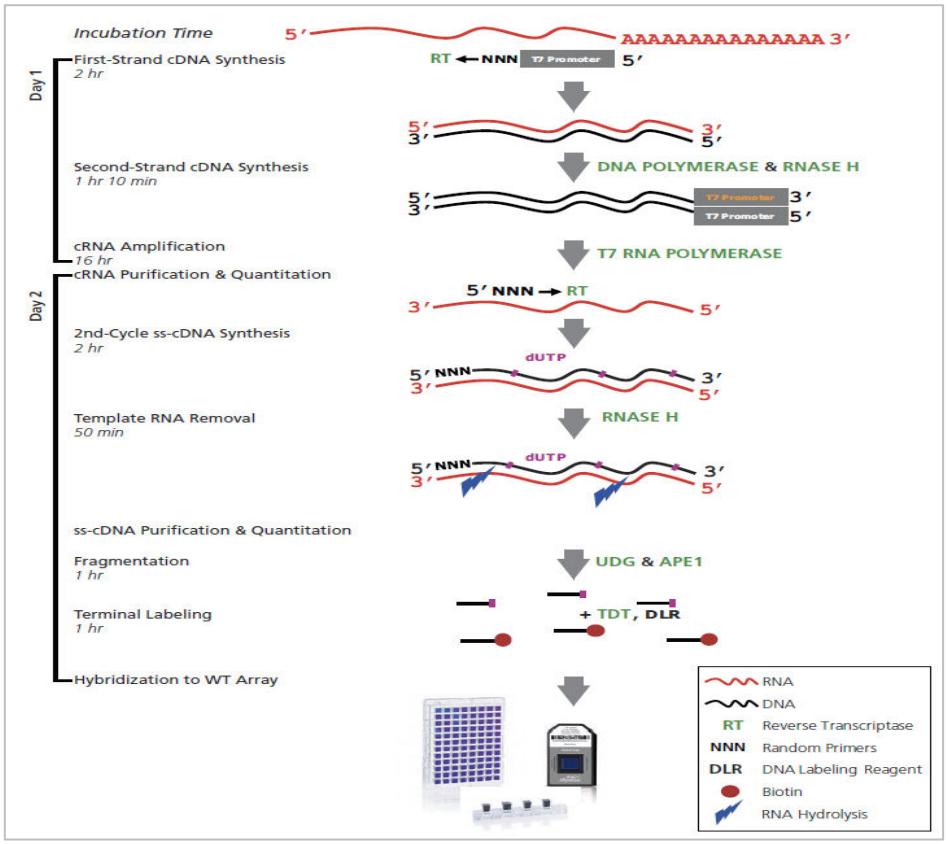


Figure 2.3: Work-flow for Affymetrix Human Transcriptome 2.0 gene expression array. Day 1 consists of first and second strand cDNA synthesis followed by cRNA amplification. Day 2 involves bead purification of the cRNA and subsequently ss-cDNA synthesis, purification followed by labelling and hybridization.

2.12.1 Reverse transcription to synthesise first-strand cDNA

All reagents were included in the WT amplification kit. Reactions were assembled on ice in 96 well plates.

First-Strand Matser mix:

Reagent	Volume/reaction
First-Strand Buffer	4 µl
First-Strand Enzyme	1 µl
Total Volume	5 µl

5 µl of the First-Strand Master Mix was then added to 5 µl of total RNA + H_2O mixture which was prepared as follows:

Reagent	Volume/reaction
Total RNA sample (100 ng)	variable
Nuclease-free water	variable
Total Volume	5 µl

Poly-A controls at this step were omitted to allow for the use of a greater volume of target total RNA (5 µl instead of 3 µl).

The two mixtures were mixed and incubated for 1 hr at 25 °C, then for 1 hr at 42 °C followed by at least 2 min at 4 °C. The lid was heated to 42 °C for the entire duration. Immediately after incubation the plate was centrifuged briefly before being placed on ice and immediately prepared for second strand cDNA synthesis.

2.12.2 Second-strand cDNA synthesis

The master mixes were assembled on ice as follows:

Reagent	Volume/reaction
Second-strand buffer	18 µl
Second-strand enzyme	2 µl
Total Volume	20 µl

20 µl of the above master mix was added to each (10 µl) first-strand cDNA sample for a final reaction volume of 30 µl. After mixing and brief centrifugation, the samples were incubated

for 1 hour at 16^{o}C, then 10 min at 65^{o}C, followed by at least 2 min at 4 °C before being placed at RT for ≥ 5 min and immediately prepared for *in vitro* transcription.

2.12.3 In Vitro Transcription (IVT) to synthesise cRNA

The IVT buffer was thawed at RT for ≥ 10 min and IVT master mix was prepared as follows:

Reagent	Volume/reaction
IVT Buffer	24 µl
IVT Enzyme	6 µl
Total Volume	30 µl

30 µl of the master mix was added to 30 µl second-strand cDNA sample for a final reaction volume of 60 µl. The mixture was vortexed briefly and incubated for 16 hr at 40 °C (lid temp: 40 ^{o}C). The samples were placed on ice and immediately prepared for purification.

2.12.4 cRNA Purification

The cRNA elution buffer (Nuclease-free dH_20) was pre-heated to 65^{o} C for at least 10 min. The purification beads were mixed thoroughly by vortexing. 100 µl/ reaction of the beads were added to 60 µl of the cRNA reaction and transferred to a 96 well plate. The reactions were mixed by pipetting up and down 10 times and incubated for 10 min. The cRNA samples bind to the beads during the incubation. The plate was transferred to a 96 well magnetic stand (Agentcourt Bioscience) to capture the beads. When the capture was completedand the beads formed pellets against the magnetic stand. The supernatant was carefully aspirated and discarded.While on the magnetic stand the beads were washed three times with 200 µl of 80 % ethanol and incubated for 30 sec. The beads were air-dried on the magnetic stand for 5 min and the cRNAwas eluted in 27 µl of preheated (65^{o}C) Nuclease-free water and quantified by ND-8000 (Nanodrop) spectrophotometer.

2.12.5 2nd-Cycle single-strand cDNA synthesis

On ice, 625 ng/µl cRNA was prepared. This is equivalent to 15 µg cRNA in a volume of 24 µl. This was combined with 4 µl of 2nd – Cycle Primers, mixed thoroughly and incubated for 5 min at 70^{o} C, then 5 min at 25^{o} C followed by 2 min at 4 ^{o}C (lid temp: 70 ^{o}C). After annealing the primers, the reactions were mixed briefly and centrifuged.

2^{nd} – Cycle ss-cDNA master mix was prepared as followed:

Reagent	Volume/reaction
2^{nd}- Cycle ss-cDNA buffer	8 µl
2^{nd} –Cycle ss-cDNA enzyme	4 µl
Total Volume	12 µl

On ice 12 µl of the master mix was added to each (28 µl) cRNA/2^{nd}-Cycle Primer reaction to get a final volume of 40 µl. After mixing briefly and centrifugation, the reactions were incubated for 10 min at 25° C, then 90 min at 42° C , 10 min at 70° C, followed by at least 2 min at 4° C (lid temp: 70° C).The samples were placed immediately on ice and prepared for Hydrolysis.

2.12.6 Hydrolyze RNA by RNase H

4 µl of RNase H was added to each (40 µl) 2^{nd}-cycle ss-cDNA sample for a final reaction volume of 44 µl. The reactions were incubated for 45 min at 37° C, then for 5 min at 95° C followed by atleast 2 min at 4° C (lid temp: 70° C) and then placed immediately on ice. 11 µl of Nuclease-free water was added to each reaction, mixed briefly and immediately prepared for purification.

2.12.7 Purification of 2^{nd}- Cycle single-strand cDNA

Nuclease-free dH_20 was pre-heated to 65° C for at least 10 min. The purification beads were mixed thoroughly by vortexing. 100 µl/ reaction of the beads were added to 55 µl of the 2^{nd}-cycle ss-cDNA sample and transferred to a 96 well plate. The samples were mixed with 150 µl of 100 % ethanol by pipetting and incubated for 20 min. The plate was placed on a magnetic stand to capture the beads. When the capture was complete (≥ 5 min), the beads formed pellets against the magnetic stand. The supernatant was carefully aspirated and discarded. While on the magnetic stand the beads were washed three times with 200 µl of 80

% (v/v) ethanol and incubated for 30 sec. The beads were air-dried on the magnetic stand for 5 min and the cRNA was eluted in 30 µl of preheated (65^{o} C) Nuclease-free water and quantified by ND-8000 (Nanodrop) spectrophotometer.

2.12.8 Fragmentation & Labelling single-strand cDNA

On ice, 176 ng/ µl of purified ss-cDNA (equal to 5.5 µg in a total volume of 31.2 µl) was mixed with 16.8 µl of fragmentation master mix which was prepared as follows:

Reagent	Volume/reaction
Nuclease-free dH_20	10 µl
10X cDNA fragmentation buffer	4.8 µl
UDG,10 U/µl	1µl
APE 1, 1,000 U/µl	1µl
Total Volume	16.8 µl

The samples were mixed gently and incubated for 1 hr at 37^{o} C, then for 2 min at 93^{o} C followed by at least 2 min at 4^{o} C (lid temp: 93^{o} C). The samples were placed immediately on ice. 45 µl of the fragmented ss-cDNA was transferred to new eppendorfs or wells for 96 well plates and mixed with 15 µl of the labelling master mix.

Preparation of Labelling master mix

Reagent	Volume/reaction
5X TdT Buffer	12µl
DNA labelling reagent, 5mM	1 µl
TdT, 30 U/µl	2µl
Total Volume	15µl

After mixing and brief centrifugation, the samples were incubated for 1 hr at 37^{o} C, then for 10 min at 70^{o} C followed by 2 min at 4^{o} C. Immediately after the incubation, the samples were briefly centrifuged and stored at -20^{o} C.

2.12.9 WT Array Hybrdization

DMSO was thawed prior to use. 20X Hybridization controls were denatured by heating for 5 min at 65^{o} C. Hybridization master mix was prepared at RT.

Preparation of Hybrdidization master mix:

Reagent	49 format Volume/reaction
Control Oligo B2 (3 nM)	3.7 µl
20X Hybridization Controls *(bioB, bioC, bioD,cre)*	11 µl
2X Hybridization Mix	110 µl
DMSO	15.4 µl
Nuclease-free water	19.9 µl
Total Volume	160 µl

The master mix was thoroughly homogenised and 160 µl was added to 5.2 µg (in 60 µl) of the fragmented and biotin-labelledss-cDNA samples. After mixing and brief centrifugation, the samples were incubated for 5 min at 99^{o} C, then for 5 min at 45^{o} C (lid temp: 99^{o} C). After incubation, 200 µl of the samples were injected and hybridized to HT 2.0 expression arrays using GeneChipR Hybridization Oven 645 (Fig: 2.4). A pipette tip was inserted into upper right septum for venting while the samples were loaded from the bottom left septum on the array.

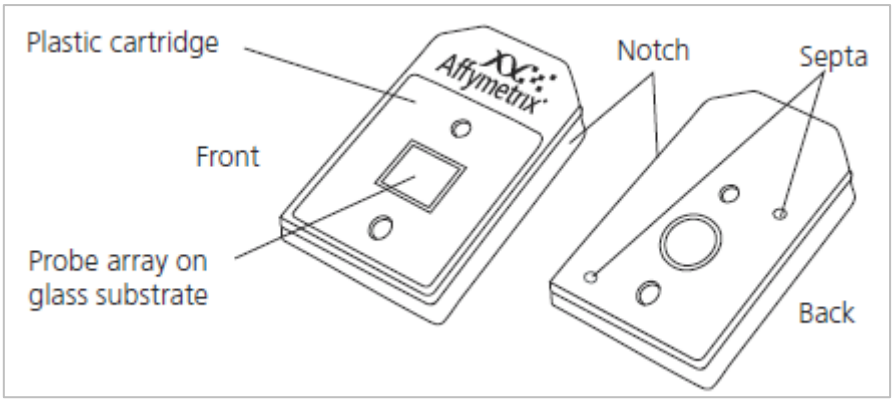


Figure 2.4: Components of the array chip. The square in the centre indicates the probe arrays.

After loading the samples, both septa were covered with tough spot stickers to minimize evaporation and to prevent leaks. The arrays were loaded on the hybridization trays and incubated with rotation at 60 rpm for 16 hr at 45^{o} C.

After 16 hours of hybridization the arrays were removed from hybridization oven and the hybridization cocktail was aspirated through the septum using a pipette. The array was refilled by 250 µl of wash solution A and placed into the GeneChipR Fluidics Station 450. Following instructions on the LCD of the fluidics station, the eppendorfs containing 600 µL Stain Cocktail 1, 600 µL Stain Cocktail 2 and 800 µL of Array Holding Buffer were placed on the holder. The needles were placed into position and the wash and stain was carried out for 90 min.

2.12.10 Expression microarray data analysis

Gene expression analysis was done under the guidance of Dr. Azim Mohamedali. The HT 2.0 arrays were scanned using a GeneChipR 3000 Scanner (7G upgrade)to capture images (DAT files). The raw data from the DAT file is converted to CEL files (based on intensities). ITERPLIER normalization was used to generate the CHP file. Outliers were filtered using Principal component analysis, area under the curve and a 1-way ANOVA was used to detect differential gene expression between the scrambled shRNA and shRNA targeting *EZH2* for the knock-down experiments. In case of *EZH2* mutations, the

gene expression in cells transfected with the mutant was compared with that of cells transfected with the empty vector. The data was filtered using a fold-change cut-off of 2 and an ANNOVA p value ≤0.05 to create a list of significantly up or down regulated genes.

2.13 Protein Blotting

2.13.1 Cell lysis

Samples of $2.5x10^5$ - $1x10^6$ MOLM13, K562 or KG1 cells were collected in 1.5 ml eppendorf tubes and centrifuged at 300 xg for 5 min and the supernatants were discarded. The tubes were centrifuged at 300 xg for a few seconds to remove the traces of media were removed with a P200 pipiette. The dry pellets were flicked to disaggregate the cells and resuspended in 25-40 µl Nupage LDS sample buffer (1X) + 10 % 2-Mercaptoethanol (preheated to 90°C) depending on the cell count. Following vigorous mixing, the samples were boiled at 100°C for 5 min. Samples could be stored indefinitely at -20°C at this stage. Samples were reheated at 100°C for 15-30 min prior to separation by SDS-PAGE to denature the proteins completely. Each tube was centrifuged at 13,000 xg for 5 min and the supernatant was utilized for SDS-PAGE separation.

2.13.2 Western blot analysis

Following protein sepperation by SDS-PAGE, western blotting was carried out on the samples. Nu-polyacrylamide gel (4-12 % Bis- Tris Gel) was placed in the electrophoresis tank which was filled with 1X running buffer. Protein samples were loaded along with Novex sharp pre-stained protein standard (Novex) to determine protein size. Electrophoresis was carried out at a constant voltage (200V) until the dye reached the bottom of the gel. Next, the NuPageBis-Tris gel was placed under a nitrocellulose-coated nylon membrane (Amersham HybondECL) and then sandwiched between blot paper (Biorad) and nylon pads soaked in NuPage transfer buffer. Proteins separated in the gel were transferred onto the nitrocellulose

membrane by electrophoresis at 20 A, 40V for 120 min. The histones contained in the gel were stained with 0.5 % (w/v) colloidal coomassie blue, 10 % (v/v) acetic acid and 10 % (v/v) methanol as positive controls for the experiment. The gel was destained (20 % (v/v) methanol, 10 % (v/v) acetic acid and 70 % (v/v) distilled water visualised in white light for histones. Where indicated γ- tubulin, or actin was used as a loading control. After the transfer, the nitrocellulose membrane was blocked with 5 % (w/v) dried skimmed milk (Marvel) diluted in PBS- 0.1 % (v/v) Tween 20 and rinsed with PBS- 0.1 % (v/v) Tween20. The primary antibodies for the various proteins were used in the dilution indicated previously in 1% (w/v) dried skimmed milk, PBS 0.1% (v/v) Tween 20 overnight at 4 °C. Membranes were subsequently washed two times in PBS-0.1% (v/v) Tween20 for 15 min. This was followed by incubation for 1 hr with the (1:1000, 1:2000, 1:5000) human anti-rabbit or anti-mouse or anti-goat secondary antibody in 1% (w/v) dried skimmed milk, PBS 0.1% (v/v) Tween 20 at RT. Membranes were washed three times in PBS- 0.1 % (v/v) Tween20. Bound antibody was detected using Amersham ECL prime WesternBlot chemilumiscence detection solution. Post-incubation for 5 min, the membranes were sealed in Saran wrap and exposed to Hyper film ECL for the preferable time and subsequently developed.

2.14 Cloning

2.14.1 Restriction Digests

Restriction digest reaction contained 2 µl of 10X restriction enzyme buffer, 0.5 µl of restriction enzyme (2000 units), 2 µl BSA (when needed), gel extracted or purified DNA (300 ng-1 µg) and dH_20 in a total volume of 20 µl. When two restriction enzymes were used simultaneously in a digestion reaction, the buffer indicating maximum activity for both enzymes was used or else a single digest was carried out initially with the enzyme requiring less salt concentration followed by a second digest in the same mixture with the other enzyme. 1 mM NaCl solution was added to adjust the salt concentration during the second

digestion step. The reactions were incubated at 37 °C for 1-2 hr and heat inactivation of enzymes was carried out 65 °C for 20 min. Digested plasmid DNA was analysed by agarose gel electrophoresis.

2.14.2 Conversion of extended DNA termini to blunt ends

T4 DNA polymerase was used for the conversion of both 3'- and 5' extended termini into blunt ends. Using a primer and a template, this enzyme catalyses the synthesis of DNA from the 5'to 3' direction. T4 DNA polymerase also has a 3' to 5' exonuclease activity that can be utilized to remove 3'- overhangs to generate blunt ends.

Reaction mix:

Reagents	Volume
2 mM dNTPs	1.5 µl
NEB 1/2/3/4	1 µl
T4 DNA polymerase	1 µl
H_20	Upto 30 µl

The mixture was incubated for 15 min at 12°C and the reaction was inactivated by heating the mixture for 20 min at 70°C.

2.14.3 Isolation and purification of DNA from agarose gels

Agarose gels were prepared at 1 % w/v agarose (Melford) to 99 ml 1XTris-acetate/EDTA electrophoresis buffer (TAE). 1X TAE was made from a stock solution of 50 X stock solution concentration. The agarose was dissolved by boiling the mixture and then ethidium bromide (Sigma) was added to a final concentration of (0.5 µg/ ml) after cooling down. The mixture was poured into an electrophoresis tray containing combs to create slots and the edges were sealed with tape. The gels were allowed to set at RT for 15 min. The combs were removed and the gel was placed in an electrophoresis tank containing 1X TAE. DNA samples were

prepared by addition of 1µl of 6X Orange loading dye (Fermentas)/ 6 µl of sample.DNA molecular weight marker (O'GeneRuler DNA Ladder Mix, 100-10,000 bp) was used as a reference for DNA fragment size. The gels were electrophoresed at 120V and visualized on a UV transilluminator. The purified fragments were treated with SuperTaq^{TM}as per the manufacturer's instructions to add dNTP "A" to the fragment.

2.14.4 Topo cloning

The gel purified fragment containing the mutation was cloned into the pCRII-TOPO vector (Invitrogen). The following reaction mixture was prepared:

Reagent	Volume
PCR fragment (14-24 ng/ µl)	2 µl
Salt	0.5 µl
pCRII- TOPO vector	0.5 µl

Table 2.15: Reaction mixture for topo cloning

The above the mixture was incubated for 5 min at RT and then transformed into DH5α^{TM}-T1^{R}*E.Coli*.

2.14.5 Transformation into DH5α-T1 *E.Coli*

Chemo-competent DH5α-T1*E.Coli*cells (Life Technologies) were used for transformations. The cells were thawed on ice for 4-5 min. 1 µl of the ligation mixture was added to cells and the cells, covered in ice and incubated for 12-30 min. The cells were heat shocked at 42°C for 30 seconds on a water bath and immediately transferred to ice for 2 min. 350 µl of LB media at RT was added to the cells and were then incubated for 37°C for 30 min- 1 hr. 100 µl of the LB media containing the cells were spread on ampicillin (100 mg/ ml) containing LB agar plates. The rest of the mixture was centrifuged briefly and resuspended in fresh 100 µl LB media and spread on a second plate.

2.14.6 Ligation

Concentrations of the gel purified components of the plasmid were determined using (Thermo Scientific) Nanodrop sphectrophotometer. In the ligation reactions a molar insert: vector ratio of (2:1) and 0.5 µl 400 U of T4 DNA ligase (New England Biolabs), 1X ligase buffer (New England Biolabs) and dH_20 in a total volume of 10 µl was used. The reactions were incubated for 1-5 hr at RT or ON at 4° C and then transformed in to DH10B competent cells.

2.14.7 Transformation into DH10B cells

MegaX DH10B™ $T1^R$Electrocomp™ Cells were used for difficult transformations. 40 µl of DH10B cells were thawed on ice for 4-5 min. 0.1-0.3 µl of the ligation mixture was added to the cells and the mixture was transferred to 0.1 cm electroporation cuvettes (Invitrogen).All seven bacterial transformation programs were tested using DH10B cells and Lonza AmaxaNucleofactor II device. Program 4 was chosen to transform DH10B cells because it produced the maximum colonies. 1 ml of LB media at room temperature was immediately added to the cells after the electroporation procedure. The media containing the cells were transferred to 1.5 ml eppendorf tubes and allowed to recover at 37°C for 1 hr. After an hour, 100 µl of the LB media containing the cells were spread on ampicillin containing LB agar plates. The rest of the mixture was centrifuged briefly and resuspended in fresh 100 µl LB media and spread on a second plate.

2.14.8 Miniprep: Alkaline lysis method

Plasmid DNA was isolated from 3 ml bacterial culture by using the Pure Yield Plasmid (Promega) alkaline lysis method as described below. Bacterial colonies were picked from LB agar plates to inoculate 3 ml LB medium (ampicilin). The bacterial cultures were grown ON onInnova 4300 incubator shaker at 37° C. The grown cultures were pelleted for 10 min at 2340 xg (Rotanta 460R Hettich centrifuge) and resupended in 600 µl of PBS. 100 µl of cell lysis buffer was added and mixed by inverting 6 times. This solution was neutralized using

350 µl of neutralization buffer (4-8° C) and mixed by inverting 6 times gently. The mixture was centrifuged at 13000 rpm for 3 min and the supernatant was transferred to PureYield™Minicolumn. The columns were centrifuged at 13000 rpm for 15 seconds. The follow through was discarded and 200 µl of Endotoxin Removal Wash was added to the column. The column was centrifuged at 13000 rpm for 15 seconds and the follow through was discarded. The column was washed with 400 µl of Column wash solution by centrifuging at 13000 rpm for 30 seconds. The plasmid DNA was eluted by incubating the column for 1 min at room temperature with 30 µl of Elution buffer followed by centrifugation at 13000 rpm for 15 seconds. The eluted DNA was stored at -20° C.

2.14.9 Miniprep: Wizard Plus SV miniprep

Plasmid DNA could also be isolated from bacterial cultures using the Wizard Plus SV miniprep (Promega) as described below. The bacterial cultures grown ON on the orbital shaker were centrifuged at 2340 xg for 5 min. The cell pellet was resuspended in 250 µl of resuspension solution. 250 µl of cell lysis buffer & 10 µl of alkaline protease were mixed with the resuspended bacterial pellet. The mixture was incubated at RT for 5 min and the reaction was neutralized by adding 350 µl of cold neutralization (4-8 °C) buffer. The solution was centrifuged at 13000 xg for 10 min. The supernatant was transferred to the Miniprep spin columns and centrifuged for 1 min at 13000 xg. The follow through was discarded and the column was washed two times with 750 µl and 250 µl of column wash solution by centrifuging for 1 min and 2 min respectively. The plasmid DNA was eluted by incubating the column for 1 min at room temperature in 100 µl of Rnase free water and centrifuging at 2340 xg for 1 min. The eluted DNA was stored at -20 °C.

2.14.10 Maxiprep: Qiagen Plasmid kit

When large amounts of high purity plasmid DNA was required, it was obtained by using Qiagen plasmid purification kit. A single colony was selected for inoculation in a 100 ml

bacterial culture in LB media (ampicilin). This culture was grown ON as previously mentioned. The bacteria were harvested by centrifugation for 10 min at 2340 xg at 4° C. The pellet was completely resuspended in 12 ml of resuspension/ Rnase A solution. Lysis solution 12 ml of was immediately mixed to the suspension and mixed gently by inverting. The mixture was incubated for 5 min at RT and then neutralized by adding 12 ml cold neutralization buffer (4-8° C). A white aggregate consisting of cell debris, proteins etc is formed, to which 9 ml of Binding solution was added. This solution is immediately poured into the barrel of a syringe and allowed to incubate for 5 min. In the meanwhile, the Genelute HP Maxiprep Binding Column was prepared by adding 12 ml of Column preparation solution and centrifuged at 3000 xg for 2 min. Cell lysate from the syringe barrel was filtered onto the binding column and centrifugation was carried out at 3000 xg for 2 min. The column was washed two times with 12 ml of Wash solution A and B by centrifugation at 3000 xg for 2 and 5 min respectively. The follow through and the collection tubes were discarded after the last wash and the column was transferred to new 50 ml collection tube. The plasmid DNA was eluted by adding 3 ml of Elution buffer to the column and centrifugation at 3000 xg for 5 min. The concentration of the purified plasmid DNA was estimated using the Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer as described previously.

2.15 Cell culture

2.15.1 Culturing non-adherent myeloid cells lines

Three cell lines were used for transfection i.e. K562 (chronic myelogenous leukemia), MOLM13 (acute monocytic leukemia) & KG1 (acute myelogenous leukemia). The cells were always kept at 2 x 10^5 viable cells/ml for KG1 and 1 x 10^6 cells/ml for MOLM13 and K562 by diluting with fresh RPMI-1640 medium supplemented with 1 % Penicillin/Streptomycin (Sigma, Aldrich) and 10 % Fetal calf serum (Sigma, Aldrich).

2.15.2 Transfection using Lipofectamine LTX reagent

MOLM13 and K562 cells were seeded in 12 or 6 well plates at a density of 2 x 10^5/ 5 x $10^5$and transfected with plasmid DNA pre-incubated with Lipofectamine LTX (4µl-12µl) and Plus reagent (1µl-2.5µl) according to the manufacturer's instructions. Lipofectamine forms lipid complexes when mixed with plasmid DNA. These complexes are taken up efficiently by the cells when added to them. The transfected cells are selected by neomycin: G418 [Sigma, stock conc: 50 mg/ml] at a concentration range of 200-800 µg/ml or puromycin 2 µg/ml for K562 and 0.2 µg/ml for MOLM13 cells.

2.15.3 RNAiMAX transfection protocol for siRNA experiment

0.25x10^6 cells/well were seeded in 24 well plate. 3 µl of RNAiMAX (Life Technologies) solution was mixed with 100 µl of OptiMEM serum free medium and siRNA at varying concentrations (50-200 nM). The mixture was incubated for 30 min at RT and added drop wise to the seeded cells.

2.15.4 Electroporation

MOLM13 and K562 cells were seeded at a density of 2 x10^6 in 6 well plates and transfected with siRNA from Dharmacon (final conc.150 nM & 200 nM) resuspended in 100 µl of MirusBio Ingenioelectroporation solution using the Nucleofactor device II (program: O-017 for MOLM13 and T-016 for K562). A non-targeting scrambled siRNA was used as a control.

2.16 Knockdown of *EZH2*

2.16.1 Constructing shRNA in pSUPER vector

The exon sequence of *EZH2* (Ensemble: ENST00000320356) was used to construct shRNAs using website (www.genelink.com).Four shRNAs were selected against *EZH2*.

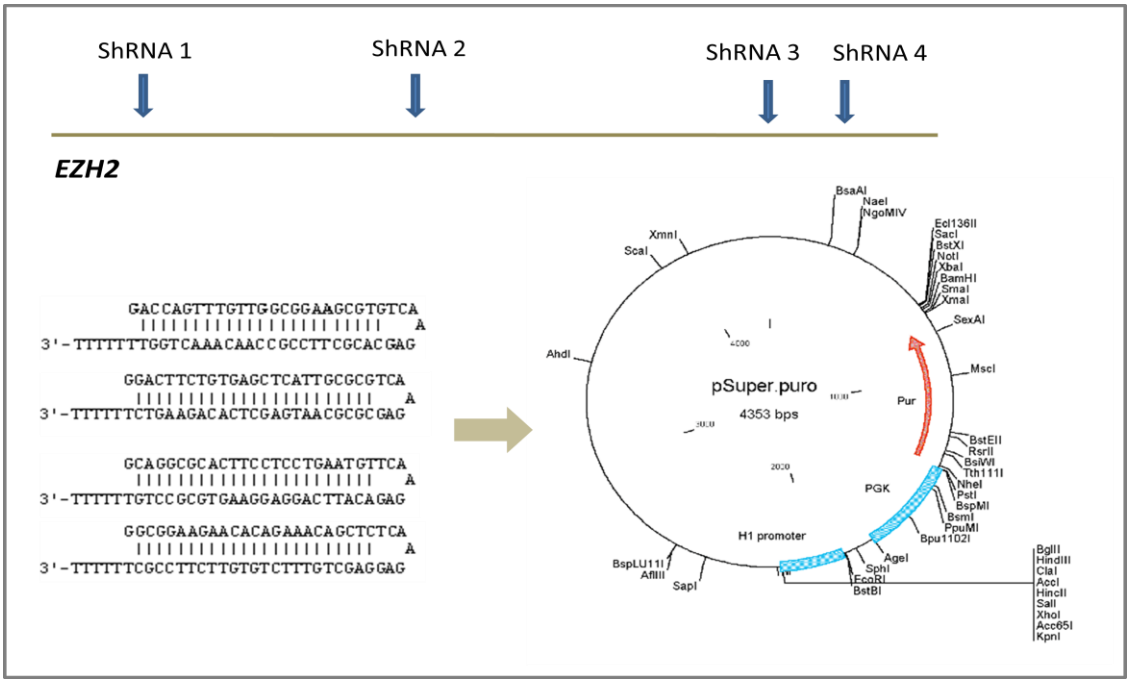


Figure 2.5: Four shRNAs were designed against different positions along the gene(Top part of the figure with arrows indicates the target region on *EZH2*). shRNA3 & 4 target the SET domain of *EZH2*. The shRNA sequences are indicated along with the plasmid map and multiple cloning sites were they were cloned.

The sequences for BglI (gatcccc) and XhoI (ctcgag) were added to the 5' and 3' end of the shRNA to promote ligation to the pSUPER vector. The pSUPER vector was gel purified using the QIAquick Gel Extraction Kit. The plasmid was digested by BglI and XhoI as given below:

Reagent	Volume
pSUPER (200 ng/ µl)	4 µl
XhoI	0.5 µl
BglI	0.5 µl
NEB3 buffer	2 µl
BSA 10X	2 µl
Total	Make up to 20 µl with dH_20

The digested product was gel and bead purified (as before). Meanwhile the hair-pin shaped shRNA sequences were linearized by mixing 5µl each of the forward and reverse primers (5µM) and incubating them in water heated to 90 $^{\circ}$C. The shRNAs were left in the water till it

cooled down to 37° C. Subsequently, they were ligated to pSUPER using 1µl of 10X ligation buffer and 1 µl of T4 ligase. The concentration of pSUPER to the shRNA insert in the ligation mixture was calculated using the formula:

$$\text{Insert concentration} = \frac{\text{Insert size} \times \text{vector concentration}}{\text{vector size}} \times 2$$

1µl of the ligation mixture was transformed into the chemo-competent DH5α™-T1^{R} *E.Coli* cells. Colonies were picked, and the insertion of thee shRNA was confirmed using M13 primers by Sanger sequencing and by XhoI-PstI digest on the maxiprep. This maxiprep was transfected into MOLM13 and K562 cells using lipofectamine or electroporation.

2.17 Lentivirus short-hairpin RNA (LVshRNA) virus particle production & concentration of the lentivirus vector

Eight shRNAs against *EZH2* were obtained of which four GFP labelled GIPZ EZH2 were targeting Lentiviral shRNAs (V2LHS_63068, V2LHS_17510 V2LHS_17507, and V2LHS_387505) from Fisher scientific and four shRNAs (TL304713) were obtained from Insight Biotechnology, UK.Lentiviral vector particles were generated by co-transfection of 293T cells with four plasmids; pMDG (envelope plasmid encodes VSVG), pRRE (packaging plasmid includes Gag & Pol), pREV (packaging plasmid also includes REV) and the shRNA against *EZH2* using a $CaPO_4$ co-precipitation method.

293T cells were seeded at a density of $1x10^7$/ 1000 cm 5 layer flasks [T1000 flasks, Millipore] using DMEM (10 % FCS, 1 % Pen/Strep) media. The cells were incubated ON at 37° C and 5 % CO_2. Four plasmid mixture containing pMDG (45 µg), pRRE (75 µg), pREV (37.5 µg) and shRNA (50 µg) were mixed and diluted to a total volume of 3.75 ml with dH_2O. 3.75 ml of 0.5 M $CaCl_2$ was added to the mixture. Subsequently 7.5 ml of 2X HEBS was added drop wise to the mixture while vortexing and incubated for 30 min. 15 ml of the total volume was diluted in 150 ml of DMEM (10 % FCS, 1 % Pen/Strep). This was added to

the 293T cells and the cells were incubated for 16 hr at 37° C and 5 % CO_2. The medium was replaced with fresh DMEM (10 % FCS, 1 % Pen/Strep) and incubated for further 24 hr. Supernatants containing the vector were collected 48 hr post transfection and concentrated at 10,000 xg over night at 8°C. The pellets were resuspended in 500 µl of RPMI-1640 medium and stored at -80° C.

2.18 Mtt cell proliferation assay

Tetrazolium salt [3-(4, 5-Dimethylthiazol-2-yl) 2,5diphenyltetrazolium- bromide] MTT assay is based on determining the metabolic activity of the cells and was used to assess cell proliferation rate. MTT (yellow) is reduced by NAD(P)H dependent oxidoreductase enzyme in metabolically active cells to generate purple insoluble formazan products. This assay was performed by seeding cells in 96 well plates (Flat bottom) at a density of 30000 cells/ 100µl of RPMI. Each sample was plated in triplicate and three wells were filled with medium alone and were used as a blank control. 20 µl of 5 mg/ml MTT was mixed with sample from each well and the plate was incubated for 2-4 hr in a humidified incubator 5% CO_2 at 37° C. After incubation, 150 µl of solubilisation buffer was added to each well (including control wells) and the plate was incubated at 5% CO_2 at 37° C in the dark for further 2 hr. The absorbance of each well was measured at 610 nm using a microtiter plate reader. Data was analysed by calculating the average value from the triplicate readings, normalizing the mean from each sample to the blank (sample - blank) and calculating the fold change compared to the sample containing the wild type cells. Cells were seeded for different time points and readings were measured at 24, 48 and 72 hr to calculate the proliferation rate of the samples.

2.19 Cell Cycle analysis

2.19.1 Cell cycle staining

1 $\times 10^{6}$ cells were centrifuged at 300 xg for 5 min and washed once with PBS. The supernatant was discarded and the cells were resuspended in 1 ml of 70 % (v/v) ice-cold ethanol by vortexing. The cells were stored at -20° C. For cell cycle analysis, the cells were pelleted by centrifugation at 300 xg for 10 min and the ethanol was carefully aspirated out. The pellet was dried for 10 min by inverting the eppendorf tubes. The cells were resuspended in 400 µl of the staining solution and incubated for 30 min at 37° C followed by analysis using flow cytometry.

Staining solution:

Reagent	Volume
Propidium Iodide (dye binds DNA)	400 µl
FITC (binds lysine residues of protein)	50 µl
RNAase A	100 µl
PBS	Upto 10 ml

2.19.2 Cell cycle analysis by flow cytometry

The relative DNA and protein content of the cells were analysed on BD FACS CantoII flow cytometer using FlowJo. Double cell were discarded from the analysis using doublet discrimination gating strategy and PI (area) histogram was generated. The cell fractions were categorized according to cell size and DNA content. Cells in G_0 have 2n DNA content, while those in G_1 are larger than G_0 but still have 2n DNA. The cells in S phase have a DNA content in the range of 2-4n while those in G2/M have 4n DNA. This difference in DNA content helped to segregate the cells into the different phases of the cell cycle and ascertain whether the KO of *EZH2* had an impact on a particular phase of the cell cycle.

2.20 Annexin V staining

Annexins are a family of calcium-dependent phospholipid-binding proteins that preferentially bind phosphatidylserine (PS). PS is predominantly located on the inner surface of the plasma membrane under normal conditions. Upon initiation of apoptosis, it is translocated to the extracellular membrane of the cells marked for phagocytosis. Annexin V staining facilitates distinction of cells in early apoptosis versus those in the late stages of apoptosis. Cells in early apoptotic phases have an intact cell membrane and therefore exclude dyes like propidium iodide (PI) making it impossible to segregate cells between early and late apoptotic phases using these methods (Koopman et al., 1994). Annexin V staining was carried out on cell pellets (2-5 $x10^5$) washed with ice cold PBS and centrifuged at 300 xg for 5 min. Next, they were washed with 1X Binding buffer, and pelleted at 300 xg. The pellet was resuspended in 250 µl of 1X Binding buffer and stained with 3-5 µl of Annexin V, based on the cell number, and incubated in the dark at room temperature for 10 min. The stained cells were washed twice with the binding buffer, centrifuged at 300 xg for 10 min and resuspended in 250-500 µl binding buffer. Before analysing the cells on the flow cytometer, the cell suspension was stained with 3 µl of fluorescently labelled 7-AAD. 7-AAD can penetrate the membrane of dying cells and stain them to discriminate between dead and live cells and in apoptotic assay it is used to distinguish necrotic cells from apoptotic cells. FACS analysis was performed in FACS Canto II and the data was analysed using FlowJo.

3 Chapter: SNP6 and Mutational Analysis of PRC1, PRC2, DNMT3A & JUMONJI genes

3.1 Introduction

3.2 Involvement of PRC subunits, DNMT3A and the Jumonji genes in MDS/AML

Aberrations of *EZH2* were first identified by Ernst et al (2010) using SNP 6.0 arrays on peripheral blood /bone marrow samples from patients with MDS to discover acquired uniparental disomy of 7q and 7q36.1 microdeletion. Mutational analysis of this region revealed monoallelic or biallelic *EZH2* mutations (detailed in section 1.7). The correlation of *EZH2* and MDS suggested the involvement of other components of PRC2; *SUZ12*, *EED* and *EZH1* in this disease. Apart from the PRC2 subunits, PRC1 complex is also involved in transcriptional regulation of target genes via ubiquitination of H2AK119 (detailed in section 1.5.2 and 1.8), infact PRC1 is recruited via H3K27me3 marks caused by PRC2. Some of PRC1 genes like *BMI-1, CBX8* are associated with MDS and MLL induced leukaemogenesis (detailed in section 1.7) but mutational analysis of the PRC1 genes (twelve) has not been carried out till date.

3.3 PRC and interacting proteins

3.3.1 DNMT3A

PRC2 interaction with *DNMT3A* was illustrated by Vire et al (2006) wherein the H3K27me3 mark caused by PRC2 function as sign-posts for the recruitment of *DNMT3A*. The SET domain of *EZH2* bind to *DNMT3A*, affecting DNA methylation. *DNMT3A* in turn is responsible for further "deep silencing" of the target regions (Vire et al., 2006). The first association between *DNMT3A* and myeloid malignancies was established by Yashimata et al (2010) using DNA sequence arrays on leukemic genome and identified heterozygous mutations of *DNMT3A* (Yamashita et al., 2010).These findings were confirmed by Ley et al (2010) when they used parallel sequencing to identify somatic mutations of *DNMT3A,* in

AML patients with a normal karyotype. 22% of the patients carried *DNMT3A* mutations and the most frequent mutations were observed in exon 23 at position R882 (59 %) leading to R882H/ R88C or P904L amino acid substitutions. Other less common mutations included exon 19 causing L737R, R771L, S770W substitutions and rarely, exon 16, 10, 6 and 7 were also involved. Their study concluded that the *DNMT3A* mutations were enriched in patients with intermediate-risk cytogenetic profile and were associated with an adverse effect on the overall survival of the patients (12.3 months vs. 41.1 months, $p < 0.001$) (Ley et al., 2010). The stability of *DNMT3A* mutations during disease evolution was demonstrated by Ewalt et al (2011) by mutational analysis on the bone marrow samples from 62 high-risk MDS patients (Ewalt et al., 2011). The incidence of mutations in the high-risk MDS patients was low (3.3 %) as compared to their occurrence in AML but the mutation at the R882 site was consistent even in this cohort, strongly implicating this mutation in disease pathogenesis.

Mutational analysis (till 2010) had not focused on *DNMT3A* mutations in relation to *EZH2* status in patients or examined the mutational overlap between the two genes which could provide an important link between transcriptional silencing by DNA and histone methylation.

3.3.2 Jumonji

The Jumonji (jmj) family of histone demethylases are α-ketoglutarate-dependent oxygenases that demethylate lysine residues in histones in a site and methylation state specific context. Interaction of the Jumonji protein (*JARID2*) with PRC2 core components was explained in section 1.3.1 where it co localizes and assists the binding of PRC2 to H3K27me3 but simultaneously also inhibits its methyltransferase activity in embryonic stem cells (ES) (Peng et al., 2009; Shen et al., 2009).*UTX* is another important member of this group as it has an opposing function to *EZH2* and causes demethylation of both di and tri-methylated H3K27 but does not act as efficiently on the mono-methylation of H3K27 (Smith et al., 2008). Lack

of *UTX* causes enhanced levels of H3K27me3, reduced mRNA levels of the target genes (Tsai et al., 2010) and has been implicated in tumorogenesis (van Haaften et al., 2009).Mutational analysis of *UTX* was carried out by Jankowska et al (2011) on 72 patients with CMML and MDS/MPN. 8 % of the cases were mutated at the *UTX* locus. These heterozygous and homozygous mutations might have an impact on the demethylation of the histones mentioned above and in turn govern the expression of target genes (Jankowska et al., 2011).

Other members of the Jumonji family have not been studied in detail in the context of MDS/AML and therefore in this study I have focused on seventeen Jumonji genes; *JARID1A-1D, JARID2, JMJD3, UTX, JMJD4, JMJD2A-2D, JMJD1A-1C, JHJDM1A-1B.*Majority of these genes are involved in the demethylation of H3K4, H3K9, H3K27 or H3K36 histone tails while others associate with the PRC proteins. Understanding their role in eukaryotic transcription, genome integrity and epigenetic inheritance would assist in comprehending the interplay of Jumonji and PRC2 proteins which fine tunes the formation and maintenance of the H3K27me3 marks (Kruidenier et al., 2012).

3.4 SNP6 Analysis

As described earlier, the involvement of *EZH2* in myeloid malignancies was discovered due to mutational analysis of a region with acquired CN (LOH) of chromosome 7q and 7q36.1 microdeletion. Following a similar approach, I have examined SNP6 array karyotyping data available for 91MDS/MPD/AML patients from previous studies to identify deletions, amplifications and CN (LOH).

3.5 Aim

- To identify aberrations in PRC1 and Jumonji genes using high density SNP data for 91 patients with MDS/MPD/AML. To determine the prevalence and clonal size of mutations affecting patients with SNP6 abnormality using 454 sequencing and to correlate them to the SNP data. This would enable high throughput analysis and importantly, help identify the frequency and prognostic significance of somatic low level mutations in PRC1 and Jumonji genes.
- To evaluate gene expression changes in Jumonji and PRC1 genes in patients with aberrations in copy number (deletion, amplification) by SNP6 analysis.
- To determine the prevalence and clonal size of mutations affecting the core components of PRC2 (*EED, SUZ12, EZH1*) and *DNMT3A* in 61 patients (out of the 91 patients) sequenced previously for *EZH2* mutations using 454 sequencing to identify the frequency of mutations affecting these genes and to examine the overlap of mutations between the different components of PRC2, *DNMT3A* and *EZH2*.

3.6 Patient samples

91 cases of MDS/MPD/AML were examined in this study for SNP6 abnormalities and mutational analysis. Patient details (available for 60/ 91 cases) are summarised in the table below.

Patient characteristics	Total
	60
Age, years	
Range	33-85
Median	63
Sex	
Male [n (%)]	33 (55%)
WHO category	
RA [n (%)]	1 (1.6 %)
RARS [n (%)]	3 (5 %)
RCMD [n (%)]	19 (31.6 %)
RAEBI/II [n(%)]	25 (42 %)
tMDS/AML [n(%)]	2 (3.3 %)
CMML & MPD/MDS-U [n (%)]	6 (10 %)
MDS with isloated 5q [n (%)]	1 (1.6 %)
Polycythaemia Rubra Vera	1 (1.6 %)
Unknown	2 (3.3 %)
Bone marrow blasts	
Median (%)	7.5
Range	0 -75
IPSS cytogenetic risk group	n= 51
Low risk (0)	8
INT-1 (0.5-1)	13
INT-2 (1.5-2)	19
High risk (> 2.5)	11
Progression to AML	
Yes [n (%)]	18 (30%)

Table 3.1: Clinical characteristics of the patients studied (n=60). Patients were stratified by age, sex, WHO classification at the time of diagnosis, IPSS score, percentage of blasts and progression to AML. "n" represents number of patients; % - represents percentage of patients.

3.6.1 Amplification of patient samples

DNA from these patients was amplified (refer section 2.5.3) for sequencing. A small amount of non-amplified sample was stored for confirmation experiments. All samples corresponded

to Lane 2 confirming good quality of the DNA and were therefore utilised for all sequencing experiments.

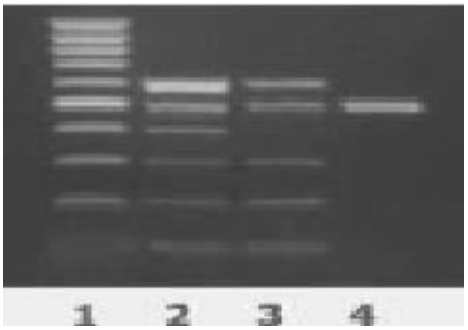


Figure 3.1: Human DNA OK results.Lane 1 – 100 bp DNA ladder, Lane 2 – Good quality intact DNA (6 bands), Lane 3 – partially degraded DNA (5 bands), Lane 4 – fully degraded DNA (only positive PCR control band is visible).*(Reference: http://www.microzone.co.uk/products/quality-testing/human-dnaok/)*

3.7 SNP6 and Chas analysis

SNP6 analysis was carried out on the 91 patients using SNP-A technology by Affymetrix (Santa Clara, CA) wherein; genomic DNA from patient samples was digested by restriction endonucleases, amplified, and labelled. The labelled patient DNA was hybridised to SNP arrays probes homologous to areas located throughout human genome. The resultant genotyping pattern helped in determination of either heterozygosity or homozygosity for each allele. At the same time, intensity of the hybridization signals allowed for determination of copy number changes at that genomic region [reviewed by (Maciejewski and Mufti, 2008)].

The SNP6 data used here was generated by Dr. Azim Mohamedali and formed part of previous studies conducted by our group. I visualised this data using CHromosome Analysis Suite (Chas) from Affymetrix for aberrations of seventeen Jumonji and twelve PRC1 genes. Examples of Chas aberrations are indicated in Fig: 3.2 c & f which show a gain in copy number from 2 to 3 suggesting segmental amplification of an allele at that locus. Fig: 3.2 b & e indicate a decrease in copy number from 2 to 1 and loss of heterozygosity as seen in Chas

while Fig: 3.2a & d indicate CN(LOH) where in there is loss of heterozygosity but the copy number remained unaltered (CN=2).

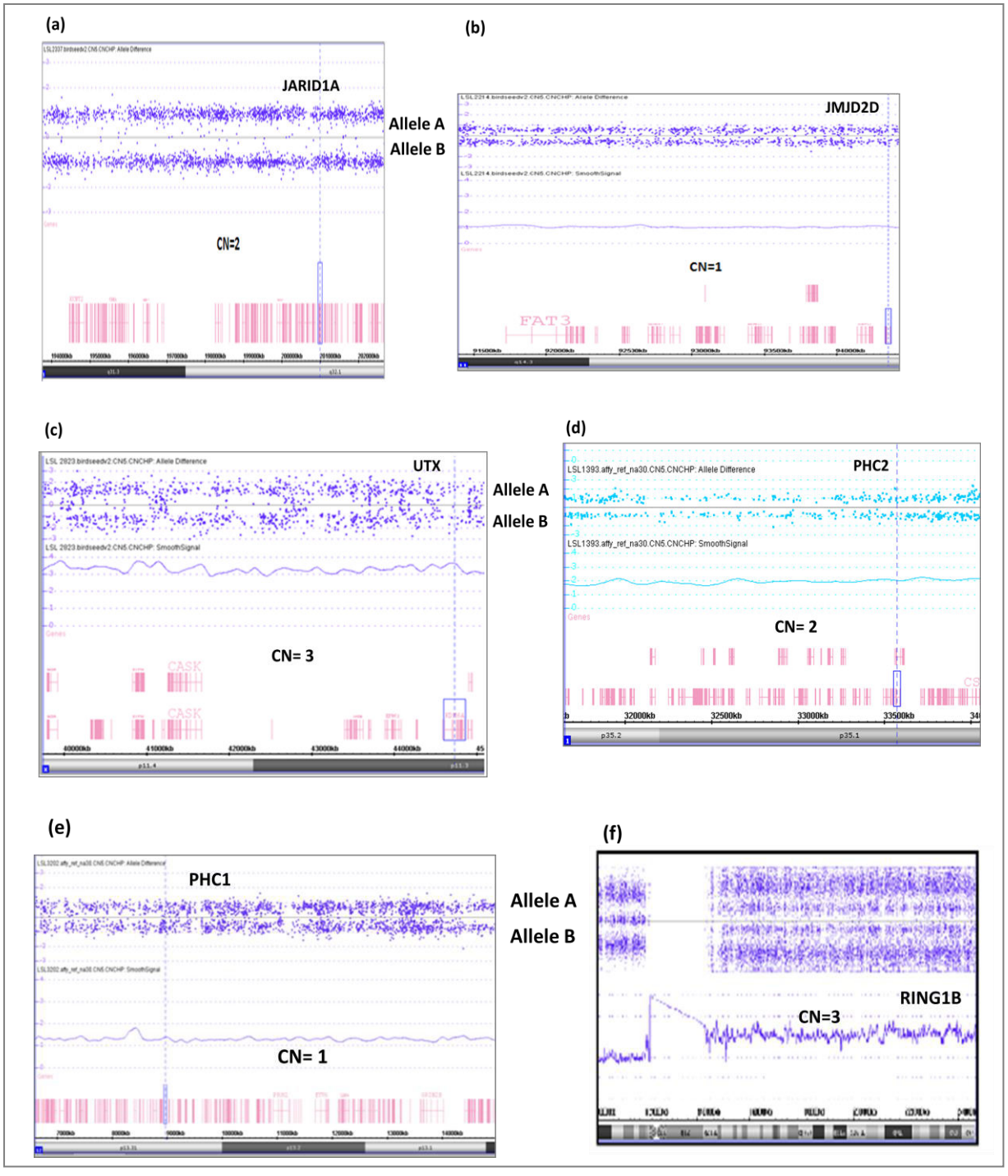


Figure 3.2: Visualisation of SNP6 data for the Jumonji and PRC1 genes by Chromosome Analysis Software (ChAS). Chromosome ideograms indicating the chromosomal location are indicated at the bottom of all the figures (grey) while the CNV probe log ratio scatter plot is seen in the middles as smooth signal panel. Dotted line indicates the location of the gene of interest (a) & (d) Illustrate CN (LOH) at *JARID1A* and *PHC2*

locus respectively indicating loss of heterozygosity but no variation in the copy number. (b) & (e) Illustrates deletion of *JMJD2D* and *PHC1* locus is seen as change in copy number from 2 to 1. (c) & (f) Indicate gain in copy number (CN=3) at *UTX* locus and *RING1B* locus.

3.7.1 PRC1

17/91 (19 %) patients were detected with SNP6 aberrations.5/12 genes exhibited either [gain/ del/ CN (LOH)] on SNP6 analysis (Figure: 3.2 d,e,f). Eleven patients showed CN (LOH) [size: ≤ 1.5 Mb] at the *BMI-1* locus (chr: 10p12.31).These patients were selected because of the high frequency of CN (LOH) at this locus and also due the association of *BMI-1* to MDS (refer section 1.8). Among the RING proteins, *RING1A* locus (chr: 6p21.32) was deleted in two patients while the *RING1B* locus (chr: 1q25.3) showed a CN gain in two patients and a CN (LOH) in one patient. In the polyhomeotic proteins, the *PHC1* (chr: 12p13.31) locus was deleted in one patient while the *PHC2* (chr:1p35.1) locus showed a CN gain in one patient and a CN (LOH) in another patient (Table: 3.2)

Patient No	MDS subtype	Cytogenetics	Gene	SNP6 aberration
LSL1105	RCMD	Normal	BMI-1	CN(LOH) (1.5)
LSL1268	RAEBII	Complex including del(7)(q22q36)	PHC2	CN=3 (12.6)
LSL1393	RARS	45,XY,t(1;6;8)(p32;p21;q12),-5,-7, +r [9]/46,XY [3]	PHC2	CN(LOH) (108)
			RING1A	del (0.2)
LSL1590	RAEBI	Complex including monosomy7,	RING1A	del (0.14)
LSL1808	RCMD	46 XY, +1, der (1:7)	BMI-1	CN(LOH) (1)
LSL1848	tMDS/AML	NA	BMI-1	CN(LOH) (1)
LSL2337	MDS/MPD	Normal	RING1B	CN(LOH) (95)
LSL2383	RCMD	NA	BMI-1	CN(LOH) (1)
LSL2470	RAEBI/II	47,XX,+1,add(1)(p11) add(7)(q11) [30]	RING1B	trisomy 1q (100)
LSL2906	RAEBI/II	NA	RING1B	trisomy 1q (100)
LSL3066	RCMD	Complex including (2;11)	BMI-1	CN(LOH) (1)
LSL3170	RAEBI/II	NA	BMI-1	CN(LOH) (1)
LSL3202	tMDS/tAML	NA	BMI-1	CN(LOH) (1)
			PHC1	del (7)
LSL3241	RAEBI/II	NA	BMI-1	CN(LOH) (1)
LSL3287	RCMD	NA	BMI-1	CN(LOH) (1)
LSL3455	CMML	NA	BMI-1	CN(LOH) (1)
LSL3532	RAEBI/II	NA	BMI-1	CN(LOH) (1.5)

Table 3.2: SNP6 aberrations [deletion, amplification and CN (LOH)] in PRC1 genes. Numbers in brackets indicate size of the aberration in Mb (NA= data not available). The size of CN (LOH) was small at the *BMI-1* locus but these patients were still included for sanger sequencing due to high frequency of occurrence of CN (LOH).

3.7.2 Sanger sequencing

17 patients (mentioned above) were selected and mutational analysis of the respective PRC1 genes was carried out using sanger sequencing (refer section 2.7) on amplified patient samples (refer section 3.2.2). Primers for all exons of the PRC1 genes were designed (refer section 2.6) using Primer 3 software (http://primer3.ut.ee/) and the PCR reactions were optimised using 10 ng of the patient samples and 5 µM of the primer (refer appendix for

primer sequence) concentration (Table 2.3). Data obtained from sanger sequencing was analyzed using Sequencing Analysis software by Applied Biosystems (Foster City, CA). Sequencing did not identify any mutations affecting the PRC1 genes concomitant with SNP6 abnormalities indicating that the PRC1 genes are not frequently mutated in MDS/AML cases. Gene expression by qPCR could not be carried out on patients with copy number variations (CNVs) due to insufficient patient samples.

3.7.3 Jumonji genes

In our cohort, 15/17 Jumonji genes exhibited either one or more of the aberrations on SNP6 analysis. 29/91 (31 %) patients showed del, gain or CN (LOH) at one or more Jumonji gene loci. The summation of loss or gain of CN was 22 out of which 16 were loss/deletions. The highest frequency of deletions (n=12) found in *JMJD1B* (chr 5q31.2) (Giagounidis et al., 2006) while *JARID2 (*chr: 6p22.3) was deleted in three patients, *JMJD3* (chr: 17p13.1), *JMJD2D* (chr: 11q21) & *JARID1A* (chr: 2p11.2) were deleted in two patients each. On the other hand, the highest frequency of gain (n= 3) was observed in *JMJD2C,* followed by *JMJD2A* (chr: 1 p34.1) and *UTX* (chr: X p11.2) (n= 2 each). Majority of the genes were either deleted or amplified but only three genes *JARID2*, *JMJD4* (chr: 1q42.13) and *JMJD2D* showed both deletion and amplification in different patients in our cohort. One patient had del of both *JMJD4* and *JARID2.* In addition to gains and deletions, the total number of CN (LOH) at different Jumonji gene loci was 12. *JMJD2D* (n=2) and *JMJD4* (n= 3) had the highest frequency of CN (LOH) followed by *JMJD3* (n= 3). The size of CN (LOH) was >20 Mb in majority of the patients (except two patients; 7 & 8 Mb). The distribution of SNP6 aberrations in the Jumonji genes are detailed in table 3.4 and pie chart (Fig: 3.3)

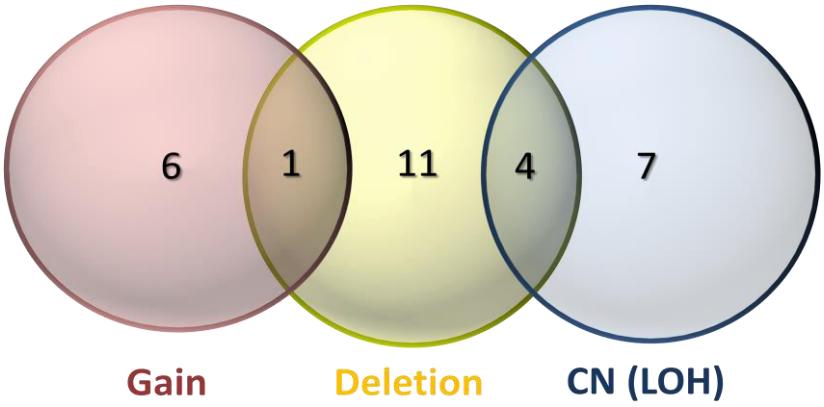


Figure 3.3: Total number of patients with SNP6 abnormalities. Deletions were observed in sixteen patients while amplifications & CN (LOH) were seen in seven and eleven patients respectively. Four patients had both deletions and CN (LOH) at different gene loci while only one patient carried both a gain and deletion. None of the patients had an overlap between gain and CN (LOH).

Patient No	MDS subtype	Cytogenetic	Gene	SNP6 aberration
LSL654	sAML	del (7)(q22q36), -17	JMJD3	del (10)
			JMJD1B	del (0.86)
LSL1268	RAEBII	Complex including del(7)(q22q36)	JMJD2A	CN=4 (12)
LSL1393	RARS	Complex including monosomy 7 and t(1;6;8)	JARID2	del (11.5)
			JMJD1B	del (37)
			JMJD2A	CN (LOH)(108)
LSL1532	RAEBI/II	46 XX, t(7:21), der 19	JMJD1B	del (1)
LSL1590	RAEBI/II	Complex including monosomy 7	JARID2	del (16.5)
			JMJD1B	del (68)
LSL1825	RCMD	del 11q(q21q25), +6, +21	JARID2	trisomy 6
LSL1874	CMML	Not available	JMJD1C	CN (LOH)(20)
LSL1898	RAEBI	Complex including del(7)(q22q34	JMJD1B	del (3)
LSL2337	MDS/MPD	Normal	JMJD4	CN (LOH) (105)
LSL2214	RAEB II	Not available	JMJD2D	del (43)
			JMJD1B	del (54)
LSL2470	RAEBII	47,XX,+1,add(1)(p11),add (7)(q11) [30]	JARID1B	trisomy 1q (120)
			JMJD4	trisomy 1q (120)
			JMJD2C	trisomy 9p (35)
LSL2823	RAEBII	45 X isodiXq13(29)	UTX	trisomy Xp (60)
			JARID1C	trisomy Xp (60)
LSL3104	RAEB II	Normal	JMJD4	CN (LOH) (95)
LSL3202	tMDS/tAML	Not available	JARID1A	del (1)
			JMJD1B	del (90)
LSL4958	RAEBI/II	Not available	JMJD2D	CN (LOH) (67)
			JHDM1A	CN (LOH) (68)

Patient No	MDS subtype	Cytogenetic	Gene	SNP6 aberration
LSL3241	RAEBI/II	Not available	UTX	CN=4 (0.2)
LSL3287	RCMD/RCMD-RS	Not available	JARID1D	CN=1.5 (24)
LSL3513	tMDS/tAML	Not available	JMJD1B	del (50)
			JMJD4	CN(LOH) (50)
LSL3532	RAEBI	Not available	JARID1A	del (0.36)
			JHDM1B	CN (LOH) (60)
LSL3711	RAEBII	Complex including del(7)(q22q36)	JMJD3	CN (LOH) (20)
			JMJD1B	del (81)
LSL3749	RAEBI/II	Not available	JMJD2A	CN=3 (50)
			JMJD3	del (8)
			JHDM1A	trisomy 11 (CN=2.5)
			JMJD1B	del (30)
			JMJD2C	trisomy 9p (35)
			JMJD2D	trisomy 11 (CN=2.5)
LSL3825	RAEBII	Not available	JMJD3	CN (LOH) (8)
LSL3826	tMDS/tAML	Not available	JMJD3	CN (LOH) (7)
LSL4272	RAEBI	45,XY,-7, del(11q),del(12p)	JMJD2D	del (12)
LSL4667	tMDS/tAML	Not available	JMJD1B	del (76)
LSL4737	RAEBI/II	Not available	JMJD4	del (15)
			JARID2	del (11.4)
LSL4855	RAEBI/II	Not available	JMJD2C	CN=4 (35)
LSL4958	CMML	Not available	JMJD2D	CN (LOH) (70)
LSL6395	RAEBI/II	Not available	JMJD1B	del (0.465)

Table 3.3: Gain/ del/ copy neutral loss of heterozygosity the Jumonji genes. Numbers in brackets indicate size of the aberration in Mb(NA= data not available).

CNVs were mutually exclusive; For e.g. Only one patient harbored a deletion of *JARID2*, *JMJD1B* and a gain at the *JMJD2A* locus. While the overlap between CN (LOH) and deletion was seen in four patients. However, none of the patients had an overlap between CN (LOH) & amplification of gene. The reason for the mutual exclusivity of the aberrations is still unclear and whether one type of aberration is favored over the other during disease progression needs further scrutiny.

Stretches of homozygozity at times harbour mutations and have been formerly associated with MDS (Mohamedali et al., 2007). Therefore mutational analysis using 454 DNA sequencing (refer section 2.8 - 2.8.17) was performed on 29 patients (mentioned above) showing telomeric stretches of [CN (LOH)] (majority; size: > 20Mb), deletions and/or copy number (CN) gain at the fifteen Jumonji genes locus.

3.7.3.1 Primer testing and PCR optimisation

All the PCR reactions were carried out initially on genomic DNA from Affymetrix, UK (stock concentration: 50 ng/ µl) and later tested on patient samples due to availability of limited amount of patient samples. The primers to sequence the Jumonji genes were designed (refer appendix for sequence) as mentioned before using primer3 software and the PCR conditions were optimised using 10 ng of the patient samples and 5 µM of the primer concentrations (Table: 2.3). Figure 3.4a shows PCR products visualized on 2 % agarose gel after the 1st round of PCR. All exons displayed a single good intensity band which corresponded with the anticipated product size on all patient samples. Subsequently, a second PCR reaction was carried out using 1or 1.3 µl of the amplified product with primers (1 µl) targeted to the aUS primers. These primers contained a 5' tag that identified forward and reverse strands for sequencing and also included a 5-base pair Multiplex Identifier (MIDs)

sequence to identify individual patients from the first round of PCR. Eight random wells/plate were selected from the 2nd round of PCR and visualised on agarose gel (Fig. 3.4b).

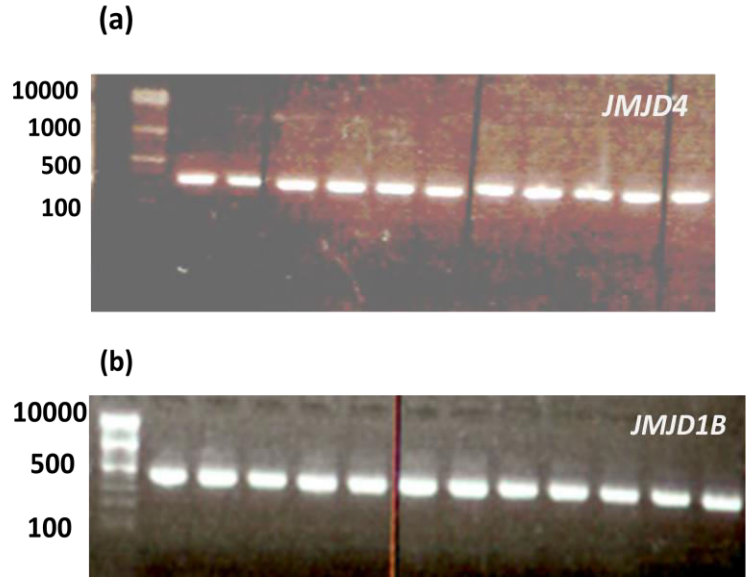


Figure 3.4: Visualisation of PCR product for *JMJD4* and *JMJD1B* on 2 % agarose for 454 sequencing. (a) Visualisation of eleven primer pairs for *JMJD4* in a patient sample after 1st round of PCR (b) Visualisation of PCR product for *JMJD1B* exon 2 for twelve patients on 2 % agarose gel after 2 nd round of PCR. All the products were in the range of 350-400 bp.

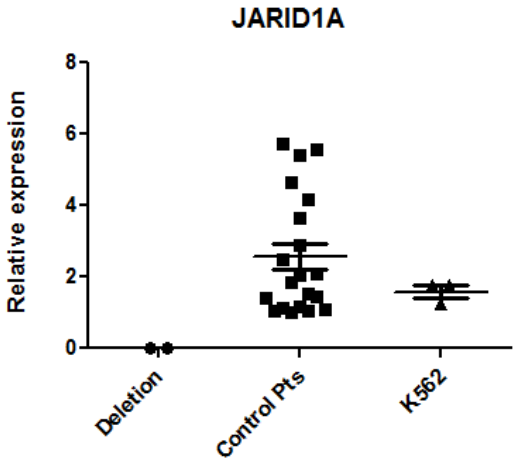
3.7.3.2 454 parallel sequencing (Jumonji genes)

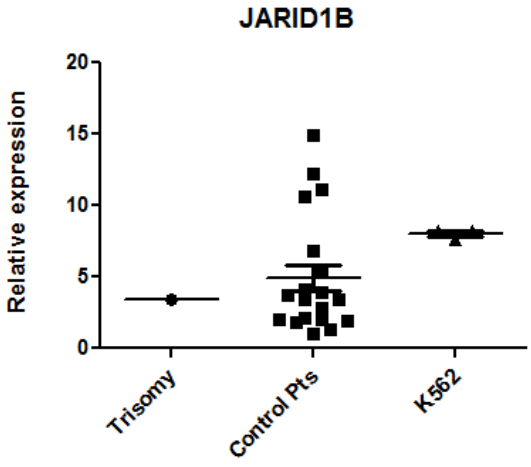
All exons were quantified using picogreen dye and were in the range of 10 to 40 ng/ µl. The exons were equalized in concentration and pooled. A second picogreen measurement was done on the pooled samples to equalise the concentration of all the patient samples/gene. All the patients were in the range of 20-40 ng/ µl, their concentrations were equalised by combining the appropriate amounts of DNA to create a library of all exons and all patients.This library was gel and bead purified using QIAquick Gel Extraction Kit and AgencourtAMPure XP magnetic beads (refer section 2.8.2-2.8.3) from Beckman Coulter (USA) and used for 454 parallel sequencing (refer section 2.8- 2.8.17). The results were analyzed using Ingenuity Variant Analysis software (http://www.ingenuity.com/products/variant-analysis). The 454 sequencer generates data in FASTA format files which also contain the quality score file which scores for each base in

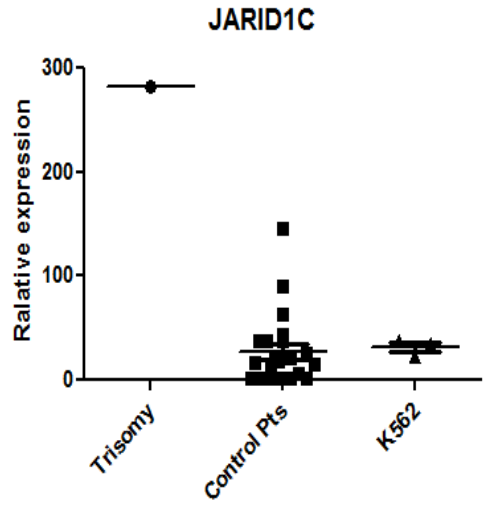
each sequence assigned to a particular region of the picotitre plate. Next the data is segregated according to different patient samples based on the nucleotide barcode tags attached to each sequence. Filtering out any low quality or ambiguous reads also takes place at this stage followed by alignment of the high-throughput reads to the input reference sequence provided by the user. The aligned reads are visualised using Vnode software to detect any variations in the sequence compared to the reference sequence provided by the user. Analysis of the sequencing data revealed that the Jumonji genes did not harbour any mutations concomitant with the SNP6 abnormalities even with a read depth of 350-450 reads/ amplicon.The absence of mutations in these genes suggests that other mechanisms and alterations in gene expression might contribute to the disease pathogenesis.

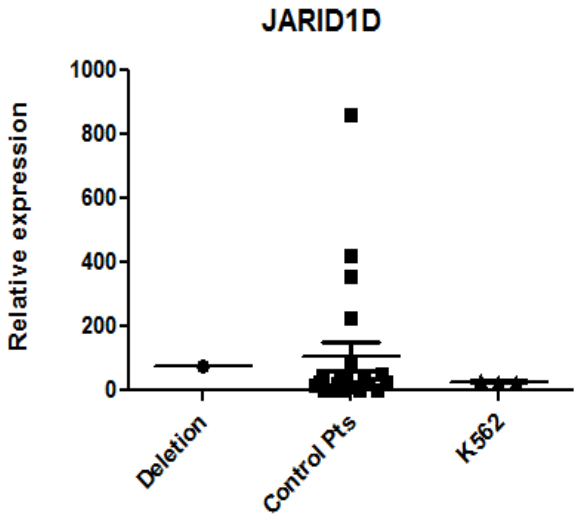
3.7.3.3 Gene expression analysis by qPCR

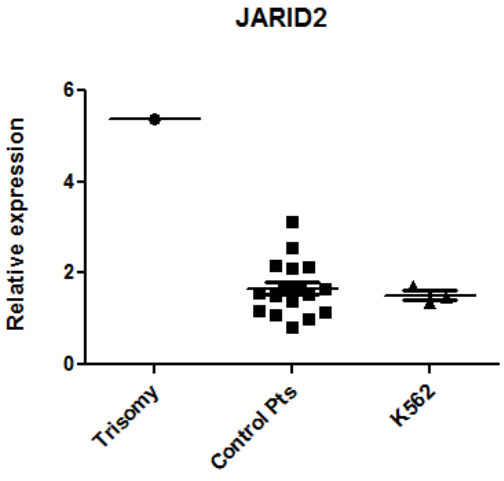
To elucidate changes in gene expression as a result of amplification/gain or deletion of genomic material, Real-time PCR (qPCR) [refer section 2.9 - 2.11] was performed on 22/29 patients for twelve Jumonji genes. Six patients were excluded because they had only CN (LOH) but no CNVs. Patients without SNP6 abnormalities at the gene loci were used as controls while K562 cells were used as a positive technical control in all experiments.

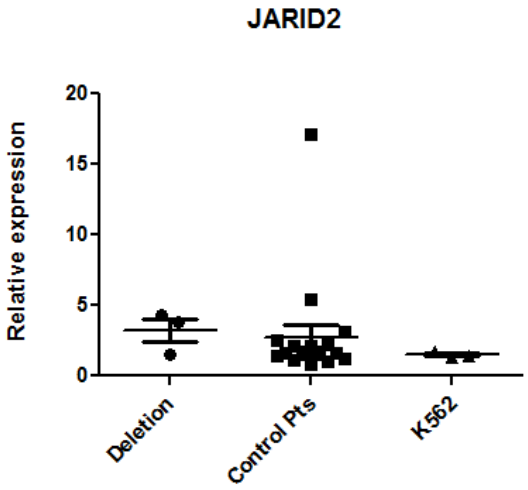

JARID1A
Relative expression
8
6
4
2
0
Deletion
Control Pts
K562

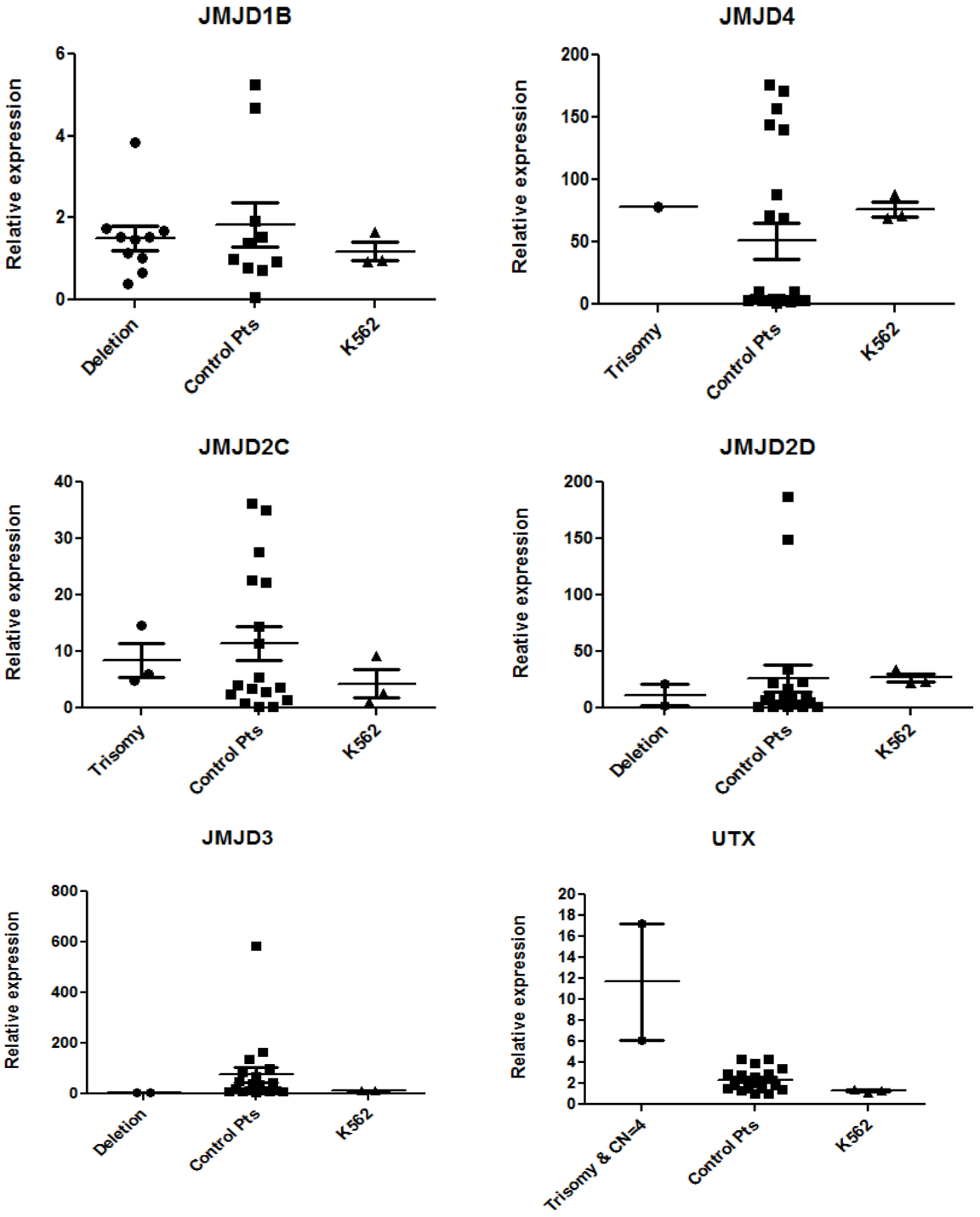

JARID1B
Relative expression
20
15
10
5
0
Trisomy
Control Pts
K562


JARID1C
Ralative expression
300
200
100
0
Trisomy
Control Pts
K562


JARID1D
Relative expression
1000
800
600
400
200
0
Deletion
Control Pts
K562


JARID2
Relative expression
6
4
2
0
Trisomy
Control Pts
K562


JARID2
Relative expression
20
15
10
5
0
Deletion
Control Pts
K562


JMJD1B
Relative expression
0
2
4
6
Deletion
Control Pts
K562
JMJD4
Relative expression
0
50
100
150
200
Trisomy
Control Pts
K562
JMJD2C
Relative expression
0
10
20
30
40
Trisomy
Control Pts
K562
JMJD2D
Reative expression
0
50
100
150
200
Deletion
Control Pts
K562
JMJD3
Relative expression
0
200
400
600
800
Deletion
Control Pts
K562
UTX
Relative expression
0
2
4
6
8
10
12
14
16
18
20
Trisomy & CN=4
Control Pts
K562

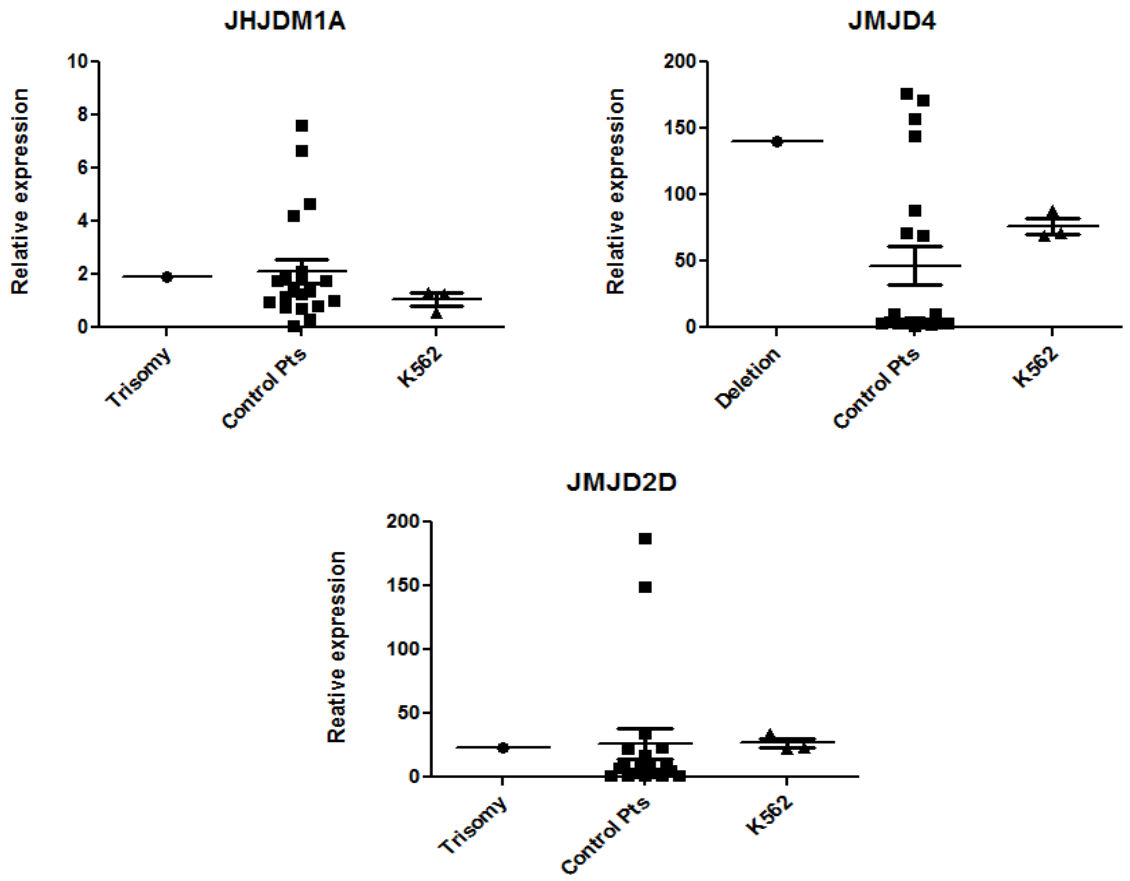


Figure 3.5: Gene expression analysis of Jumonji genes by qPCR. There was no change in the gene expression of ten Jumonji genes concomitant with CNVs. While complete loss of *JARID1A* (p value <0.0001) expression was observed in two patients with deletion of a region on chr 12p.One patient with trisomy Chr Xp showed concomitant increase in expression of *UTX* and *JARID1C*. Trisomy at the *JARID2* locus was associated with increased expression of the gene. Patients without SNP6 aberrations at the Jumonji gene locus were used as controls and mRNA from K562 cell line was used as a technical control in all the experiments. Error bars calculated using SEM.

UTX showed over expression (p value < 0.0001) concomitant with CNVs. Two patients had *UTX* aberrations on SNP6 analysis; trisomy in one patient and copy number of 4 in the other which compare with up regulation of the *UTX* expression seen on qPCR.

In addition to *UTX*, other genes with deregulated expression were *JARID1A*, *JARID1C* and *JARID2*. *JARID1A* expression was reduced (p value < 0.0001) in two patients with deletion of the *JARID1A* locus as compared to the control samples (n=20) while *JARID1C* and

JARID2 was over expressed in one patient with trisomy. This confirms that CNVs detected by SNP6 analysis can be correlated to changes in gene alterations in these genes.

3.8 PRC2 and DNMT3A

EZH2 was previously sequenced by our group on 61 out of 91 patients (mentioned earlier) and therefore the same cohort of patients were selected for mutational analysis of the other core components of PRC2; *SUZ12*, *EED*, *EZH1*and *DNMT3A* to identify mutations in these genes and to examine the mutational overlap of these genes with *EZH2*.

3.8.1 Primer testing and PCR optimisation (PRC2)

As mentioned above, the PCR reactions were carried out initially on genomic DNA from Affymetrix, UK (stock concentration: 50 ng/ µl) and later tested on patient samples due to availability of limited amount of the samples. The PCR reaction was optimised at primer and template concentrations of 10 µM and 10 ng/µl, respectively. The PCR amplified product was visualized on 2 % agarose gel and the band size of the products were matched with the expected product size from the primer design software (Primer3, reference: http://primer3.ut.ee/). Primers for all the exons of *SUZ12* worked except for exon 1 (a & b) (Fig: 3.6c blue arrows). These two primers were redesigned and optimised at a primer and template concentration of 10 µM & 10 ng/µl, respectively (refer appendix for primer sequences). Similarly, all primers for the *EZH1* gene were also optimised at 10 µM primer & 10 ng/µl of template concentrations except for exon 4 which showed primer-dimer formation at these concentrations (Fig: 3.6a) and therefore PCR for this exon was carried out using 8 ng/µl template and 5 µM primer. *EED* primers were tested and optimised at 5µM primer concentration and 10 ng/µl of genomic DNA (Fig: 3.6b). All primers worked except exon 1 which was redesigned and optimised at 5 µM primer concentration. Subsequently all primers were verified on amplified DNA from patients with similar results.

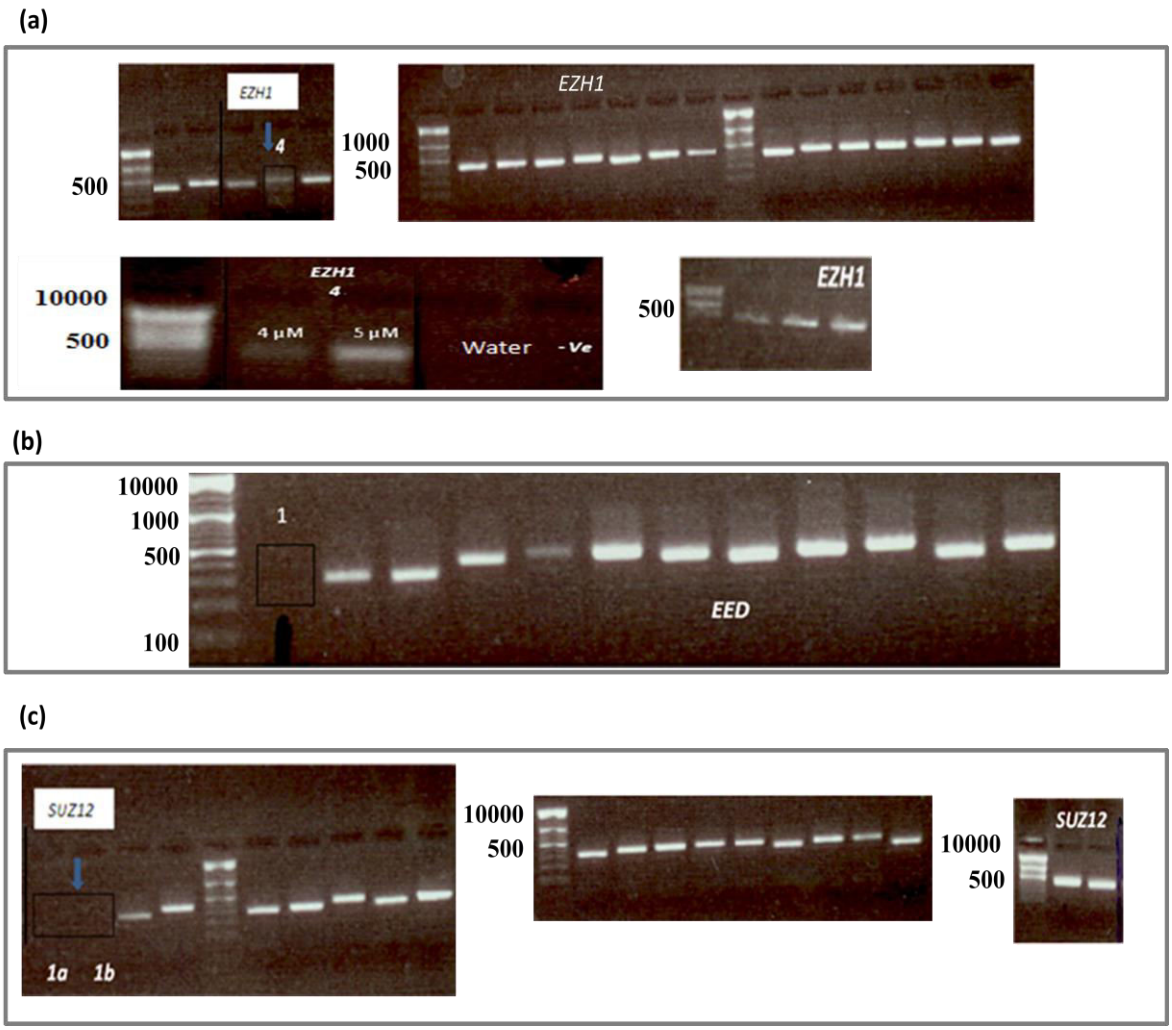


Figure 3.6: Primer optimisation for *EZH1*, *EED* & *SUZ12* using different primer and template DNA concentrations. The PCR products were visualized on 2 % TAE agarose geland blue arrows/black boxes indicate failed PCR reactions. Exon 1 of *EED* and *SUZ12* were redesigned while all the other exons worked with a template concentration between 8-10 ng/µl and primer concentrations of 5-10 µM. A negative control containing water instead of the DNA template was used to rule out contamination (a).

3.8.2 454 parallel sequencing (PRC2)

1st round of PCR was carried out on 61 amplified patient samples using the optimised primer concentrations for the three genes (Fig. 3.6 a,b,c) and 4 random wells/sample were verified on 2% TAE agarose gel. All exons displayed a single good intensity band which

corresponded with the anticipated product size. 0.8 to 1.3µl of product from the 1st round of PCR was utilised for the 2nd PCR depending on the intensity of the product on gel.

Eight random wells/plate were selected from the 2nd round of PCR for all the genes; *EZH1*, *SUZ12* and *EED* (gel shown below) visualised on 2 % agarose gel (Fig. 3.7). PCR for exon 1 and exon 3 (one well each) did not work. The 2nd round of PCR was repeated successfully for these exons for the respective patients keeping all the conditions constant.

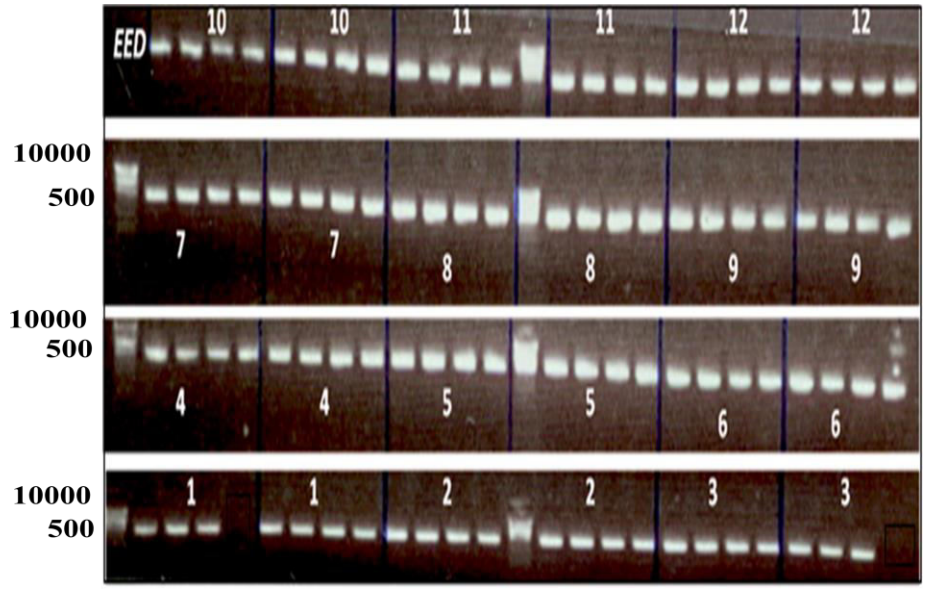


Figure 3.7: Visualisation of the PCR product for *EED* after second round of PCR.8 samples /exon were visualised on 2 % TAE agarose gel for the *EED* gene. Two patients (exon 1 and exon 3) showed no product and therefore the PCR reactions were repeated for these patients keeping the PCR conditions constant. The PCR worked well for these two patients when the experiment was repeated.

After the second round of PCR, all the exons for each gene were quantified using picogreen dye on the Rotor-Gene 6000 Multiplexing System. *EZH1*, *EED* and *SUZ12* exons were in the range of 8 to 61 ng/ µl, 11 to 51 ng/ µl and 22 to 51 ng/ µl respectively. This measurement was used to equalize the concentration of all exons/gene. A second picogreen measurement was performed to equalize concentration of all the patient samples/gene. All samples of

EZH1, *EED* and *SUZ12* were in the range of 23-98 ng/ µl, 19-66 ng/ µl and 22-76 ng/ µl concentrations respectively. These samples were equalized and pooled to form a single mixture of all the samples for each gene (3 libraries).

These (three) libraries were visualised on a 2 % agarose gel for 45 min at 120V to obtain a single demarcated band as illustrated in Fig. 3.9 This band was purified two times using QIAquick Gel Extraction Kit and AgencourtAMPure XP magnetic beads (refer section 2.8.2-2.8.3) from Beckman Coulter (USA) (Fig. 3.8) and the yields were verified on 2% agarose gel.

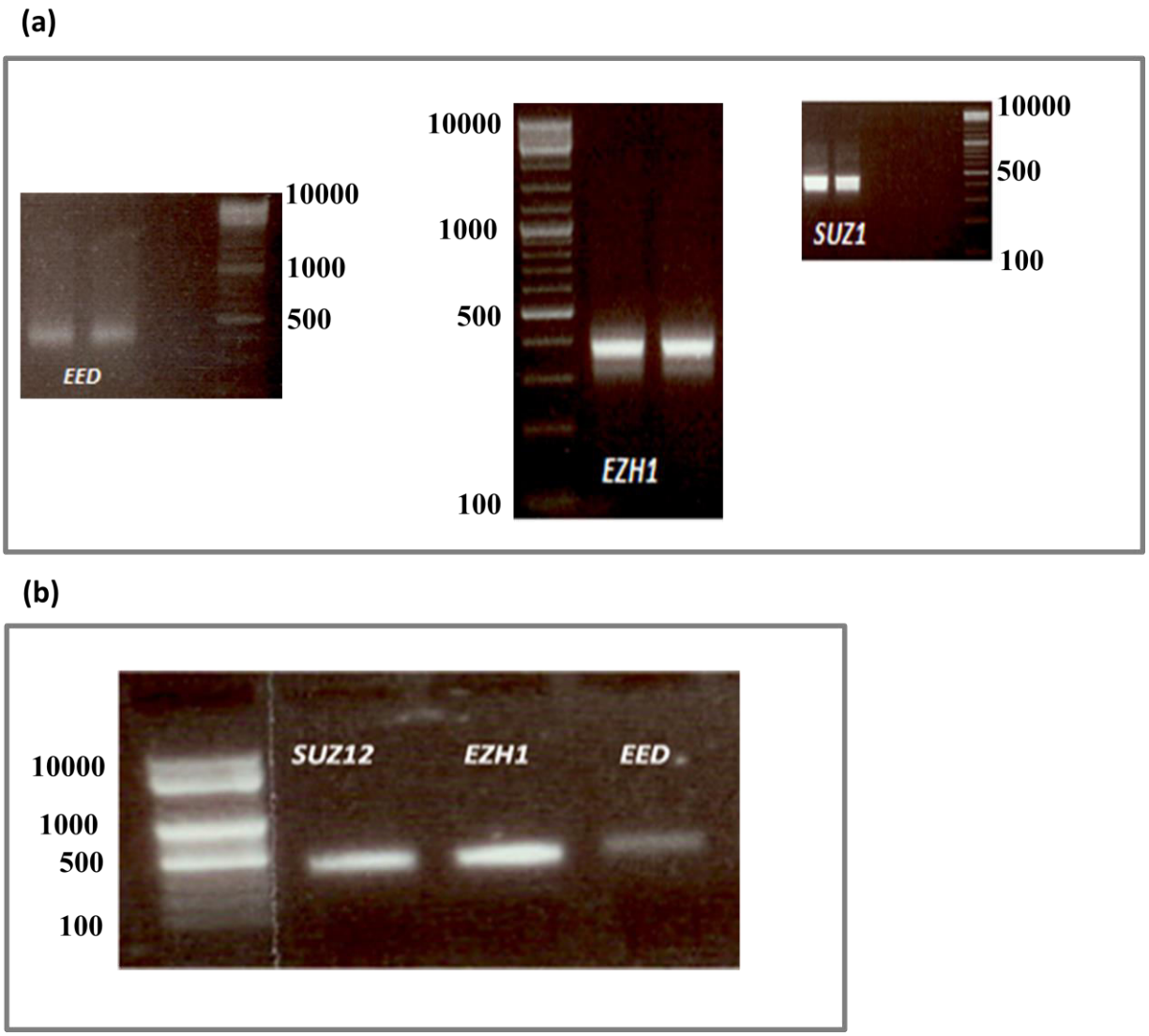


Figure 3.8: *SUZ12, EED, EZH1* library purification. (a) Libraries for *SUZ12, EED, EZH1*were visualized on 2 % agarose gel for purification. All the products were in the range of 350-400 bp in size. (b) Libraries visualized on gel after bead purification.

These libraries were sequenced on Roche 454 sequencer (refer section 2.8.5 to 2.8.17) and the data was analyzed. Data analysis revealed that PRC2 genes namely; *SUZ12*, *EZH1* and *EED* were not mutated in the 61 MDS/AML patients. Lack of mutations in these genes indicated that *EZH2* is exclusively targeted for mutations in MDS as compared to the other components of PRC2.

3.8.3 Primer testing and PCR optimisation (DNMT3A)

Similar to the genes of PRC2, PCR reactions were also set up for sequencing *DNMT3A* using 5 µM of the primer concentration and 10 ng of genomic and patient DNA, abundant primer-dimer (Fig. 3.9 red box) formation was observed.

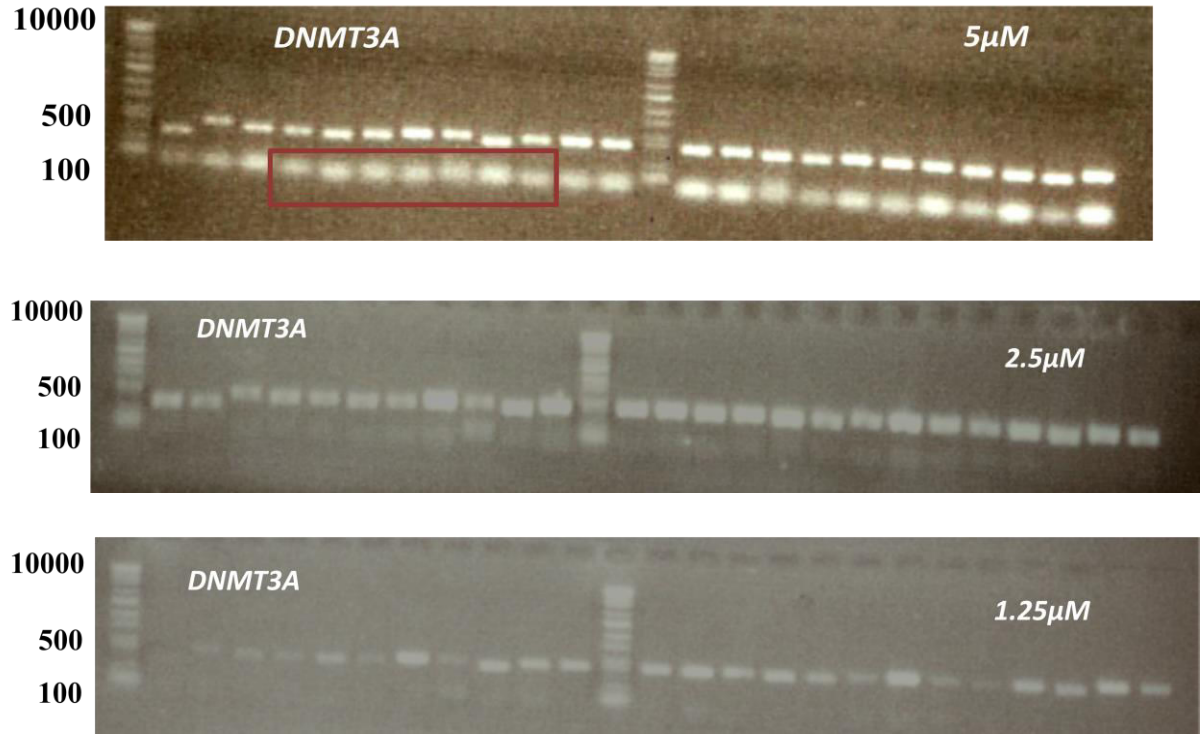


Figure 3.9: Optimisation of *DNMT3A* primers. 5 µM, 2.5 µM and 1.25 µM concentrations of primers were tested. Primer-dimer was observed (red box) with 5 µM concentration of the primer while 1.25 µM conc was inadequate for some exons. All the experiments were therefore carried out with 2.5 µM primer concentration.

Therefore, the reactions were repeated with 2.5 µM and 1.25 µM of primer concentrations. The bands were faint when the primer concentration was reduced to 1.25 µM. Hence 2.5 µM primer concentration was used for all subsequent experiments.

1st round of PCR was conducted on 61 amplified samples using the optimised primer concentrations and 4 random wells/sample were verified on 2% TAE agarose gel. All exons displayed a single good intensity band which corresponded with the anticipated product size.

Next, four random wells/plate were selected from the 2nd round of PCR and visualized on agarose gel (Fig. 3.10a).

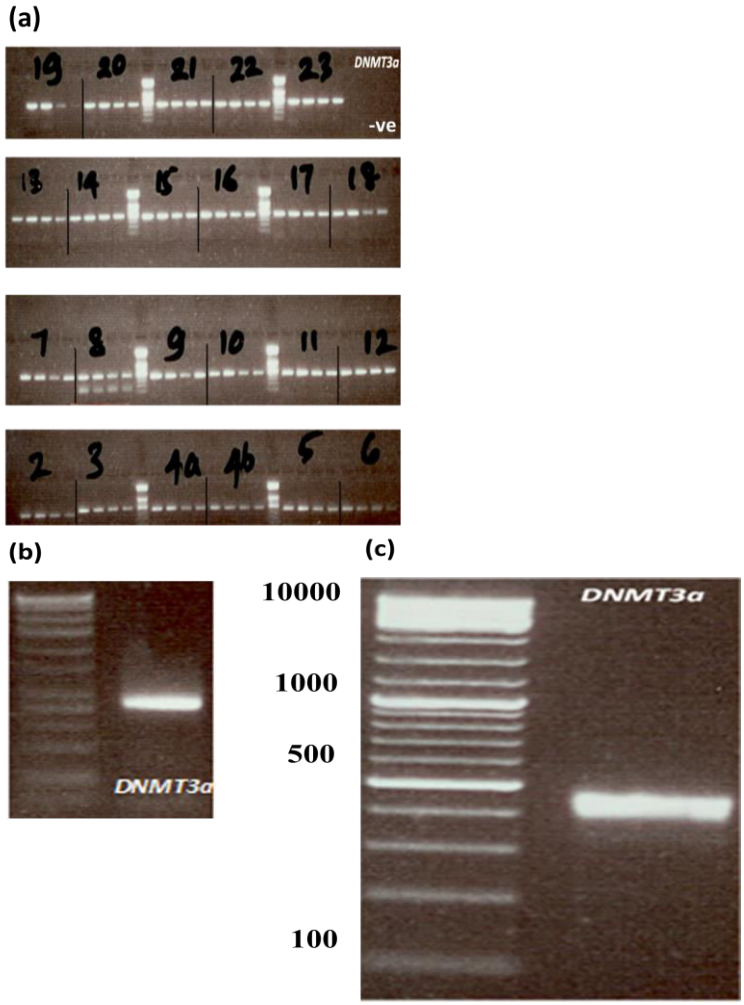


Figure 3.10: *DNMT3A* library preparation. (a) Visualisation of 4 samples for each exon of *DNMT3A* after 2nd round of PCR. Exon 8 showed two bands suggesting contamination but the negative control containing water instead of template DNA showed no product. Error during sample loading on the gel could be the reason for the double bands. (b) Visualisation of the *DNMT3A* library on 2 % agarose gel for purification. All the products were in the range of 350-400 bp (c) Library visualised on gel after bead purification.

Exon 8 showed two bands for all the four patients suggesting contamination but the negative control containing water instead of template DNA indicated no product. Error during sample loading on the gel was thought to be the probable cause and therefore the gel was repeated with new samples and contamination was ruled out.

3.8.4 454 parallel sequencing (DNMT3A)

All exons were quantified using picogreen dye and were in the range of 12 to 43 ng/ µl. The exons were equalized in concentration and pooled. A second picogreen measurement was done on the pooled samples to equalise the concentration of all the patient samples/gene. All the patients were in the range of 16-40 ng/ µl and their concentrations were equalised by combining appropriate amounts of DNA to create a library of all exons and all patients. This library was visualised on 2 % agarose gel after 45 min at 120V to obtain a single demarcated band as illustrated in Fig. 3.10. This band was purified two times using QIAquick Gel Extraction Kit and AgencourtAMPure XP magnetic beads from Beckman Coulter (USA) (Fig. 3.10 b & c) and the yield was verified on a 2% agarose gel and used for 454 parallel sequencing (refer section 2.8.5- 2.8.17). The results were analysed using Ingenuity Variant Analysis software (http://www.ingenuity.com/products/variant-analysis) as described previously and the aligned reads are visualised using Vnode to detect any variations in the sequence compared to the reference sequence provided by the user. Analysis of the sequencing data identified heterozygous mutations of *DNMT3A* in 10/61 (16 %) patients (clone size 20-44%) which are summarised in Table: 3.4 and all the mutations were confirmed by Sanger on unamplified DNA from patient samples. Sequencing depth for the amplicons was in the range of 400-600 reads. There was no significant difference in the median survival of monosomy 7 patients with or without *DNMT3A* mutations (p value = 0.6). Interestingly, patients with normal karyotype did not have any *DNMT3A* mutations, whereas, the frequency of these mutations was high in patients with monosomy 7 (n=6) with a WHO classification of RAEB or RAEBII implicating that *DNMT3A* mutations are more common in advanced stage of the disease. The R882 site (Fig: 3.11) was mutated in 2/10 patients and matched the findings of other research groups which have reported high frequency of this

mutation in their cohort. The mapping of the mutations to different domains of *DNMT3A* and their functional impact on the enzyme are detailed below.

Patient no.	WHO Classification	Cytogenetics	Clone size	Mutation	Sanger confirmation	Survival**
LSL1898	RAEB	Complex	42%	E30A	Y	0
LSL1597	RAEB	mon7	22%	R326H	Y	1
LSL4014	RAEB	mon7	22%	R366H	Y	1
LSL2823	RAEBII	isodiXq13	37%	A376T	Y	1
LSL6395	RAEB	mon7	43%	*Exon18 & Intron 18-19	Y	1
LSL1667	RCMD	mon 7	44%	A741V	Y	0
LSL3532	RAEB	Complex	>22%	T835A	Y	1
LSL4062	RAEB	mon 7	34%	R882C	Y	1
LSL3241	RAEB	isodiXq13	34%	R882H	Y	0
LSL1280	RAEB	mon 7	>20 %	P904S	Y	1
			25%	F751V	Y	

Table 3.4: Mutational analysis of *DNMT3A* in ten patients with MDS/tAML using 454 sequencing. Confirmation of these mutations by Sanger was carried out on unamplified DNA as indicated by "Y". The R882 site was mutated in 2/10 patients and matched the findings of other research groups which have reported high frequency of this mutation in their cohort. All the mutations identified in this study are heterozygous (clone size 20-44 %) with 6/10 patients having monosomy7 as the cytogenetic abnormality. "*" = Splice site junction, ** Survival: 0 = surviving and 1= dead.

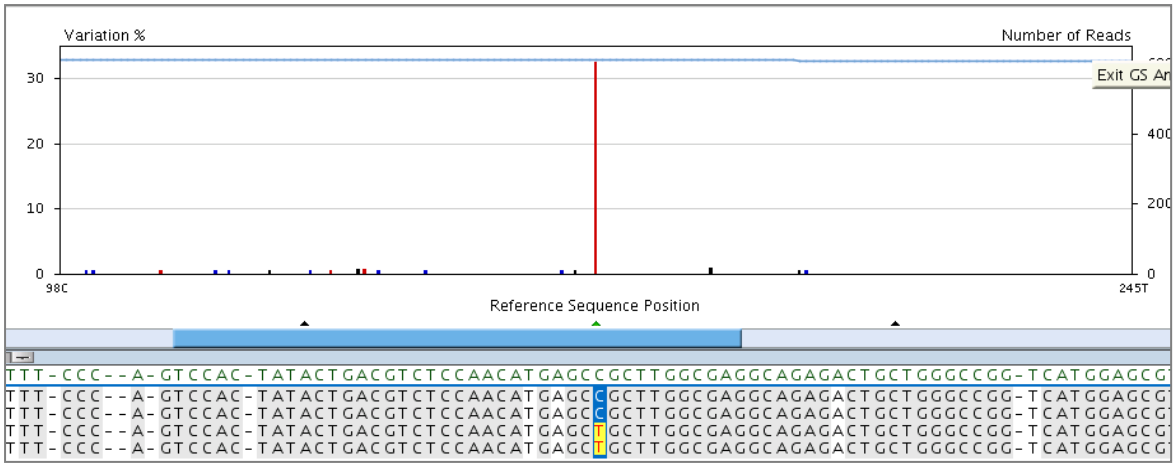


Figure 3.11: *DNMT3A* (R882C) mutation in patient (LSL4062) visualised by Vnode. The top row illustrates the reference sequence (Ensembl) of wild type *DNMT3A* to which the high throughput reads generated by 454 sequencing were aligned. In case of a base pair mismatch, a red line indicating the site of the mutation (C to T) was observed. The mutant clone size was 32 % (indicated on the left as variation %) and sequencing depth was 600 (indicated on the right).

3.8.5 Mapping of *DNMT3A* mutations

DNMT3A mutations were mapped to the different domains of the protein (Fig. 3.12).

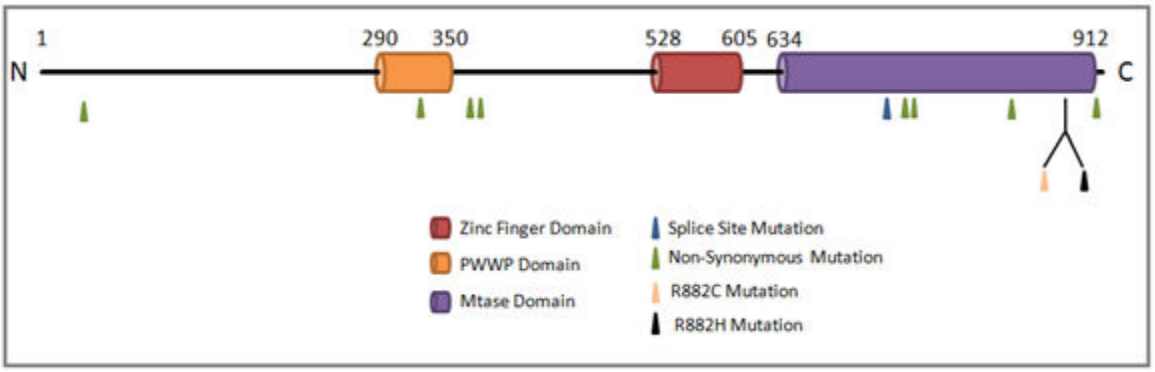


Figure 3.12: Mapping of mutations to the different domains of the *DNMT3A* protein.

Nine non-synonymous mutations (green) were identified across the entire length of the gene. Five non-synonmyous mutations (including P904) were localized in the Mtase domain which is responsible for the methyltransferase activity of the enzyme. A splice site mutation between exon 18 and the intron preceding exon 19 was also observed in this domain.

Mutation at the R882 locus identified frequently by other research groups was observed in 2/10 patients. The PWWP domain demonstrated one non-synonymous mutation. This domain is concerned with directing *DNMT3A* to heterochromatin. Three mutations were dispersed between the coding domains.

3.8.6 Correlation with EZH2 mutations

1/10 patients (LSL4062) showed mutations of both *EZH2* and *DNMT3A* in this cohort. This patient had mutation at the R882 site (clone size = 34 %) along with a mutation in exon 16 of *EZH2* at V626M (clone size = 27 %). Apart from one patient, mutations of *EZH2* and *DNMT3A* did not show any specific overlap or exclusivity in their occurrence indicating that though *EZH2*, *DNMT3A* proteins interact with each other, the mutational status of these genes are isolated events suggesting independent functionality.

3.9 Discussion

3.9.1 SNP6 and mutational analysis of PRC1 and Jumonji genes

Aberrations of *EZH2*, a core component of PRC2 were first identified by Ernst et al (2010) using SNP 6.0 arrays on peripheral blood /bone marrow samples from patients with MDS to discover acquired mutations of this gene associated with uniparental disomy of 7q and 7q36.1 microdeletion (Ernst et al., 2010). In this study a similar approach was used to identify aberrations of the other polycomb and Jumonji genes. SNP6 data for 91 patients with MDS generated for earlier studies by our group was examined for CNVs (amplification/deletion) and CN (LOH) affecting the PRC1 and Jumonji genes. 17 (19 %) patients showed aberrations of 5 genes belonging to PRC1 while 29 patients (31 %) showed SNP6 abnormalities at 15 Jumonji gene loci. Interestingly, this demonstrated that the frequency of SNP6 aberrations affecting the Jumonji genes was higher than the PRC1 genes. In the Jumonji genes the highest frequency of deletions was observed in *JMJD1B* (chr 5q31.2) [n=12] while in PRC1, *RING1A* (chr 6p21.32) was deleted with high frequency (n= 2). Patients with deletion of *JMJD1B* had 5q deletion which is a common cytogenetic abnormality associated with MDS. Whether loss of *JMJD1B* contributes to the 5q syndrome phenotype is unknown. *RING1A* on the other hand is essential to reduce self-polyubiquitination of *RING1B* and thus prevent its degradation (Wong et al., 2007). Whether reduced levels of *RING1A* due deletion of the *RING1A* locus affects *RING1B* mediated repression of transcription needs further clarification. Apart from amplifications, the occurrence of CN (LOH) was also noted in this cohort wherein the highest frequency was seen in *JMJD4* (n= 3) and *BMI-1* (n=11) genes. Literature on role of *JMJD4* in myeloid cells and its exact mechanism of action is not available and studies need to be undertaken to understand the effects of modulation of *JMJD4* expression in detail. In the case of *BMI-1*,the size of the CN (LOH) observed at this locus

were very small (≤ 1 Mb) raising a question of whether they were real but mutational analysis of these patients for *BMI-1* was still carried out due their high frequency of incidence (n= 11).

In addition to the presence of deletions and CN (LOH) in the PRC1 and Jumonji genes, amplifications/ gain were also observed in our patient cohort. The highest frequency of gain or amplifications (n= 3) were observed in *JMJD2C* and in the PRC1 genes at the *RING1B* locus [chr: 1q25.3] (n= 2). *JMJD2C* is a histone demethylase involved in removal of H3K9me3 marks. This enzyme has been studied in the context of embryonic stem cells (ES) wherein KO of *JMJD2C* levels followed by gene expression profiling showed gene signature similar to that of differentiated ES cells. Restoring the levels of *JMJD2C* in ES enhanced their growth and self renewal capacity.*JMJD2C* targets identified by both gene expression and ChIP seq overlapped with the H3K27me3 marks, *EZH2* and *SUZ12* (Das et al., 2014). It would be interesting to observe whether modulation of *JMJD2C* levels would interfere with PRC2 regulation of target genes and whether this connection between Jumonji and PRC2 is prevalent in even in myeloid cells. On the other hand, *RING1B* is a key component in transcriptional regulation because of its role in H2AK119Ub and its restraining effect on RNA polymerase II at the promoters of target genes. Inactivation of *RING1B* resulted in global loss of H2AK119Ub marks and release of RNA pol II enzyme (Stock et al., 2007). As explained in section 1.8, *RING1B* KO is linked to increased proliferation of myeloid progenitor cells by up regulation of cyclinD2 and cdc6. There is concomitant up regulation of *p16INK4A* causing decrease in the number of mature myeloid and lymphoid precursor cells (Cales et al., 2008). However it is unclear whether *RING1B* over expression has an opposite effect of reduced myeloid progenitor cells, increased mature myeloid precursor cells and whether it affects pathogenesis of MDS/AML.

Mutational analysis of all the patients with SNP6 abnormalities was carried out by sanger and 454 parallel sequencing but no mutations were identified in the Jumonji and PRC1 genes indicating that (a) the SNP6 status was not predictive of the presence of mutations in these genes and (b) the frequency of mutations of the Jumonji and PRC1 genes if present at all was very low in MDS/AML. Similar findings were reported by Score et al (2012) when they identified *JARID2* mutations (R717Q, D744N, and I1052V) in peripheral blood and bone marrow of 3/148 cases of MDS/MPN but did not find any association between CN (LOH) of chr 6p with these mutations. Therefore it can be concluded that mutations of PRC1 and Jumonji genes are rare in MDS/AML and are not associated with SNP6 abnormalities of these genes.

Next, I examined changes in gene expression of the Jumonji genes as a result of CNVs detected by SNP6 at these gene loci by qPCR. Due to insufficient quantities of patient samples qPCR could not be carried out for the PRC1 genes.

Altered expression of Jumonji proteins have been reported previously in carcinogenesis; Up regulation of *JHDM1B* in leukaemia facilitates proliferation of hematopoietic progenitor cells via demethylation of H3K36me2 and affects target *p15INK4B* (He et al., 2011). Another Jumonji gene associated with leukaemia is *JMJD1B* which is up regulated in acute lymphoblastic leukaemia (ALL) and causes repression of cancer cell differentiation (Kim et al., 2012).Besides leukaemia, over expression of *JMJD3* and *JMJD1A* have been shown to correlate with advance disease stage of prostate cancer (Xiang et al., 2007) and bladder cancer respectively. In bladder cancer, *JMJD1A* knock out (KO) caused G1 arrest of cancer cells suppressing proliferation (Cho et al., 2012).This indicates that the modulation of Jumonji gene expression can contribute to carcinogenesis and therefore these genes can be utilised as potential targets in cancer therapy (Sroczynska et al., 2014). In this study it was found that the expression of four Jumonji genes; *JARID2*, *JARID1A*, *JARID1C* and *UTX* was

modulated concomitant with SNP6 abnormality of these genes. *JARID2* expression was upregulated in one patient with trisomy of the *JARID2* locus while *JARID1A* locus (chr 12p13.33) was deleted (CN= 1) in two patients while a single patient had trisosmy (CN= 3) of *JARID1C* (Xp11.22), *UTX* (chr Xp11.3). CN= 4 was also observed at *UTX* locus in one patient. Deletion of *JARID1A* locus was linked with reduced gene expression (p value <0.0001) in both the patients while increased expression of *UTX* and *JARID1C* due to gain in copy number was seen in a single patient. *UTX* levels were also increased 5 fold concomitant with CN= 4 in another patient. Inactivating mutations and deletions of *UTX* have been reported in AML and CML cases (van Haaften et al., 2009) but over expression of *UTX* is a novel finding. Studies to understand the mechanistic effect of *UTX* have used small molecule inhibitors of *UTX* in human promyelocytic leukemic cell line with promising effects on reducing proinflammatory cytokine production by human primary macrophages (Kruidenier et al, Nature, 2012). However, downstream effects of *UTX* inhibition need further clarification. Chromosomal translocation of *JARID1A* is implicated in leukaemia with Nuclear pore complex protein (*NUP98)* as its binding partner (de Rooij et al., 2013;van Zutven et al., 2006; Wang et al., 2009). Functional analysis of the fusion protein demonstrated arrested in haematopoietic differentiation (Wang et al., 2009). This effect might be due to inability of the fusion protein to bind to its target H3K4me3 affecting the downstream targets of this histone mark. Whether *JARID1A* deletion mimics the fusion protein in its mechanism of leukemogenesis needs to be ascertained. On the other hand, *JARID1C* is implicated in mental retardation, renal carcinoma and cervical cancer. Mutations of this gene (30%) are associated with mental retardation (Blair et al., 2011). However its link to MDS has not been explored till date.

3.9.2 Mutational analysis of PRC2 and DNMT3A

EZH2 was previously sequenced by our group on 61 out of 91 patients (mentioned earlier) and therefore the same cohort of patients were selected for mutational analysis of the other core components of PRC2; *SUZ12*, *EED*, *EZH1* and *DNMT3A* to identify mutations in these genes and to examine the mutational overlap of these with *EZH2*. PRC2 genes were not mutated in our cohort of patients suggesting that the frequency of mutation of the other components of PRC2 is low compared to *EZH2.* Subsequently, similar findings were published by two research groups; Score et al (2012) and Brecqueville et al (2011) wherein the first group reported only 1.4 % (2/148) frequency of *SUZ12* mutation in their cohort with all the mutations affecting the VEFS domain. However, their study included cases with MDS, Myeloproliferative neoplasms (MPN) as well as myelofibrosis in their cohort. One of the mutations in *SUZ12* was identified in patient with MPN and was associated with CN (LOH) of 17q. This group also analyzed *EED* and identified 1% (1/148) frequency of mutation in this gene. They demonstrated that mutations of both *SUZ12/EED* caused loss of H3K27me3 levels (Score et al., 2012). On the other hand, Brecqueville et al (2011) focused entirely on *SUZ12* in MPN, chronic myelomonocytic leukaemia and polycythemia vera (PV) cases and found three missense mutations (2 MPN & 1 PV cases) /186 patients (1.6 %) of this gene (Brecqueville et al., 2011). Therefore it can be conclude that mutations of *EZH2* are more frequent in cases with MDS than those of *SUZ12*, *EED* or *EZH1* and therefore the focus needs to be on the functional analysis of *EZH2* in the context of myeloid cells. Since mutations in any component of PRC2 resulted in loss of H3K27me3, it emphasizes the importance of this mark and suggests that deregulation of the epigenetic control of gene expression plays an important role in the development of myeloid disorders.

In contrast mutational analysis of *DNMT3A* in the same cohort identified 10/61 patients with heterozygous mutations (clone size 20-44 %) of this gene. Interestingly, these mutations were seen predominantly (n= 6) in patients with monosomy 7/del 7q and not observed in patients with normal karyotype but were classified in RAEBI/II of the WHO classification suggesting the association of this gene with advanced stage of the disease. The precise role of these mutations in leukaemogenesis remains elusive but a number of studies have investigated their potential prognostic significance in AML. For this purpose, Shivarovet al (2013) examined 4582 AML patients and derived that, *DNMT3A* mutations appeared to be an independent adverse prognostic factor in younger patients with normal cytogenetics AML [Overall Survival (OS) [p = 0.01], Relapse Free Survival (RFS) [p= 0.0005]] and therefore can be used for risk stratification of AML patients in connection to integrated mutational profiling (Shivarov et al., 2013).More recently, Tie et al (2014) reviewed 12 studies (n= 6377) focusing on implications of *DNMT3A* mutations on OS in de novo AML and concluded that patients with *DNMT3A* mutations had slightly shorter OS (HR = 1.60; 95% CI, 1.31–1.95; p <0.001), as compared to wild-type carriers. In addition, mutant *DNMT3A* predicted inferior OS (HR = 2.30; 95% CI, 1.78–2.97; p = 0.862) in patients with unfavourable genotype abnormalities but there was no significant prognostic impact on the OS in the favourable genotype subgroup. They also looked at event, relapse-free survival and found similar results (Tie et al., 2014), (Ribeiro et al., 2012). In contrast, there was no significant difference in the median survival of monosomy 7 patients with or without *DNMT3A* mutations (p value = 0.6) in our cohort with inconclusive impact on prognosis of the patients with mutated *DNMT3A* based on the patients that survived.

3.9.3 Mutational overlap

A single patient was mutated for both *EZH2* and *DNMT3A* in this study. This patient had mutation at the R882 site (clone size = 34 %) along with a mutation in exon 16 of *EZH2* at V626M (clone size = 27 %) site. The functional significance of these overlapping mutations is not clearly understood and have not been studied so far. Apart from one patient, mutations of *EZH2* and *DNMT3A* did not show any specific overlap or exclusivity in their occurrence. Similar results were observed by Haferlach et al (2014) when they sequenced 944 patients belonging to various subtypes of MDS and found *DNMT3A* to be amongst the commonly mutated genes (> 10 %) but they did not find any particular link between *EZH2* and *DNMT3A* mutations. *DNMT3A* mutations in their study showed overlap with *SF3B1* (Splicing Factor 3b) mutations but were mutually exclusive with *SRSF2* and *ASXL1* mutations (Haferlach et al., 2014). *DNMT3A* mutations in AML are often found with mutations of Nucleophosmin *(NPM1)*, Fms-Related Tyrosine Kinase 3 (*FLT3)*, *IDH1* or *IDH2* (isocitrate dehydrogenase1/2) but never with *RUNX1* (Runt-Related Transcription Factor) (Shah and Licht, 2011).The functional consequence of these overlapping mutations and their individual or cumulative effect on the pathogenesis of MDS needs to be ascertained.

3.10 Conclusion

SNP6 karyotyping facilitated identification of CNVs (deletions, amplifications) and CN (LOH) in Jumonji and PRC1 genes. The frequency of SNP6 abnormalities in the Jumonji genes was high (approx double) as compared to PRC1. Though SNP6 abnormalities were not associated with mutations of the Jumonji and PRC1 genes, gene expression modulations were identified in the patients with CNVs at the Jumonji loci suggesting changes in gene expression as a important mechanism for disease pathogenesis.

Sanger and 454 parallel sequencing of *EZH1, SUZ12, EED* and twelve genes belonging to PRC1 did not identify mutations associated with these genes confirming that among the polycomb proteins, *EZH2* is the most important link to MDS/AML. **Therefore the next part of my thesis is focused on examining the functional consequences of *EZH2* mutations, knock down and over expression of this gene in the context of myeloid cell lines.**

16 % patients in this study were mutated for *DNMT3A*. The high frequency of mutations associated with this gene identified by us and other research groups, their adverse impact on the prognosis of haematological malignancies and the site specificity of the mutation (R882) is evidence that the enzymatic activity of this gene gets altered/ compromised in myeloid malignancies making it a crucial target for therapies designed to treat MDS. **Therefore a part of this thesis is based on constructing the R882 mutation in plasmid vectors and studying their effect on myeloid cell lines.**

4 Chapter: Role of EZH2 mutations (R690C & R690H) in Myeloid Malignancies

4.1 Introduction

As discussed in Chapter 1 (refer section 1.7), the incidence of *EZH2* mutations is 8 % in patients with MDS/MPN and 2 % in AML. In contrast to patients with lymphomas where the Y641 site of *EZH2* is most frequently mutated, there is no hotspot for mutation observed in myeloid malignancies. Truncation mutations occur throughout the gene while missense mutations were commonly found in the conserved SET domain (Ernst et al., 2010b). Mutations affecting the R690 site in the SET domain of *EZH2* have been reported in many studies (Ernst et al., 2010; Jerez et al., 2012; Makishima et al., 2010; Nikoloski et al., 2010) including by us though the frequency of occurrence of this mutation in a single cohort was not high. Two mutations R690C and R690H are identified and frequently correlate with the absence of H3K27me3 mark (Ernst et al., 2010) but the precise downstream targets affected by these mutations have not been examined till date. Their strategic location in the SET domain of *EZH2* which is responsible for the histone methyltransferase activity of the enzyme makes them an ideal candidate to study the role of *EZH2* in myeloid malignancies.

4.2 Aim

- To construct R690C and R690H mutations of *EZH2* in a plasmid vector. To generate Flag-tagged wild type *EZH2* in order to over express the wild type. Transfect both mutants and the Flag-tagged wild type *EZH2* in myeloid cell lines.
- To examine the effect of the two mutations and over expression of *EZH2* on its target; H3K27me3 protein levels in myeloid cell lines.
- To identify and compare the gene signature as a consequence of the mutations and over expression of the gene by microarray gene expression profiling to identify *EZH2* target genes.

4.3 Constructing the R690C & R690H mutations in pCMV sport6 vector

Wild type *EZH2* cDNA cloned in pCMV-SPORT6 (IRATp970F1210D) was obtained from Source Bioscience and was used for all experiments (refer appendix). This *EZH2* cDNA sequence was verified using different primer pairs (refer appendix) by Sanger sequencing.

To introduce the R690C and R690H mutations in the wild type *EZH2* cDNA (Fig: 4.1), the plasmid was digested by SalI-NotI restriction enzymes (refer 2.14.1) which generated two fragments (size: 4396 & 2500 bp) as visualized on 1 % agarose gel (Fig: 4.2a). The large fragment (4396 bp) was excised under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions while the small fragment was discarded. Secondly, pCMV-SPORT6 was digested with SalI-EcoRI restriction enzymes which generated two fragments (size: 5000 & 2335 bp) (Fig: 4.2b) of which the small fragment was gel purified while the large fragment was discarded.

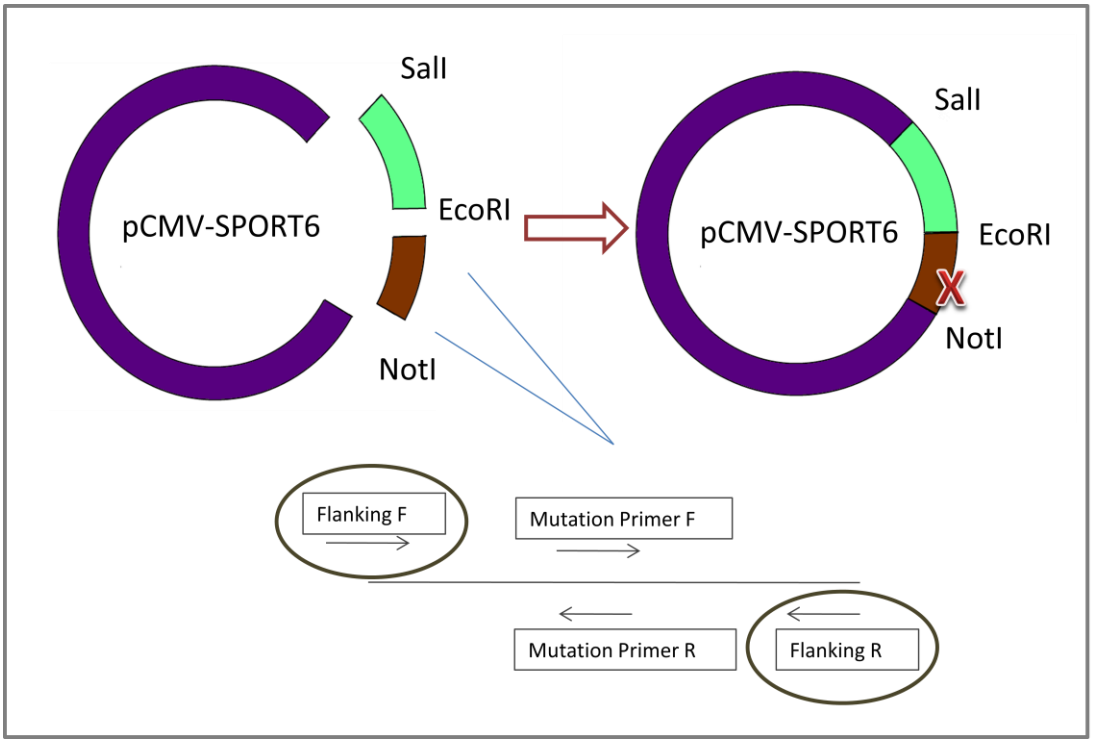


Figure 4.1: Cloning strategy for introducing the R690C and R690H mutation in pCMV-SPORT6. Red cross indicates the site of the mutation. Generation of the mutant in the EcoRI-NotI fragment has been illustrated using two pairs of primers (of each pair one primer flanked the mutational site and a second primer overlapped the mutation site and introduced the mutation), using a PCR based approach (See text).

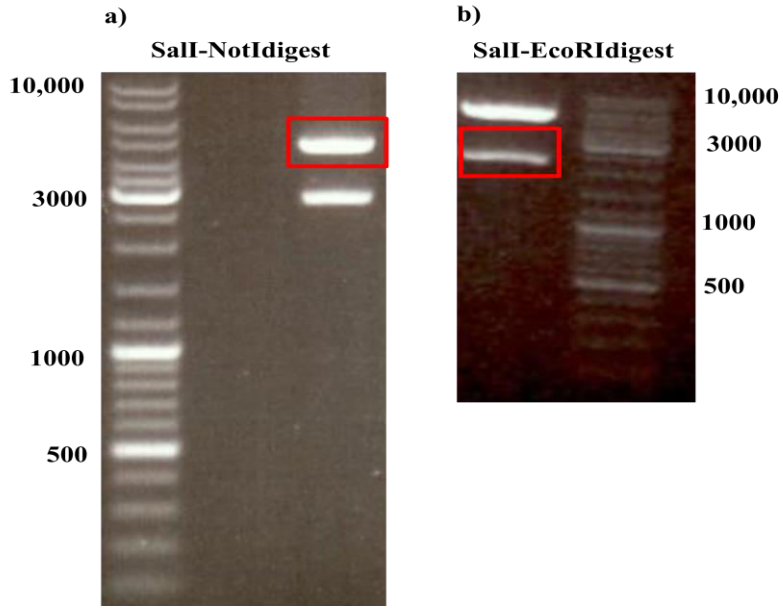


Figure 4.2: Digestion of pCMV-SPORT6 vector containing *EZH2* cDNA(a) pCMV-SPORT6 digested by SalI-NotI to generate the empty vector (4396 bp). (b) pCMV-SPORT6 digested by SalI-EcoRI to generate a fragment of the *EZH2* cDNA (size: 2335 bp)

4.3.1 Generation of fragment containing R690C & R690H mutation

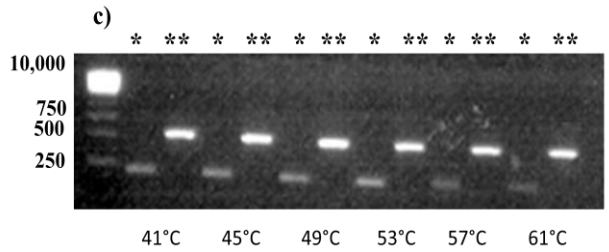
The mutation was generated using two sets of primers; the first pair of primers overlapped the position 690 [mutational primer F (MF) and mutational primer reverse (MR)]. This primer pair was different for the R690C and R690H mutation. The change in the base pair to be introduced was positioned exactly at the centre of the primer. The second pair of primer or flanking primers [flanking forward (FF) and flanking reverse (FR)] were same for both the mutations and were designed just outside EcoRI-NotI restriction enzyme sites. PCR amplification using these primers was carried out on wild type pCMV-SPORT6 vector at a range of temperatures (For R690C: 41˚C, 45˚C, 49˚C, 53˚C, 57˚C & 61˚C and for R690H: 57˚C & 60 ˚C) using the following conditions:

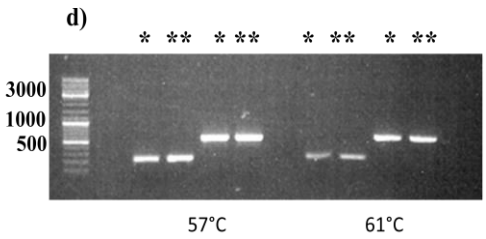
a)

Reagent	Volume
Promega GoTaq (colourless)	10 µl
Primer (5 µM)	0.5 µl FF + 0.5 µl MR OR 0.5 µl FR + 0.5 µl MF
EZH2 cDNA plasmid (10 ng/ µl)	1 µl
H_20	8 µl
Total	20 µl

b)

Cycles	Temperature	Time	Comment
1	95 °C	2 min	Enzyme activation
35	95 °C	30 sec	Denaturation
	R690C: 41°C, 45°C, 49°C,53°C,57°C, 61°C	30 sec	Primer annealing
	R690H: 57°C,61°C		
	72 °C	30 sec	Elongation
1	72 °C	5 min	Inserting the missing nucleotides





e)

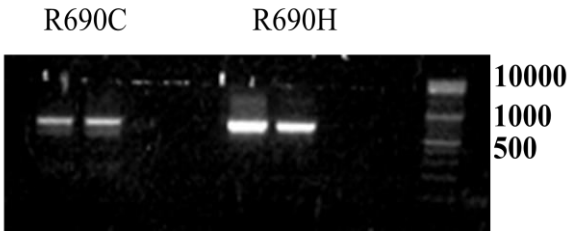


Figure 4.3: Generation of mutational fragment R690C and R690H. Table (a) and (b) show reaction mixture and the PCR conditions used to generate the fragment of *EZH2* cDNA containing either R690C or R690H mutations. Gel picture (c) for R690C and (d) for R690H indicate visualisation of the PCR products at different temperatures (41-61˚C) using forward flanking primer + reverse mutational primer (*) and reverse flanking primer + forward mutational primer (**). Visualisation of the second round of PCR product using only the flanking primers for both mutations is shown in (e). Gene O'Ruler ladder was used in all the gel pictures except for (c) were the gene ruler 1 kb ladder was utilised.

The product from the second PCR (R690C & R690H: 61˚C) was gel purified, the fragments were treated with SuperTaq and incubated for 30 min at 72 ˚C to add dNTP "A" to the fragment.

Reagents	Volume
PCR fragment (24.34 ng/ µl)	12 µl
dNTPs (2 mM)	1 µl
Super Taq	0.25 µl
10X buffer	2 µl
H_20	upto 20 µl

Table 4.1: Reaction mix for treatment of plasmid DNA with SuperTaq to add dNTP "A" to the fragment.

Subsequently, 2 µl of the above mixture was cloned in to pCRII- TOPO vector and transformed into DH5α-T1 *E.Coli* cells (refer 2.14.4). After over night incubation of the LB agar + 100 mg/ml ampicilin +Xgal (20 µg/ml) plates, five colonies were selected and miniprep was carried out using the Wizard Plus SV kit (Promega, UK). All the colonies showed correct band size (700 bp) (Fig: 4.4) when verified by digesting the plasmid with

EcoRI restriction enzyme (refer 2.14.1). The presence of the mutations (R690C and R690H) was confirmed by sanger sequencing of the respective colonies using M13 primers (data not shown).

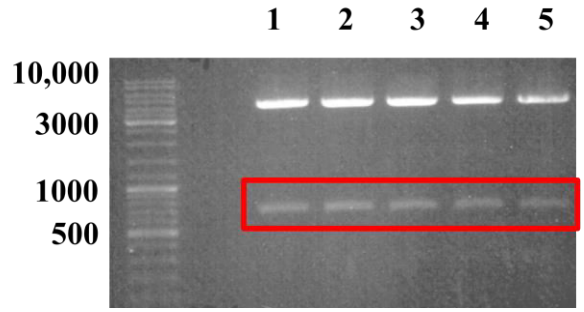


Figure 4.4: Verification of the correct colony after cloning the PCR fragment in pCRII-TOPO vector. Five colonies were digested by EcoRI restriction enzyme and all the colonies showed the correct band size (700 bp)

A single colony was verified by EcorI digest and sanger sequencing for each mutant and was next digested by EcoRI-NotI, the fragment (700 bp) was gel purified and stored for ligation.

4.3.2 Three-way ligation

The concentrations of the plasmid DNA for all three fragments generated by restriction enzymes; SalI-NotI, SalI-EcoRI (both these fragments were common for the R690C and R690H mutations) and the fragment containing the mutation (EcoRI-NotI) (separate for each mutant) were measured using Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer and ligated using T4 DNA ligase. The control reaction contained only the empty vector but none of the other fragments. This control was used to check the number of background non-specific colonies. The ligation was carried out over night at 4 ˚C.

Reagents	Volume (For 50 ng)
Empty vector (SalI-NotI: 4390 bp) [16.2 ng/ µl]	3.1µl
SalI-EcoRI: 2335 bp (18.25 ng/ µl)	2.9µl
EcoRI-NotI: 700 bp [fragment containing R690C (12.19 ng/ µl) or R690H mutation (22.10 ng/ µl]	R690C (1.3µl) OR R690H (0.72 µl)
10X Ligase buffer	1 µl
T4 DNA ligase	0.5 µl
H_20	upto 10µl

Table 4.2 :Ligation mixture for three fragments (SalI-NotI, SalI-EcoRI & EcoRI-NotI) of pCMV-SPORT6. Empty vector: Insert ratio for ligation was 1:2.

0.2 µ1 of the ligation mixture was transformed into MegaX DH10B T1 Electrocomp cells (refer 2.14.7) and plated on LB agar + 100 mg/ml ampicilin plates for 16 hr at 37°C. Four colonies for R690C and three for R690H mutants were selected and accurate ligation was confirmed by SalI-NotI digest (band size: 2500 bp) and sanger sequencing using M13 primers. All colonies showed the correct size fragment on SalI-NotI digest and contained either the R690C or R690H mutant. One colony for each mutant was selected and maxiprep was carried out using Qiagen Maxi Prep plasmid kit (refer 2.14.10).

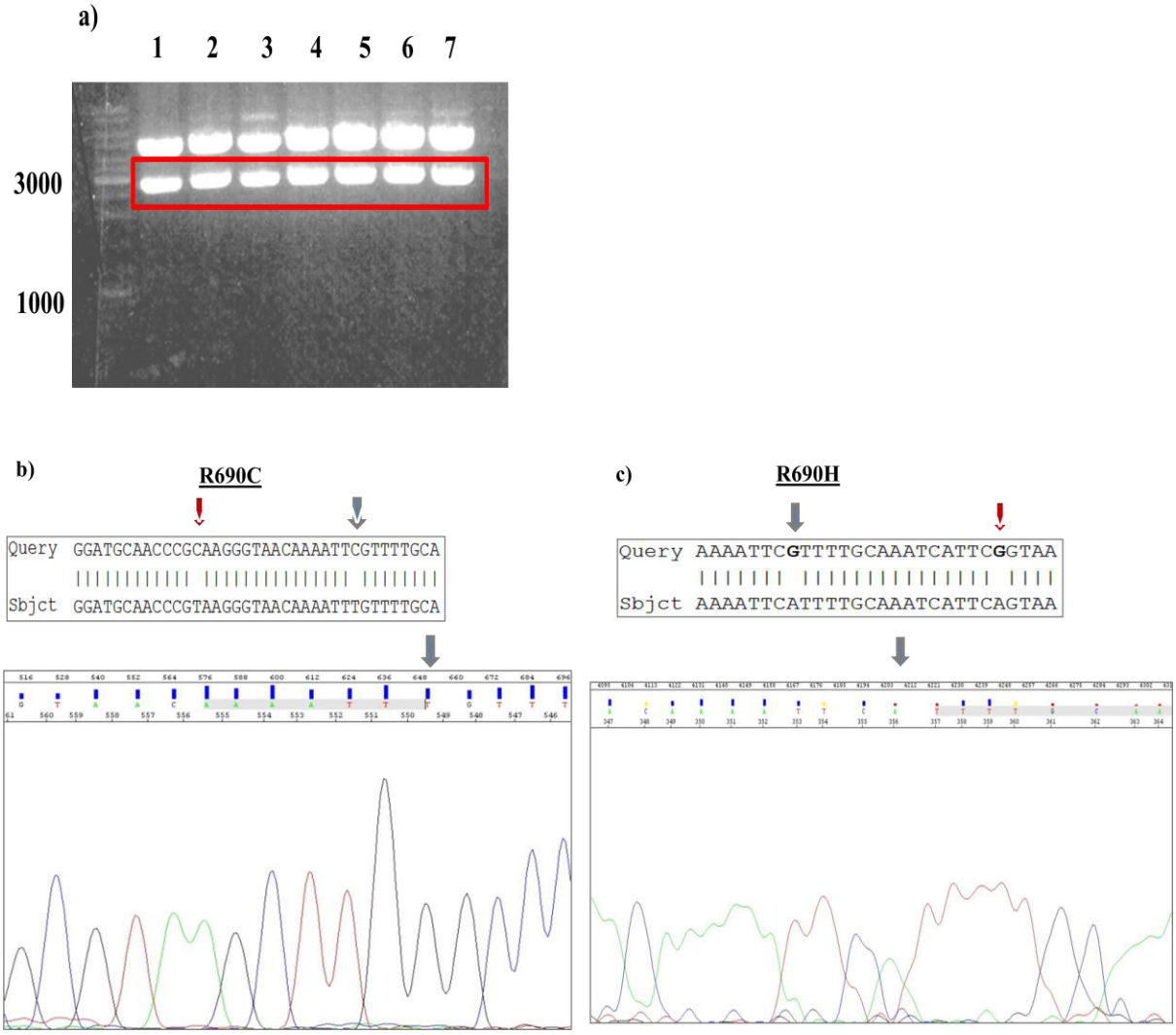


Figure 4.5: Confirmation of R690C and R690H mutant in pCMV-SPORT6 vector. (a) Digestion of seven colonies (four for R690C and three for R690H) by SalI-NotI restriction enzymes illustrate correct band size of the insert (2500 bp) for all the colonies. Sanger sequencing of one colony per mutant [R690C: (b)] and [R690H: (c)] confirm the change in amino acid [grey arrows] from Arginine (CGT) to Cysteine (TGT) for R690C and for R690H from Arginine (CGT) to Histidine (CAT). Two silent mutations were observed, one in the fragment carrying the R690C mutation [CGC to CGT (Arginine to Arginine)] and the second in the fragment carrying the R690H mutant [TCG to TCA (Serine to Serine)] (red arrows).

4.4 Cloning R690C & R690H mutants into p3XFLAG-myc-CMV-26 expression vector

A selectable marker (neomycin) and a Flag tag were attached to the R690C and R690H *EZH2* mutants to help distinguish it from the endogenous *EZH2* in the cells. Therefore *EZH2* cDNAs containing R690C or R690H in pCMV sport 6 vector were cloned into p3XFLAG-myc-CMV-26 Expression Vector (Sigma Aldrich, UK) (Fig: 4.6). **To study the effect of over expressing wild type *EZH2* in myeloid cells, *EZH2* cDNA without the mutations was also cloned into p3XFLAG-myc-CMV-26 Expression vector.**

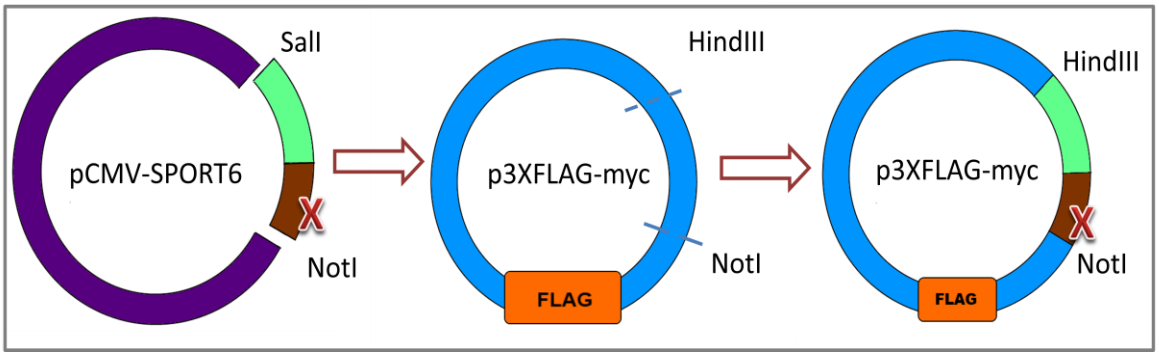


Figure 4.6: Cloning Strategy. Cloning strategy for inserting *EZH2* cDNA containing R690C/ R690H and the Flag tagged wild type gene from pCMV-SPORT6 to pX3FLAG-myc. The insert containing the mutations and the Flag tagged wild type *EZH2* were isolated from pCMV-SPORT6 by restriction enzymes SalI-NotI. p3XFLAG-myc was linearised using HindIII-NotI and the insert containing *EZH2* was ligated into this vector by joining the NotI digested ends of both the plasmids.

4.4.1 Linearise the p3XFLAG-myc-CMV-26 expression vector for ligation

p3XFLAG-myc-CMV-26 expression vector (Sigma, UK) was digested by restriction enzyme HindIII for 1 hr at 37° C (refer section 2.14.1) followed by conversion of DNA termini to blunt ends (refer section 2.14.2) and digested by NotI restriction enzyme for 1hr at 37° C. The

linearised plasmid was visualized on 1 % agarose gel (Fig: 4.7a), the band was excised under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions.

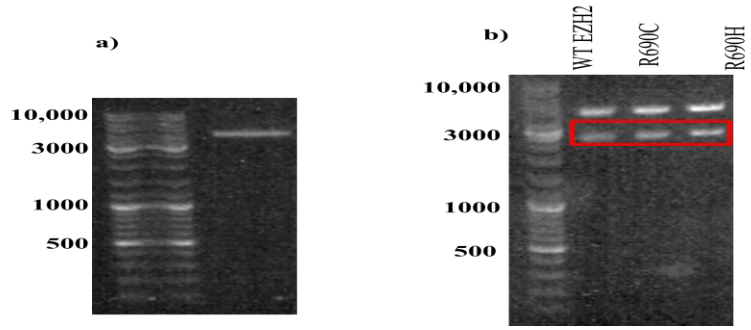


Figure 4.7: Cloning EZH2 cDNA into p3XFLAG-myc-CMV. (a) p3XFLAG-myc-CMV vector was linearised using HindIII-NotI restriction enzymes and visualised on 1 % agarose gel. (b) *EZH2* cDNA containing R690C/R690H or wild type *EZH2* were isolated from pCMV sport6 using restriction enzymes SalI-NotI (size = 2500 bp).

4.4.2 Isolation of EZH2 cDNA from pCMV sport6 vector

pCMV sport6 vectors containing either R690C/R690H or wild type *EZH2* were digested using restriction enzyme SalI for 1 hr at 37 °C (refer section 2.14.1) followed by conversion of DNA termini to blunt ends (refer section 2.14.2) and digested next by NotI restriction enzyme for 1hr at 37 °C to isolate *EZH2* cDNA from pCMV sport6 plasmid (Fig: 4.7b). The band (2400 bp) was visualized on 1 % agarose gel, excised under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions.

4.4.3 Ligation of EZH2 cDNA and p3XFLAG-myc-CMV-26 expression vector

Gel purified *EZH2* cDNA (R690C/R690H/wild type) was ligated to the linear p3XFLAG-myc-CMV-26 using the conditions mentioned in table 4.3. 0.2 µl of the ligation mixture was transformed into MegaX DH10B T1 Electrocomp cells (refer 2.14.7) and plated on LB agar +

100 mg/ml ampicilin plates for 16 hr at 37°C. Five colonies each for the two mutants and the wild type *EZH2* were selected and accurate ligation was confirmed by HindIII-NotI digest (Fig: 4.8). Sanger sequencing was carried out using primers designed against *EZH2* cDNA to verify the presence of the mutations. One accurate colony for R690C, R690H and wild type *EZH2* each was selected for maxiprep using the Qiagen Maxi Prep plasmid kit (refer 2.14.10).

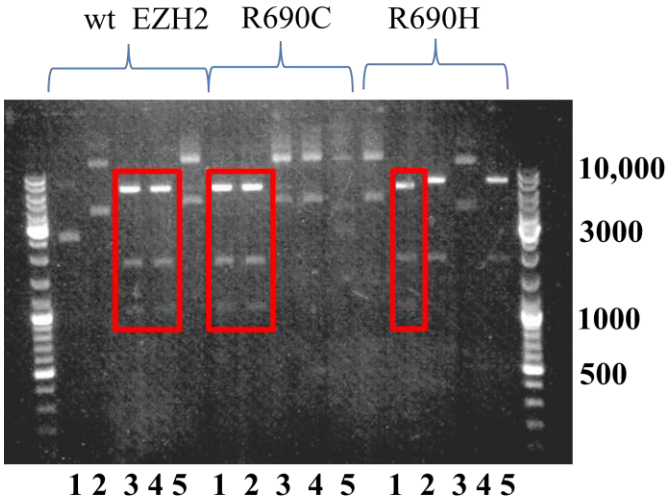


Figure 4.8: Verification of accurate ligation of *EZH2* cDNA into p3XFLAG-myc-CMV plasmid. Five colonies per mutant R690C/R690H and wild type *EZH2* were selected and digested by HindIII-NotI restriction enzymes. The correct colonies carrying wild type *EZH2*, R690C and R690H mutants are indicated by red box and contained the correct band sizes (1000 bp, 1800 bp and 6000 bp).

Reagents	Volume (For 50 ng)
Linear p3XFLAG-myc-CMV-26 (23 ng/µl)	2.17µl
cDNA *EZH2* (R690C: 8.6 ng/ µl, R690H: 8.3 ng/ µl, wild type *EZH2*: 8.4 ng/ µl)	R690C: 5.4 µl, R690H: 2.8µl, wild type: 5.4 µl
10X Ligase buffer	1 µl
T4 DNA ligase	0.5 µl
H_20	upto 10µl

Table 4.3 :Ligation of EZH2 cDNA (R690C /R690H/ wild type EZH2) into p3XFLAG-myc-CMV-26 expression vector. Linear p3XFLAG-myc-CMV-26: Insert ratio for ligation was 1:2.

4.5 Transfection of R690C/R690H & wild type EZH2 into myeloid cell lines

Five myeloid cell lines; U937, KG1, NB4, K562 and MOLM13 were screened for *EZH2* expression levels. RNA was extracted from the cell lines, converted to cDNA using the Superscript Vilo kit and qPCR (refer section 2.9-2.11) was performed to estimate *EZH2* mRNA levels (Fig: 4.9).

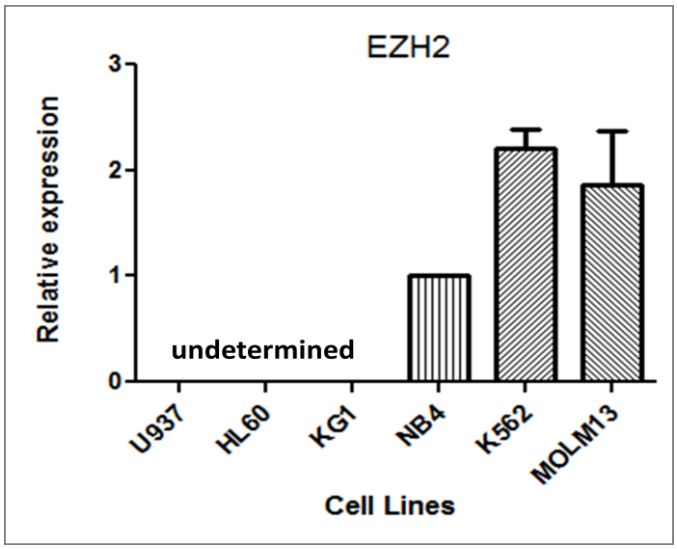


Figure 4.9: Endogenous expression of *EZH2* in myeloid cell lines (U937, KG1, NB4, K562 and MOLM13) by qPCR analysis. K562, NB4 and MOLM13 cell lines showed a higher expression of *EZH2* compared to U937, HL60 and KG1 which did not express endogenous *EZH2*. Therefore MOLM13 (wild type *TP53*) and K562 (*TP53* null) cell lines were selected for all experiments.(Y axis: Fold change compared to endogenous control *GAPDH*) Error bars= SEM in two experiments.

K562, NB4 and MOLM13 cell lines showed a higher expression of *EZH2* compared to U937, HL60 and KG1 which did not express *EZH2*. *TP53* expression levels are deregulated in K562 (insertion in exon 5 of *TP53* (Law et al., 1993)), NB4 (missense mutation in codon 319 and 344) (Fleckenstein et al., 2002) but MOLM13 cells carry wild type *TP53* (Kojima et al., 2005) therefore K562 (*TP53* null) and MOLM13 (wild type *TP53*) cell lines were selected for all subsequent experiments. The presence of wild type *EZH2* in both K562 and MOLM13

cells was confirmed by sanger sequencing using primers designed against *EZH2* cDNA sequence.

4.5.1 Electroporation& Lipofectamine transfection methods

p3XFLAG-myc-CMV-26 carrying the R690C/ R690H/ wild type *EZH2* were transfected using lipofectamine (refer section 2.15.2) and electroporation (refer section 2.15.4) methods into both K562 and MOLM13 cell lines. The experimental workflow is detailed in the Figure 4.10. 95% of the MOLM13 and 65% K562 cells were dead at 48 hr as estimated by Trypan blue staining which could be attributed to electroporation since similar percentage of cell death was also observed in the mock-transfected cells. Therefore Lipofectamine LTX was used for all subsequent transfections and cells were selected using neomycin (400 µg/ml) 48 hr after transfection.

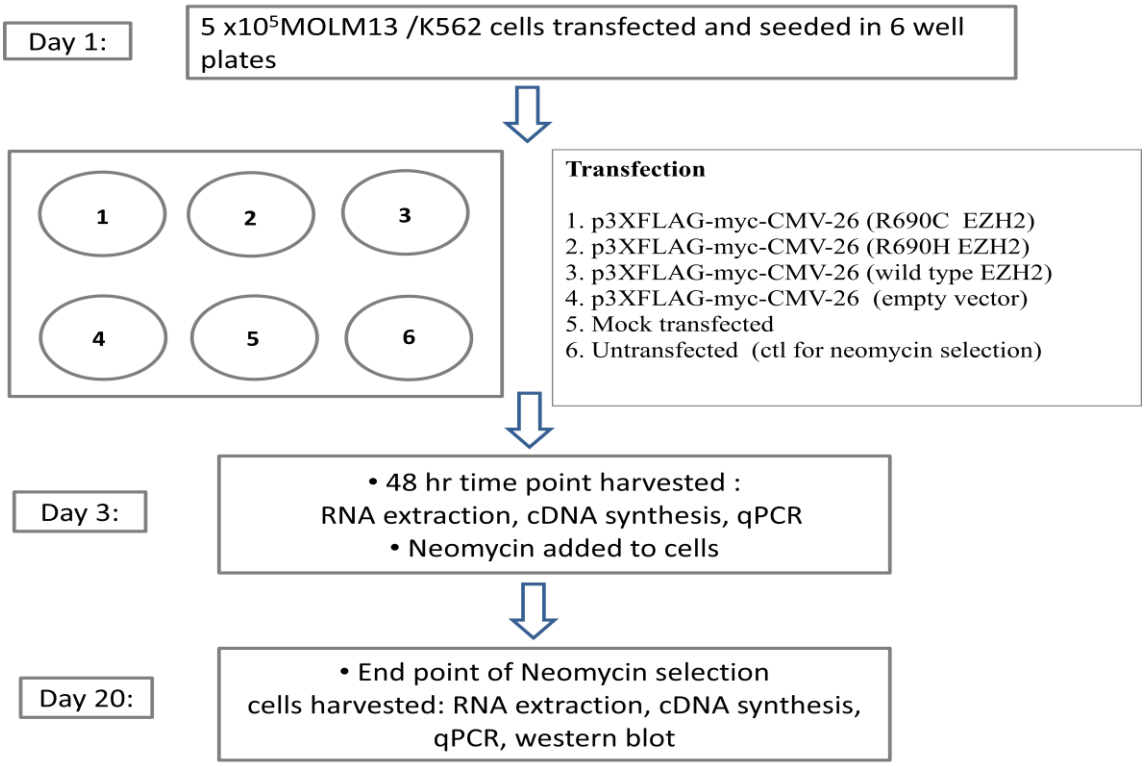


Figure 4.10: Experimental workflow for transfection by Lipofectamine LTX.MOLM13/K562 cells were seeded at a density of 5 x10^5 in 6 well plates and transfected with 2 µg of p3XFLAG-myc-CMV-26 vector carrying either R690C or R690H mutant or

wild type *EZH2* using Lipofectamine LTX solution as per the manufacturer's instructions (refer section 2.16.2). Mock transfected cells were used as a ctl for alterations caused due to the transfection procedure. The untransfected cells were used as a ctl to estimate the end point of neomycin selection for the cells.

Neomycin selection was continued for 18-20 days till all the untransfected cells (ctl) were dead by trypan blue dye exclusion method. Cells were harvested at 48 hr after transfection before neomycin selection and on day 21 after neomycin selection for qPCR (Fig: 4.11).

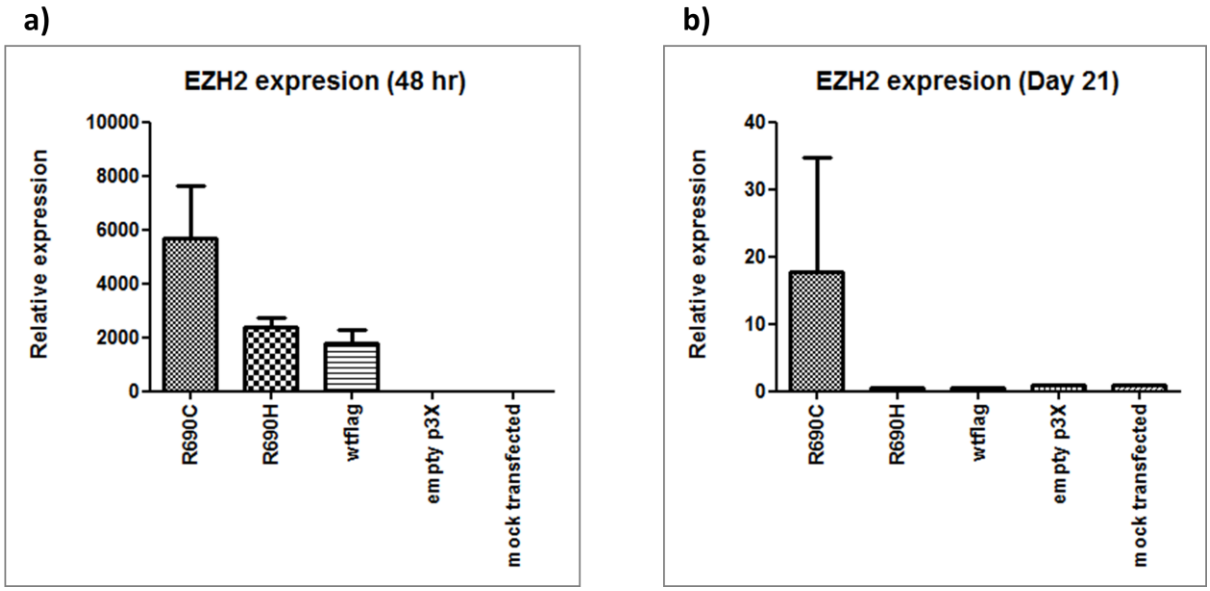


Figure 4.11: qPCR analysis of *EZH2* expression in MOLM13 cells, 48 hr and 21 days after transfection.*EZH2* mRNA levels were high (1000-5500 fold) in cells transfected with R690C, R690H or flag tagged wild type (wt flag) *EZH2* compared to those transfected with empty p3XFLAG-myc-CMV-26 without *EZH2* cDNA and mock transfected cells at 48 hr. However, these levels reduced and were variable (0.56 to 17.5 fold) at day 21 after neomycin selection indicating CMV promoter deactivation. Error bars= SEM in three experiments.

EZH2 mRNA expression was high (1000-5500 fold) in cells transfected with R690C, R690H or FLAG tagged wild type (wt flag) *EZH2* compared to those transfected with empty p3XFLAG-myc-CMV-26 without *EZH2* cDNA and mock transfected cells at 48 hr. The expression level of *EZH2* in cells selected by neomycin on day 21 was variable ranging between 0.56 to 17.5 fold compared to the mock tranfected cells in both cell lines (Fig: 4.11,

data not shown for K652). The probable cause for reduced *EZH2* expression over time might be due to its promoter inactivity. Cytomegalovirus (CMV) promoter drives the expression of *EZH2* in p3XFLAG-myc-CMV-26. Although the CMV promoter allowed for a very strong short-term expression (48 hr after transfection) of *EZH2*, it declined within 3-4 weeks and in some samples was barely detectable at the end of neomycin selection (day 21-24). CMV promoter can undergo methylation of CpG islands after few weeks of transfection making it inactive. To test whether CpG methylation of the CMV promoter was responsible for decreased *EZH2* expression, cells transfected with wt flag and R690H (since *EZH2* expression in these cells was lower than endogenous *EZH2* expression levels in mock transfected cells) were treated with 5-Azacytidine (500 nM, 1 µM, 5 µM) and the expression of *EZH2* was examined 48 and 72 hr after drug treatment by qPCR (Fig: 4.12). These time points were chosen to allow the cycles to undergo cell cycle so that the drug could incorporate into the replicating DNA to block DNA methyltransfearse resulting in demethylation of the CpG islands.

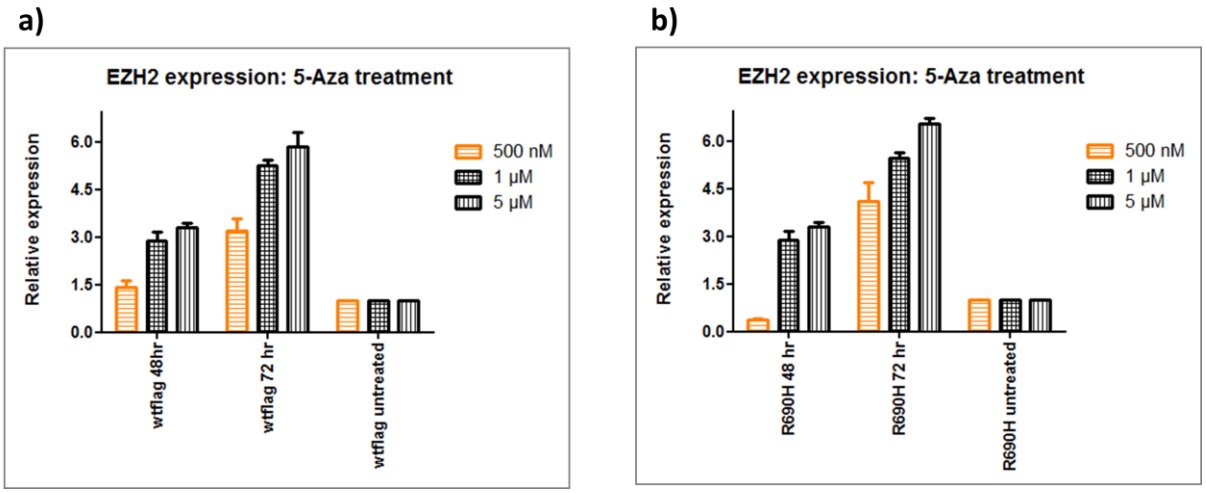


Figure 4.12: qPCR analysis of *EZH2* expression in MOLM13 cells, 48 and 72 hr after 5-Azacytidine treatment. Three concentrations (500 nM, 1 µM and 5 µM) of the drug were used to treat MOLM13 cells transfected with Flag tagged wild type *EZH2* (a) and R690H mutant (b). *EZH2* mRNA levels (wt flag) increased from 0.5 fold (Fig: 4.11) in untreated cells to 1.4-5.8 fold in cells treated with 5-Azacytidine. The time points were chosen to allow the cells to undergo cell cycle and for the 5-Azacytidine to incorporate into the DNA. Higher

concentrations of the drug (1 µM and 5 µM) were associated with better recovery of *EZH2* expression levels indicating a dose dependent response. Similar association between *EZH2* expression and drug dose was also observed in cells transfected with R690H mutant except for cells treated with 500 nM of 5-Azacytidine which showed reduced *EZH2* expression at 48 hr. This was observed in two independent experiments and the cause of this remained unclear. Error bar= SEM in two independent experiments, Yaxis: ΔΔ CT Fold change compared to untreated cells containing wtflag.

The expression levels of *EZH2* recovered after 5-Aza treatment at both 48 and 72 hr time points in cells transfected with Flag tagged wild type *EZH2* confirming that the reduced expression of the transgene was due its promoter methylation. The level of *EZH2* at 48 hr in cells carrying the R690H mutation was lower than the untreated cells which could not be explained even after repeated experiments and still remains unclear. However the expression of R690H *EZH2* recovered similar to the wt flag *EZH2* at both the time points at 1 and 5 µM concentrations. Therefore to avoid long selection periods, I cloned all (R690C, R690H and wt flag *EZH2*cDNA) into p3XFLAG-myc-CMV-26 containing GFP and puromycin markers instead of neomycin. GFP and puromyin markers were isolated from pBABE puromycin vector (gifted by Dr. Joop Gaken, Department of haematological medicine, KCL) and ligated into empty pX3FLAG-myc-CMV (26) (data not shown). This empty vector containing GFP and puromycin was used for cloning wild type and mutant *EZH2*cDNA as detailed below.

4.6 Cloning EZH2 cDNA into p3XFLAG-myc-CMV-26 containing GFP and puromycin

4.6.1 Construction of wild type EZH2

Wild type *EZH2* present in p3XFLAG-myc-CMV-26 was digested by; 1) restriction enzyme HindIII for 1 hr at 37 ˚C (refer section 2.14.1) followed by conversion of DNA termini to blunt ends (refer section 2.14.2) and digestion by BstEII restriction enzyme for 1hr at 60 ˚C. 2) The plasmid was also digested by HindIII for 1 hr at 37 ˚C (refer section 2.14.1) followed by conversion of DNA termini to blunt ends (refer section 2.14.2) and digestion by BsiWI restriction enzyme for 1hr at 55 ˚C and 3) wild type *EZH2* was digested by BstEII at 60 ˚C for 30 min followed by BsiWI at 55 ˚C for 30 min. Fragment size of 380 bp (HindIII-BstEII), 4400 bp (HindIII-BsiWI) and 4200 bp (BstEII-BsiWI) were visualized on 1 % TAE agarose gel, excised under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions (Fig: 4.13).

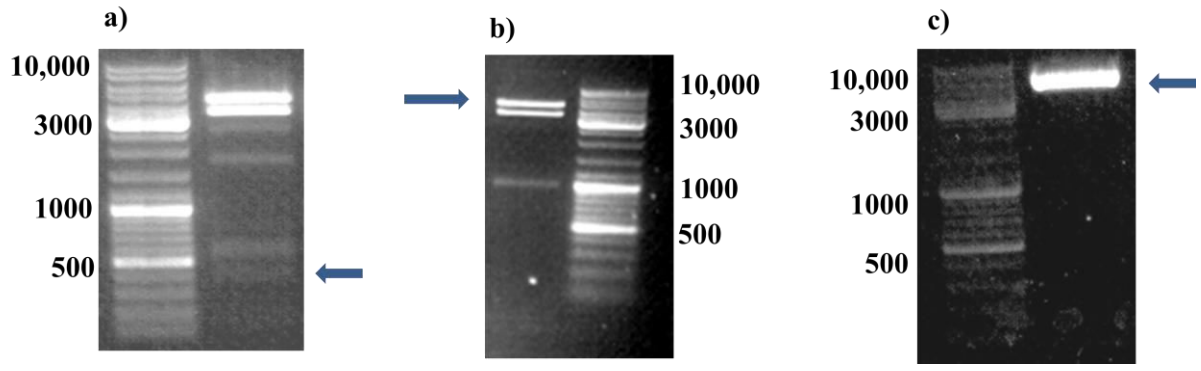


Figure 4.13: Cloning *EZH2* cDNA into p3XFLAG-myc-CMV-26 containing GFP and puromycin. Wild type *EZH2* in p3XFLAG-myc-CMV-26 was digested by:(a) restriction enzyme HindIII for 1 hr at 37 ˚C followed by conversion of DNA termini to blunt ends and next by BstEII restriction enzyme for 1hr at 60 ˚C. The band at 380 bp (blue arrow) was gel purified. (b) The plasmid was also digested by HindIII for 1 hr at 37 ˚C, DNA termini converted to blunt ends and digested by BsiWI restriction enzyme for 1hr at 55 ˚C (band size; 4400 bp). Lastly,wild type *EZH2* in p3XFLAG-myc-CMV-26 was digested (c)by BstEII at 60 ˚C for 30 min followed by BsiWI at 55 ˚C for 30 min to isolate a fragment of 4200 bp (blue arrow). The three fragments (380 bp, 4400 bp and 4200 bp) were gel purified and used for ligation.

The three gel purified fragments were ligated under following conditions:

Reagents	Volume (for 50 ng)
Fragment (HindIII-BstEII): 380 bp	1µl
Fragment (HindIII-BsiWI): 4400 bp	3.5 µl
Fragment (BstEII-BsiWI): 4200 bp	2.7 µl
10X T4 Ligase buffer	1 µl
T4 DNA ligase	0.5 µl
H_20	upto 10µl

Table4.4: Ligation of wild type *EZH2* into p3XFLAG-myc-CMV-26 expression vector containing GFP + puromycin markers. The fragments (HindIII-BsiWI, BstEII-BsiWI): Insert (HindIII-BstEII) ratio for ligation was 1:2.

2 µl of the ligation mixture was transformed into DH5α-T1 *E.Coli* (refer 2.14.5) chemo-competent cells and plated on LB agar + 100 mg/ml ampicilin plates for 16 hr at 37°C. Eleven colonies were selected and accurate ligation was confirmed (Fig: 4.14 - 4.15) by HindIII digest and sanger sequencing using primers designed against *EZH2* cDNA to verify that the Flag tag and the gene were inframe (Expasy translator :http://web.expasy.org/translate/). One accurate colony was selected and maxiprep was carried out using Qiagen Maxi Prep plasmid kit (refer 2.14.10).

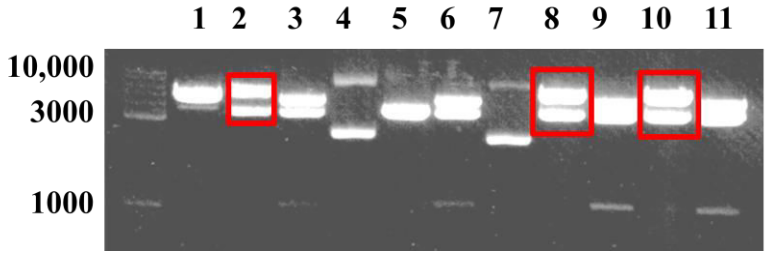


Figure 4.14: Confirmation of accurate ligation of wild type *EZH2* cDNA into p3XFLAG-myc-CMV plasmid. HindIII digest on eleven colonies showed three colonies (red box) with accurate band size. One of the three colonies was selected for maxiprep.

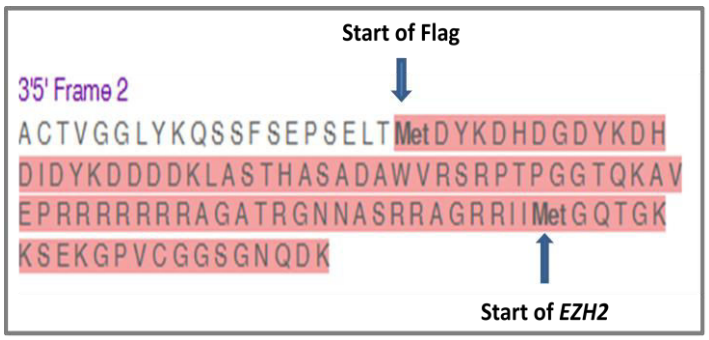


Figure 4.15: Confirmation of the FLAG and *EZH2* sequences by Sanger sequencing. The read out obtained by sanger sequencing was used as an input sequence in Expasy translator online software to confirm that both the Flag and *EZH2* sequences (blue arrows) were inframe (3'5').

4.6.2 Construction of R690C and R690H EZH2 mutants

Cloning strategy for generating the R690C and R690H mutants was based on using the wild type *EZH2* cloned in p3XFLAG-myc-CMV-26 containing GFP and puromycin as a starting point. This plasmid was digested by BsaBI and SfiI restriction enzymes and the fragment at 1600 bp was discarded (Fig: 4.16). p3XFLAG-myc-CMV-26 with neomycin containing the R690C or R690H mutation (generated earlier) was also digested using restriction enzymes BsaBI and SfiI which isolated the fragment containing the mutation (1600 bp) which was gel purified on 1 % TAE agarose gel and ligated into the p3XFLAG-myc-CMV-26 (GFP and puromycin) wild type *EZH2*.

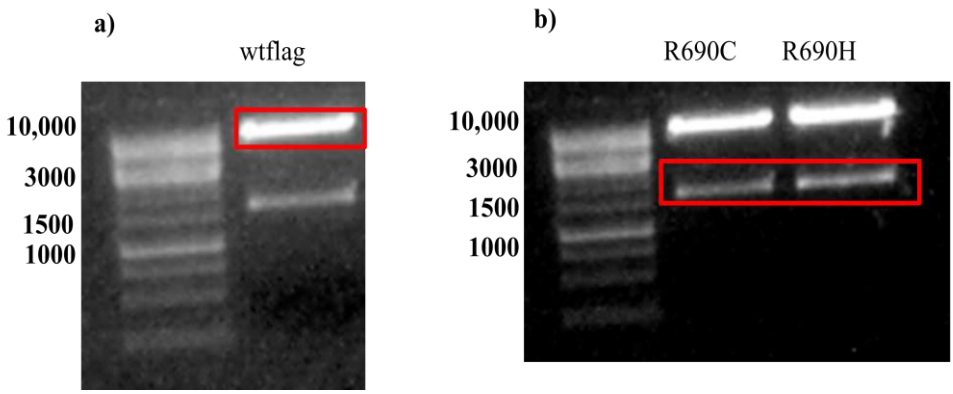


Figure 4.16:Ligation of fragment containing R690C or R690H mutant in to wt Flag p3XFLAG-myc-CMV (GFP+puromycin). (a) Flag tagged wild type *EZH2* cloned into p3XFLAG-myc-CMV was digested by restriction enzymes BsaBI-SfiI to isolate and discard the fragment at 1600 bp. The remaining plasmid (red box) was gel purified and used for ligation (b) The mutants R690C and R690H in p3XFLAG-myc-CMV (neomycin) were digested by BsaBI-SfiI restriction enzymes to isolate and purify the fragment containing the mutation (red box: 1600 bp). The plasmid isolated from wtflag *EZH2* and the fragments containing the mutations were ligated using T4 DNA polymerase (Ladder: Gene ruler 1kb ladder).

Reagents	Volume (for 50 ng)
Fragment R690C & R690H (BsaBI-SfiI): 1600 bp	1µl
wild type *EZH2* p3XFLAG-myc-CMV-26 containing GFP and puromycin (BsaBI-SfiI): 7000 bp	3.5 µl
10X T4 Ligase buffer	1 µl
T4 DNA ligase	0.5 µl
H_20	upto 10µl

Table 4.5: Ligation of EZH2 cDNA (R690C /R690H/ wild type EZH2) into p3XFLAG-myc-CMV-26 expression vector containing GFP+Puromycin. Vector**:** Insert ratio for ligation was 1:2.

2 µl of the ligation mixture was transformed into DH5α-T1 *E.Coli* (refer 2.14.5) chemo-competent cells and plated on LB agar + 100 mg/ml ampicilin plates for 16 hr at 37˚C. Five colonies per mutant were selected and accurate ligation was confirmed by BsaBI-SfiI digest and sanger sequencing using primers designed against *EZH2* cDNA to verify the presence of the mutant and also confirm that the Flag tag and the gene were inframe (Expasy translator: http://web.expasy.org/translate/). Sanger traces for the mutations and results from Expasy to

show that the gene and the flag are in the same frame have not been shown to avoid repetition because they look the same as shown in previous sections. One accurate colony for each mutant was selected and maxiprep was carried out using the Qiagen Maxi Prep plasmid kit (refer 2.14.10).

4.7 Transfection of R690C/R690H/wild type EZH2 in p3XFLAG-myc-CMV-26 (GFP + Puromycin) into myeloid cell lines

p3XFLAG-myc-CMV-26 (GFP+ puromycin) containing R690C/ R690H/ wild type *EZH2* were transfected using lipofectamine into both K562 and MOLM13 cell lines as previously described in Figure 4.10. The percentage of GFP+ cells were analyzed (FACS BD Fortessa) after 48 hr which reflected the transfection efficiency. MOLM13 cells showed poor (0.5- 2 % GFP+ cells) while K562 showed moderate transfection efficiency (15-20 % GFP+ cells) (data not shown). Puromycin 0.2 µg/ml and 2 µg/ml was added after 48 hr of transfection to MOLM13 and K562 cells respectively. 98 % cell death was observed by trypan blue staining in MOLM13 cells during selection which matched the tranfection efficiency calculated by the percentage of GFP+ cells and the remaining 2 % took > 6 weeks to recover. *EZH2* expression in these cells was reduced compared to mock transfected cells, similar to that observed after neomycin selection therefore I discontinued using this cell line for subsequent experiments.

Similar to MOLM13 cells, K562 cells also showed a long duration of recovery (4 weeks) when selected by puromycin and therefore I FACS sorted GFP+ cells (BD Facs Aria) to isolate cells containing the transgene from the untransfected cells. The time point for sorting cells was estimated by analyzing the percentage of GFP+ cells (BD Fortessa) on each day after transfection. There was a sudden drop in GFP+ cells observed on day 4 after transfection (5-7%) and therefore cells were collected on day 3 for sorting (Fig: 4.17).

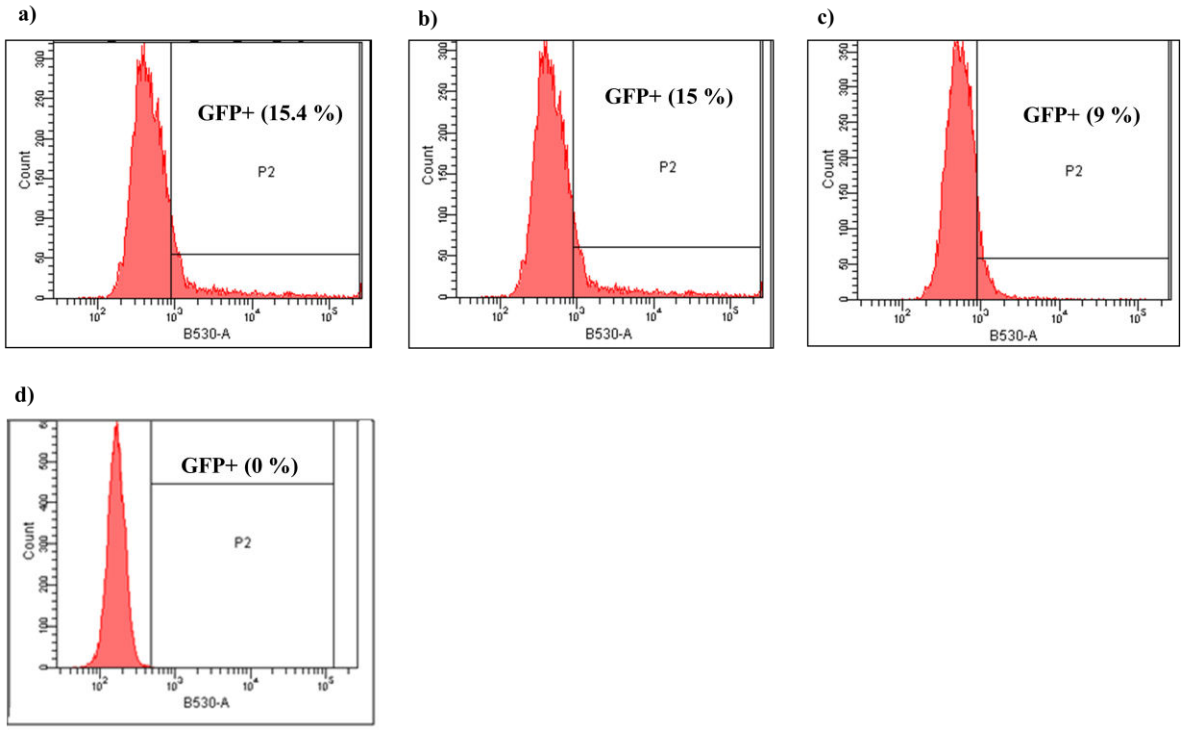


Figure 4.17: FACS BD Fortessa analysis of GFP+ K562 cells transfected with R690C mutant. (a) Illustrates GFP+ cells (15.4 %) 48 hr after transfection while (b) and (c) indicate GFP+ cells (15%) and (9%) on 72 and 96 hr post transfection respectively. (d) Specify mock transfected cells with plasmid DNA (0% GFP). There was reduction GFP+ cells at 96 hr post transfection therefore GFP+ cells were sorted at 72 hr for all experiments.

4.8 Impact of EZH2 mutation on H3K27me3

K562 cells were collected at day 3 post FACs sort for GFP+ cells to evaluate the functional effect of the R690C/ R690H/ over expression of wild type *EZH2* on the protein levels of *EZH2* and on its target H3K27me3 by western blot.

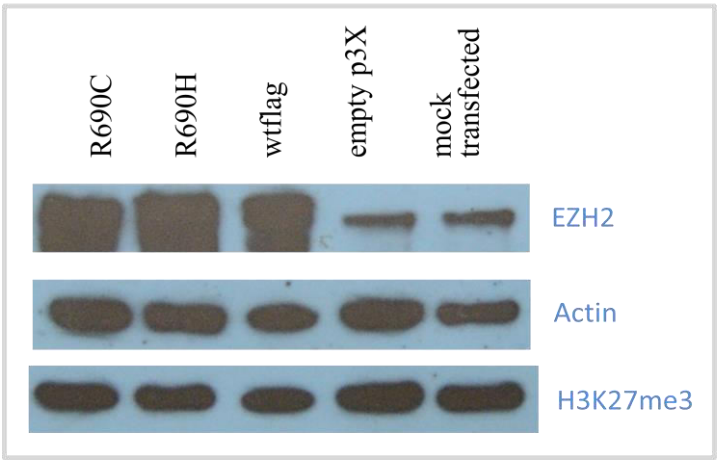


Figure 4.18: Western blot analysis at 72 hr post transfection. Top row illustrates *EZH2* expression in cells transfected with R690C mutant, R690H mutant, Flag-tagged wild type *EZH2* (wtflag), empty p3XFLAG-myc-CMV-26 (GFP+Puromycin) and mock transfected cells. As expected there was over expression of *EZH2* transgene (first three lanes) as compared to the cells transfected with the empty vector and the mock transfected cells. However, H3K27me3 levels remained unaltered in all samples suggesting that these levels were not affected by the transgene expression levels but were maintained by the endogenous *EZH2* present in the cells.

Western blot analysis showed elevated protein levels of mutant *EZH2* (R690C and R690H) as well as Flag tagged wild type *EZH2* compared to empty p3XFLAG-myc-CMV-26 (GFP+ puromycin) and mock transfected cells. *ACTIN* was used as a loading control. Surprisingly, there was no change in H3K27me3 protein levels though elevated levels of both wild type and mutant *EZH2* proteins were evident. Previously two studies (Ernst et al., 2010; Makishima et al., 2010) have shown a decrease in H3K27me3 levels in the presence of *EZH2* mutation (R690C and R690H). Makishima et al, examined and reported decreased H3K27me3 levels in one patient with homozygous R690H *EZH2* mutation [CN(LOH)

7q31.33] while Ernst et al, (2010) conducted immunoprecipitation of *EZH2* on Sf9 insect cells transduced with baculovirus expressing Flag tagged *EZH2* mutant (R690C) and confirmed the presence of all components of PRC2 (*SUZ12,EED* and *EZH2*) but decreased H3K27me3 levels. The system used in this study reflects a heterozygous mutational status of both R690H and R690C *EZH2* mutants since the background endogenous levels of *EZH2* in K562 cells could not be eliminated. Western blot findings in this study suggest that (a) homozygous state of *EZH2* mutants are essential to observe their effects on H3K27me 3 levels or (b) The time point used in this study needs to be revised since the half life of H3K27me3 protein is long and therefore its degradation may not be detected at 72 hr post transfection or (c) The mutant proteins and Flag tagged wild type *EZH2* over expressed in K562 cells were incapable of interacting with H3K27me3 due to modification in the structure of the protein as a consequence of the cloning procedure [p3XFLAG-myc-CMV-26 (GFP+puromycin)] and therefore H3K27me3 levels detected on western blot analysis are only a reflection of the endogenous *EZH2* levels in K562 cells and not related to the transgene expression.

To rule out defects in *EZH2* cDNA (deletion of a part of the gene) cloned in p3XFLAG-myc-CMV-26 (GFP+puromycin) the entire gene was sequenced using primers designed along the whole length of the gene, expression levels of *EZH2* at 72 hr post sort were analyzed by qPCR. The RNA extracted from these cells was used to generate cDNA which was sequenced a second time to validate the presence of the mutation. The sequencing results in both instances showed presence of both the entire gene as well as the mutation in p3XFLAG-myc-CMV-26 (GFP+puromycin). The qPCR results showed 12 to 14 fold increase in mRNA levels of *EZH2* ruling out defects introduced in the gene during cloning.

Due to time constrain, it was not possible to repeat the experiments at variable time points to evaluate long term effects of mutant and wild type *EZH2* over expression on H3K27me3 protein levels.

4.9 Microarray analysis of *EZH2* mutants and over expressed wild type *EZH2* reveals novel *EZH2* targets

Though H3K27me3 levels evaluated on western blot analysis remained unaltered after over expression of both mutant and wild type *EZH2* in K562 cells, I examined the changes in gene signature as a consequence of elevated mRNA levels of *EZH2* to evaluate if mechanisms (other than modulation of H3K27me3) were at work. To examine the effect of *EZH2* mutations (R690C and R690H) and over expression of *EZH2* on transcription, microarray gene expression profiling was carried out on K562 cells sorted for GFP + on day 3. RNA form these experiments was extracted and the presence of mutations was confirmed a second time by sequencing cDNA generated from the RNA. Samples confirmed for *EZH2* mutations were used to probe gene expression microarrays (Affymetrix Human Transcriptome 2.0 arrays). Data analysis identified (refer section 2.10.12), a list of genes either induced or down regulated in cells with R690C or R690H *EZH2* mutants or wild type *EZH2* versus those containing the empty p3XFLAG-myc-CMV-26 (GFP + Puromycin). 69 and 42 genes were down and up regulated (>2 average fold difference, $p < 0.05$, ANOVA) in cells transfected with R690C mutant compared to empty p3XFLAG-myc-CMV-26 vector respectively. Cells transfected with R690H mutant 46 and 35 genes were down and up regulated when compared to empty p3XFLAG-myc-CMV-26 vector while those transfected with Fag tagged wild type *EZH2* showed inhibition of 71 genes and enhanced expression of 34 genes (Table:4.6).

a)

Gene Symbol	Fold Change	p-value (690c vs. empty)
LIPT2	4.83	0.013483
MLLT10	3.62	0.0355
TYRP1	2.29	0.032842
PHF17	-2.03	0.030244
HEPACAM2	-2.03	0.039919
TNFRSF9	-2.04	0.025053
GSR	-2.06	0.042063
EML5	-2.06	0.040441
CCL4	-2.06	0.015373
MAGI2-AS2	-2.07	0.016422
CELF2	-2.08	0.032546
PLA2G4C	-2.13	0.006493
CDH11	-2.14	0.038518
RIMS1	-2.15	0.027053
MIR552	-2.16	0.038518
GALNT13	-2.17	0.037307
MYNN	-2.19	0.019809
MAP1LC3B2	-2.19	0.022536
IL26	-2.22	0.043248
SLITRK1	-2.25	0.014988
ATF3	-2.26	0.009051
MIR432	-2.28	0.007388
CBR1	-2.3	0.002382
OR5AP2	-2.35	0.012135
RNASEH1	-2.38	0.001732
ZCCHC13	-2.51	0.036682
SCLT1	-2.53	0.000826
SLITRK5	-2.68	0.008916
LIG4	-2.93	0.037784
FPGT	-2.99	0.035472
SEMA3A	-3.09	0.000703
MIR515-1	-3.23	0.015623
MIR515-2	-3.23	0.015623
MIR4266	-3.74	0.048431
BEST1	-4.62	0.004488

b)

Gene Symbol	Fold Change	p-value (690h vs. empty)
LIPT2	4.81	0.00022
MLLT10	4.69	0.004707
ACTRT1	2.51	0.020135
MT1A	2.43	0.035517
MT1G	2.33	0.027381
DLGAP2	2.32	0.02309
MUC20	2.1	0.015925
CLLU1	2.04	0.003531
MT1B	2.04	0.039413
MT1F	2.02	0.038223
MIR583	-2.08	0.005463
PRDM9	-2.08	0.003931
MIR552	-2.09	0.042168
SEMA3A	-2.1	0.019244
PLA2G4C	-2.12	0.009438
SCN2A	-2.13	0.03316
LECT1	-2.13	0.001891
GALNT13	-2.14	0.012031
MIR3199-2	-2.2	0.012892
FAM84B	-2.37	0.043775
SNORD115-2	-2.42	0.043292
MIR4768	-2.43	0.038326
ATF3	-2.46	0.011086
MMP1	-2.88	0.048788
SNORD115-30	-4	0.019045
BEST1	-4.36	0.002728
MIR4503	-4.54	0.000012
NKAIN2	-4.61	0.02876

c)

Gene Symbol	Fold Change	p-value (wtflag vs. empty)
ALKBH8	-2.02	0.045186
OPN3	-2.02	0.043913
LYSMD3	-2.03	0.047898
ZNF23	-2.03	0.036535
ZNF429	-2.04	0.042857
LACE1	-2.05	0.031688
KIF3A	-2.06	0.04697
ODF2L	-2.06	0.027571
OR10X1	-2.06	0.006979
EXD2	-2.07	0.043868
CCL4L1	-2.08	0.01042
DHX57	-2.08	0.0371
INTS6	-2.08	0.021098
SEMA3A	-2.08	0.006543
SLC35D1	-2.08	0.033395
SCN2A	-2.09	0.013443
NAT1	-2.11	0.042346
PLCB4	-2.11	0.024082
RNF20	-2.11	0.046718
FBXO21	-2.12	0.015772
EPT1	-2.14	0.049051
GABRG1	-2.14	0.003088
GPHN	-2.14	0.030876
MTHFD1L	-2.14	0.03587
ATP7A	-2.16	0.024671
FPGT-TNNI3K	-2.16	0.023018
ULBP3	-2.16	0.015435
LRRC28	-2.17	0.030789
NEK3	-2.17	0.015698
XBP1	-2.18	0.027967
CSTF3	-2.19	0.037384
MAPK6	-2.2	0.038302
FAM185A	-2.22	0.039883
TMEM26	-2.22	0.045321
ITPR3	-2.23	0.034441
GPR183	-2.25	0.036653

Gene Symbol	Fold Change	p-value (wtflag vs. empty)
MTMR6	-2.26	0.039634
PEX12	-2.26	0.019967
ATF6	-2.27	0.026472
NUBPL	-2.27	0.017355
PTPLB	-2.27	0.008965
NUDT2	-2.3	0.049146
ADAM9	-2.31	0.033381
ASCC3	-2.31	0.041902
SNX13	-2.35	0.041864
MIRLET7C	-2.4	0.017915
WDR19	-2.41	0.041815
GAS2L3	-2.44	0.043646
TNFSF4	-2.46	0.048582
MIR2052	-2.49	0.016142
ELTD1	-2.54	0.045997
LMLN	-2.58	0.048924
ULBP1	-2.59	0.008235
SNORD115-31	-2.61	0.010197
KCTD9	-2.65	0.0498
ATF3	-2.67	0.001905
METTL21B	-2.67	0.041785
TULP3	-2.67	0.049277
MBTPS2	-2.69	0.042893
DDX60L	-2.72	0.04772
RHOH	-2.72	0.019758
MIR379	-2.85	0.029877
TLDC1	-2.86	0.044648
SEPHS2	-2.87	0.049575
PIGN	-2.9	0.029635
TNFRSF9	-3.1	0.015765
MIR515-1	-3.39	0.033914
MIR515-2	-3.39	0.033914
MIR4696	-4.4	0.047037
BEST1	-5.34	0.003953
MIR1-1	-8.16	0.001045

d)

Gene Symbol	Fold Change	p-value (Wtflag vs. empty)
LIPT2	4.13	0.00076
MIR374C	3.49	0.048986
DNAH6	2.98	0.020855
MIR523	2.94	0.006925
MIR513B	2.63	0.040622
OR56A5	2.61	0.010178
CLLU1	2.6	0.002826
TNFAIP3	2.43	0.025021
RHOG	2.36	0.036901
PTH	2.35	0.020579
DISP1	2.34	0.023018
MLLT10	2.23	0.038961
MT1F	2.22	0.047923
NLRP8	2.19	0.034632
FANC	2.18	0.043105
GAS1	2.17	0.017134
WNT16	2.17	0.035242
PML	2.16	0.008881
NAV3	2.15	0.028588
EGFLAM	2.12	0.037886
SEC61B	2.12	0.01111
ACTRT1	2.09	0.005839
CHD9	2.09	0.005786
SLITRK1	2.08	0.034375
SNORD115-15	2.08	0.036645
CCDC140	2.07	0.022275
VTA1	2.07	0.023057
CNTNAP5	2.04	0.02977
CACNA1C	2.03	0.043119
GPD1	2.02	0.048848
PLCXD3	2.02	0.031544
CTPS	2.01	0.010357
SNORD115-10	2.01	0.002024

e)

R690C, R690H & Wtflag	R690C & R690H	R690C & Wtflag	R690H & Wtflag
SEMA3A	LIPT2	TNFRSF9	CLLU1
ATF3	MLLT10	MIR515-1	
BEST1	PLA2G4C	MIR515-2	
LIPT2	MIR552		
MLLT10	GALNT13		
	ATF3		
	SEMA3A		
	BEST1		

Table 4.6: List of genes modulated by over expression of wild type and mutant *EZH2* by micro array gene expression profiling. (a) Genes over or under expressed in cells transfected by R60C versus empty vector, (b) R690H versus empty vector and (c) Flag-tagged wild type *EZH2* and (d) List of common genes between both the mutants and the flag tagged wild type *EZH2* (column 1), in the two mutants (column 2), between R690C and flag tagged wild type *EZH2* (column 3) and (last column) in R690H and flag tagged wild type *EZH2*.

Microarray gene expression profiling showed 105, 111 and 81 genes modulated by Flag tagged wild type *EZH2*, R690C and R690H respectively. Table 4.6 illustrates the filtered gene list based on gene functionality. Non coding RNAs, genes not coding for proteins and pseudogenes have been excluded from this list. Table 4.6 indicate genes common between the two mutants (R690C & R690H), in all three (R690C, R690H and wtflag) and between each mutant and the cells transfected with wtflag. *MLLT10* (Mixed lineage leukaemia translocated to 10) gene expression was enhanced in cells transfected with R690C (3.62 fold), R690H (4.69 fold) as well as wild type *EZH2* (2.23 fold). This gene is located on chromosome 10p12 and encodes a transcription factor. It is recognised as a partner gene in multiple gene translocations resulting in leukaemogenesis (discussed later).

Other genes associated with haematological malignancies which were enhanced or inhibited by over expressing wild type *EZH2* but not affected by mutant *EZH2* were *PML* (promyelocytic leukemia) [up regulated by 2.16 fold] and *FANCL* (Fanconi Anemia, Complementation Group L) [up regulated by 2.18 fold] (discussed later). Amongst the down regulated genes, none of the genes had an association with MDS and none of them were known *EZH2* target genes. Pathways Analysis was carried out using Metacore but the p values were high suggesting that the mutations as well as the wild type over expressed gene did not have a strong effect on a single pathway (data not shown). Such networks are indicative and need experimental validation.

4.10 Discussion

Mutations of *EZH2* were first discovered by Ernst et al (2010) using SNP 6.0 arrays on peripheral blood /bone marrow samples (n=614 MDS/MPN patients) and were associated with UPD7q and 7q36.1 microdeletion. One patient (WHO: aCML, cytogenetics: not available) without del 7q36.1 or UPD 7q carried a heterozygous mutation at site R690 (R to C) (Ernst et al., 2010). These findings were confirmed by Nikoloski et al (2010) wherein *EZH2* mutations were identified in a cohort 102 MDS patients. Both heterozygous and homozygous mutations were reported including a heterozygous mutation (R690C) occurring in the SET domain of *EZH2* in one patient (WHO: RCMD, IPSS: low, cytogenetics: normal karyotype) (Nikoloski et al., 2010).Truncation mutations were identified throughout the gene while missense mutations were commonly found in the conserved SET domain (detailed in section 1.7) responsible for the methyltransferase activity of *EZH2*. Homozygous mutations affecting the R690 site have also been identified in patients (WHO: RCMD, CMML) with UPD 7q31.33 where there is conversion of arginine to histidine (R to H) (Makishima et al., 2010). However the frequency of this mutation was not high (n=1 or 2) enough to qualify it as a hot spot of *EZH2* mutation. But since it was reported in all cohorts of patients analyzed so far for *EZH2* mutations and has a strategic location in the SET domain of the enzyme, these two mutations (R690C and R690H) were selected to understand their role in myeloid malignancies.

These mutations had been previously examined in insect cells Sf9 transduced with baculovirus expressing Flag tagged *EZH2* mutant (R690C) and also in patient samples carrying homozygous R690H mutation (Ernst et al., 2010; Makishima et al., 2010). My study is the first attempt reported so far to construct these mutations in p3XFLAG-myc-CMV-26 (GFP+ puromycin) and transfect in to human myeloid cell lines.

4.10.1 Cloning R690C/R690H and Flag tagged wild type *EZH2* into p3XFLAG-myc-CMV-26 (neomycin)

R690C and R690H mutations were generated by using a PCR based cloning method involving two primer pairs (one flanking primer pair and second pair overlapping the mutation). Both mutations and Flag tagged wild type *EZH2* were cloned into p3XFLAG-myc-CMV-26 vector containing the neomycin selectable marker and transfected into myeloid cell lines. qPCR analysis of five myeloid cell lines (U927, NB4, K562, MOLM13 and KG1) indicated high expression of *EZH2* in K562 (chronic myelogenous leukemia) and MOLM13 (acute monocytic leukemia) cells. In addition to *EZH2, TP53* expression levels were also assessed in order to consider for a appropriate cell line selection since expression of both *EZH2* and *TP53* are closely linked; high levels of *EZH2* are linked to presence of *TP53* mutations in breast cancer (Pietersen et al., 2008) while in oral squamous carcinoma aberrant *EZH2* expression was associated with *TP53* alteration (Shiogama et al., 2013). *TP53* expression is deregulated in K562 (insertion in exon 5 of *TP53)*(Law et al., 1993) cells but MOLM13 cells carry wild type *TP53* (Kojima et al., 2005) therefore K562 (*TP53* null) and MOLM13 (wild type *TP53*) cell lines were used for the transfection experiments and selected using neomycin. The expression level of *EZH2* in cells selected by neomycin on day 21 was variable ranging between 0.56 to 17.5 fold compared to the mock transfected cells in both cell lines. The predicted cause for reduced *EZH2* expression over time might be due to its promoter inactivity. Cytomegalovirus (CMV) promoter drives the expression of *EZH2* in p3XFLAG-myc-CMV-26. The CMV promoter contains at least four types of repetitive sequence elements; 17, 18, 19, and 21 bp which form complexes with nuclear proteins (Ghazal et al., 1987), transcription factors therefore enhancing the transcription of the transgene, in this case *EZH2* (Loser et al., 1998).The 18 and 19 bp repeats contain consensus binding sites for NFκB and CREB/ATF, respectively, and enhance CMV promoter activity

(Sambucetti et al., 1989). While the 21bp repeat binds to negative regulators like YY1 and is suggested to repress CMV promoter-dependent transcription (Sinclair et al., 1992). In addition to different binding partners and their effect on the CMV promoter activity, studies have shown that the promoter undergoes methylation of CpG islands after few weeks of transfection making it inactive (Brooks et al., 2004; Herbst et al., 2012). The effect of methylation on the CMV promoter is independent of the vector system, the method of tranfection and might attract inhibitory proteins that bind to the methylated CpG or those that may interfere directly with binding of transcription factors. Although the CMV promoter allowed for a very strong short-term expression (48 hr after transfection) of *EZH2*, it declined within 3-4 weeks and in some samples was barely detectable at the end of neomycin selection (day 21-24). To test whether CpG methylation of the CMV promoter was responsible for decreased *EZH2* expression, neomycin selected cells were treated with 5-Azacytidine (500 nM, 1 µM, 5 µM). 5-Azacytidine is a DNA methyltransferase inhibitor that is routinely used in the treatment of Intermediate or high risk subtypes of MDS. Its mechanisms of action depends on conversion of Azacitidine to 5-aza-2′-deoxycytidine-triphosphate which acts as a substrate and is incorporated into host DNA as azacytosine instead of cytosine. These residues are recognised by DNA methyltransferases as natural substrates and therefore bind to it initiating the process of DNA methylation. The formation of this covalent bond blocks the DNA methyltransferase at that site resulting in loss of methylation marks during DNA replication in cells treated with 5-Azacytidine (Stresemann and Lyko, 2008). In my experiments, *EZH2* expression recovered (1.4-5.8 fold) in all cells (R690H) and (Flag tagged wild type *EZH2*) at both 48 and 72 hr post 5-Azacytidine treatment. This time point was chosen to allow cells to undergo cell cycle and for the 5-Azacytidine to incorporate into the DNA. Recovery of *EZH2* expression suggests that reduced expression of the transgene was due to methylation of its CMV promoter. To avoid long duration of selection, *EZH2* mutants

and Flag-tagged *EZH2* were cloned into p3XFLAG-myc-CMV-26 containing GFP + puromycin markers and transfected into MOLM13 and K562 cell lines. Due to poor transfection efficiency of MOLM13 cells (0.5-2% GFP + cells) even after extensive optimization using a range of plasmid DNA concentrations, both lipofectamine, electroporation methods and verifying GFP content at different time points, this cell line was discontinued.

K562 cell line transfected with the mutant (R690C and R690H) and Flag tagged wild type *EZH2* were sorted at 72 hr post transfection for GFP+. Western blot analysis of these cells showed elevated protein levels of mutant *EZH2* (R690C and R690H) as well as Flag tagged wild type *EZH2* compared to empty p3XFLAG-myc-CMV-26 (GFP+ puromycin) and mock transfected cells. Surprisingly, there was no change in H3K27me3 protein levels with increase in *EZH2* expression. Two reports have (Ernst et al., 2010; Makishima et al., 2010) carried out studies on *EZH2* mutants (R690C and R690H) wherein they have shown decrease in H3K27me3 levels with presence of the mutation. However, Makishima et al (2010) examined the effect of homozygous R690H mutation on the levels of H3K27me3 in a single patient by western blot analysis while in my thesis, a heterozygous expression of *EZH2* R690H mutant was achieved since the endogenous *EZH2* present in the cells could not be eliminated (Makishima et al., 2010). This suggests that the heterozygous mutation does not have a dominant negative effect on the H3K27me3 levels in the transfected cells and that these levels are maintained by the endogenous wild type *EZH2*. The second study focused on R690C mutant infected in Sf9 insect cells wherein; expression of *EZH2* protein remained unaltered but its catalytic activity was reduced (H3K27me3). This system mimics the experimental design presented in my work wherein the expression of the mutant is heterozygous rather than homozygous. However, it is important to note that the model system used in both studies is different i.e insect cells versus human myeloid cells (K562). To date, it

is not known whether there are any species related changes in modulation of H3K27me3 levels by *EZH2* and more work needs to be done in this area.

In addition to the possibility of species related differences or the effect of homozygous versus heterozygous *EZH2* expression, the half life of H3K27me3 protein levels needs to be considered and the time point of sample collection (72 hr post transfection) needs to be revised. This time point was chosen based on the maximum number of GFP positive cells that could be sorted after transfection. There was a decline in GFP+ cells from 12-15 % on day 3 to 9 % on day 4 post transfection. However, the reduction in GFP+ cells at day 4 might be due to the absence of puromycin selection pressure on the transfected cells. One way to circumvent this problem would be continue selection of cells by puromycin after sorting at 72 hr. This may provide more time to examine the effect on H3K27me3 protein levels. However, due to insufficient time, it was not possible to repeat the experiments at variable time points using the above strategy to evaluate long term effects of mutant and Flag tagged wild type *EZH2* over expression on H3K27me3 protein levels.

Though H3K27me3 levels evaluated on western blot analysis remained unaltered after over expression of both mutant and wild type *EZH2* in K562 cells was at work. Four genes implicated previously in leukaemogenesis were identified with enhanced or reduced expression levels. *MLLT10* (Mixed lineage leukaemia translocated to 10) gene expression was enhanced in cells transfected with R690C (3.62 fold), R690H (4.69 fold) as well as wild type *EZH2* (2.23 fold). The *MLLT10* gene is located on chromosome 10p12 and is implicated in the pathogenesis of AML (adult as well as paediatric) [reviewed by (Marschalek, 2011; Meyer et al., 2013). The protein produced by this fusion gene is predicted to interact with and enhance H3K79me2/3 marks via *DOT1L* histone methyltransferase. Recent reports on T-cell lymphoblastic leukaemia have established that the fusion proteins of *MLLT10* assist in up regulating the expression of the *HOXA* genes (Brandimarte et al., 2013).In this study,

expression of *MLLT10* gene is enhanced by over expression of both mutant (R690C/R690H) and Flag tagged wild type *EZH2* suggesting that the mutations do not affect the function of the wild type *EZH2* in regulating this gene. *HOXA* genes are known targets of *EZH2* and are up regulated in *EZH2* mutated leukemic cell line like SKM-1 (Khan et al., 2013). There is an inverse relation of *EZH2* and *HOXA* expression levels (especially *HOXA9* and *HOXA10*). The association between *EZH2-MLLT10-HOXA* genes need further investigation.

Unlike *MLLT10, PML* (promyelocytic leukaemia) [up regulated by 2.16 fold] and *FANCL* (Fanconi Anaemia, Complementation Group L) [up regulated by 2.18 fold] genes were only over expressed in Flag tagged wild type *EZH2* cells. *FANCL* forms a part of the core Fanconi anaemia complex and posses a strong E3 ligase activity essential for ubquitination of *FANCD2*. Monoubiquitination of *FANCD2* is induced by DNA damage and is predicted in transcriptional activation of tumour suppressor genes (Park et al., 2013) and activation of the Fanconi anaemia pathway (Huang et al., 2014). On the other hand the promyelocytic leukaemic (*PML*) gene is causally linked to acute promyelocytic leukaemia caused due to fusion of the N-terminus of *PML* to the C terminus of the retinoic acid receptor-a (*RARa*) transcription factor resulting in a balanced t (15; 17) translocation.The disease is defined by (a) morphological features (hypergranular cells with low white blood cell count (FAB M3)], (b) coagulopathy associated with hypofibrinogenaemia and (c) presence of *PML-RARα* (Degos and Wang, 2001).*RARa* is a nuclear receptor which in the absence of its ligand retinoic acid (*RA*) reduces the transcription of target genes by recruiting histone deacetylases and co-repressor molecules. In the presence of the fusion gene (*PML-RARα*) there is ineffectual effect on transcription of the downstream target genes (Dos Santos et al., 2013). Studies in granulocytes have shown that *RARα* unbound to retinoic acid negatively regulates granulocyte differentiation while *RARα* bound to *RA* promotes granulocyte differentiation (Mistry et al., 2003). Both *PML-RARα* and *FANCL* are not directly implicated in the

pathogenesis of MDS/AML but it would be important to verify the findings of the micro array gene expression profiling on patient samples with R690C or R690H mutation to evaluate the effect of the mutations on the expression of the two genes.

4.11 Conclusion

Over expression of wild type *EZH2* and its mutants (R690C & R690H) did not affect the target H3K27me3 levels in K562 cells. It was predicted to be due to heterozygous expression of the transgene suggesting that the heterozygous mutation does not have a dominant negative effect on the H3K27me3 levels. The second explanation could be that the half life of H3K27me3 protein is long therefore 72 hr post transfection was not an ideal time point to observe any alterations in its levels.

Gene expression profiling identified *MLLT10* gene to be up regulated by both mutant and wild type *EZH2*. This finding could provide an insight into *EZH2* dependant regulation of *HOXA* genes via modulation of *MLLT10* gene expression. *PML* and *FANCL* genes implicated in acute promyelocytic leukaemia and fanconi anaemia were up regulated in cells over expressing the wild type but not the mutant *EZH2* suggesting that the mutation might alter the capacity of the wild type *EZH2* to modulate expression of some its target genes.

5 Chapter: Role of EZH2 inhibition in Myeloid malignancies

5.1 Introduction

5.2 Knock out (KO) of EZH2 in cancer

The role of *EZH2* varies in context of different types of tissues for e.g. in ovarian cancer increased expression levels of *EZH2* correlates with more aggressive histological grade of the tumour. In fact the association is robust enough to qualify *EZH2* as an independent prognosticator for predicted shortened survival (Rao et al., 2010). *EZH2* over expression also predicts poor survival in patients with prostate (Bryant et al., 2007; Varambally et al., 2002), breast (Mu et al., 2013), renal (Zhang et al., 2011) and colon cancer with strong correlation to tumour metastasis(Yu et al., 2012).

Knock out (KO) of *EZH2* using siRNA (small interfering RNA) in ovarian cancer demonstrated G1 phase cell cycle arrest, reduced cell proliferation and cell migration (Rao et al., 2010).p21 was established as an important protein in G1 to S phase check point inducing cell cycle arrest. *EZH2*-p21 association has been validated by ChIP studies.Western blot analysis after KO of over expressed *EZH2* demonstrate recovery of typically down regulated p21 expression in ovarian cancer (Seward et al., 2013). In prostate cancer, *EZH2* KO using siRNA in cancer stem cells inhibited cell cycle and progression to apoptosis (Li et al., 2013) while in breast cancer, *EZH2* KO had no impact on apoptosis but decreased proliferation with delayed G2/M cell-cycle transition and up regulation of *BRCA1* (Gonzalez et al., 2009). Other studies on prostate cancer indicate similar results with reduced cell invasion (Mu et al., 2013) (seen on Matrigel) and proliferation (observed on MTT assay) (Bryant et al., 2007). In renal cancer, cell proliferation and invasion caused by down regulation of *EZH2* is due to the regulation of the WNT/β-catenin signalling pathway (Yuan et al., 2012). Interestingly, similar impact on proliferation and cell invasion was observed in endometrial cancer wherein down regulation of *EZH2* was concomitant with reduced expression levels of β-catenin (Eskander

et al., 2013) and *WNT3A* (Yuan et al., 2012). The impact of *EZH2* on cell proliferation can be utilised for therapeutic effect as demonstrated by Chen et al (2007) who showed that intratumor injection of lentiviral shRNA (short hairpin RNA) or siRNA targeting *EZH2* resulted in tumour regression of hepatocellular carcinoma (Chen et al., 2007). In subsequent studies they also showed that *EZH2* inhibition reduced pancreatic tumour growth and liver metastasis (Chen et al., 2010).

EZH2 has a specific role in haematological malignancies. Missense mutations (Y646) of *EZH2* observed in lymphoma are associated with up regulation of gene expression and siRNA mediated KO results in G1/S phase cell cycle arrest (Velichutina et al., 2010). Results on natural killer/T-cell lymphoma (NKTL) implicate *EZH2* as an oncogene with reduction in its levels closely linked to cancer cell growth retardation (Yan et al., 2013). While *EZH2* acts as an oncogene in lymphoma, in myeloid malignancies it has a tumour suppressor role. Cross et al (2010) and Makishima et al (2011) showed that both the R690C & R690H mutations seen in MDS/AML patients reduced the protein levels of H3K27me3 on western blot (Cross et al, 2010; Makishima et al, 2011). However, further analysis of *EZH2* down regulation on the downstream targets has not been studied till date.

5.3 RNA interference

The discovery of RNAi revolutionised the approach to gene silencing and our understanding of post transcriptional gene regulation. RNAi technology has improved tremendously since its advent and now provides a simple and effective way to reduce the expression levels of the gene of interest (Fig: 5.1) (Bantounas et al., 2004; Leung et al, 2005).

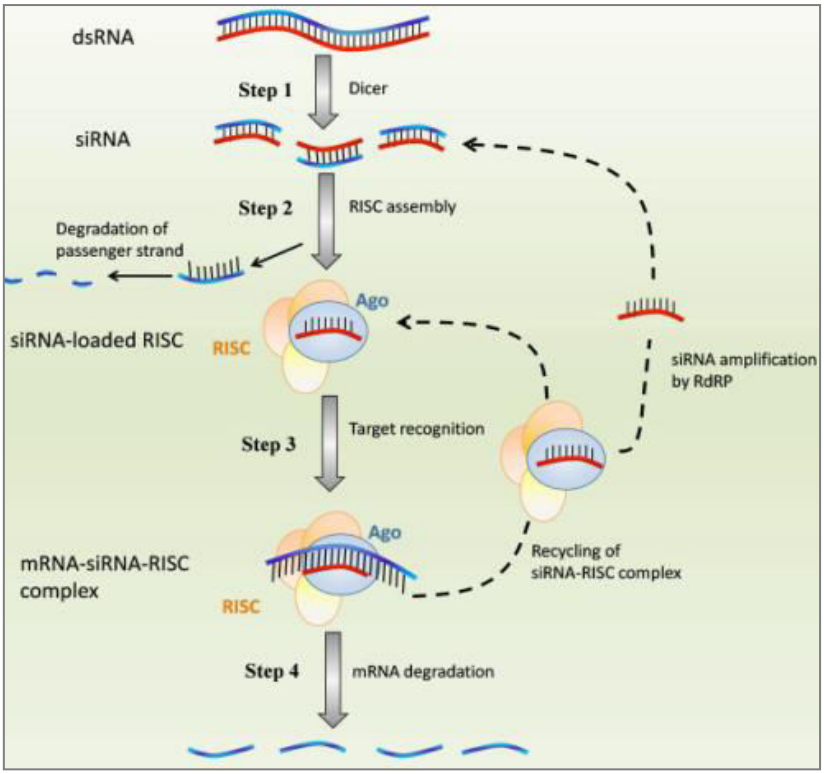


Figure 5.1: Schematic representation of RNA interference in cells. Step 1: double stranded RNA (dsRNA) is cleaved by enzyme Dicer into fragments of 21-23 base pairs (siRNA). Step 2: siRNAs combines with the multiprotein complex: RNA Induced Silencing Complex (RISC) and one strand (the passenger strand) is degraded, while the guide strand remains attached to RISC and serves as a template for the silencing reaction. Step 3: the guide strand assembled into the siRNA-RISC complex targets the complementary mRNA. Step 4: The target mRNA is degraded and is dissociated from the siRNA. The siRNA-RISC complex is released to process further mRNA targets (Cuccato et al., 2011).

A major problem in RNAi experiments is the widespread off target effects despite sequence specific targeting (Jackson et al., 2003; Lin et al., 2005; Sledz et al., 2003), (Birmingham et al., 2006; Jackson et al., 2006). Off target effects can be sequence related and specific to individual siRNA/shRNA or sequence independent and caused by siRNA/shRNA in general. Introducing double stranded RNA; dsRNA (one step prior to siRNA) into the cells can mimic a viral threat and trigger an interferon response (IFN) with widespread consequences on cellular processes such as cessation of protein synthesis (Bantounas et al., 2004). However the use of synthetic siRNA has helped circumvent IFN to a large extent. However some studies suggest that the vectors carrying shRNA can have effects that are similar to the

shRNA (Jackson and Linsley, 2004). This response is not always activated and avoiding high concentrations of siRNA/shRNA is an effective control against it (Jackson and Linsley, 2004; Sledz et al., 2003).

Off target effects of RNAi are sequence dependant and caused by non-specific binding to partially complementary mRNA transcripts. Initially it was believed that siRNA/shRNA contained a 7nt seed region (between base positions 2-8) at its 5'end which could bind and silence only those mRNAs which had the exact complementary sequence to this region. But many studies show that this is may not be the case (Birmingham et al., 2006; Lin et al., 2005), (Jackson et al., 2006). Imperfect matches between the siRNA/shRNA and the mRNA transcripts have been observed resulting in silencing of undesired genes. These off target effects do not always result from transcript cleavage but may be caused by inhibition of translation like miRNAs (Zeng et al., 2003). Sequence homology to a 7nt siRNA seed region elsewhere in the genome is unavoidable making sequence specific off target effects almost inevitable and difficult to predict. Despite the high incidence of off target effects, RNAi technology is still extremely informative as long as the limitations are considered and carefully controlled for. In the knockdown experiments carried out in this study, possible non-specific effects as a result of the procedure or the addition of siRNA/shRNA are controlled for by using mock transfections and negative controls.

5.4 Aim

- Achieve KO of *EZH2* expression in myeloid cells using siRNA or shRNA and identify the downstream targets of *EZH2* using microarray gene expression profiling.
- Examine the effect of *EZH2* KO on cell viability, proliferation, cell cycle kinetics and compare it with observations made in other cancers.

- Validate the gene signature obtained in myeloid cell lines in patient samples. Evaluate whether R690H/C mutation of *EZH2* commonly seen in patients with MDS mimic the KO of *EZH2*.

5.5 Transient knockdown of EZH2

Two pre-designed siRNAs were obtained from Dharmacon (ON-Target Plus SMARTPool) and Santacruz targeting *EZH2* mRNA transcripts. A negative control siRNA (scrambled) was utilised which did not have any known targets in the transfected cells (Refer Appendix). Mock transfected cells were used to control for variations caused as a result of the procedure. MOLM13 and K562 cell lines were used for all experiments for the reasons mentioned in the previous chapter.

5.5.1 Silencing Efficiency of siRNA (Dharmacon) in MOLM13 & K562 cells using RNAiMAX

Cell density and confluence at the time of transfection is important for transfection efficiency and cell viability during the experiment. Cell density of 0.25x 10^6 cells/well was found to achieve suitable confluence for transfection with good cell viability. siRNA concentrations (60 nM, 300 nM and 600 nM) were tested for a maximal silencing effect by RNAiMAX (refer section 2.15.3 for protocol). Samples were collected at 24, 48 (Fig: 5.2), 72 hr and examined for changes in *EZH2* mRNA levels using qPCR.

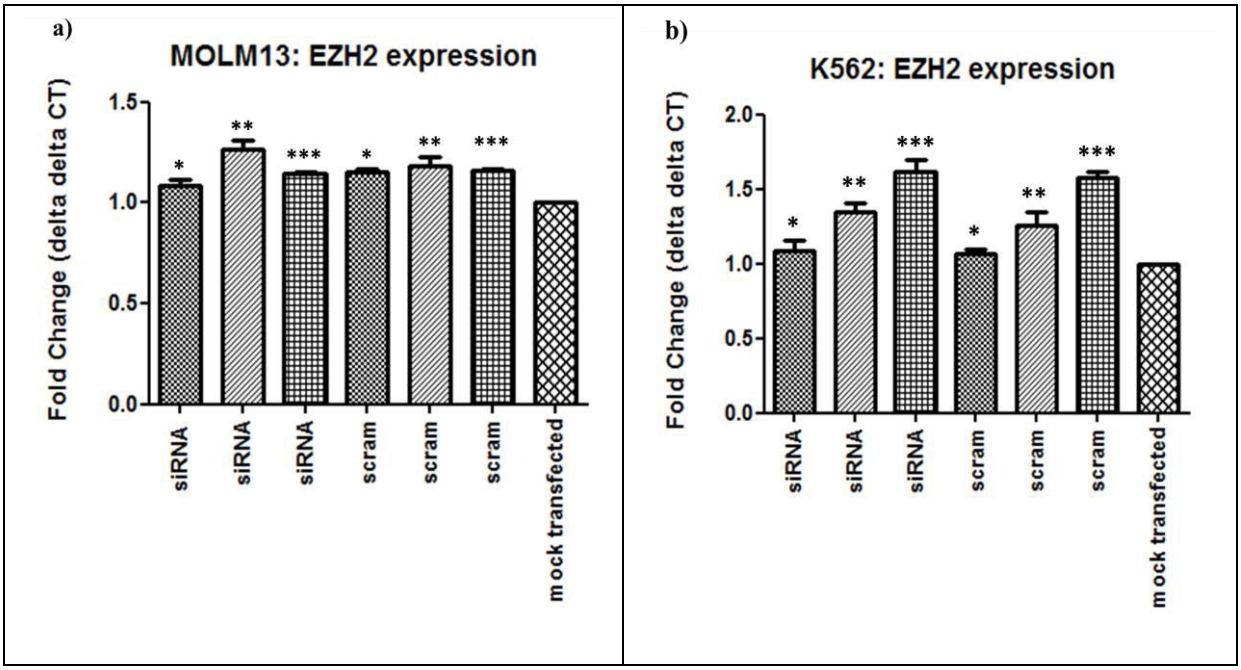


Figure 5.2: *EZH2* expression levels at 48 hr by qPCR. *EZH2* expression levels in MOLM13 (a) and K562 (b) cells calculated relative to the mock transfected cells (Y axis: ΔΔCT) at 48 hr post transfection by siRNA. *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls and the error bars mark the standard error between three independent experiments. The mRNA levels of *EZH2* were unaltered post siRNA transfection. Note: '*'= 60 nM, '**'= 300 nM and '***'= 600 nM concentration of the control and siRNA against *EZH2*.

Down regulation in *EZH2* expression was not observed in both MOLM13 and K562 cells at 24, 48 (Fig: 5.2) or 72 hr indicating either a poor transfection efficiency by RNAiMAX protocol or an ineffective siRNA. Data for the 24 hr and 72 hr time point is not shown because none of the samples showed down regulation of *EZH2* expression.

To improve the transfection efficiency, electroporation was used for subsequent experiments.

5.5.2 Silencing Efficiency of siRNA (Dharmacon) in MOLM13 & K562 cells using electroporation

Electroporation results in massive cell death and therefore cell seeding density was increased to 2 x 10^6 cells/well to allow for sufficient live cells for post transfection assays. siRNA concentration of 150 nM was tested in both K562 and MOLM13 cells (refer section 2.15.4

for protocol) and samples were collected at 24, 48 and 72 hr followed by assessment of *EZH2* silencing by qPCR.

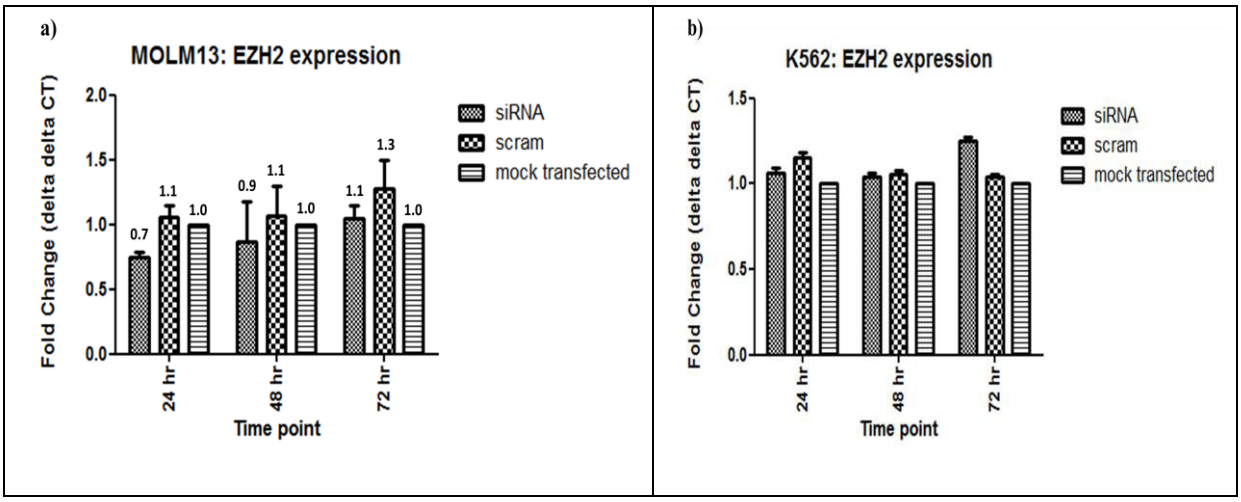


Figure 5.3: *EZH2* expression levels by qPCR using 150 nM siRNA (Dharmacon) concentration. *EZH2* expression in MOLM13 (a) and K562 (b) cells was calculated relative to the mock transfected cells (Y axis: ΔΔCT) at 24, 48 and 72 hr post transfection by siRNA. *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between three independent experiments. mRNA levels of *EZH2* remained unchanged in K562 cells while in MOLM13 35% *EZH2* down regulation was observed in siRNA transfected cells compared to cells transfected with the scrambled (24 hr) with gradual recovery of *EZH2* levels (10 %) at 48 hr.

35 % reduction in *EZH2* mRNA level was observed in MOLM13 cells (24 hr) by electroporation confirming no error in siRNA functionality and transfection efficiency dependent improvement in *EZH2* KO. Inhibition of *EZH2* was maximal at 24 hr compared to 48 and 72 hr. However, further reduction in *EZH2* levels was required to correlate it to KO levels caused by *EZH2* mutations which can be both heterozygous and homozygous.

In K562 cells, *EZH2* levels remained unchanged at all time points indicating that the siRNA was ineffective in these cells at 150 nM concentration. To improve the level of *EZH2* KO in both the cell lines, the siRNA concentration was increased to 200 nM while other conditions were kept constant. *EZH2* levels were analyzed by qPCR at 24, 48 and 72 hr (Fig 5.4).

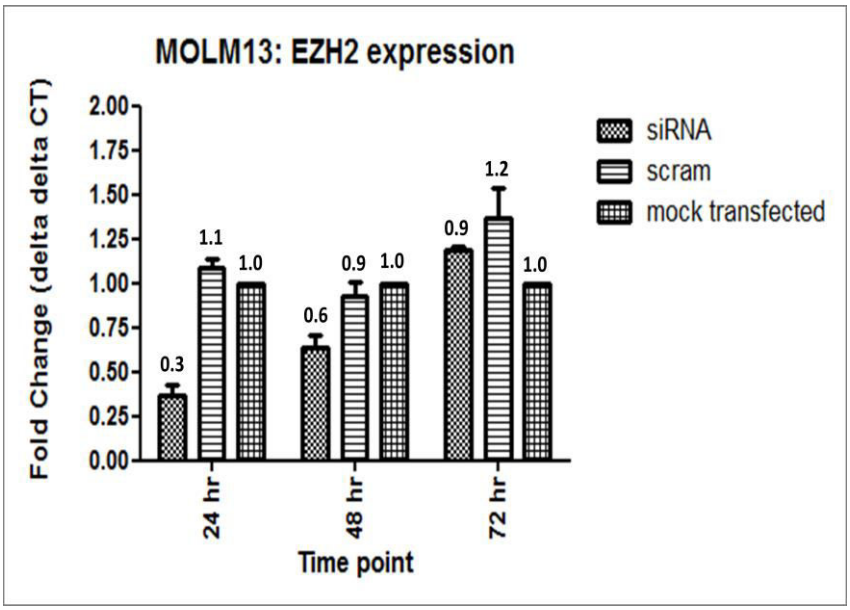


Figure 5.4: Measurement of *EZH2* expression levels by qPCR using 200 nM concentration of siRNA (Dharmacon).*EZH2* expression in MOLM13 cells was calculated relative to the mock transfected cells (Y axis:ΔΔCT) at 24, 48 and 72 hr post transfection by siRNA. *GAPDH*, *B2M* and *TUBULIN* (data shown) were used as endogenous controls. Error bars mark the standard error between three independent experiments. Maximal KO was observed at 24 hr (65-70 %) in cells transfected with siRNA versus those transfected by scrambled siRNA with gradual recovery of *EZH2* levels at the subsequent time points.

Increasing the siRNA concentration to 200 nM increased the level of *EZH2* KO (65-70 %) at 24 hr with gradual recovery of *EZH2* levels (35 %) at 48 hr. In contrast, *EZH2* levels remained unaltered in K562 cells (data not shown) even at 200 nM concentration and therefore I changed to shRNA as an alternative method of gene Knock out.

5.6 Impact of EZH2 KO on H3K27me3

MOLM13 cells were collected at 24 hr and 48 hr post siRNA transfection (200 nM) for evaluating the effect of the KO on the protein levels of *EZH2* and on its target H3K27me3 by western blot.

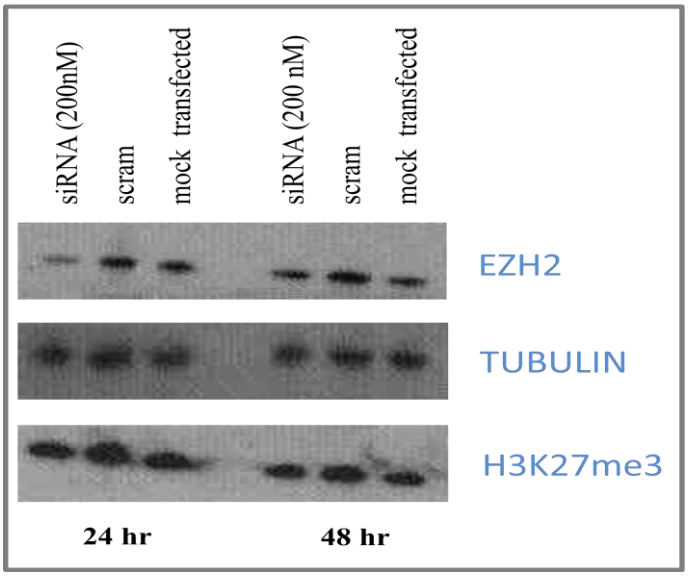


Figure 5.5: Western blot analysis of siRNA mediated KO of *EZH2* in MOLM13 cells at 24 and 48 hr.65 % inhibition of *EZH2* levels was observed at 24 hr compared to the scrambled and mock transfected cells with recovery of *EZH2* levels at 48 hr. However, there was no concomitant decrease in H3K27me3 levels.Tubulin was used as a loading control.

Western blot results confirmed the findings of qPCR wherein the maximal KO (approx 65 %) of *EZH2* expression was observed at 24 hr post transfection with gradual recovery in the *EZH2* levels at 48 hr (KO 40 %) indicating that 24 hr is the optimal time point to carry out subsequent experiments. Interestingly, the H3K27me3 levels remained unaltered at both 24 and 48 hr suggesting that a stronger KO of *EZH2* expression is required to perceive modifications in the target expression or that the half life of H3K27me3 protein is long and therefore protein degradation is not observed at 24 or 48 hr.

5.7 Cell Cycle analysis

EZH2 is closely linked to cell proliferation and proteins associated with cell cycle (refer section 5.2) therefore the kinetics of cell cycle entry after *EZH2* KO was examined by flow cytometric analysis (PI/FITC) (section 2.19) and the percentage of cells in each cell cycle phase as well as the apoptotic cells (sub G0) were studied on samples collected at 24 hr in MOLM13 cells (Fig: 5.6).

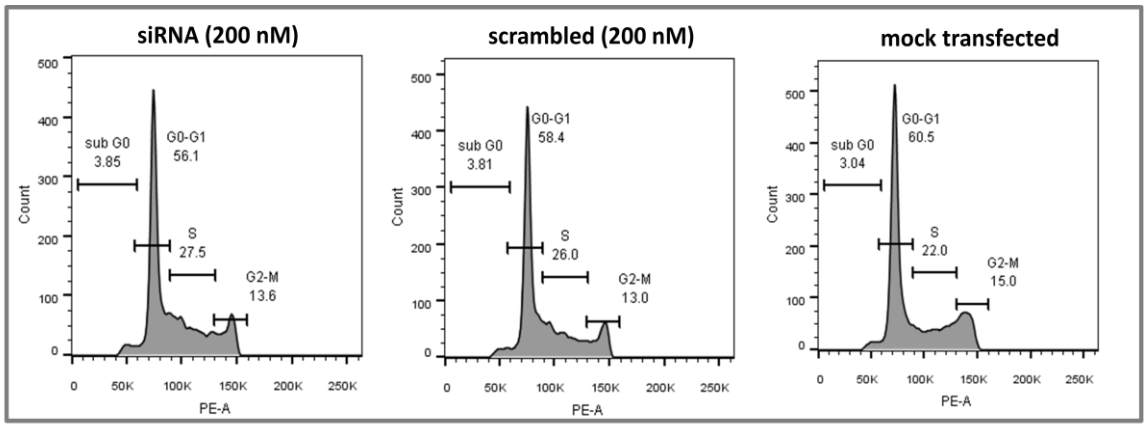


Figure 5.6: Cell cycle profile of MOLM13 at 24 hr post transfection. Samples were fixed (70% ethanol v/v) and stained with PI (DNA content) and FITC (protein content).The percentage of cells in sub G0, G0/G1, S and G2/M phases of the cell cycle remained unaltered in the cells transfected with the siRNA versus those transfected with the scrambled or the mock transfected cells. Data represents three independent experiments.

There were no variations in the percentage of cells in the four phases of the cell cycle (sub G0, G1/M, S and G2/M) between cells transfected with the siRNA and those transfected with the scrambled or the mock transfected cells. This suggests that either a stronger KO of *EZH2* expression is required in order to detect changes in the cell cycle kinetics or else *EZH2* expression levels did not affect cell cycle proteins in MOLM13 cells. To test this hypothesis, higher concentrations of the siRNA were employed to accomplish a stronger KO of *EZH2* and its target; H3K27me3 in MOLM13 cells. Increasing concentrations of siRNA did not lead to further down regulation of *EZH2* or H3K27me3 but increased the risk of off target

effects caused by non-specific binding of the siRNA to partially complementary mRNA transcripts. Therefore observations noted at 200 nM concentration of the siRNA (Dharmacon) were considered as final results and an additional siRNA or shRNA mediated KO *EZH2* was utilised to confirm the findings of the first siRNA (Dharmacon).

5.8 Silencing Efficiency of siRNA (Santacruz) in MOLM13 & K562 cells using electroporation

As previously described, MOLM13 cells were seeded at a density of 2x 10^6 cells/well and transfected by electroporation (refer section 2.15.4) with siRNA from Santacruz Biotechnology at varying concentrations (50-200 nM). Samples were collected at 24 (Fig: 5.7), 48 and 72 hr followed by assessment of *EZH2* silencing by qPCR. There was no noticeable KO of *EZH2* at 48 and 72 hr and therefore the data has not been shown.

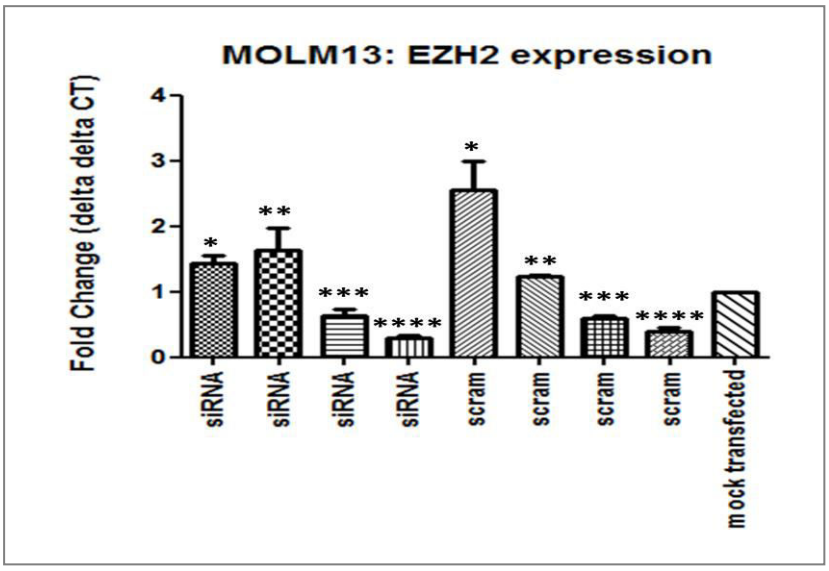


Figure 5.7: *EZH2* expression levels at 24 hr after transfection with siRNA from Santacruz biotechnology by qPCR. *EZH2* expression in MOLM13 cells was calculated relative to the mock transfected cells (Y axis: ΔΔCT). *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between two independent experiments. Note: '*'= 50 nM, '**'= 100 nM and '***'= 150 nM and '****'= 200 nM concentration of the control and siRNA against *EZH2*.The expression of *EZH2* varied markedly with increasing concentrations of the scrambled. At 150 and 200 nM of scrambled siRNA, >60% KO of *EZH2* was noted which was unfavourable.

At 24 hr, increasing concentrations of the siRNA (50 to 200 nM) correlated with down regulation of the *EZH2* levels. However, this effect was also observed in cells transfected with the scrambled (-ve control) which showed > 60 % KO at 150 and 200 nM concentrations. This unfavourable effect of the scrambled siRNA was absent at low concentrations (50 & 100 nM). However, lowering the concentration (50 or 100 nM) of the siRNA against *EZH2* to match that of the scrambled resulted in absence *EZH2* KO. Therefore the experimental strategy was changed to using shRNA mediated KO of *EZH2*.

5.9 Stable knockdown of EZH2 using shRNA in pSUPER

To achieve long term down regulation of *EZH2* to assess the effect on H3K27me3 protein levels, a stable system of gene suppression was required. This was accomplished by adapting the method of RNAi from siRNAs to the more sustainable short hairpin RNA (shRNA) mediated approach. shRNA is expressed in the cell from a DNA construct encoding the sense and antisense strands of the siRNA separated by a short loop sequence. The RNA transcripts fold at the loop and self anneal to form dsRNA stem loop structure: a short hairpin RNA. The shRNA is processed like a miRNA by the RNAi machinery to mediate target silencing. Conventionally RNA polymerase III promoters U6 and H1 are used for shRNA expression, generating transcripts lacking a long poly (A) tail that may hinder subsequent RNAi processing machinery. In this study four shRNAs against *EZH2* were designed and cloned in to the pSUPER vector (refer section 2.16.1) which is driven by H1 promoter (Amarzguioui et al., 2005; Brummelkamp et al., 2002; Gou et al., 2003). The plasmid was transfected into MOLM13 and K562 cells using both eletroporation and lipofectamine methods. The experimental work flow is detailed in Fig: 5.8. Post puromycin selection, RNA was extracted, cDNA was synthesised and qPCR was performed on the samples.

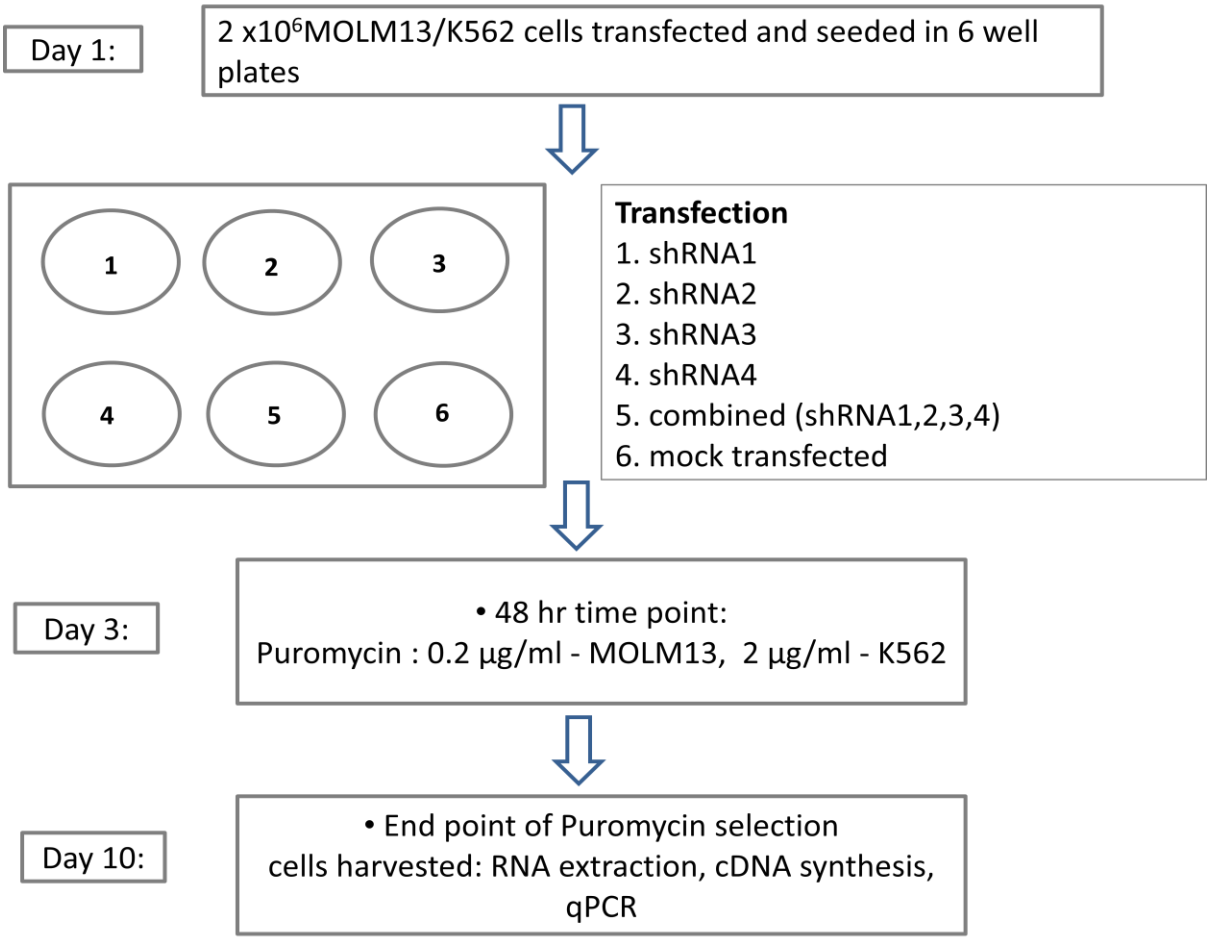


Figure 5.8: Experimental workflow for transfection of pSUPER containing shRNA against *EZH2* into MOLM13 and K562 cells. The cells were seeded in 6 well plates and transfected with either with 1 µg/well of individual shRNA or 1,3 and 5 µg/well of combination of all four shRNAs resuspened in 100 µl of MirusBio Ingenio electroporation solution using the Nucleofactor device II (program: O-017 for MOLM13 and T-016 for K562). Mock transfected cells were technical control for alterations caused due to the transfection procedure. Untransfected cells were exposed to puromycin to estimate the end point of puromycin selection for the cells.

qPCR results for both electroporation and lipofectamine showed similar level of *EZH2* KO in both cell lines and therefore data for MOLM13 has been illustrated. shRNA4 and combination of all four shRNAs showed more KO as compared to shRNA1, shRNA2 or shRNA3 and combination of all at 5 µg/well (Fig: 5.9). Three independent experiments were conducted using shRNA4 and the combination of all the shRNAs (Fig: 5.10) using 1 and 3 µg/well of the plasmid. However the level of *EZH2* KO achieved after puromycin selection

was only 40-50 %. To improve the level of *EZH2* KO, shRNAs cloned in lentiviral vector were used in the subsequent experiments.

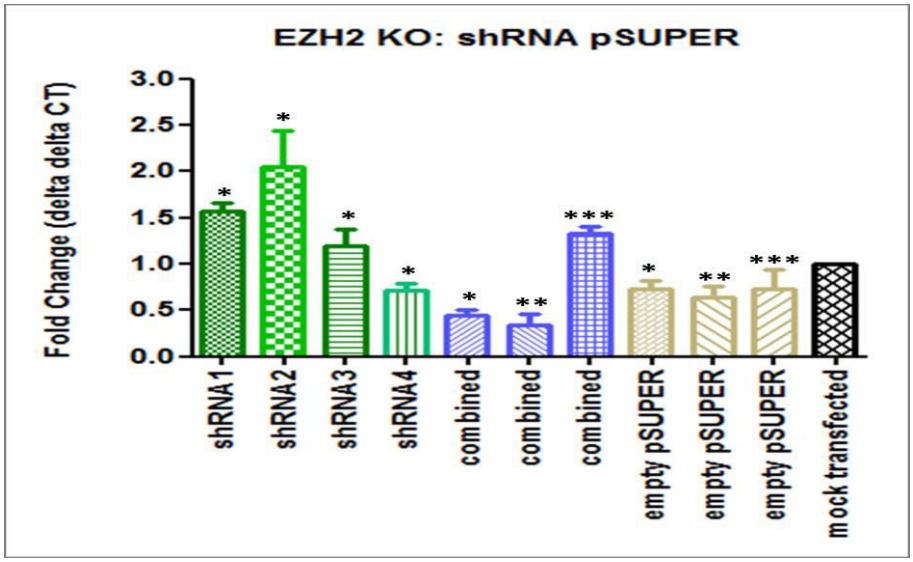


Figure 5.9: *EZH2* expression levels after puromycin selectionin cells transfected with shRNA in pSUPER vector by qPCR. *EZH2* expression levels in MOLM13 cells were calculated relative to the mock transfected cells (Y axis: ΔΔCT). *GAPDH*, *B2M* and *TUBULIN* (data shown) were used as endogenous controls. Error bars mark the standard error between two independent experiments. Note: '*'= 1 µg/well, '**'= 3 µg/well and '***'= 5 µg/well concentration of the control and shRNA against *EZH2*. shRNA4 and combination of all four shRNAs at 1 and 3 µg/well concentrations (blue bars) showed higher levels of KO (40-50 %) as compared to shRNA1,2,3.

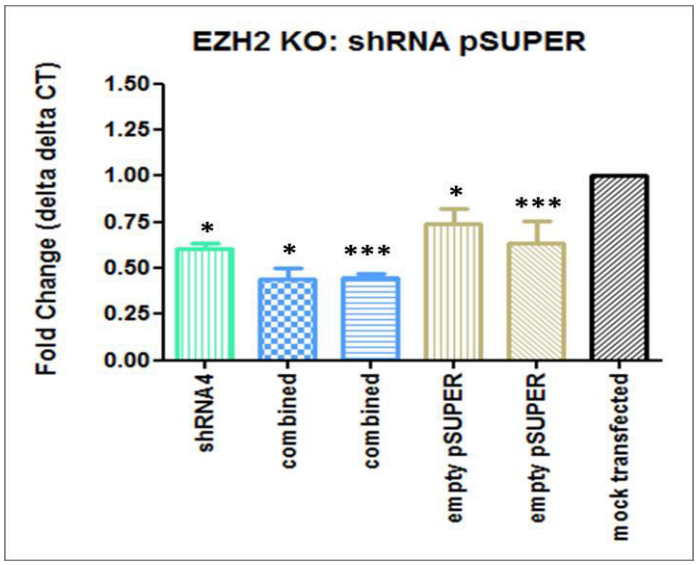


Figure 5.10: qPCR analysis of *EZH2* expression after puromycin selectionin cells transfected with shRNA in pSUPER vector. *EZH2* expression in MOLM13 cells was calculated relative to the mock transfected cells (Y axis: ΔΔCT). *GAPDH*, *B2M* and

TUBULIN (data shown) were used as endogenous controls. Error bars mark the standard error between three independent experiments. Note: '*'= 1µg/well and '**'= 3 µg/wellconcentration of the control and shRNA against *EZH2*. shRNA4 and combination of all four shRNAs at 1 and 3 µg/well concentrations (blue bars) showed 40-50 % KO of *EZH2*.

5.10 Stable knockdown of EZH2 using shRNA in lentivirus

5.10.1 shRNA (Insight Biotechnology)

Four shRNAs (TL304713) were obtained from Insight Biotechnology (UK) and lentiviral particles were generated (refer section 2.17). K562 and MOLM cells were transduced with the virus containing the shRNA and *EZH2* expression was analyzed by qPCR at 48 hr and subsequently on each day till day7. Downregulation of *EZH2* was not observed in both the cell lines at any of time points on the contrary the levels of *EZH2* were elevated at day 7 post transduction (Fig: 5.11). Two independent experiments were conducted on both cell lines with similar results confirming that the shRNAs were ineffective against the target gene and were discontinued for further experiments.

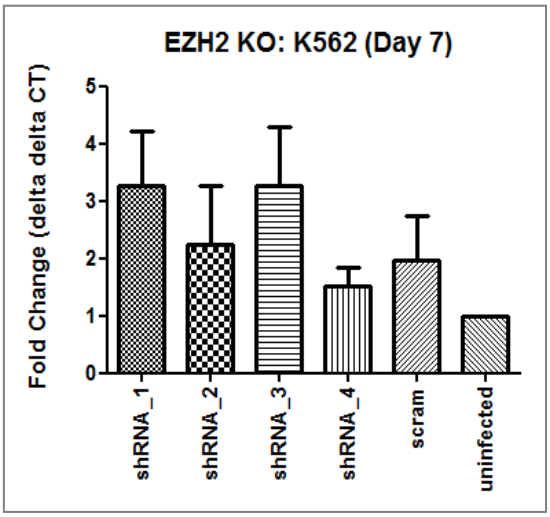


Figure 5.11: qPCR analysis of *EZH2* expression levels at day7 post infection by shRNA (Insight Biotechnology).*EZH2* expression levels were calculated relative to the untransduced cells (Y axis: ΔΔCT). *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between two independent experiments. The expression of *EZH2* was up regulated in cells transduced with all shRNAs including the scrambled confirming that they were ineffective in knocking down *EZH2*.

5.10.2 shRNA (Fisher Scientific)

Four shRNAs were obtained from Fisher Scientific (labelled A-D) and lentiviral particles containing these shRNAs were generated (refer section 2.17). K562 (Fig: 5.12) and MOLM13 cells were transduced with the virus carrying the shRNA and the transduced cells were selected by puromycin. Cells subjected to puromycin selection, after 48 hr of transduction showed 80-90 % cell death with very slow recovery of the remaining cells (15-18 days). This was observed in spite of 40-50 % of the transdcued cells being GFP positive (BD LSRFortesssa) at 48 hr post transduction indicating moderate infection efficiency. But repeated experiments indicated that GFP positivity and puromycin resistance were not concomitant in all instances. The GFP expression in the transduced cells was more consistent at a given time point for a particular concentration of the shRNA in majority of the experiments and therefore this parameter was used for cell selection instead of employing puromycin selection. GFP positive cells were sorted [BD Fluorescence-activated cell sorter FACS) Aria I] at day 7 to obtain pure population of cells transduced by the shRNA. In MOLM13 cell line the percentage of GFP positive cells at all time points (24 hr – 7 day) was very low rendering it difficult to sort enough cells for subsequent assays. The level of *EZH2* KO in MOLM13 was negligible (10 %) which correlated with the percentage of GFP positive cells at all time points.

In K562 cells, *EZH2* expression was analyzed by qPCR at 48 hr post infection (a) and on subsequent days with the maximal KO observed at day 7 (b). This time point was used to sort (BD Facs Aria I) GFP positive cells using a 100 μM nozzle, sheath pressure of 20 psi and post sort purity of 97-99 % to segregate cells carrying the shRNA from the untransduced cells. This enabled collection of pure population of GFP positive cells carrying the shRNA for downstream assays.

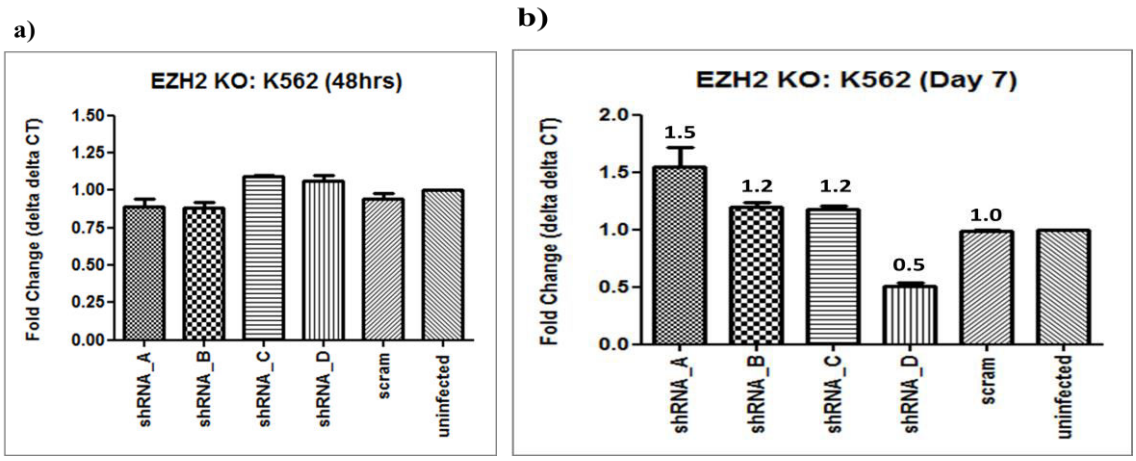


Figure 5.12: qPCR analysis of *EZH2* expression levels at 48 hr and 7 days after transduction with lentivirus containing shRNA (Fisher Scientific) in K562 cells. *EZH2* expression levels were calculated relative to the untransduced cells (Y axis: ΔΔCT). *GAPDH*, *B2M* (data shown) and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between two (a) and three (b) independent experiments. shRNA_D showed 50 % KO at day 7 and therefore this shRNA was used for subsequent experiments.

The level of *EZH2* KO was minimal at 48 hr with all four shRNAs but gradual increase in the down regulation of *EZH2* was observed in cells transduced with shRNA_D with the maximal KO noted at day 7 post transduction. To increase the level of *EZH2* KO the virus (shRNA_D) was concentrated by centrifugation and different quantities (range 10-100 µl) of the concentrated virus were tested on both the cell lines. MOLM13 cells transduced with the concentrated virus showed no change in *EZH2* expression by qPCR while the level of KO increased when the viral titre was increased in K562 cells (Fig: 5.13). The poor transduction efficiency of the MOLM13 cells by lentivirus made it difficult to achieve a significant down regulation of *EZH2* in this cell line and therefore it was discontinued in future experiments.

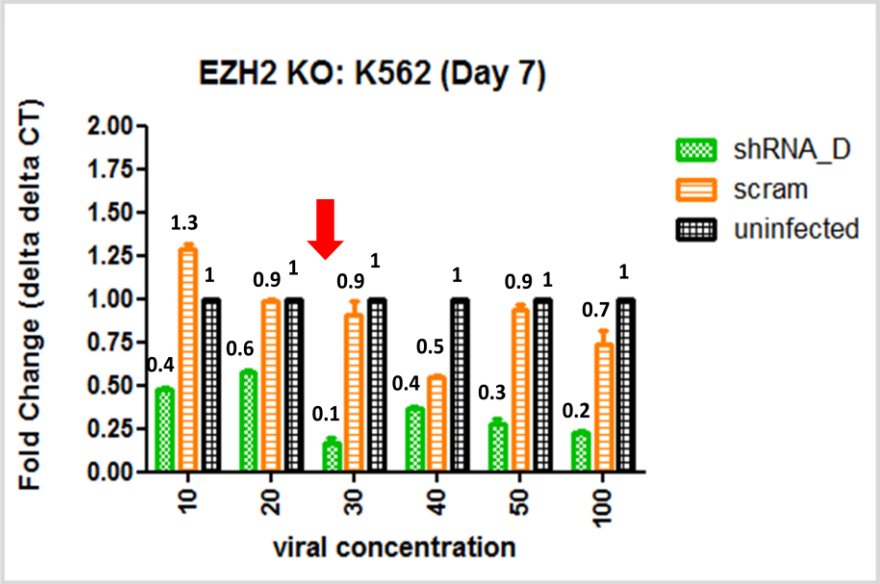


Figure 5.13: qPCR analysis of *EZH2* expression at day 7 after transduction with the concentrated lentivirus containing shRNA (Fisher Scientific) in K562 cells. *EZH2* expression levels were calculated relative to the untransduced cells (Yaxis: ΔΔCT). *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between three independent experiments. Maximal KO was observed when 100 µl of the virus was used. However with both 50 and 100 µl of the virus, massive cell death was observed due to viral toxicity. Using 40 µl of the virus resulted in significant KO of *EZH2* expression in the scrambled shRNA, therefore 30 µl (red arrow) of the viral concentration was used for all subsequent experiments which achieved 80-85 % down regulation of *EZH2* mRNA levels.

Maximal KO (80%) was observed when 100 µl of the virus was used. However using large quantities of the virus (50 & 100 µl) resulted in markedly increased cell death due to viral toxicity. Lowering the concentration to 40 µl of the virus resulted in the KO of *EZH2* expression in the scrambled shRNA and therefore 30 µl (red arrow) of the viral concentration was chosen for all subsequent experiments which achieved 80-95 % down regulation of *EZH2* mRNA levels.

5.11 Effect of shRNA_D on EZH1 expression

EZH1 is a paralogue of *EZH2* with 65% and 94% conservation of amino acids and SET domainsequence respectively between the two (refer 1.4.3). Therefore it was important to

evaluate the effect of shRNA_D on *EZH1* expression levels in addition to its effect on *EZH2*. In the event of both *EZH2* and *EZH1* down regulation by shRNA_D it would be difficult to interpret the individual effect of *EZH2* KO in downstream assays. Therefore, it was important that shRNA_D regulated KO was limited only to *EZH2* gene and not *EZH1*. This was evaluated by qPCR using primers against the *EZH1* gene.

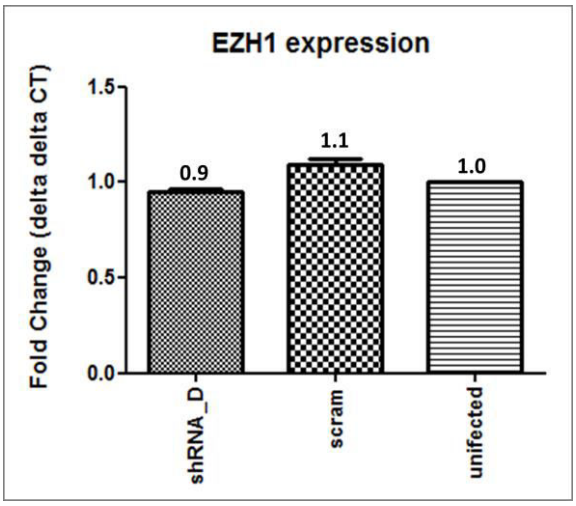


Figure 5.14:qPCR analysis of *EZH1* expression levels at day 7 after transduction with shRNA_D in K562 cells.*EZH1* expression levels were calculated relative to the uninfected/ untransduced cells (Y axis: ΔΔCT). *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls. Error bars mark the standard error between three independent experiments. The expression of *EZH1* was not altered by shRNA_D (<10 % KO) confirming that shRNA_D was specific against *EZH2*.

qPCR analysis illustrated that shRNA_D did not alter *EZH1* levels at day 7 post transduction in GFP + sorted cells. *EZH2* levels decreased >85 % at the same time point in all the samples confirming that shRNA_D specifically targeted *EZH2* expression and could be reliably used to assess the effect of *EZH2* KO on downstream targets.

5.12 Impact of EZH2 KO on H3K27me3

K562 cells were collected at day 7 post FACs sort for GFP positive cells to evaluate the functional effect of the KO on the protein levels of *EZH2* and on its target H3K27me3 by western blot.

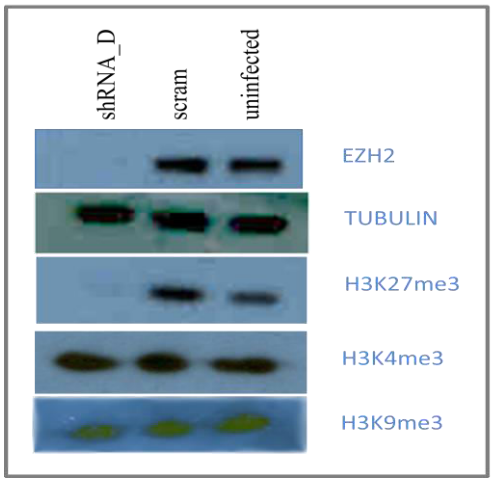


Figure 5.15: Western blot analysis of shRNA mediated KO of *EZH2* in K562 cells on day 7 post FACs sorting for GFP positive cells. 95-100 % KO in the *EZH2* and H3K27me3 protein levels were observed in cells transduced with shRNA_D compared to those transduced with the scarmbled shRNA and the wild type cells. The impact of H3K27me3 downregulation was evaluated on other histone 3 related marks (H3K4me3 & H3K9me3) and no change was observed in their protein levels. Tubulin was used as a loading control. Note: All data are representative of findings from five independant experiments.

Western blot analysis confirmed the findings of qPCR wherein shRNA_D down regulated the expression levels of *EZH2*. In fact the KO of *EZH2* at the protein level was higher (95-100%) than that observed at the mRNA level (85%). The reason for the difference in *EZH2* mRNA and protein levels of was not clear and was observed in multiple experiments. Western blot results suggested that a robust *EZH2* down regulation was required to down regulate its functional target H3K27me3. The samples were also collected at a later time point (day 7 versus 24 hr for the siRNA experiment) compared to siRNA experiment therefore allowing sufficient time for degradation of H3K27me3 protein. The loss of H3K27me3 had no impact

on other histone 3 marks (H3K9me3 and H3K4me3) suggesting that their levels were independent of each other.

5.13 Microarray gene expression analysis of shRNA transduced cells reveals novel *EZH2* targets

EZH2 is directly linked with modulation of gene transcription by targeting H3K27 trimethylation. As shown in the previous section, shRNA mediated *EZH2* down regulation resulted in loss of H3K27me3 levels in K562 cells. To examine the effect of *EZH2* loss on gene expression profile, microarray analysis was carried out on K562 cells transduced with the shRNA, scramble shRNA and the untransduced cells sorted for GFP + on day 7. RNA from these experiments was extracted and used to probe gene expression microarrays (Affymetrix Human Transcriptome 2.0 arrays). Data analysis identified (refer section 2.12.10), a list of genes either induced or down regulated in the cells carrying shRNA against *EZH2* versus those carrying the scrambled shRNA (Table: 5.1). 10 genes were up regulated and 66 were down regulated (>2 average fold difference, $p < 0.05$, ANOVA) in the *EZH2* shRNA transduced cells. The downstream effectors of *EZH2* classified functionally into four major groups encoding cell adhesion, Immune response & Apoptosis and survival. Top enriched pathways and processes were determined Metacore Analysis and the following p values were obtained for top three pathways: Cell adhesion_Chemokines and adhesion (p value 0.001), Cell adhesion_ECM remodeling (p value 0.0017), Apoptosis and survival_p53-dependent apoptosis (p value 0.007) and process; Cell adhesion_Cell-matrix interactions (p value 0.001), Signal Transduction_BMP and GDF signaling (p value 0.005), Development_Blood vessel morphogenesis (p value 0.009). The p values are high suggesting that *EZH2* KO did not have a strong effect on a single pathway but affected genes belonging

to different pathways (Fig: 5.16). Such networks are indicative and need experimental validation.

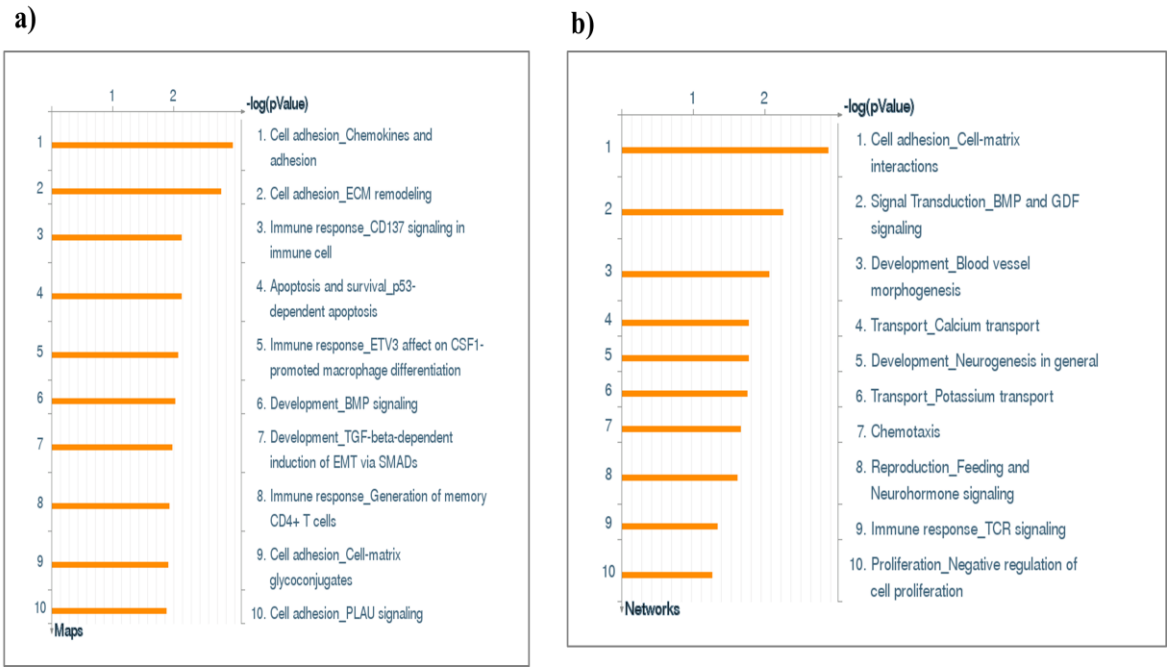


Figure 5.16: Illustrates schematic representation of top enriched pathways (a) and processes (b) for all up and down regulated genes after *EZH2* KO. Indicated pathways and processes were determined by Metacore Analysis. The log p value for each pathway is mentioned at the top. Genes involved in cell adhesion were ranked at top followed by those involved in Immune responses and apoptosis. However the p values were high suggesting that *EZH2* KO does not have a strong effect on a single pathway but affects genes belonging to different pathways.

Upregulated Gene Symbol	p-value (shrna vs. scr)	Fold Change
SLC16A7	0.034	4.77
C5orf42	0.002	3.1
MIR1302-5	0.012	2.44
MIR4653	0.038	2.38
HBZ	0.022	2.23
HBM	0.012	2.16
RASL10A/RRP22	0.047	2.05
TRIM66	0.038	2.04
SLC25A48	0.036	2.04
NHLH1	0.026	2.02

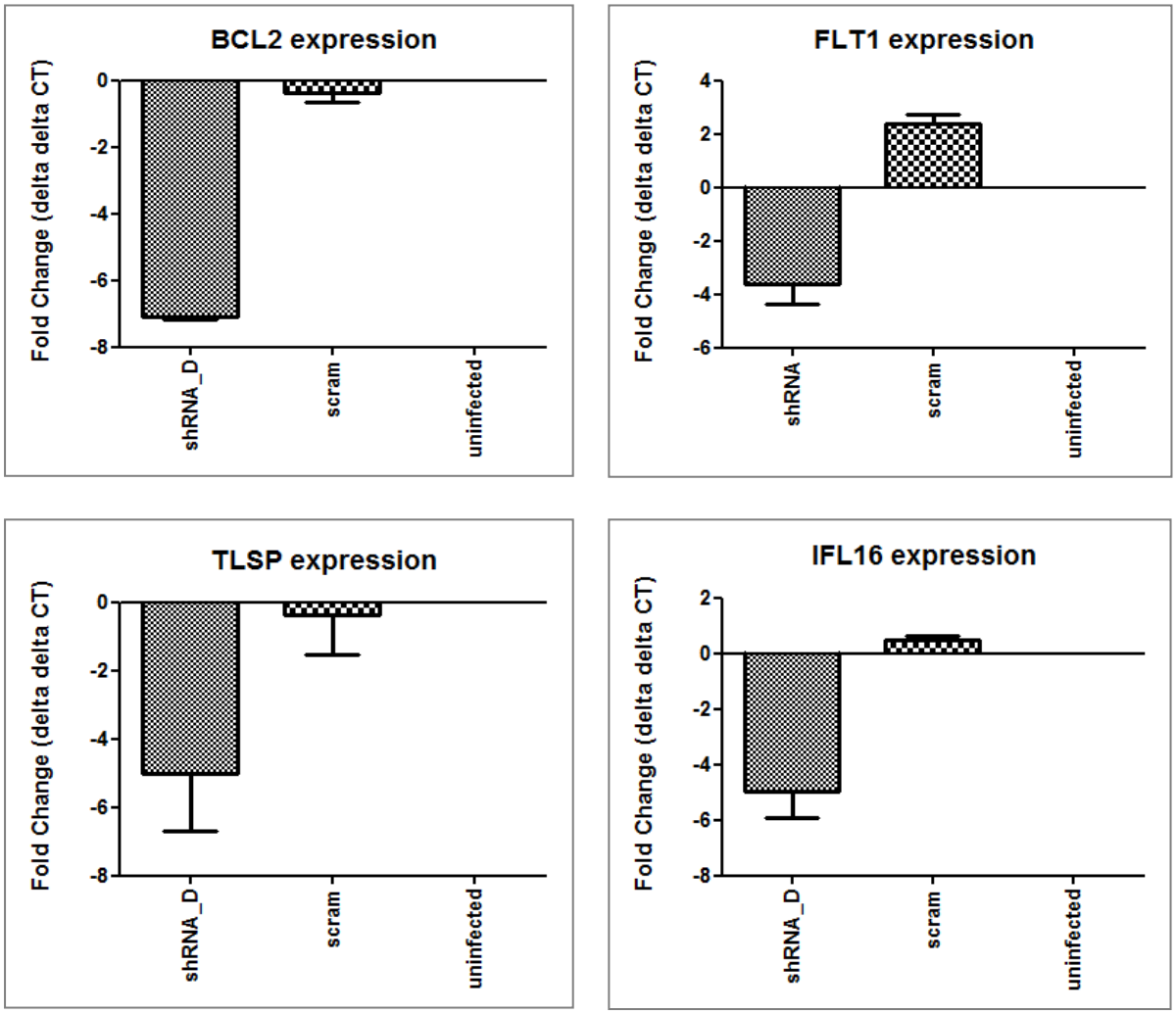
Down regulated Gene Symbol	p-value (shrna vs. scr)	Fold Change
CD44	0.040	-8.12
SNORD45	0.002	-6.37
OLAH	0.004	-5.99
MIR4791	0.022	-5.74
MIR676	0.044	-4.96
COL6A3	0.011	-4.69
UPP1	0.023	-4.54
FLT1	0.038	-4.03
SLC22A15	0.029	-4
MIR4432	0.020	-3.67
CCDC152	0.018	-3.67
PAG1	0.030	-3.37
TSLP	0.023	-3.24
ETS1	0.001	-3.21
IFI16	0.014	-3.11
MIR101-2	0.009	-3.1
NRP1	0.023	-3.09
GPR65	0.030	-2.86
ANXA3	0.001	-2.67
TGFBR2	0.008	-2.64
AREG, AREGB	0.020	-2.64
SNORD115-6	0.025	-2.62
CSTA	0.019	-2.55
PLAU	0.020	-2.55
CD3D	0.016	-2.53
SMAD7	0.036	-2.53

Down regulated Gene Symbol	p-value (shrna vs. scr)	Fold Change
ARHGAP25	0.005	-2.47
SRPX	0.002	-2.46
CALB1	0.033	-2.42
AREG, AREGB	0.015	-2.36
SUCNR1	0.048	-2.36
HVCN1	0.009	-2.33
SLC24A2	0.001	-2.33
SLC9A7	0.043	-2.32
ZCWPW1	0.039	-2.3
SLC27A6	0.011	-2.29
ARHGAP20	0.003	-2.28
MPZL3	0.005	-2.28
AMIGO2	0.020	-2.28
SDC4	0.009	-2.28
PPIC	0.043	-2.25
C7orf23	0.043	-2.24
HOXA10	0.005	-2.2
RAP1B	0.002	-2.18
BCL2	0.001	-2.14
SERPINB8	0.002	-2.13
MTX3	0.034	-2.12
GADD45B	0.011	-2.11
CD83	0.001	-2.1
ATF3	0.019	-2.1
SERPINE2	0.009	-2.1
NEK3	0.022	-2.09
EPS8	0.005	-2.08
CXorf66	0.042	-2.07
C1orf198	0.010	-2.06
TTC39B	0.017	-2.05
NRIP3	0.002	-2.05
C10orf55	0.036	-2.03
MIR376B	0.014	-2.02
ST6GALNAC3	0.008	-2.01
LGALS3	0.005	-2.01
CYLD	0.016	-2.01
ACRC	0.020	-2.01
TNFRSF9	0.021	-2.01

Table 5.1: List of genes modulated by loss of *EZH2* expression.Ten genes were up regulated and 66 were down regulated (>2 average fold difference, p <0.05, ANOVA). Nine genes (highlighted) were chosen for qPCR based validation because of their most pronounced change in expression and prior correlation to haematological malignancies.

5.13.1 Validation of *EZH2* targets

qPCR was used to confirm the results obtained by microarray gene expression analysis (Fig: 5.17). The genes investigated (highlighted in the table) were; *BCL2, FLT1, HOXA10, CD44, CD83, TSLP, IFI16* and *PAG1*. These genes were chosen based on the degree of fold change [*CD83, TLSP, PAG1, IFI16*], known correlation to haematological malignancies [*BCL2, FLT1,CD44*] or for their association with *EZH2* [*HOXA10*].



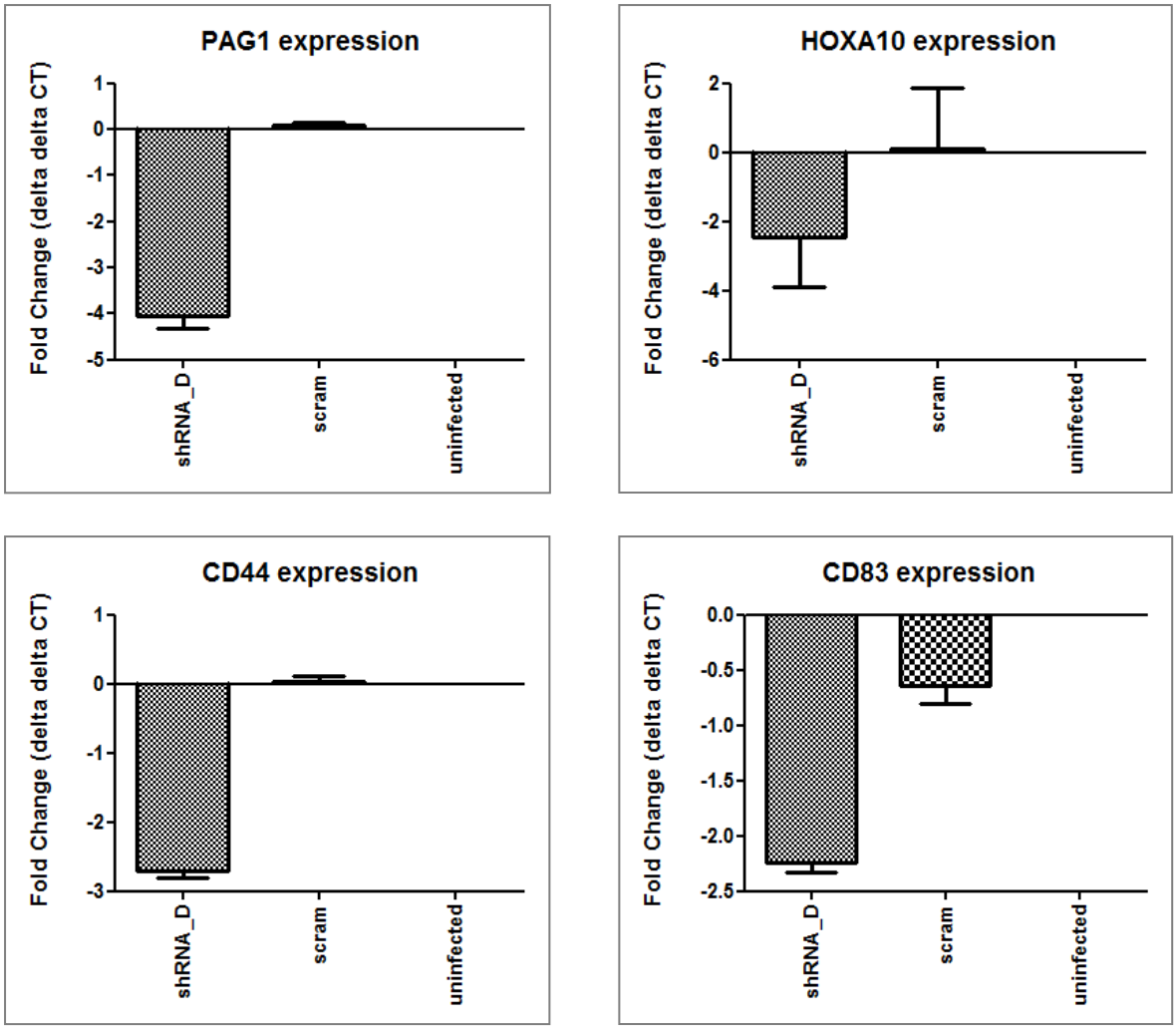


Figure 5.17: qPCR validation of genes modulated due to loss of *EZH2* in K562 cells.Gene expression levels were calculated relative to the transduced cells (Y axis: ΔΔCT). *GAPDH* (data shown), *B2M* and *TUBULIN* were used as endogenous controls and the error bars mark the standard error between three independent experiments.The level of KO mentioned in the text is compared to the cells infected with the scrambled shRNA.The genes associated with apoptosis (*BCL2, PAG1 & IFI16*) showed -4 to -6 fold reduction in their expression while those linked to immune modulation (*CD44, CD83, TLSP*) were down regulated to -2.5 to -7 fold and *FLT1* and *HOXA10* expressions were reduced to -4 fold change.

mRNA used for qPCR based validation of genes was collected from three independent experiments and was separate from mRNA collected for microarray gene expression studies. The deregulation of genes due to loss of *EZH2* observed in microarray gene expression experiment were confirmed, *FLT1* and *HOXA10* (associated with cell differentiation) expressions showed a direct association to *EZH2* levels with loss of one leading to reduced

expression of the other. Other genes were categorised as; a) those associated with immune modulation (*CD44, CD83, TLSP*) which were down regulated to -2.5 to -7 fold change and b) genes associated with apoptosis (*BCL2, PAG1 & IFI16*) which showed a fold change of -4 to -6 of *BCL2* and *PAG1*expression causeing decreased cell proliferation while reduced expression of *IFI16* induces cell proliferation (discussed later). This paradoxical effect of *EZH2* modulated genes on cell proliferation can contribute to disease pathogenesis of MDS. MDS is closely linked to apoptosis/deregulated cell proliferation leading to peripheral blood cytopenias with paradoxical bone marrow hypercellularity. Cytopenias are generally classified as deficiency or immune mediated, induced by BM failure or idiopathic. Cytopenias in MDS are caused due to bone marrow failure and are important or sometimes are the only clinical indication to diagnose this disease in its early stages (Valent, 2012). Apart from their role in disease detection, they are also critical in predicting the prognosis of the disease. International Prognostic Scoring system (IPSS) evaluates prognosis of the patients based on the number of lineages (anaemia, thrombocytopenia, neutropenia) affected by cytopenias. Increasing clinical stages of MDS, shortened OS and time to evolution were observed in patients with severe cytopenias (Kao et al., 2008). The importance of cell proliferation and apoptosis in this disease needs further clarification and to this end I examined the effects of *EZH2* KO on apoptosis (Annexin V staining), cell proliferation (MTT assay, trypan blue staining) and cell cycle kinetics (PI/FITC staining) in K562 cells.

5.14 Impact of EZH2 KO on cell viability & proliferation

5.14.1 Trypan blue staining

To assess cell viability and proliferation, Trypan blue dye exclusion method was utilised. Trypan blue does not stain live cells (intact cell membrane) but it traverses the cell membrane of dead cells staining them with a distinctive blue colour under the microscope.

Cell suspension was inoculated (triplicates) in 12-6 well plates and on each day post transduction with lentivirus containing shRNA_D, 10 µl of the cell suspension was mixed with 10 µl of 0.4% Trypan blue solution (w/v) (Lonza). The mix was incubated at room temperature for 3 min and the cells were counted using a dual chamber haemocytometer and a light microscope. The cell count/ml was calculated using formula:

$$\frac{\text{Total number of cells (four corner squares of haemocytometer)}}{4} \times \text{dilution factor} \times 10^4$$

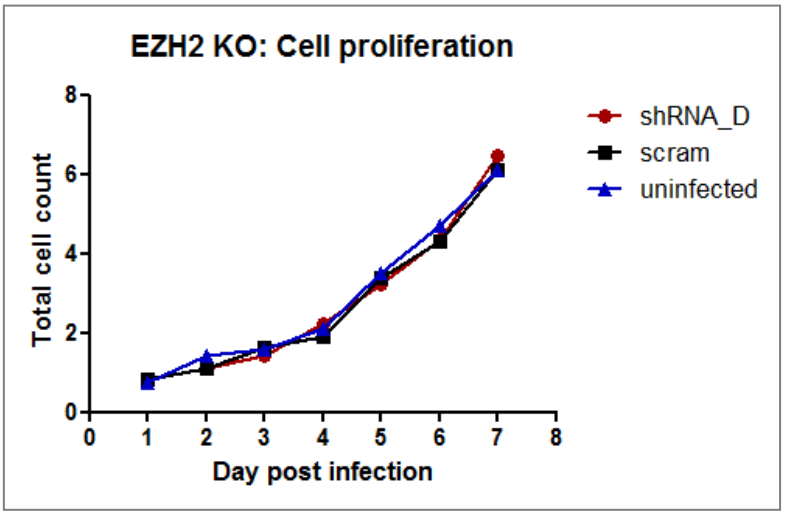


Figure 5.18: Cell count of K562 cells using Trypan blue staining post shRNA lentiviral infection. X axis = number of days post transduction, Y axis = Total cell count (x 10^6). Very few (< 0.5 %) dead cells (staining blue) were observed in the three experimental conditions at all time points. There was no difference in the proliferation rates of cells transduced with shRNA_D against *EZH2* (red), scrambled (black) and the wild type cells (blue) suggesting that *EZH2* KO in K562 cells did not affect cell viability and proliferation.

Trypan blue dye exclusion method indicated nearly identical cell counts in all three experimental conditions at a given time point. The number of dead cells (staining blue) were low (<0.5 %) in all the samples suggesting that there was no difference in cell viability and proliferation rate in cells transduced with shRNA_D against *EZH2* versus the cells transduced with the scrambled shRNA and the wild type cells.

5.14.2 MTT tetrazolium cell proliferation assay

To confirm the effect of *EZH2* inhibition on cell proliferation, K562 cells were transduced with shRNA _D and scrambled shRNA. At day 7 (95 % KO), the cells containing GFP were sorted (BD FACS Aria) and were seeded at a density of 30,000 cells/ well in 100 µl RPMI media for different time points (24, 48 and 72 hr). All samples were seeded in triplicates including wild type cells. Three wells without cells but with media (RPMI) were used as control (blank) to negate background signal. MTT assay was performed (refer section 2.18) and readings were measured at 570 nm using a plate reader. The data was analysed by calculating the average for the triplicates in each sample and normalizing this mean value with the average of the blank (sample-blank). Fold change was determined by dividing the normalized average of the samples by the normalized average of the calibrator sample (wild type cells). Though wild type cells are used as calibrator, the change in the proliferation rate and the p vaules are mentioned compared to the cells transduced with scrambled shRNA. There was no significant (p =0.1 - 0.4) reduction in the metabolic activity of the cells transduced with shRNA_D versus those transduced with scrambled shRNA or wild type cells confirming that *EZH2* KO has a minimal, non significant effect on myeloid cell proliferation (Fig: 5.19).

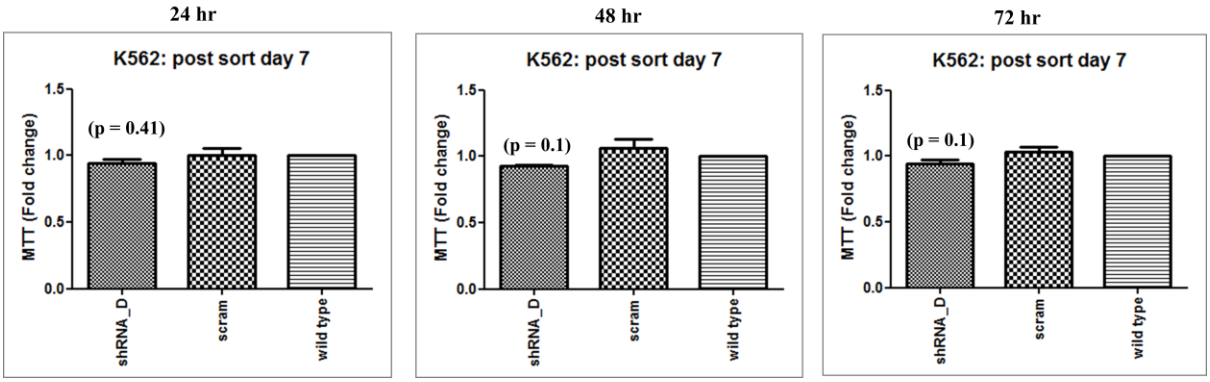


Figure 5.19:Cell proliferation assessment by MTT assay on day 7 after transduction in K562 cells. Cells were seeded at a density of 30,000/well in 100 µl RPMI for three time points (24, 48 and 72 hr). Cell proliferation by MTT assay revealed minimal reduction in proliferation of cells (p = 0.1 to 0.4) with *EZH2* inhibition (95 %) at all time points. Results are expressed as fold change relative to the wild type cells while the p values are calculated relative to the scrambled. Error bars represent standard error for five independent experiments each performed in triplicate.

The results from trypan blue dye exclusion method and MTT assay confirm that KO of *EZH2* (95 %) does not have a significant effect on cell proliferation in K562 cells.

5.14.3 Annexin V staining

To examine the effect of *EZH2* inhibition on cell viability and to confirm the findings of Trypan blue dye exclusion method, cells transduced with shRNA_D (against *EZH2*) were stained with Annexin V (refer section 2.20). Annexin V staining provided an advantage in this aspect wherein in it allowed detection of cells in early apoptotic phase.

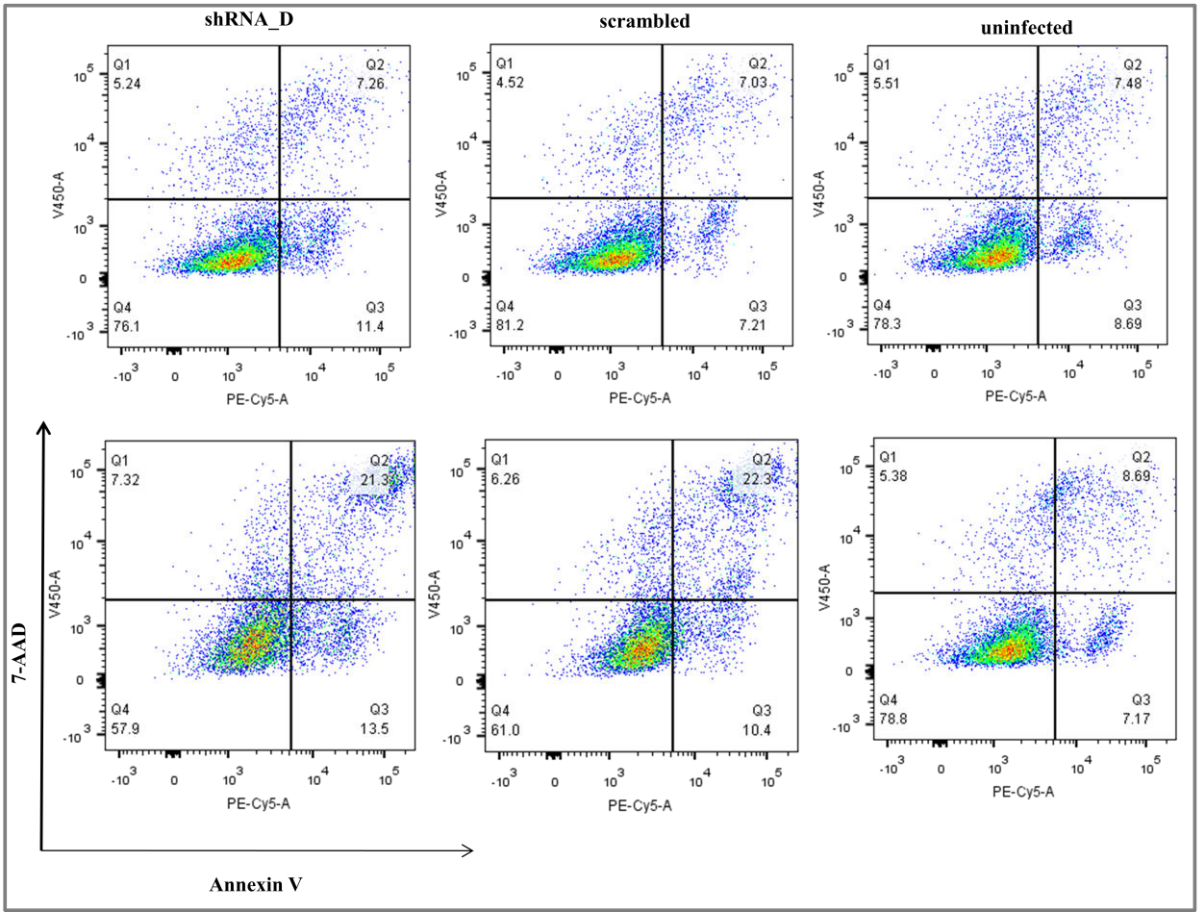


Figure 5.20: Annexin V and 7-AAD staining of K562 cells at 48 hr post shRNA transduction (top panel) and on day 7 post sorting for GFP positive cells (bottom panel). The difference between the percentage of live (Q4 = double -ve for annexin and 7-AAD), early apoptotic (Q3 = + ve annexin and - ve 7-AAD) and late apoptotic cells (Q2 = double +ve for annexin and 7-AAD) was not significant between cells transduced by shRNA_D versus those transduced by scrambled shRNA and wild type cells indicating that *EZH2* KO does not affect cell viability in myeloid neoplasms. Note = data represents three independent experiments.

Fig: 5.20 illustrate Annexin V and 7-AAD staining in K562 cells at two time points (48 hr post shRNA transduction and day 7 post sorting for GFP positive cells). There was no difference in the percentages of live, early apoptotic and late apoptotic cells transduced by shRNA_D versus those transduced by scrambled shRNA and wild type cells. The results imply that loss of *EZH2* expression does not affect cell viability in K562 cells confirming the findings of cell count by trypan blue.

5.14.4 Cell Cycle analysis

The close association of *EZH2* to cell cycle has been shown in various types of cancers and has been discussed in section 5.2. To study the change in kinetics of cell cycle entry after *EZH2* inhibition, flow cytometric analysis (PI/FITC) (section 2.19) and the percentage of cells in each cell cycle phase as well as apoptotic cells (sub G0) were studied in GFP positive K562 cells sorted on day 7 (Fig: 5.21).

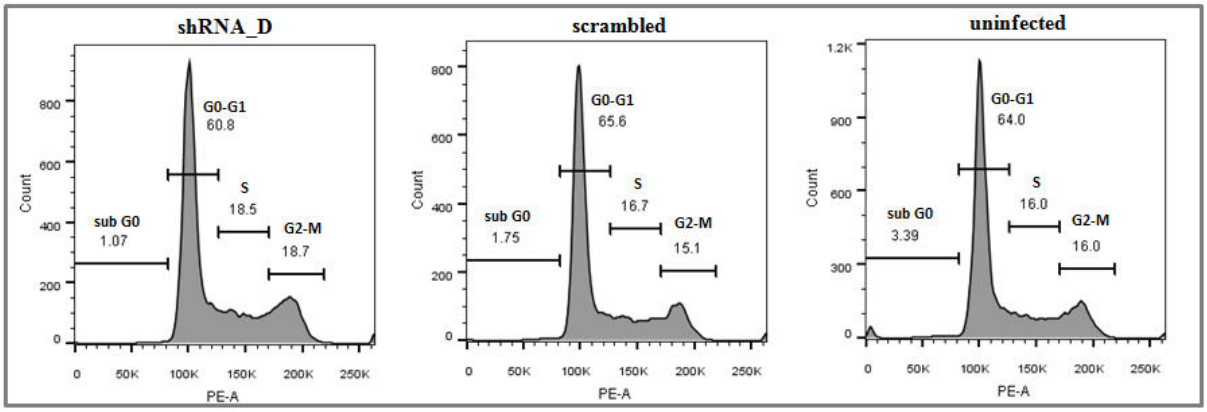


Figure 5.21:Cell cycle profile in K562 cells on day 7 after sorting for GFP positive cells. Samples were fixed (70% ethanol v/v) and stained with PI (DNA content) and FITC (protein content).The percentage of cells in sub G0, G0/G1, S and G2/M phases of the cell cycle were determined by flow cytometry and remained unaltered in the cells transduced with the shRNA (KO 95 %) versus those transduced with the scrambled or wild type cells. The percentage of apoptotic cells (sub G0) varied marginally between the wild type cells (3.39 %) and those transduced with the shRNA (1.07 %) and the scrambled (1.75 %) but the difference was variable between the experiments and not significant. Note = data represents findings from three independent experiments.

Down regulation of *EZH2* (95 %) in K562 cells did not alter the protein content (i.e. cell size) or the DNA content (PI stain) in myeloid cells (K562), a finding consistent with results obtained by siRNA mediated *EZH2* KO (65 %) in MOLM13 cell line (refer section 5.7). The percentage of cells in the different phases of the cell cycle (G1/M, S and G2/M) as well as the apoptotic cells (sub G0) remained unaltered after *EZH2* KO. **The role of *EZH2* in cell cycle kinetics of myeloid cells (MOLM13 and K562) was not as a enhancer of proliferation as**

compared to its functionality in other cancers (ovarian cancer G1 phase arrest and prostate cancer G2/M phase delay) wherein down regulation of *EZH2* had profound impact on cell proliferation and cell invasion indicating an extremely cell type specific role of this gene.

5.15 Confirmation of gene signature in patient samples

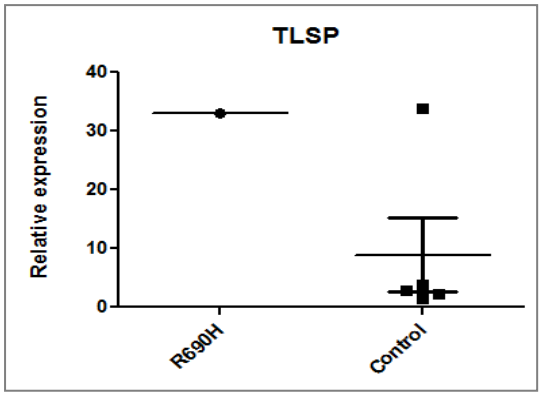
Bone marrow samples from 61 patients were previously sequenced by our group (refer chapter 3) and nine patients were identified with *EZH2* mutations. The previous chapter was focused on constructing and studying *EZH2* R690H mutation located in the SET domain of this enzyme. One of the nine patients in our cohort carried the R690H mutation (Table: 5.2).

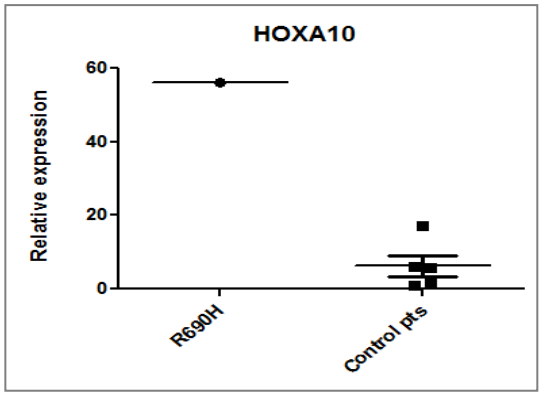
Patient No	Cytogenetics	Evolution to AML	WHO classification	IPSS category	% bone marrow blasts
LSL/001326 (R690H)	Normal	No	RCMD	Low	2 %
LSL/001240	Normal	No	RCMD	Int-I	<5 %
LSL/000961	complex	No	RAEB 1	Int-II	2 %
LSL/002025	Monosomy 7	No	RCMD	Int-II	2%
LSL/002029	Normal	No	RCMD	Int-I	1%
LSL/001904	Normal	No	RARS	Int-I	4%

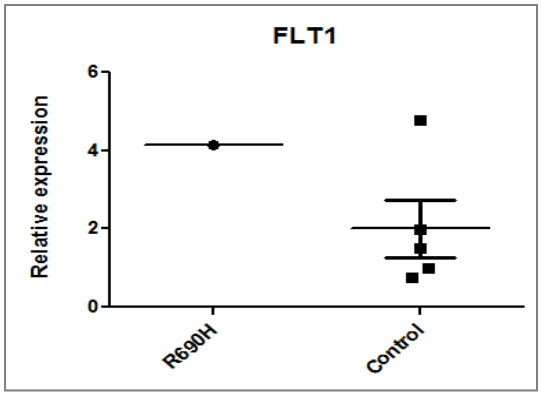
Table 5.2: Characteristics of patients used for confirmation of gene signature obtained by microarray gene expression profiling. LSL/001326 patient is mutated for *EZH2* (R690H) while the other five patients contain wild type *EZH2*.

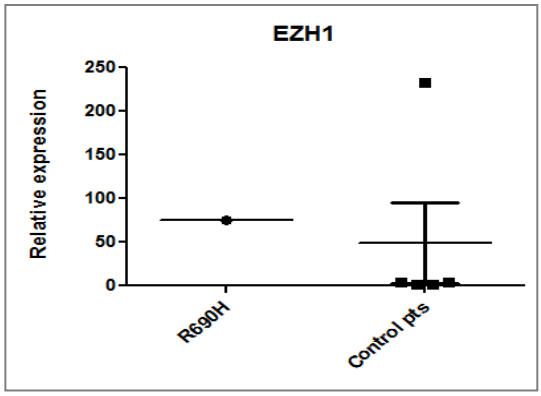
To evaluate whether this mutation mimics loss of *EZH2* in patients, qPCR was carried out on RNA extracted from this patient and compared to five control patients (Table: 5.2) without *EZH2* mutation. Genes identified by microarray analysis were tested (*BCL2, PAG1, IFI16, CD44, CD83, HOXA10, RAP1B, FLT1*and *TSLP*) on patient samples. Two additional genes *EZH1* and *CDKN2A* were also examined due to their association with *EZH2*. Statistical tests

to calculate significance (p value) could not be carried out for this experiment since the sample size in each category was very small [n=1 (R690H mutant)]. Five genes *HOXA10*, *FLT1*, *PAG1B*, *EZH1* and *TLSP* (Fig: 5.22) were up regulated in the patient carrying the R690H mutant compared to the control patients while other genes showed no change in their expression levels. **Interestingly, these results were opposite to the findings observed with inhibition of *EZH2* expression using shRNA in lentivirus where these genes were down regulated suggesting that *EZH2* mutation does not mimic loss of function of the gene.**









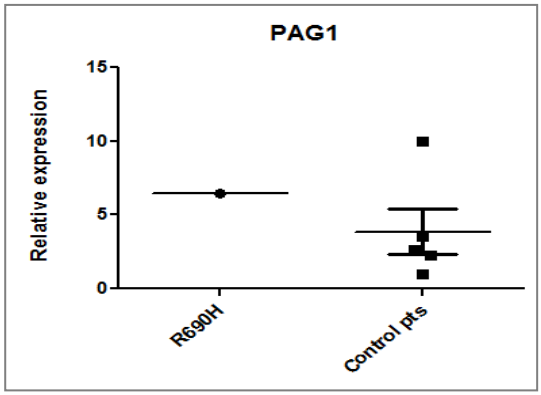


Figure 5.22: qPCR analysis of patient samples with (R690H) or without *EZH2* mutation. Gene expression levels were calculated relative to *GAPDH* (data shown) and *B2M* (Y axis: ΔCT) which were used as endogenous controls. 5/11 genes showed change in their expression in patient with *EZH2* mutation (R690H) (n=1) versus the control patients (n=5). Interestingly, these results were reverse to the findings observed with inhibition of *EZH2* expression using shRNA_D suggesting that *EZH2* mutation does not mimic loss of function of the gene. Statistical tests for significance (p value) could not be calculated due to small sample size.

5.16 Discussion

Target genes and binding partners of histone methyltransferase *EZH2* are extremely tissue specific and vary depending upon the stage of cell differentiation. Over expression of *EZH2* has been confirmed in ovarian, renal, colorectal, breast, prostate cancer and the level of *EZH2* expression has been correlated negatively to disease prognosis and relapse free survival. In contrast, its role in haematological cancers differs in lymphoid malignancies were mutations of this gene were linked to it's over expression while in myeloid malignancies they were associated with loss of its methyltransferase activity. The involvement of *EZH2* in H3K27 methylation makes it a key regulator of transcription.

5.16.1 Transcription and gene expression profiling

Transcription is an intricate process which is regulated by both genetic and epigenetic factors. Epigenetic marks, such as DNA methylation and modification of histone tails are crucial in transcriptional regulation. As mentioned above, *EZH2* modifies H3K27 tails and hence has widespread influence on gene expression. Therefore it is not surprising that this gene is disrupted in cancers including MDS/AML. Understanding the deregulation of *EZH2* and its influence on transcriptional regulation of target genes is at the heart of understanding the role of this gene in the pathogenesis of MDS/AML(Young et al., 2011). The data presented here show that inhibition of *EZH2* expression in myeloid cells caused loss of H3K27me3 marks as seen on western blot analysis. Transient KO of *EZH2* by siRNA (Dharmacon) in MOLM13

cells reduced the expression of *EZH2* by 65% but a concomitant loss of H3K27me3 was not perceived at both 24 and 48 hr time points indicating that either a more effective KO of *EZH2* was essential or H3K27me3 had a long half life and therefore the degradation of this protein is not observed at 24 hr (Budhavarapu et al., 2013). Therefore lentiviral based shRNA (against *EZH2*) transduction was carried out resulting in 95 % KO of both *EZH2* and H3K27me3 expression levels in K562 cells at Day7 after sorting for GFP positive cells. The effect of H3K27me3 loss on gene expression can be multi-dimensional as studies using ChIP-sequencing on murine embryonic fibroblasts and human embryonic cells have generated genome wide maps of H3K27me3 marks and found that the position of these marks across the gene dictated the resultant transcriptional inhibition (H3K27me3 marks across the body of the gene), transcriptional activation (H3K27me3 marks at the promoter) or both (H3K27me3 marks at the transcriptional start site). H3K27me3 marks at the transcriptional start site are bivalent and usually related to H3K4me3 at these sites (Young et al., 2011). Apart from individual gene occupancy, these marks can also be found in large domains that spread over hundreds of kilobases and regulate entire gene families. To define a gene signature caused by loss of H3K27me3 as a result of *EZH2* inhibition in K562 myeloid cells, microarray gene expression profiling was carried out which identified *BCL2, FLT1, HOXA10, CD44, CD83, TLSP, IFI16*and *PAG1* genes to be down regulated. Gene expression profile obtained in this study was compared to gene signatures data available from studies conducted in other types of cancers (Table: 5.3) to examine overlapping genes signatures. As expected, there was no overlap of *EZH2* targets between the data sets confirming that *EZH2* target genes differ in the context of cell/ tissue type. The exact reason for the varying mechanisms in *EZH2* mediated transcriptional regulation in different cell types is still unclear though the effect on cell proliferation and differentiation was invariable between epithelial ovarian cancer, epidermal progenitor cells,neural stem cells and Ewing tumour.

Top genes	Tissue type	EZH2 KO (Functional effect)	Reference
Up regulated: *ALDH1A1*, *SSTR1* and *DACT3*	Epithelial ovarian cancer (EOC)	Regulates undifferentiated state of stem cells in EOC cells by altering *ALDH1AL* levels.	(Li et al., 2012)
Down regulated: *EMP1, GFAP, GAP43, EPHB2, ALCHAM, NGFR*	Ewing tumour	Down regulation of *EZH2* prevents oncogenic transformation of cells and reexpression of neuroectodermal and endothelial differentiation genes.	(Richter et al., 2009)
Up regulation: *LCELC,EIF2S3Y,FLG,LCELD, LCELM*	Epidermal progenitor cells	*EZH2* controls proliferation of basal progenitors by repressing the *INK4A* locus. Loss of *EZH2* increased epidermal differentiation.	(Ezhkova et al., 2009)
Expression alters according to differentiation status of NSC: *TAL1, SIX1, TLX3, PHOX2B, OTP, PDGFR-A, NKX2-2*	Neural stem cells (NSC)	Decreased *EZH2* is associated with loss of stem cell characteristics in NSC, blocked proliferation and increased apoptosis.	(Sher et al., 2012)

Table 5.3: Gene expression profiling in different cell types/ cancers after KO of *EZH2*.
As expected, there was no overlap of gene signatures in different cancers suggesting extremely tissue specific role of *EZH2* in transcriptional regulation. However, the next effect on stem cell differentiation and proliferation was constant in these cells.

HOXA10 is a transcription factor associated with cellular differentiation and closely linked to T, B and Natural killer cells. It showed 2-fold down regulation after KO of *EZH2* in K562 cells. *HOXA10* drives the differentiation of hematopoietic precursor cells into myeloid or lymphoid lineages and over expression of this gene affects differentiation of human stem cells wherein it causes severe deregulation of the lymphoid (T, B, and NK) cells development. Some studies show that *HOXA10* deregulation affects terminal differentiation

in myeloid cells (Buske et al., 2001) while others indicate an enhanced monocytic differentiation with inhibition of only granulocytic precursor cells differentiation towards polymorphonuclear leukocytes (Taghon et al., 2002). Dysregulation of *HOXA10* expression is also correlated with abnormal expansion of granulocyte macrophage progenitor cells (GMP) (Bei et al., 2012). Increased expression of *HOXA10* is associated with poor prognosis in AML with increased myeloproliferation (Bjornsson et al., 2001). Decreased *HOXA10* levels after *EZH2* KO might indicate a tumour suppressor role of *EZH2* in myeloid malignancies.

In addition to cell differentiation, *EZH2* KO also modulated proteins involved in immune responses like *CD44*, *CD83* and *TLSP*. *CD44* is a cell adhesion molecule expression of which is elevated in MDS and AML. Soluble *CD44* retain sbiological functions and competes with cell bound *CD44* in ligand binding, cellular activation, induction of apoptosis and inhibition of tumour growth. Serum concentrations of soluble *CD44* increase with increasing grade of the disease. Differential expression in various FAB or WHO subtypes, display the lowest levels in del5q MDS and the highest levels in RAEB, CMML and in secondaryAML evolved from MDS. The levels are also high in patients with relapsed disease and low in patients with remission. Therefore serum *CD44* has been suggested as a marker for monitoring response to treatment and disease progression to acute leukaemia (Loeffler-Ragg et al., 2011). Cell bound *CD44* molecule exhibit a trend similar to the serum soluble *CD44* with high levels (15 to 35 %) observed in MDS. Cell bound *CD44* molecule was not expressed in healthy controls (Xie et al., 2010). In addition to *CD44*, *CD83* molecule expressed on the cell surface of mature dendritic cells was also down regulated along with the Thymic stromal lymphopoietin (*TSLP*) which mainly impacts myeloid cells, induces the release of T cell attracting chemokines and enhances the maturation of myeloid dendritic cells. Within the thymus *TSLP* activation of both myeloid and plasmacytoid (CD123+) dendritic cells results in the production of

regulatory T cells (Hanabuchi et al., 2010). The downstream targets and the effect of their modulation on the pathogenesis of MDS/AML need further clarification.

Apart from genes involved in cell differentiation and immune modulation, reduced *FLT1* expression was observed after *EZH2* KO. *FLT1* also known as *VEGFR1* binds to *VEGFB* and placental growth factor (*PIGF*). *PlGF* signals directly through *FLT1* in various cell types including angiogenesis-competent myeloid progenitors thereby promoting tumour angiogenesis and formation of premetastatic niche. *PIGF* expression is not crucial for vascular development in healthy cells but contributes to angiogenic and inflammatory switch in cancer cells. Myeloid cells confer resistance to therapies that target VEGF by secreting more pro-angiogenic factors. Monoclonal antibody against *PlGF* enhances the responsiveness to VEGF targeted therapies by inhibiting macrophage recruitment and therefore this antibody can be used against *PIGF* to eliminate the source of angiogenic factors that contributes to anti-angiogenic escape of tumours (Fischer et al., 2008). High expression of *FLT1* gene is associated with poor prognosis and/or unfavorable outcome in MDS but does not correlate with peripheral blood blast counts (Kracmarova et al., 2008). *FLT1* is regulated by mir145 which is down regulated in 5q- syndrome seen in MDS. Whether down regulation of mir145 causes concomitant up regulation of *FLT1* is unclear. It is unknown whether down regulation of *FLT1* modulates mir145 expression or influences the 5q- phenotype of MDS and further research is required to decipher the association of *FLT1* to MDS.

5.16.2 Effect of EZH2 on cell cycle, cell viability and cell cycle kinetics

Three genes related to cell proliferation; *BCL2*, *PAG1* and *IFI16* were down regulated by *EZH2* inhibition. Reduction of *BCL2* (-7 fold) and *PAG1* (-4.5 fold) causes decreased cell proliferation while decreased expression of *IFI16* (-6 fold) induces cell proliferation *PAG1* (Glycosphingolipid-enriched microdomains 1) is a transmembrane adaptor protein which has

been studied in B-cell non-hodgkin's lymphoma wherein KO of *PAG1* is associated with reduced cell proliferation. Reduced *PAG1* expression is also linked to abnormal response to antigen stimulation in germinal centre derived B-cell non-hodgkin's lymphoma (Rubbi et al., 2011). *IFI16* on the other hand is a nucleoprotein expressed during differentiation of myeloid precursor cells. It is expressed inherently in T and B cell lines but also inducible in myeloid cell in response to Interferon γ stimulation (Dawson and Trapani, 1995). *IFI16* controls cellular proliferation by binding and modulating p53 and the retinoblastoma protein pRb. The nucleoprotein is expressed in monocytic lineage but is down regulated in other myeloid precursor cells (Dawson et al., 1998). Similar to *PAG1, BCL2* also enhances cell proliferation. This gene and its association with apoptosis/ proliferation has been studied in detail compared to *IFI16* and *PAG1* due its role in *BCR-ABL* induced chronic myelogenous leukaemia. *BCL2* over expression can block apoptosis leading to improved peripheral blood counts and prevents disease progression to AML (Slape et al., 2012). A recent study confirmed that high *FLT3* + KIT/*BCL2* ratio was an independent prognostic factor for poor event free survival (EFS) and overall survival (OS). Interrelationship rather than individual expression levels of the proliferative and antiapoptotic transcripts are critical in predicting the outcome in paediatric AML patients(Sharawat et al., 2014).MDS is closely linked to apoptosis/deregulated cell proliferation leading to peripheral blood cytopenias with paradoxical bone marrow hypercellularity. The two major pathways of apoptosis are the extrinsic pathway linked with death receptors and the intrinsic or mitochondria associated pathways. The extrinsic pathway is triggered by death receptor (For e.g. Fas, TNFR1 etc) engagement with activation of caspase-8 which in turn initiates both caspase-3 activation and the release of cytochrome c by mitochondria. Caspase-3 activation causes degradation of cellular proteins necessary for cell survival. In the intrinsic pathway regulated by the *BCL2* family of proteins, is triggered by various apoptotic stimuli to release cytochrome c from the

mitochondria independent of caspase-8 activation. Cytochrome c interacts with *APAF1* and caspase-9 to promote the activation of caspase-3 with downstream degradation of cell maintenance and cell survival proteins (reviewed in (Youle and Strasser, 2008)).

BCL2 prevents apoptosis by interacting with and inhibiting *BH123* proteins on the mitochondrial outer membrane. In the presence of an apoptotic stimulus, BH3-only proteins areactivated and bind to *BCL2* proteins so that they can no longer interact with *BH123*. bh123 proteins can then aggregate on the outer mitochondrial membrane and promote the release of intermembrane mitochondrial cytochrome c. Data gathered from gene expression study in my thesis show that *EZH2* modulated reduction of *BCL2* gene expression in K562 (chronic myeloid leukemic cell line) was mild (2 fold) but still an interesting finding because this cell line expresses the *BCR-ABL* fusion gene which is known to induce expression of *BCL2* evading apoptosis of cancerous cells. Modulation of cell proliferation, viability related genes in gene expression profiling and the clinical, prognostic importance of cytopenias in MDS and myeloid malignancies made examination of cell viability, proliferation and cell kinetics a focal point of investigation in this study. Therefore trypan blue dye exclusion method, MTT assay, Annexin V staining and cell cycle kinetics were examined for alterations in apoptosis, proliferation and cell cycle dynamics respectively. Trypan blue dye exclusion method at day 1 to day 7 after transduction and Annexin V staining at day 7 after sorting for GFP+ cells showed no alterations in the cell viability of cells transduced with shRNA against *EZH2* compared to wild type cells or cells transduced with scrambled. Similarly, cell proliferation measured by MTT assay at three time points (24, 48 and 72 hr) after sorting GFP+ K562 cells at day 7 showed minimal but not significant ($p = 0.1$ to 0.4) reduction in the proliferation rate of the cells with *EZH2* KO. In addition to examining cell viability and proliferation, cell cycle kinetics after *EZH2* and H3K27me3 down regulation was assessed due to previously reported connection of *EZH2* KO to G1 phase cell cycle arrest in ovarian cancer, delayed

G2/M phase transition in prostate cancer. Additionally, there is a close association between H3K27me3 epigenetic marks and cell cycle. The propagation of these marks to the next generation of cells takes place via cell cycle. Cell cycle kinetics examined in synchronised Hela cells indicated that formation of H3K27 methylation begins during post replication S phase of the cell cycle and gets completed during the G1 phase of the subsequent cell cycle. H3K27 methylation change within approximately 10% in abundance across cell cycle but modifications in *EZH2* levels during the phases of cell cycle have not been reported. H3K27 dimethylation occurs via two pathways; processive methylation and nonprocessive methylation. Processive methylation occurs when methyltransferase binds to unmodified substrate and carries on with the next methylation step without releasing the histone. While nonprocessive dimethylation occurs when the methyltransferase releases the histone after transferring a methyl group and binds a new substrate to form a dimethylated peptide (Zee et al., 2012). In this study 95 % inhibition of H3K27me3 marks was observed after loss of *EZH2* and the different phases of the cell cycle (sub $G_{0,}$ G0/G1, S and G2/M) were examined by staining the cells with PI (DNA content) and FITC (protein content).The percentage of apoptotic cells (sub G0) varied marginally between the wild type K562 cells (3.39 %) and those transduced with the shRNA (1.07 %) and the scrambled (1.75 %) but the difference was variable between experiments and not significant suggesting that *EZH2* and H3K27me3 inhibition did not affect the cell cycle entry of K562 cells. *EZH2* KO in MOLM13 cells showed similar results as K562 wherein in the cell cycle profile remained unaltered after 65 % inhibition of *EZH2*. Therefore it can be concluded that the role of *EZH2* in cell cycle regulation, cell viability and proliferation in myeloid cells is distinct compared to its functionality in other cancers wherein down regulation of *EZH2* had profound impact on cell proliferation, cell invasion and in some cancers was linked with G1 phase arrest or delayed G2/M phase of the cell cycle indicating an extremely cell type specific role of this gene.There

might be other mechanisms at work like a) substitution of *EZH2* in PRC2 by its paralogue *EZH1* compensating for the loss of *EZH2* b) *EZH2* dependent modulation of cell proliferation reducing (*IFI16)* and proliferation enhancing (*BCL2* and *PAG1*) genes balancing out the net effect on proliferation or c) absence of cell cycle regulator; *p16INK* a target of *EZH2* in K562 cell line (deletion of *CDKN2A* locus) because of which effect of *EZH2* loss on cell proliferation is not evident in this cell line. In MOLM13 cells, the *CDKN2A* locus is present which suggests that either *EZH2* has no effect on the cell cycle profile or else a stronger KO of *EZH2* is required to detect changes in the cell cycle profile. Unfortunately further KO of *EZH2* levels in this cell line was not achievable using either siRNA or shRNA methodologies and hence it was not possible to conclude the reason behind these findings.

5.16.3 Gene signature in patient samples

EZH2 histone methyltransferase gene is mutated in 8 % patients with MDS/MPN and 2 % in AML. These mutations are discussed in detail in the previous chapter wherein it was observed that there is no hot spot of mutation in *EZH2* associated with myeloid malignancies. In the previous chapter, R690H mutation in *EZH2* was constructed and studied. To evaluate whether this mutation mimics the gene signature obtained by loss of *EZH2* in patient samples, six patients (one with R690H and five control samples with wild type *EZH2*) were selected and qPCR was carried out on RNA extracted from these patients for expression levels of *BCL2,PAG1, IFI16, CD44, CD83, HOXA10, RAP1B, FLT1, TSLP*, *EZH1* and *CDKN2A*. The selected patients were age and AML evolution matched. Statistical tests to calculate significance (p value) could not be carried out for this experiment since the sample size in each category was very small (number of patients with R690H mutant was one). Five genes *HOXA10*, *FLT1*, *PAG1B*, *EZH1* and *TLSP* were up regulated in the patient with R690H mutant compared to the control patients while the other genes showed no change in their

expression levels. **Interestingly, these results contrast with findings (except for *EZH1*) observed with inhibition of *EZH2* expression using shRNA in lentivirus where *HOXA10*, *FLT1*, *PAG1B* and *TLSP* were all down regulated. This incongruity in the gene signature between patient sample with R690H mutations and K562 cell line with 95 % KO of *EZH2* suggest that *EZH2* mutation does not mimic loss of function of the gene.**

In contrast to other genes, *EZH1* mRNA levels remained unchanged with loss of *EZH2* in K562 cells while they were elevated in patients carrying *EZH2* R690H mutation. *EZH1* mRNA levels have not been examined in the context of *EZH2* mutations before. Nikosolski et al (2010) examined the effect of R690H mutation introduced in Sf9 insect cells on the levels of *SUZ12*, *EED* and found no change in their expression. But concurrent *EZH1* levels were not evaluated (Nikoloski et al., 2010). Up regulation of *EZH1* in this study indicate that *EZH1* may compensate for loss *EZH2* function. However decrease in H3K27me3 levels have been reported with *EZH2* mutations and were also observed in this study after decrease in *EZH2* levels. This study suggests a different mechanism of transcriptional regulation of *EZH1* which is distinct from its H3K27 trimethylation activity. Immunoprecipitation studies of flag tagged *EZH1* in Sf9 cells demonstrated that *EZH1* forms part of the PRC2 complex but KO studies of *EZH1* indicate that it does not affect H3K27me3 levels as drastically as loss of *EZH2*. In fact *EZH1*-PRC2 H3K27 methyltransferase activity was 20 fold less then *EZH2*-PRC2. Analysis of mouse kidney cells illustrates that *EZH1*-PRC2 compacts chromatin in the absence of the methyltransferase cofactor SAM indicating that H3K27me3 is not the main target of *EZH1*-PRC2 (Margueron et al., 2008). These findings are in agreement with the data presented here wherein *EZH1* is not associated with H3K27me3 levels while *EZH2* expression is closely linked to this mark. Western blot analysis of the patient samples was carried out to confirm the status of H3K27me3 with increased *EZH1*

expression but the results on several attempts were inconclusive (cellular proteins were degraded).

5.17 Conclusion

Loss of 95 % of *EZH2* expression in myeloid cell line K562 (TP53 null) did not alter cell viability, cell proliferation or cell cycle kinetics. Observations in MOLM13 cell line (TP53 wild type) were similar to findings in K562 cells in the context of cell cycle kinetics. Gene signature obtained by microarray gene expression profiling demonstrated down regulation of genes involved in (a) Apoptosis and proliferation: *BCL2, PAG1, IFI16* (b) Immune modulation: *CD44, CD83, TLSP* (c) Cell differentiation: *HOXA10* and *FLT1*. Qpcr analysis of this gene signature in patient samples (R690H versus *EZH2* wild type) show up regulation instead of down regulation of *HOXA10*, *FLT1*, *PAG1B*, *EZH1* and *TLSP* genes confirming that *EZH2* KO does not mimic *EZH2* mutations. Assessment of cell viability and proliferation showed no alteration in cells with 95% inhibition of *EZH2* compared to the scrambled shRNA suggesting the role of compensatory mechanism. Up regulation of *EZH1* in patient with R690H suggest substitution of *EZH1* to compensate for mutation in *EZH2* causing reduction in H3K27me3.

6 Chapter: Role of DNMT3A (R882H) mutation in Myeloid malignancies

6.1 Introduction

DNMT3A mutations have been detailed in chapter 3 in the context of their incidence in MDS/AML, the mutational hotspot R882, prognostic value of these mutations and their correlation to other genes mutated in MDS/AML.Gene expression profiling of AML patients with/without *DNMT3A* mutations revealed that patients carrying the R882 mutation showed under expression of *DNMT3A* suggesting that the mutation knocks down the enzymatic activity of the gene (Huang et al., 2013). However, this mutation has not been studied in detail in the context of MDS/AML.

6.2 Aim

- To construct the R882H mutation commonly seen patients with *DNMT3A* mutation in p3XFLAG-myc-CMV26 vector.
- To transfect and examine the effect of this mutation in myeloid cell line.

6.3 Constructing the R88H mutant in pCMV sport6 vector

Wild type *DNMT3A* cDNA cloned in pCMV-SPORT6 (IRATp970A0473D) was obtained from Source Bioscience and was used for all experiments (refer appendix).The *DNMT3A* cDNA sequence was verified using different primer pairs (refer appendix) by Sanger sequencing.

To introduce the mutation in the wild type *DNMT3A* cDNA (Fig: 6.1), the plasmid was digested by NruI-NotI restriction enzymes (refer section 2.14.1) which generated two fragments (size: 1800 & 8000 bp) as visualised on 1 % agarose gel. The large fragment (8000 bp) was isolated under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions (refer section 2.14.3) while the small fragment was discarded.

Secondly, pCMV-SPORT6 was digested with XbaI-NotI restriction enzymes to generate two fragments (size: 8000 & 600 bp) (Fig: 6.2) of which the small fragment was gel purified while the large fragment was discarded.

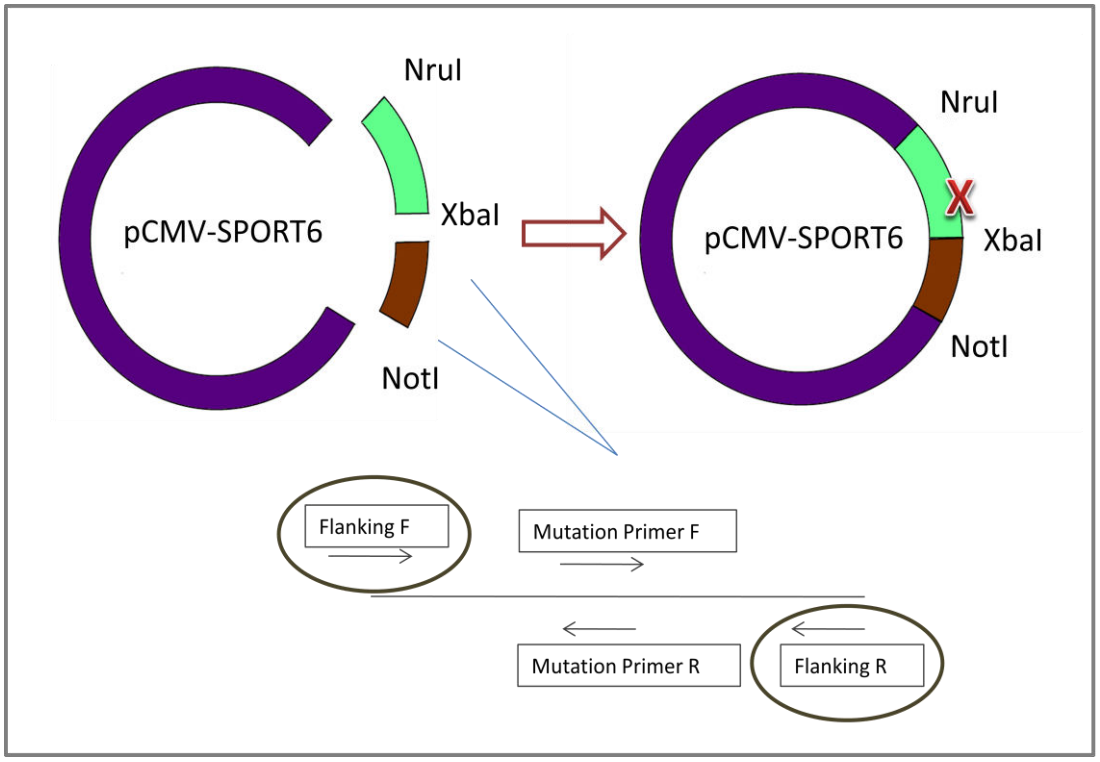


Figure 6.1: Cloning strategy for introducing the R882H mutation in pCMV-SPORT6. Generation of the mutant in the NruI-XbaI fragment has been illustrated using two pairs of primers (of each pair; one primer flanked the mutational site and a second primer overlapped the mutation site and introduced the mutation) using a PCR based approach. Red cross indicates the site of the mutation in pCMV-SPORT6 vector.

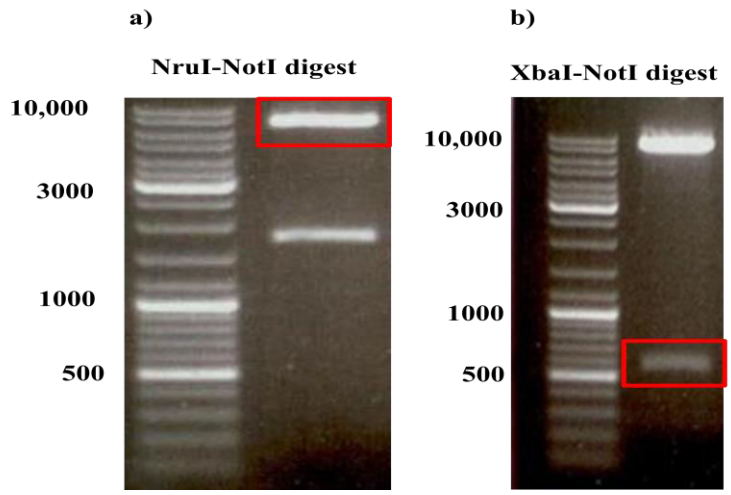


Figure 6.2: Digestion of pCMV-SPORT6 vector containing *DNMT3A* cDNA.(a) pCMV-SPORT6 was digested by NruI-NotI restriction enzymes to generate the empty vector (8000 bp).(b) pCMV-SPORT6 was digested by XbaI-NotI restriction enzymes to generate a fragment of the *DNMT3A* cDNA (size: 600 bp). These framents were gel purified and used for ligation later.

6.3.1 Generating the fragment containing the R882 mutation

The mutation was generated using two sets of primers; the first pair of primers overlapped the position 882 [mutational primer F (MF) and mutational primer reverse (MR)]. The change in the base pair to be introduced was positioned exactly at the centre of the primer. The second pair of primer or flanking primers [flanking forward (FF) and flanking reverse (FR)] were designed just outside XbaI-NruI restriction enzyme sites. PCR amplification using these primers was carried out on wild type pCMV-SPORT6 vector at a range of temperatures (52 $^{\circ}$C, 56°C & 60°C) [Fig: 6.3].

a)

Reagent	Volume
Promega GoTaq (colourless)	10 µl
Primer (5 µM)	0.5 µl FF + 0.5 µl MR OR 0.5 µl FR + 0.5 µl MF
EZH2 cDNA plasmid (10 ng/ µl)	1 µl
H_20	8 µl
Total	20 µl

b)

Cycle	Temp	Time	Comment
1	95°C	2 min	Enzyme Activation
35	95°C	30 sec	Denaturation
	52,56,60°C	30 sec	Primer annealing
	72°C	45 sec	Elongation
1	72°C	5 min	Fill missing nucleotides

c)

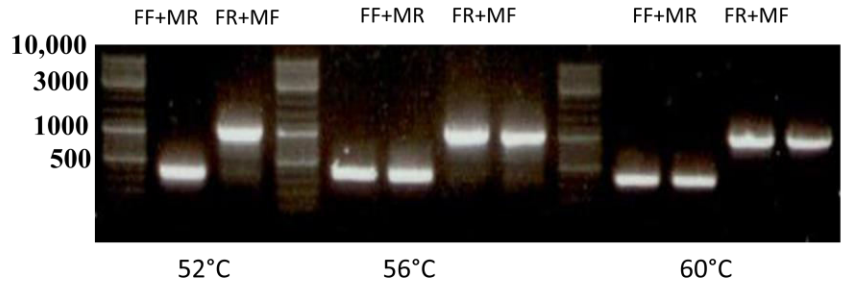


Figure 6.3: PCR based generation of the mutational (R882H) fragment in *DNMT3A* cDNA (a) Table indicates reaction mix for pcr containing either flanking primer forward plus mutational primer reverse or flanking primer reverse plus mutational primer forward (b) Illustrates PCR conditions on the thermocycler to generate these fragments (c) Visualisation of the PCR products at a range of temperatures on 1 % agarose gel. Products obtained at 60 ˚C were used for a second PCR reaction. O'Gene Ruler DNA ladder (Fermentas, UK) was used to evaluate correct size of the product.

The pcr product was analysed on 1% TAE agarose gel alongside O'Gene Ruler DNA ladder (Fermentas, UK). Samples were electrophoresed at 120V for 40 min in 1X TAE buffer. The bands were isolated under UV light and purified by Qiagen gel purification kit as per the manufacturer's instructions. The products from the first round PCR (60 ˚C) were used for a second PCR using only the flanking primers and the elongation time of the PCR reaction was increased from 30 to 45 seconds due to the large size of the fragment (1200 bp). Due to the extensive overlap of the products from the above PCR reaction the products will anneal together and will be amplified by the flanking primers. Four different temperatures (gradient

PCR) were used (51°C, 55°C, 57°C & 60°C). The rest of the PCR conditions were kept constant:

a)

Reagents	Volume
Promega Go Taq (Colourless)	10 µl
Product (FF+MR)	10 µl
Product (FR+MF)	5 µl
Primer Flanking forward	0.5 µl
Primer Flanking reverse	0.5 µl
H_20	4.0 µl
Total	**30 µl**

b)

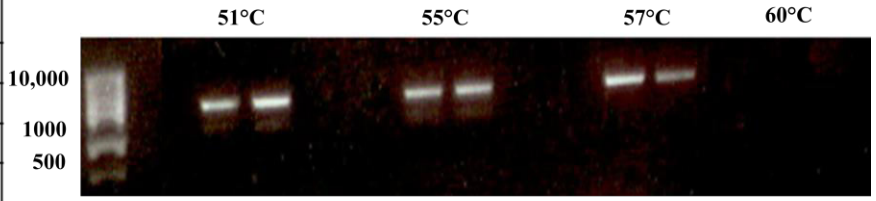


Figure 6.4: Second PCR to generate the mutational (R882H) fragment in *DNMT3A* cDNA(a) Illustrates reaction mix for second round of pcr using the product from the first PCR reaction. PCR amplification in this instance was carried out using only the flanking primers. (b) Visualisation of the PCR product (size: 1200 bp) on 1 % agarose gel.

The product from the second PCR (57°C) was gel purified as before and the fragments were treated with SuperTaq and incubated for 30 min at 72 °C to add dNTP "A" to the fragment.

Reagents	Volume
Plasmid DNA (37 ng/ µl)	10 µl
dNTPs (2 mM)	1µl
Super Taq	0.25 µl
10X buffer	2 µl
H_20	upto 20 µl

Table 6.1: Reaction mix for treatment of plasmid DNA with SuperTaq to add dNTP "A" to the fragment. The reaction mix was incubated for 30 min at 72°C.

Subsequently, 2 µl of the above the mixture was cloned in to the pCRII- TOPO vector and transformed into DH5α-T1 *E.Coli* cells (refer section 2.14.5). After overnight incubation of

the LB agar + 100mg/ml ampicillin +Xgal (20 µg/ml) plates, four colonies were selected and miniprep was carried out using the Wizard Plus SV kit (refer section 2.14.9). Colonies with the correct band size (1200 bp) were verified by digesting the plasmid with XbaI-NruI restriction enzymes (refer method) and sanger sequencing was used to confirm the presence of the R882H mutation (data not shown).

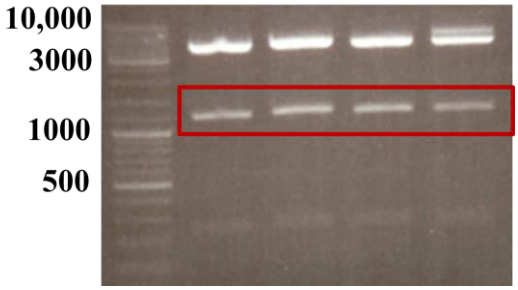


Figure 6.5 Visualisation of four colonies on 1 % agarose gel aftertreatment with XbaI-NruI restriction enzymes. The verification was carried out after Topo cloning of the PCR fragment. All four colonies showed the correct band size (1200 bp)

6.3.2 Ligation of mutation into pCMV-SPORT6

The concentration of plasmid DNA for all three fragments generated by restriction enzymes; NruI-NotI, XbaI-NotI and the fragment containing the mutation (XbaI-NruI) were measured using Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer and ligated using T4 DNA ligase. The control reaction contained only the empty vector but none of the other fragments. This control was used to check the number of background non-specific colonies. The ligation was carried out for 2-6 hr at room temperature.

Reagents	Volume
Empty vector (NruI-NotI) [21.66 ng/ µl]	2.31 µl
XbaI-NotI (12.35 ng/ µl)	1.3 µl
XbaI-NruI (fragment containing R882H mutation) [7.65 ng/ µl]	4.4 µl
10X Ligase buffer	1 µl
T4 DNA ligase	0.5 µl
H_20	0.49 µl
Total volume	10 µl

Table 6.2: Ligation mixture for three fragments (NruI-XbaI, XbaI-NotI&NruI-NotI) of pCMV-SPORT6. Vector :Insert ratio of 1:2 was used for the ligation.

0.2 µ1 of the ligation mixture was transformed into MegaX DH10B T1 Electrocomp cells (refer section 2.14.7). No colonies were observed after incubating the LB agar + 100 mg/ml ampicilin plates for 16 hr at 37°C. Therefore the ligation reaction was repeated and the duration of ligation was increased to over night (ON) at 4 °C. As done previously the plasmid was transformed into MegaX DH10B T1 Electrocomp cells and eight colonies were selected after 16 hr incubation at 37°C. Accurate ligation of all the fragments was confirmed by digesting the plasmid with NruI-XbaI (Fig: 6.6a) restriction enzymes and the R882H mutation was verified by sanger sequencing using M13 primers and primers designed against *DNMT3A* cDNA (refer appendix). The first colony was selected based on results obtained by sanger sequencing (Fig: 6.6b), digestion by restriction enzyme results and maxiprep was carried out using the Qiagen Maxi Prep plasmid kit (refer section 2.14.10).

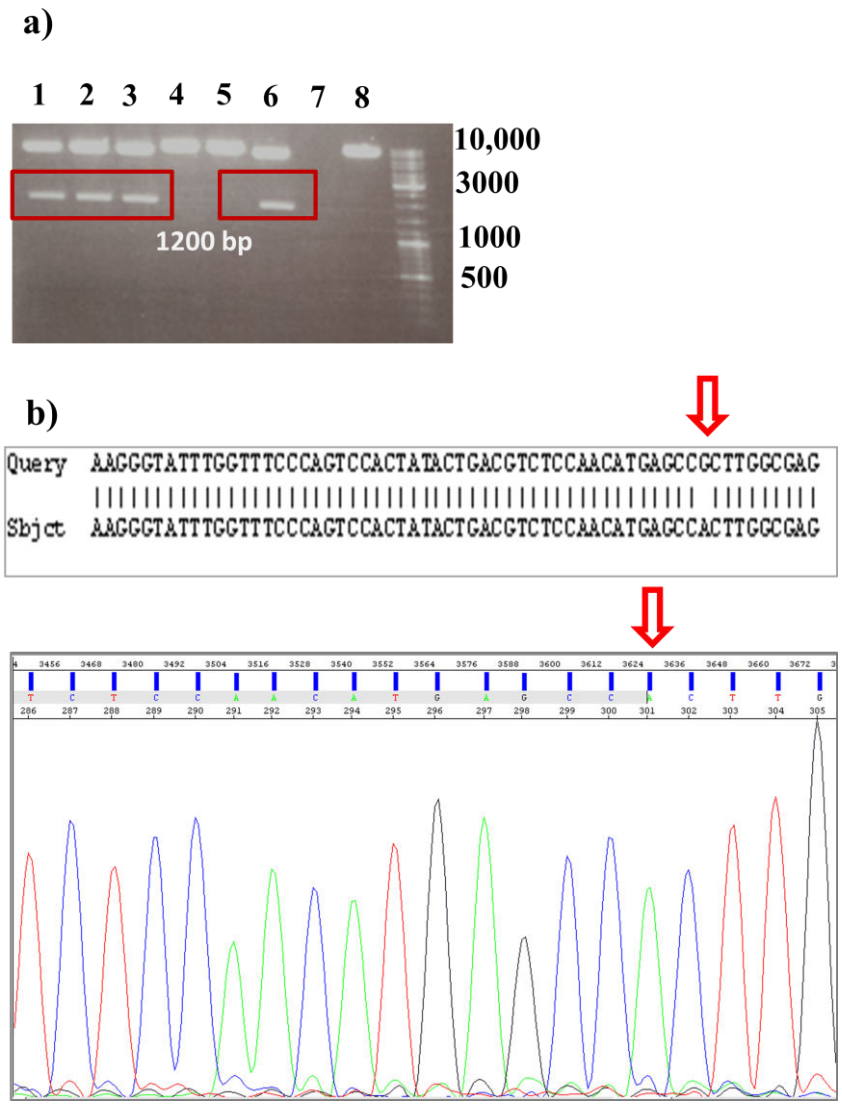


Figure 6.6: Confirmation of the presence of R882H mutant in pCMV-SPORT6 vector using sanger sequencing. (a) Digestion of eight colonies by NruI-XbaI restriction enzymes. Four of the colonies (1, 2, 3 & 6) indicate correct band size (1200 bp). (b) Sanger sequencing of colony 1 illustrated the change in amino acid CGC (Arginine) to CAC (Histidine) confirming the presence of the R882 mutation in the colony.

6.3.3 Cloning R882H mutation into p3XFLAG-myc-CMV-26 expression vector

To insert a selectable marker (neomycin) and to attach a Flag tag to the R882H *DNMT3A* mutant to distinguish it from endogenous *DNMT3A* in cells post transfection, *DNMT3A* cDNA from pCMV-SPORT6 vector was cloned into p3XFLAG-myc-CMV-26 vector (Sigma

Aldrich, UK). To achieve this, two primers incorporating the SalI (GTCGAC) and HindIII (AAGCTT) restriction sites were designed:

Primer name	Sequence	Function
*DNMT3A*_flag_F	5' - AAGTCG ACAAGC TTA TGC CCG CCA TGC CCT CC - 3'	To get the gene in frame by removing the 5' UTR before start codon
*DNMT3A*_flag_R	5' - CTT TGG TGG CAT TCT TGT CC - 3'	To get the gene in frame by removing the 5' UTR before start codon

PCR using this primer pair was carried out on wild type *DNMT3A* cDNA in pCMV Sport6. The PCR product was visualized on 1 % agarose gel as seen in Fig: 6.7a and purified as mentioned previously.

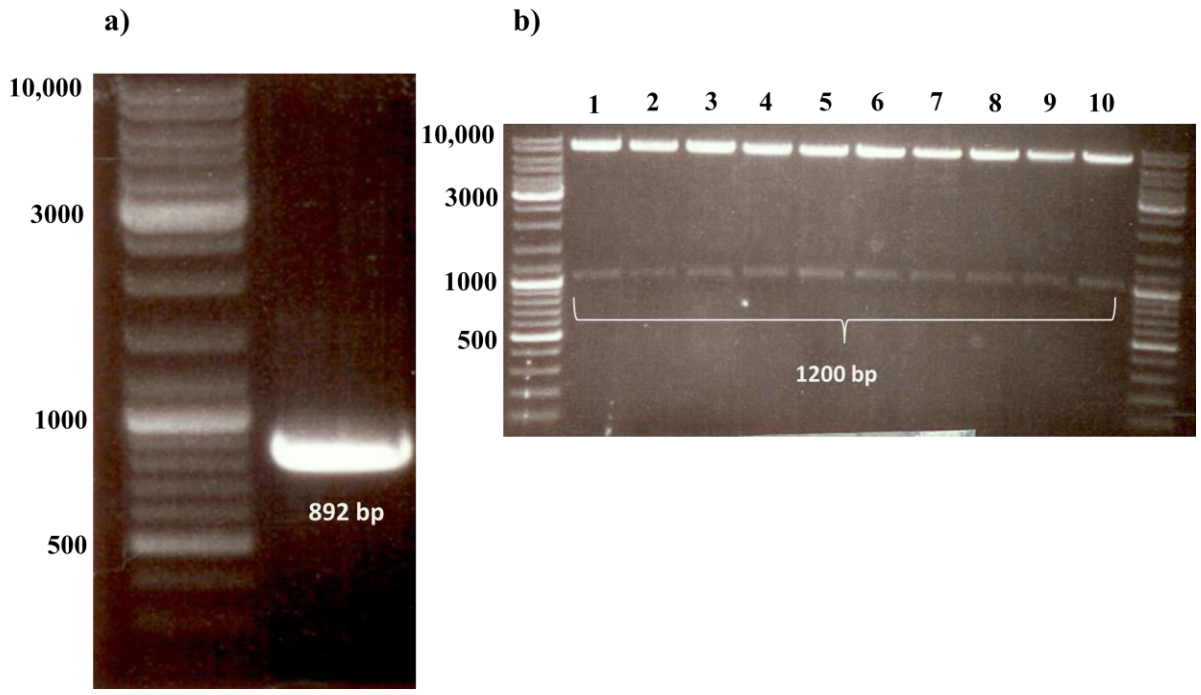


Figure 6.7: Cloning the R882H *DNMT3A* mutation into p3XFLAG-myc-CMV-26 expression vector. (a) PCR fragment using the primers designed to introduce SalI (GTCGAC) and HindIII (AAGCTT) enzyme sites in the sequence. (b) Confirmation of accurate colonies obtained after TOPO cloning the PCR fragment by digesting the plasmid by SalI-Bpu10I restriction enzyme to generate the correct band (size: 1200 bp).

The gel purified product was treated with SuperTaq, cloned into pCRII- TOPO vector and transformed into DH5α-T1 *E.Coli*. Ten colonies were selected and the correct colony was verified by digesting the plasmid with SaI-Bpu10I (Fig: 6.7b) restriction enzymes. The fragment (size: 1200 bp) was gel purified and concentration of the plasmid DNA was measured using Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer. Secondly, pCMV-SPORT6 plasmid containing the R882H mutant was digested using SalI-Bpu10I (Fig: 6.8) restriction enzymes and the large fragment (size: 8000 bp) was gel purified. The two fragments; size 1200 bp and 8000 bp were ligated using T4 DNA ligase as mentioned previously and transformed into DH5α-T1 *E.Coli*.

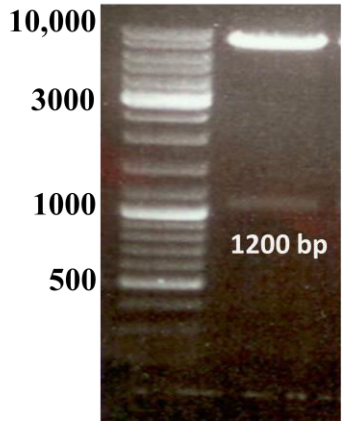


Figure 6.8: Digestion of pCMV-SPORT6 containing R882H mutant by SalI-Bpu10I restriction enzymes.

However, no colonies were observed post ligation. The ligation reaction was repeated several times by increasing the concentration of both the fragments and increasing the duration of ligation. The plasmid DNA obtained from the colonies after each attempt was digested using HindIII-NotI (Fig: 6.9) but a single band was obtained indicating absence of the smaller fragment.

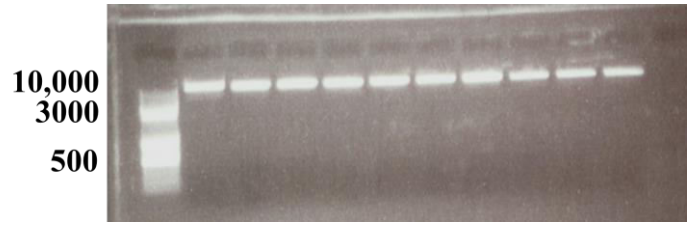


Figure 6.9: Digestion of colonies obtained after ligation of inframe *DNMT3A* cDNA into pCMV-SPORT6 expression vector by HindIII-NotI restriction enzymes. The ligation failed since the band indicating the presence of the insert was not obtained.

Therefore cloning of the PCR product (size: 892 bp) into the pCRII- TOPO vector and transformation into DH5α-T1 *E.Coli* was repeated. Five colonies were selected, plasmid DNA digested by EcoRI restriction enzyme (Fig: 6.10a) showed band of the correct size; 892 bp in colony 1, 4 and 5. Sequencing these colonies confirmed the presence of the insert (data not shown).

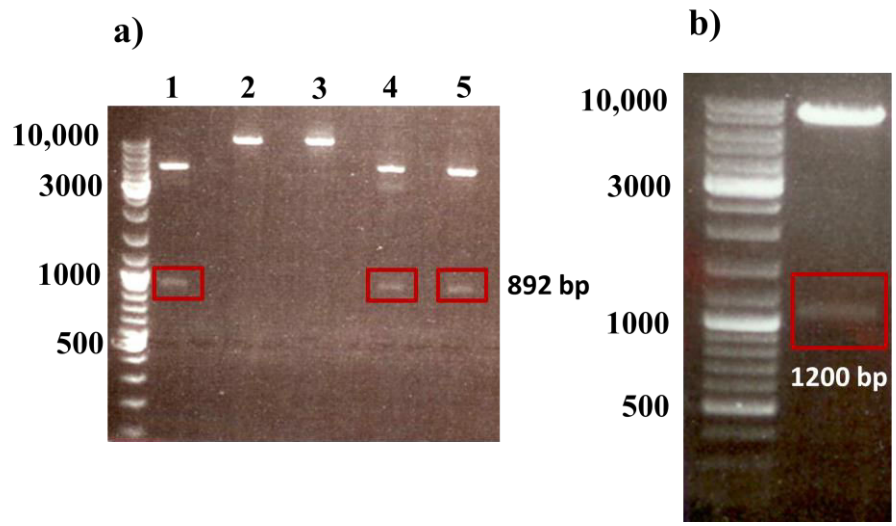


Figure 6.10: Second attempt to ligate *DNMT3A* into p3XFLAG-myc-CMV-26 plasmid.(a) The colonies were digested using EcoRI restriction enzyme to confirm the presence of the insert in five colonies after topo cloning. Three colonies (1, 4 and 5) showed the correct size bands (red boxes) while colony 2& 3 were false positives. (b) One of the positive colonies (1, 4 or 5) was digested by SalI-Bpu10I and 1200 bp fragment was used for ligation.

The plasmid DNA from the colonies was digested using SalI-Bpu10I (Fig: 6.10b) and the band at 1200 bp was gel purified. Ligation of the two bands (size: 1200 and 8000 bp) was repeated as before and six colonies were selected. Plasmid DNA after miniprep from these colonies was digested first by ScaI-NotI followed by HindIII partial digest (Fig: 6.11).Three out of six colonies generated bands of the correct size; 1200, 3000 & 4300 bp. The fragment at size 4300 bp (*DNMT3A* gene) was gel purified.

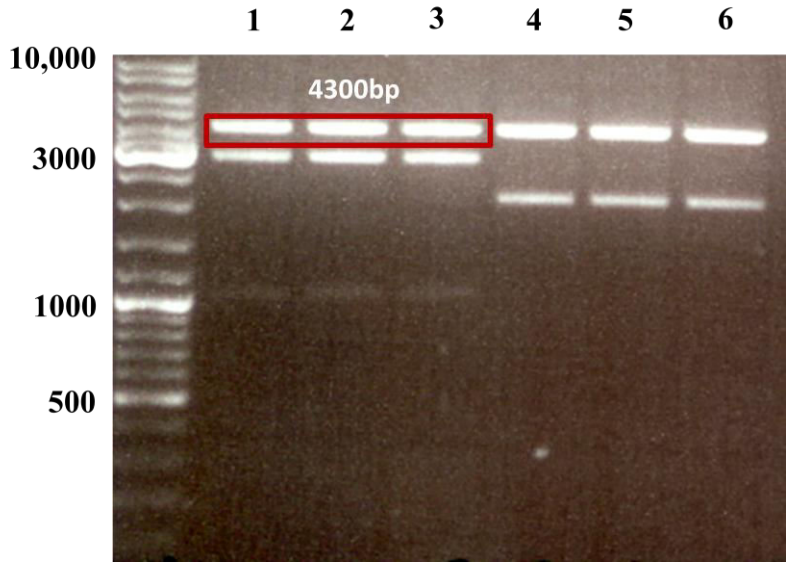


Figure 6.11: Plasmid DNA from six colonies was digested by ScaI-NotI restriction enzymes followed by partial digest by HindIII. Partial digest by HindIII enzyme:Five eppendorfs containing all the reagents required for restriction digest except the enzyme were prepared. A separate eppendorf with the same reagents without plasmid DNA was prepared and all the contents from this were mixed into the first of the five eppendorfs. 20 µl of this mix was added to the second eppendorf and so on making a serial dilution of the enzyme. After adding the mix to the fifth eppendorf, 20 µl was taken and added to a clean eppendorf. All the mixtures were incubated for 3 min at 37°C and then at 65°C for 20 min to inactivate the enzyme.

6.3.4 Ligation of DNMT3A cDNA and p3XFLAG-myc-CMV

The p3XFLAG-myc-CMV vector was linearlized by digesting the plasmid with HindIII-NotI restriction enzymes and gel purifying the fragment.

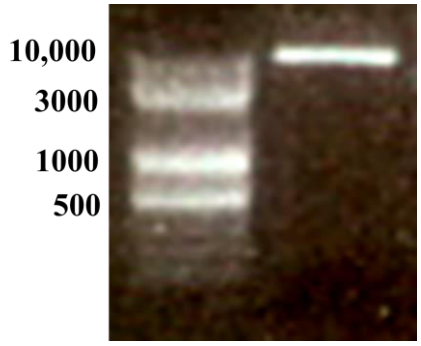


Figure 6.12: Digestion of p3XFLAG-myc-CMV-26 plasmid by HindIII-NotI restriction enzymes.

The linear plasmid was ligated with *DNMT3A* cDNA generated above (size: 4300 bp) using T4 DNA ligase as mentioned previously but the fragments failed to ligate. The ligation reaction was optimised by increasing the DNA concentrations of both the insert and the vector, increasing the duration of ligation from 3-4 hr at room temperature to ON at 4°C and treating the fragment after HindIII-NotI digest with CIP (0.1 µl, incubate 37°C for 30 min) to prevent self-ligation of the digested ends. But none of these modifications resulted in successful ligation of *DNMT3A* into p3XFLAG-myc-CMV vector and therefore, the ligation strategy was reconsidered with a view to simplify the procedure by ligating fragments of smaller size.

6.3.5 Ligation of Flag tag into pCMV-SPORT6

A different approach was used wherein; the Flag sequence was removed from p3XFlag-myc-CMV vector and ligated into pCMV-SPORT6 containing R882H mutation (Fig: 6.13). 2 µg of p3XFLAG-myc-CMV vector was digested using SnabI-HindIII restriction enzymes and

visualised on 1 % agarose gel (Fig: 6.14a). The fragment at 320 bp was purified and the DNA concentration was measured using Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer.

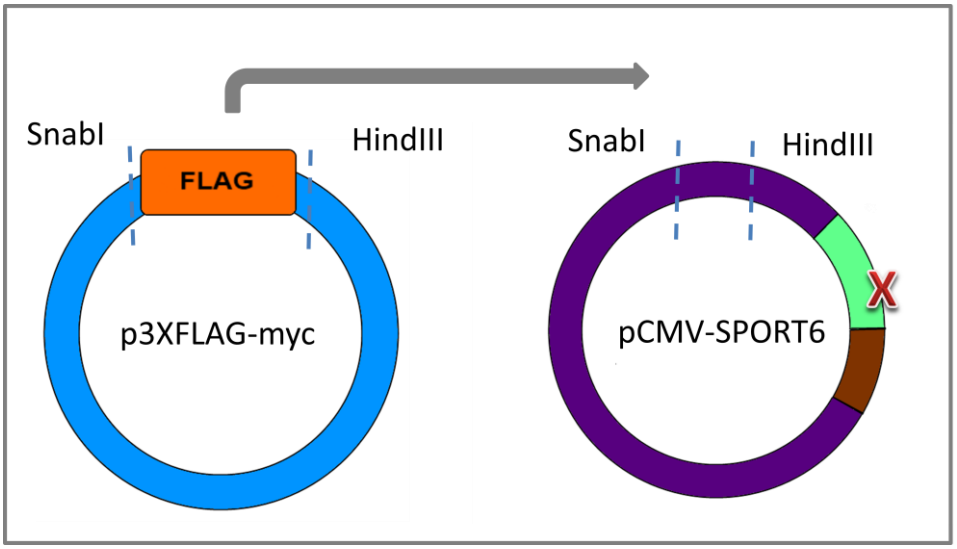


Figure 6.13: Cloning strategy for inserting the FLAG tag from pX3FLAG-myc into pCMV-SPORT6.

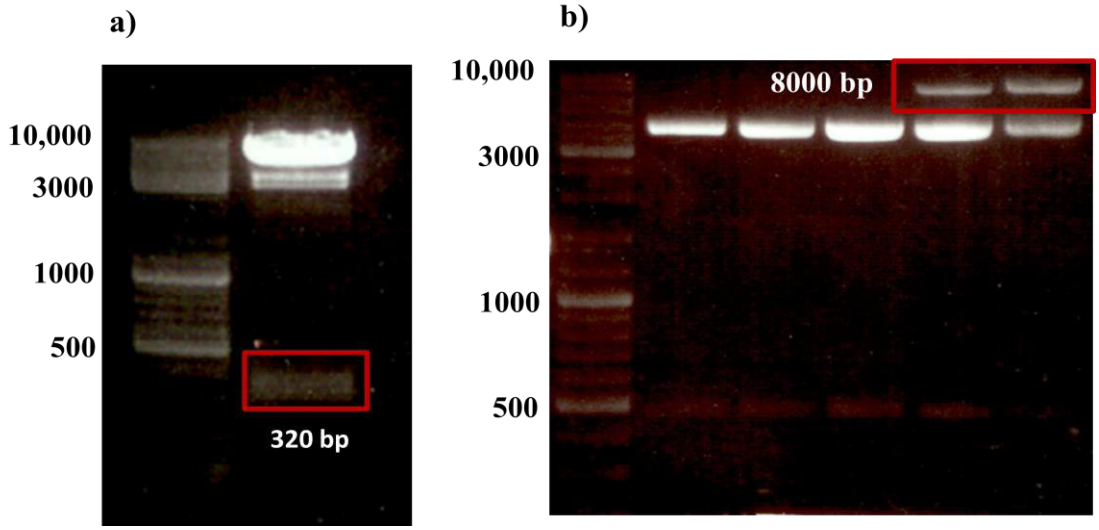


Figure 6.14: (a) Digestion of p3XFLAG-myc-CMV by SnabI-HindIII restriction enzymes (b) SnabI digest followed by HindIII partial digest on pCMV-SPORT6.

Secondly, R882H in pCMV-SPORT6 was digested using first SnabI for 1 hr at 37°C followed by a partial digest by HindIII (Fig: 6.14b) to generate a fragment of size 8000 bp. This fragment was purified, quantified and used for ligation. Two fragments (320 & 8000 bp) were ligated using T4 DNA ligase as previously described and transformed into DH5α-T1 *E.Coli* and the colonies were verified using sanger sequencing. Similar to R882H, wild type *DNMT3A* cDNA was also ligated with Flag tag to study the effect of over expressing the wild type gene. Empty pCMV-SPORT6 vector without cDNA of *DNMT3A* was used as a control for all experiments.

6.4 Transfection of R882H-mutant into myeloid cell lines

MOLM13 and KG1 (acute myelogenous leukaemia) cell lines were used for all experiments. The reasons for using MOLM13 cells were described in the previous chapter. The second cell line; KG1 was utilized because it contains wild type *DNMT3A* (Tiacci et al., 2012) and the expression levels of *DNMT3A* were easily detected by qPCR facilitating easy monitoring of variations in *DNMT3A* gene expression due to the R882H mutation.

6.4.1 Electroporation

pCMV-SPORT6 does not contain a mammalian selectable marker therefore the plasmid was transfected along with pBABE-puro (gifted by Dr. Joop Gaken) which contains puromycin. The experimental workflow is detailed in figure 6.15. 95% of the MOLM13and 65% KG1 cells were dead at 48 hr as estimated by Trypan blue staining which could be attributed to electroporation since similar percentage of cell death was also observed in the mock-transfected cells. Therefore Lipofectamine LTX was used for all subsequent transfections.

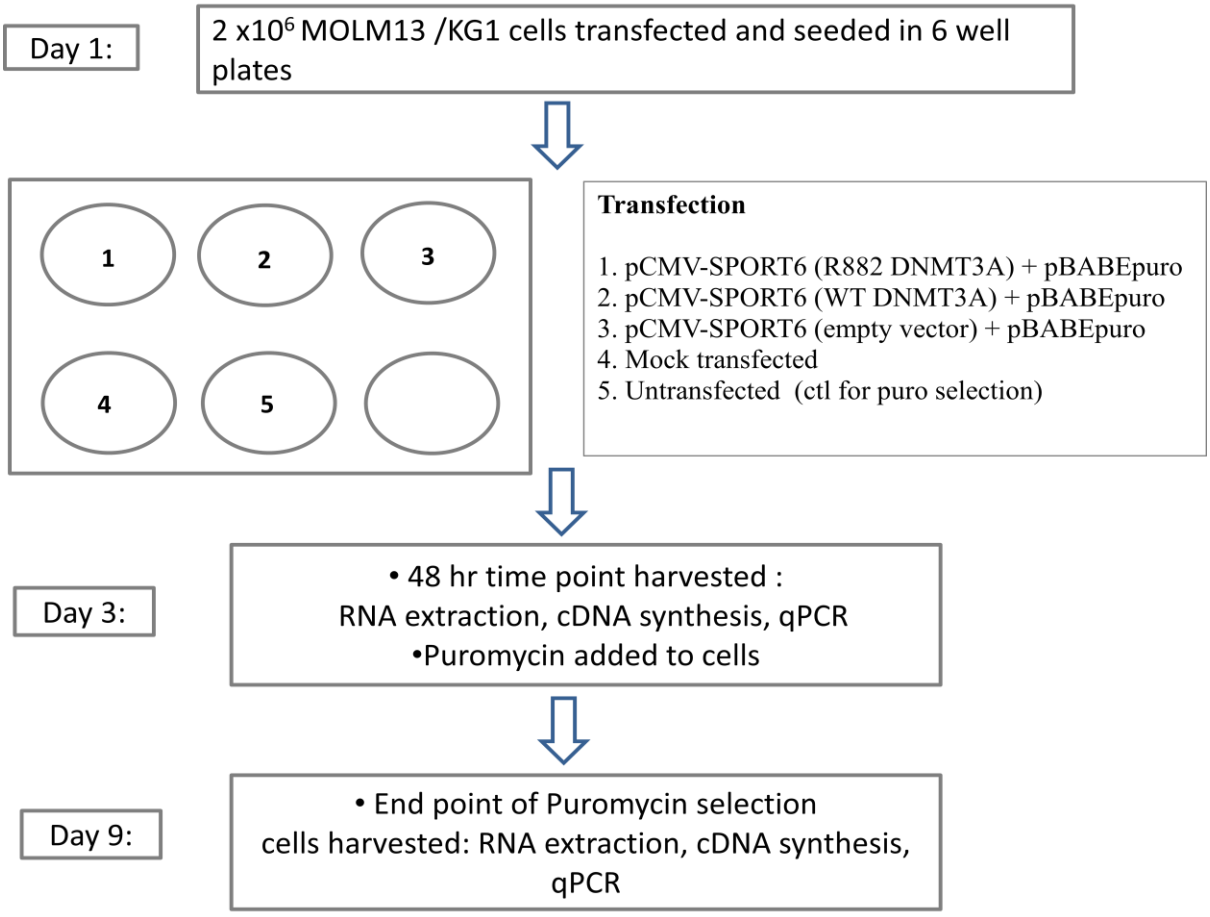


Figure 6.15: Experimental workflow for transfection by electroporation. MOLM13/Kg1 cells were seeded at a density of 2 x10^6 in 6 well plates and transfected with 1.8 µg of pCMV-SPORT6 R882H + 0.2 µg of pBABE-puro vector per well resuspened in 100 µl of Mirus BioIngenio electroporation solution using the Nucleofactor device II (program: O-017 for MOLM13 and V-0001 for KG1). Wild type *DNMT3A* and empty pCMV-SPORT6 vector + pBABE-puro were transfected in separate wells. Mock transfected cells were used as ctl for alterations caused due to the transfection procedure. Untransfected cells were used as a ctl to estimate the end point of puromycin selection for the cells.

6.4.2 Lipofectamine LTX reagent

The transfection methodology was changed to Lipofectamine LTX while other experimental conditions were kept constant. The transfection protocol was optimised using 24, 12 and 6 well plates as (http://www.lifetechnologies.com/uk/en/home/references/protocols.htm and combining the transfection protcol for K562 suspension cell line) as per manufacturer's instructions. The transfected cells are selected by puromycin at a concentration at 0.2 µg /ml

for MOLM13 cells and 2 µg /ml for KG1 cells for six days. *DNMT3A* expression was measured by qPCR at 48 hr and day 9 after transfection (Fig: 6.16).

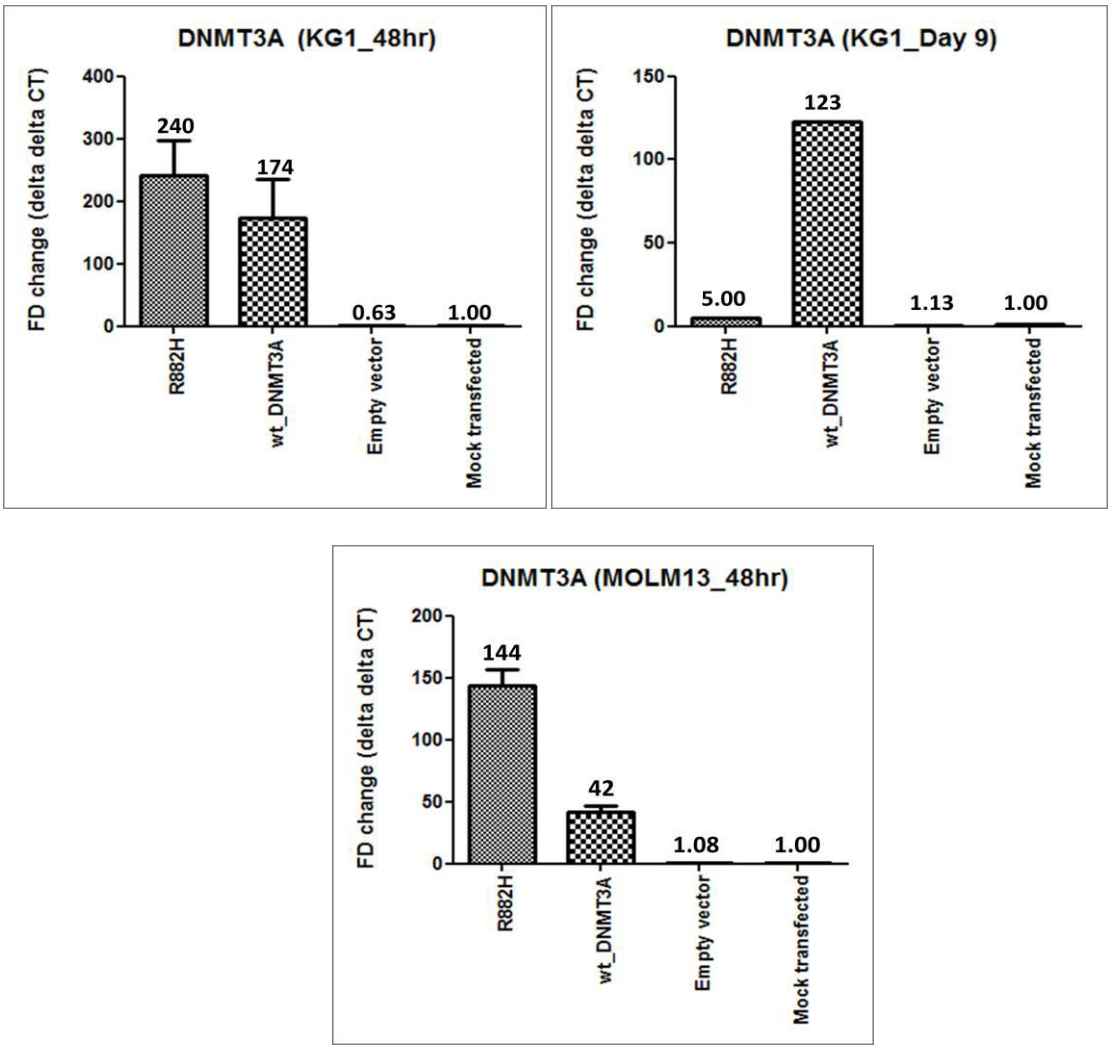


Figure 6.16: *DNMT3A* expression in MOLM13 (48hr post transfection) and in KG1(48hr/ Day 9 post transfection) cell lines.MOLM13 cells did not survive the puromycin selection and therefore the day 9 qPCR profile for this cell line is not available. Both the cell lines were seeded at a density of 2 x 10^5 - 5 x 10^5cells/well and transfected with plasmid DNA pre-incubated with Lipofectamine LTX (4µl-12µl) and Plus reagent (1µl-2.5µl) according to the manufacturer's instructions. Data are represented as mean ± SEM; n = 3.

Post-selection *DNMT3A* gene expression was highly variable between experiments in KG1 cell line (a example is shown in figure 6.17). Despite conducting multiple experiments the expression of either wild type *DNMT3A* or R882H was low (below the *DNMT3A* expression level in the untransfected cells). It is unclear whether the CMV promoter driving gene transcription of *DNMT3A* in the plasmid was switched off or other biological compensatory mechanisms were at work.The time taken to conduct these experiments and the degree of uncertainty involved was a concern. Having already spent considerable time on cloning and transfecting the *DNMT3A* mutation, after careful evaluation it was decided to discontinue work on this mutation.

6.5 Discussion

Studying the functional consequences of the R882H mutation has been attempted by various research groups including our group in this study. Large size of the *DNMT3A* gene provided numerous technical difficulties in cloning this gene into p3XFLAG-myc-CMV-26 vector to allow for easy detection of the FLAG tagged transgene expression in the transfected cell lines. Ligation was attempted many times by varying the concentration of vector: insert ratio, duration of the ligation (2-6 hr at RT to over night at 4°C) and treating the fragment with CIP (0.1 µl, incubate 37°C for 30 min) to prevent self-ligation of the digested ends and reduce the background. However, all the strategies to clone cDNA of *DNMT3A* into p3XFLAG-myc-CMV-26 were unsuccessful and therefore the FLAG tag from p3XFLAG-myc-CMV-26 was isolated and ligated into pCMV-SPORT6. This plasmid was transfected along with pBABE (containing puromycin marker) into myeloid cell lines but extremely poor transfection efficiency was observed in both MOLM13 and KG1 cells (more so in MOLM13 cells) resulting in long recovery period of the cells post puromycin selection. This resulted in variable expression levels (lower than mock transfected in some instances) of *DNMT3A*

mRNA. CMV promoter drives the expression of *DNMT3A* in pCMV-SPORT6 vector and as discussed in chapter 4, inactivity of the CMV promoter due to methylation of CpG islands in its promoter could be the cause for the variable expression of *DNMT3A*. However, due to insufficient time, this hypothesis could be tested and therefore work on R882H *DNMT3A* mutation was discontinued.

6.5.1 Literature survey

Ley et al (2010) first examined global methylation levels in *DNMT3A*-mutant and wild-type AMLs using liquid chromatography-tandem mass spectrometry (LC-MS) and found no difference in global 5-mC levels in patient samples. DNA methylation profiling in R882H-mutant AMLs using MeDIP-chip showed 182 regions that were differentially methylated between R882H versus wild-type AMLs, all of which were hypomethylated in the patients carrying R882H (Ley et al., 2010). Similar findings were reported by Yan et al (2011) and Kim et al (2013) when they examined contribution of aberrant *DNMT3A* activity to the pathogenesis of acute monocyticleukemia (Yan et al., 2011) and the in vivo effects of R878H mutation in murine (equivalent to R882H in humans) embryonic stem cells and found reduced global methylation as compared to wild type *DNMT3A* respectively (Kim et al., 2013).In addition to changes in global DNA methylation patterns, researchers have also examined gene expression patterns of R882H mutant versus wild type *DNMT3A*. These studies have revealed1,323 up regulated and 563 down regulated genes in patients carrying the R882 mutant. Among the genes with altered expression, *IDH1* (Yan et al., 2011) and *RUNX1* (Challen et al, 2012) have been linked to MDS/AML previously. Rest of the genes were involved either in cell adhesion, proliferation & migration. However an obvious patient cluster clearly linking to mutational status of the gene is not evident. microRNA profiling, on younger and older patients has been carried out which did not reveal a gene signature

associated with *DNMT3A* mutations in the younger group. In contrast, a signature consisting of 12 microRNAs associated with *DNMT3A* mutations was evident in the older patients (Marcucci et al., 2012).

In vivo research on loss of *DNMT3A* on HSC function suggests that it has no effect no HSC differentiation potential (Tadokoro et al., 2007). Recently, Challen et al (2012) looked at *DNMT3A* knockout mouse model combined with serial transplantation assays and found an increased expansion of the HSC compartment in the serial transplants with impairment of HSC differentiation over serial transplantation but an overt phenotype was not observed and the mice did not develop spontaneous leukaemia (Challen et al., 2012). In contrast, Xu et al (2014) examined the transformation ability of the R882H mutant on murine hematopoietic stem/progenitor cells (HSPCs) and found that the mutant gene can induce aberrant proliferation of hematopoietic stem/progenitor cells. At 12 month post bone marrow transplant, all mice developed chronic myelomonocyticleukaemia with thrombocytosis (Xu et al., 2014).

Structural studies on interactions of the mutant protein with wild type *DNMT3A* to understand their differences in functionality, supports the theory of dominant-negative effect of mutant. The interaction of the R882H mutant with wild type *DNMT3A*, *3B* or *3L* remains unchanged but R878H mutant protein antagonized wild type *DNMT3A* and *DNMT3B* in MDS/AML(Kim et al., 2013).Similar observations were made by Russler-Germainet al (2014) were they found that the enzymatic activity of *DNMT3A* was reduced by 80 % in the cells carrying the mutant and they attributed this to the failure of the R882H mutant *DNMT3A* to form homotetramers (Russler-Germain et al., 2014).However, further genome-wide and functional studies are needed to address the exact role of mutant *DNMT3A* inaberrant DNA methylation and leukaemogenesis.

6.6 Conclusion

In spite of genome wide studies to understand the role of *DNMT3A* mutations, additional functional studies both in vitro and in vivo are essential to discover the precise of role of this gene in connection to its downstream targets and its contribution to the pathogenesis and prognosis of MDS.

7 Chapter: General Conclusions and Future work

7.1 Identification of SNP6 abnormalities and mutations affecting Polycomb genes and their binding partners (*DNMT3A* and Jumonji)

The importance of polycomb protein *EZH2* in epigenetic regulation of transcription by its histone methyltransferase activity and the mutational analysis of this gene in haematological malignancies, including MDS has been addressed by previous studies as described in the introduction. Mutations of *EZH2* have been associated with SNP6 abnormalities like 7q CN (LOH) and 7q36.1 microdeletion (Ernst et al., 2010). However, SNP6 and mutational analysis of other components of the polycomb complex (PRC1 and PRC2) in MDS have not been reported and therefore the aim of my PhD was to study the PRC2 components (*EZH1, EED, SUZ12)* and twelve genes belonging to PRC1 to identify genomic aberrations affecting them using SNP6 arrays and to link these findings to the mutational status of these genes. In addition, I have also sequenced *DNMT3A* and seventeen Jumonji genes which have been reported to interact with the polycomb proteins and have been discussed in the Introduction chapter.

SNP6 data used in this study was generated by Dr. Azim Mohamedali and formed part of previous research projects conducted by our group. I visualised this data using CHromosome Analysis Suite (Chas) from Affymetrix and identified five PRC1 genes in 17/91 (19 %) patients with either CNVs or CN (LOH). Eleven patients showed CN (LOH) [size: ≤ 1.5 Mb] at the *BMI-1* locus. Though the size of the CN (LOH) indicated strongly that these could be not real these patients were still selected for mutational analysis based on the high frequency of CN(LOH) (n= 11) and due to the association of *BMI-1* with MDS discussed in detail in chapter 1. Among the ring proteins, *RING1A* was deleted in two patients while *RING1B* showed CN gain in two patients and a CN (LOH) in one patient. The polyhomoetic gene, *PHC1* was deleted in one patient while *PHC2* showed CN gain in a single patient and a CN

(LOH) in another patient. Sanger sequencing of patients with SNP6 aberrations for the five PRC1 genes did not identify any mutations associated these genes.

Similar to the PRC1 genes, seventeen Jumonji genes were also investigated for SNP6 aberrations in the same cohort of patients. 29 patients (31 %) showed either CNVs or CN (LOH) at 15 Jumonji gene loci. Interestingly, the frequency of SNP6 aberrations affecting the Jumonji genes was higher than the PRC1 genes. The highest frequency of deletions was observed in *JMJD1B* [n=12] while *JMJD4* and *JMJD2C* locus showed maximum number of CN (LOH) [n= 4] and amplifications [n= 3] respectively. These 29 patients were sequenced using 454 parallel sequencing for mutational analysis of the fifteen Jumonji genes. There were no mutations detected in these patients suggesting that mutations of PRC1 and Jumonji genes are rare in MDS and not associated with genominc abnormalities of these genes. Similar findings were reported by Score et al (2012) when they identified only 2 % patients with *JARID2* mutations in the peripheral blood and bone marrow of patients with MDS/MPN but did not find any association between CN (LOH) of chr 6p with these mutations (Score et al., 2012).

To elucidate changes in gene expression as a result of amplification or deletion of genomic material, qPCR was performed on 22/28 patients for thirteen Jumonji genes. Patients without SNP6 abnormalities at the gene loci were used as controls while K562 cells were used as a positive technical control in all experiments. Due to insufficient patient material, gene expression studies were not carried out on PRC1 genes. The expression of three Jumonji genes; *JARID1A*, *JARID1C* and *UTX* was modulated concomitant with CNVs of these genes. Deletion of *JARID1A* locus was associated with reduced gene expression (p value <0.0001) in two patients while trisomy of *JARID1C* and *UTX* in a single patient was associated with increased expression of both the genes. One patient with CN=4 at the *UTX* locus showed 5

fold increase in its expression. Altered expression of Jumonji proteins have been reported previously in carcinogenesis and discussed in chapter 3

EZH2 was previously sequenced by our group on 61 out of 91 patients and therefore the same cohort of patients were selected for mutational analysis using 454 parallel sequencing of the other core components of PRC2; *SUZ12*, *EED*, *EZH1* and *DNMT3A* to identify mutations and to examine their mutational overlap with *EZH2*. PRC2 genes were not mutated in this cohort of patients. Subsequently, similar findings were published in two reports wherein *SUZ12* and *EED* mutations were identified in only 1.4 % and 1% of patients respectively (Brecqueville et al., 2011; Score et al., 2012).This confirmed that among the polycomb and Jumonji family of proteins, mutations of *EZH2* were most frequent establishing the importance of *EZH2* in MDS. Therefore the second aim of this work was to focus on the functional analysis of *EZH2* in the context of myeloid cells. Since mutations in any component of PRC2 resulted in loss of H3K27me3, it emphasizes the importance of this mark and suggests that deregulation of the epigenetic control of gene expression plays an important role in the development of myeloid disorders.

In contrast to PRC1 and Jumoji genes, sequencing of *DNMT3A* in the same cohort of patients identified 10/61 patients with heterozygous mutations (clone size 20-44 %).Interestingly, these mutations were seen predominantly (n= 6) in patients with monosomy 7/del 7q, not observed in patients with normal karyotype and were classified as RAEBI/II according to the WHO classification suggesting the association of this gene with advanced stage of the disease.Nine non-synonymous mutations were identified throughout the gene; four of which were localized to the Mtase domain responsible for the methyltransferase activity of *DNMT3A*. A splice site mutation between exon 18 and the intron preceding exon 19 was also observed in this domain. Mutation at the R882 locus identified as a mutational hot spot by other research groups was observed in 2/10 patients. The Zinc finger domain and the PWWP

domain demonstrated one non-synonymous mutation each while three mutations were dispersed between the coding domains. The PWWP domain directs *DNMT3A* to heterochromatin while the Zinc finger domain helps it to bind firmly to the DNA sequence. Only one patient with the R882H mutations was also mutated for *EZH2* suggesting that there was very little association between mutations of these genes, a finding which was confirmed by other reports(Haferlach et al., 2014).Studies have observed connection between *DNMT3A* mutations and poor prognosis in patients but our data showed no significant difference in the median survival of monosomy 7 patients with or without *DNMT3A* mutations (p value = 0.6) cohort with inconclusive impact on prognosis of the patients with mutated *DNMT3A*. The high frequency of mutations affecting this gene, their adverse effect on the prognosis of the disease and the identification of a mutational hot spot (R882) makes *DNMT3A* an important gene for functional analysis and therefore in addition to functional analysis of *EZH2*, I have also examined the functional effect of *DNMT3A* mutations in my project.

7.2 Functional analysis in myeloid malignancies: R690C/R690H mutations, over expression and knock down of EZH2.

Functional analysis of *EZH2* was carried out by constructing the R690C, R690H mutations and Flag tagged wild type *EZH2* in p3XFLAG-myc-CMV-26 (GFP+ puromycin) vector as described in chapter 4.These two mutations were investigated because they were reported in many studies including in our cohort and located in the SET domain of the enzyme. They have been previously examined in insect cells and in patient samples carrying the homozygous form of the mutation. This is the first attempt to study the consequence of the R690C/R690H mutations and the over expression of Flag tagged *EZH2* in human myeloid cell lines. To achieve this, plasmid containing the mutants (R690C/R690H) and Flag tagged *EZH2* were transfected in myeloid cell line K562 using Lipofectamine LTX according to the manufacturer's instructions and GFP+ cells were sorted at 72 hr. Western blot analysis at this

time point, to evaluate alterations in *EZH2* proteins levels and their effect on its target H3K27me3 illustrated high expression of both R690C/R690H mutants and Flag tagged wild type *EZH2* compared to cells transfected with the empty vector and the mock transfected cells. Surprisingly, the H3K27me3 protein levels remained unaltered in all samples. These findings were in contrast to two studies; one reported decreased H3K27me3 levels in the presence of homozygous R690H mutation in patient samples from MDS/AML (Ernst et al., 2010; Makishima et al., 2010) and the second looked at heterozygous effect of R690C mutation in insect cells respectively. This system mimics the experimental design presented in my thesis wherein the expression of the mutant is heterozygous rather than homozygous. However, the model system used in both the studies is different i.e. insect cells versus human myeloid cells (K562) and it is unknown whether H3K27me3 modulation by *EZH2* has species related alterations.Another explanation for these findings was a long half-life of H3K27me3 protein making it impossible to detect changes in its levels at 72 hr post transfection. Time restrictions did not allow further experiments to optimise additional time points for sample collection to evaluate long term effects of mutant and Flag tagged wild type *EZH2* over expression on H3K27me3 protein levels.

To investigate the effects of elevated *EZH2* (both mutant and wild type) on transcription, microarray gene expression analysis using Affymetrix Human Transcriptome 2.0 arrays was carried out in K562 cells (72 hr time point) which identified up regulation (2.2 fold in wild type and 3.6 - 4.6 fold in R690C/R690H mutants) of *MLLT10* gene in mutated as well as wild type *EZH2* suggesting that the mutations do not affect the function of the wild type *EZH2* in regulating this gene. Reports on increase of *HOXA* gene expression by *MLLT10* fusion proteins (Brandimarte et al., 2013) as well as an inverse correlation of *EZH2* and *HOXA* gene expression have been published (Khan et al., 2013). It would be important to examine the *EZH2-MLLT10-HOXA*connection in myeloid cells to ascertain whether regulation of *HOXA*

genes by *EZH2* occurs via modulation of *MLLT10*. Unlike *MLLT10* which was up regulated in both wild type and mutant *EZH2* genes, *PML* (promyelocytic leukaemia) [up regulated by 2.16 fold] and *FANCL* (Fanconi Anaemia, Complementation Group L) [up regulated by 2.18 fold] genes were modulated exclusively in cells over expressing the Flag tagged wild type *EZH2*.Both *PML* and *FANCL* are not related to the pathogenesis of MDS but are implicated in promyelocytic leukaemia and Fanconi anaemia respectively. It would be important to examine the above mentioned genes in patient samples with R690C or R690H mutation to validate the results observed in the myeloid cell line.

Effects of *EZH2* knock down (KO) in myeloid cells assessed by transfecting MOLM13 cells with siRNA (200 nM) and transducing K562 cells with lentiviral based shRNA as described in chapter 5.In MOLM13 cells, 65% KO (24 hr post transfection) while in K562 cells 95 % KO (Day 7 after transduction) of both mRNA and protein levels of *EZH2* were observed by qPCR and western blot analysis respectively. Interestingly, western blot analysis demonstrated unaltered H3K27me3 levels in MOLM13 cells at 24 hr post transfection while 95-100 % reduction in H3K27me3 levels were observed at day 7 post transduction in K562 cells.This inconsistency in the effect of *EZH2* KO on H3K27me3 in MOLM13 and K562 cells suggested that either a stronger KO of *EZH2* (> 65%) expression is required to detect H3K27me3 modulation or that the half life of H3K27me3 protein is long and therefore protein degradation is not observed at 24 hr. Since a strong reduction (>95 %) in both *EZH2* and H3K27me3 expression was observed in K562 cells, microarray gene expression profiling using Affymetrix Human Transcriptome 2.0 arrayswas carried out as per the manufacturer's instructions. *BCL2* (70%), *FLT1* (40%), *HOXA10* (30 %), *CD44* (30%), *CD83* (30%), *TLSP* (50%), *IFI16* (50%) and *PAG1* (50%) genes were down regulated in cells transduced with shRNA against *EZH2* compared to the scrambled and wild type K562 cells. Comparison of this gene signature to data available in other types of cancers by different research groups did

not identify any overlap between *EZH2* targets confirming that *EZH2* target genes vary considerably between different cell types (Ezhkova et al., 2009; Li et al., 2012; Richter et al., 2009; Sher et al., 2012). These were categorised by their functions in to genes associated with cell differentiation (*HOXA10*), cell proliferation (*BCL2, IFI16* and *PAG1*) and Immune response (*CD44*, *CD83* and *TLSP*). A strong association between *EZH2* and cell proliferation/viability has been studied in various cancers as detailed in chapter 5 and therefore the next aim of my project was to study the effect of *EZH2* KO on cell proliferation, viability and cell cycle kinetics. Effects on cell viability were assessed by trypan blue dye exclusion method and Annexin V staining of GFP+ K562 cells sorted on day 7. Difference between the percentage of live, early apoptotic and late apoptotic cells stained by Annexin V was not significant between cells transduced by shRNA against *EZH2* compared to scrambled shRNA and wild type cells confirming that *EZH2* KO does not affect cell viability in K562 myeloid cells.MTT assay performed at 24, 48 and 72 hr on cells sorted for GFP+ K562 cells to inspect altered cell proliferation rates, revealed minimal effect on cell proliferation ($p = 0.1$ to 0.4) with *EZH2* inhibition (95 %) at all time points. In addition to cell viability and proliferation, changes in the kinetics of cell cycle entry after *EZH2* inhibition was examined by flow cytometric analysis using PI/FITC staining. There was no difference in the percentage of cells in each cell cycle phase as well as apoptotic cells (sub G0) between cells with *EZH2* inhibition compared to wild type K562 cells. These findings are in contrast to reports published in other cancers wherein *EZH2* KO reduces cell proliferation, affects cell cycle kinetics and could be attributed to the lack of *CDKN2A* locus in K562 cells which is the major downstream target of *EZH2* through which it governs the entry into cell cycle.

The gene signature obtained in K562 cells was verified in patient samples containing the R690H mutation (n= 1) and compared to control patients (n = 5) with wild type *EZH2*. Five genes *HOXA10*, *FLT1*, *PAG1B*, *EZH1* and *TLSP* were up regulated in the patient with R690H

mutant compared to the control patients while the other genes showed no change in their expression levels. Interestingly, these results are in contrast to the findings (except for *EZH1*) observed with *EZH2* inhibition where *HOXA10*, *FLT1*, *PAG1B* and *TLSP* were down regulated. Elevated levels of *EZH1* in patient sample suggest that *EZH1* may compensate for loss *EZH2* function. However this has not been shown in K562 cells. This incongruity in the gene signature between patient with R690H mutation and K562 cell line with 95 % KO of *EZH2* suggest that *EZH2* mutation does not mimic loss of function of the gene.

7.3 Functional analysis of DNMT3A in myeloid malignancies

Investigating the functional consequences of the R882H mutation has been attempted by various research groups described in chapter 6. The aim of this project was to construct R882H *DNMT3A* mutation and transfect it into myeloid cell lines to observe the in vitro effect of this mutation. However, large size of the *DNMT3A* gene provided numerous technical difficulties in cloning this gene into p3XFLAG-myc-CMV-26 vector to allow for easy detection of the FLAG tagged transgene expression in the transfected cell lines. Ligation was attempted several times, however all strategies to clone *DNMT3A*cDNA into p3XFLAG-myc-CMV-26 were unsuccessful. This problem was circumvented by isolating the FLAG tag from p3XFLAG-myc-CMV-26 and ligating it into pCMV-SPORT6 vector containing the R882H mutation. This plasmid was transfected along with pBABE (containing puromycin marker) into myeloid cell lines but very poor transfection efficiency was observed in both MOLM13 and KG1 cells resulting in long recovery period of cells after puromycin selection probably causing inactivation of the CMV promoter which drives the expression of the transgene in the vector. Variable expression levels (lower than mock transfected in some instances) of *DNMT3A* mRNA were observed by qPCR analysis. Due to insufficient time the

hypothesis regarding CMV promoter methylation and further work on this mutation was not conducted.

7.4 Future Work

The results obtained in this thesis raised questions, which can be addressed in future studies.

- To investigate alterations in the known targets of *UTX*, *JARID1A* and *JARID1C* by qPCR or western blot analysis due to changes in gene expression of these genes (Chapter 3) in MDS patients.
- To optimise and prolong the time point for sample collection in cells transfected with R690C/R690H mutants and to evaluate the long term effects of these mutations on H3K27me3 expression. Perform gene expression profiling on samples collected at a later time point if it is associated with alterations in H3K27me3 levels and verify this gene signature with the one mentioned in chapter 4 to determine overlapping genes.
- To KO *EZH2* in other myeloid cell line like KG1 (with wild type *CDKN2A*) and observe its effect of cell proliferation/viability and cell cycle kinetics. Compare these results to the findings obtained in K562 cells in chapter 5.
- Validate the findings of the first shRNA mentioned in chapter 5 using another shRNA to rule out off target effects.

Acquaviva, C., Gelsi-Boyer, V., and Birnbaum, D. (2010). Myelodysplastic syndromes: lost between two states? Leukemia *24*, 1-5.
Akasaka, T., Takahashi, N., Suzuki, M., Koseki, H., Bodmer, R., and Koga, H. (2002). MBLR, a new RING finger protein resembling mammalian Polycomb gene products, is regulated by cell cycle-dependent phosphorylation. Genes to Cells *7*, 835-850.
Akasaka, T., Tsuji, K., Kawahira, H., Kanno, M., Harigaya, K., Hu, L., Ebihara, Y., Nakahata, T., Tetsu, O., Taniguchi, M.*, et al.* (1997). The role of mel-18, a mammalian Polycomb group gene, during IL-7-dependent proliferation of lymphocyte precursors. Immunity *7*, 135-146.
Amarzguioui, M., Rossi, J.J., and Kim, D. (2005). Approaches for chemically synthesized siRNA and vector-mediated RNAi. FEBS Lett *579*, 5974-5981.
Amat, R., and Gudas, L.J. (2011). RARgamma is required for correct deposition and removal of Suz12 and H2A.Z in embryonic stem cells. Journal of Cellular Physiology *226*, 293-298.
Bains, A., Luthra, R., Medeiros, L.J., and Zuo, Z. (2011). FLT3 and NPM1 mutations in myelodysplastic syndromes: Frequency and potential value for predicting progression to acute myeloid leukemia. Am J Clin Pathol *135*, 62-69.
Bantounas, I., Phylactou, L.A., and Uney, J.B. (2004). RNA interference and the use of small interfering RNA to study gene function in mammalian systems. J Mol Endocrinol *33*, 545-557.
Beck, S.A., Falconer, E., Catching, A., Hodgson, J.W., and Brock, H.W. (2010). Cell cycle defects in polyhomeotic mutants are caused by abrogation of the DNA damage checkpoint. Developmental Biology *339*, 320-328.
Beerman, I., Bock, C., Garrison, B.S., Smith, Z.D., Gu, H., Meissner, A., and Rossi, D.J. (2013). Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem Cell *12*, 413-425.
Beguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T.*, et al.* (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell *23*, 677-692.
Bei, L., Shah, C., Wang, H., Huang, W., Roy, R., and Eklund, E.A. (2012). beta-Catenin activates the HOXA10 and CDX4 genes in myeloid progenitor cells. J Biol Chem *287*, 39589-39601.
Bejar, R. (2014). Clinical and genetic predictors of prognosis in myelodysplastic syndromes. Haematologica *99*, 956-964.
Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R.L., Neuberg, D.*, et al.* (2011). Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med *364*, 2496-2506.
Ben-Saadon, R., Zaaroor, D., Ziv, T., and Ciechanover, A. (2006). The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. Molecular Cell *24*, 701-711.
Bennett, J.M., Brunning, R.D., and Vardiman, J.W. (2002). Myelodysplastic syndromes: from French-American-British to World Health Organization: a commentary. Blood *99*, 3074-3075.
Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006a). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.
Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.
Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J.*, et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat Methods *3*, 199-204.
Bjornsson, J.M., Andersson, E., Lundstrom, P., Larsson, N., Xu, X., Repetowska, E., Humphries, R.K., and Karlsson, S. (2001). Proliferation of primitive myeloid progenitors can be reversibly induced by HOXA10. Blood *98*, 3301-3308.

Blair, L.P., Cao, J., Zou, M.R., Sayegh, J., and Yan, Q. (2011). Epigenetic Regulation by Lysine Demethylase 5 (KDM5) Enzymes in Cancer. Cancers (Basel) *3*, 1383-1404.
Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J *22*, 5323-5335.
Brandimarte, L., Pierini, V., Di Giacomo, D., Borga, C., Nozza, F., Gorello, P., Giordan, M., Cazzaniga, G., Te Kronnie, G., La Starza, R.*, et al.* (2013). New MLLT10 gene recombinations in pediatric T-acute lymphoblastic leukemia. Blood *121*, 5064-5067.
Brecqueville, M., Cervera, N., Adelaide, J., Rey, J., Carbuccia, N., Chaffanet, M., Mozziconacci, M.J., Vey, N., Birnbaum, D., Gelsi-Boyer, V.*, et al.* (2011). Mutations and deletions of the SUZ12 polycomb gene in myeloproliferative neoplasms. Blood Cancer J *1*, e33.
Brinkman, A.B., Gu, H., Bartels, S.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A.*, et al.* (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res *22*, 1128-1138.
Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P., and Rubanyi, G.M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. J Gene Med *6*, 395-404.
Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science *296*, 550-553.
Bryant, R.J., Cross, N.A., Eaton, C.L., Hamdy, F.C., and Cunliffe, V.T. (2007). EZH2 promotes proliferation and invasiveness of prostate cancer cells. Prostate *67*, 547-556.
Budhavarapu, V.N., Chavez, M., and Tyler, J.K. (2013). How is epigenetic information maintained through DNA replication? Epigenetics Chromatin *6*, 32.
Buske, C., Feuring-Buske, M., Antonchuk, J., Rosten, P., Hogge, D.E., Eaves, C.J., and Humphries, R.K. (2001). Overexpression of HOXA10 perturbs human lymphomyelopoiesis in vitro and in vivo. Blood *97*, 2286-2292.
Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D.*, et al.* (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Mol Cell *49*, 571-582.
Cales, C., Roman-Trufero, M., Pavon, L., Serrano, I., Melgar, T., Endoh, M., Perez, C., Koseki, H., and Vidal, M. (2008). Inactivation of the polycomb group protein Ring1B unveils an antiproliferative role in hematopoietic cell expansion and cooperation with tumorigenesis associated with Ink4a deletion. Molecular and Cellular Biology *28*, 1018-1028.
Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Molecular Cell *20*, 845-854.
Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol Cell Biol *28*, 1862-1872.
Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.
Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y.*, et al.* (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet *44*, 23-31.
Chandler, H., Patel, H., Palermo, R., Brookes, S., Matthews, N., and Peters, G. (2014). Role of polycomb group proteins in the DNA damage response - a reassessment. PLoS One *9*, e102968.
Chang, C.J., Yang, J.Y., Xia, W., Chen, C.T., Xie, X., Chao, C.H., Woodward, W.A., Hsu, J.M., Hortobagyi, G.N., and Hung, M.C. (2011). EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-beta-catenin signaling. Cancer Cell *19*, 86-100.
Chen, Y., Lin, M.C., Yao, H., Wang, H., Zhang, A.Q., Yu, J., Hui, C.K., Lau, G.K., He, M.L., Sung, J.*, et al.* (2007). Lentivirus-mediated RNA interference targeting enhancer of zeste homolog 2 inhibits hepatocellular carcinoma growth through down-regulation of stathmin. Hepatology *46*, 200-208.

Chen, Y., Xie, D., Yin Li, W., Man Cheung, C., Yao, H., Chan, C.Y., Xu, F.P., Liu, Y.H., Sung, J.J., and Kung, H.F. (2010). RNAi targeting EZH2 inhibits tumor growth and liver metastasis of pancreatic cancer in vivo. Cancer Lett *297*, 109-116.
Chiang, P.K. (1998). Biological effects of inhibitors of S-adenosylhomocysteine hydrolase. Pharmacol Ther *77*, 115-134.
Chiyomaru, T., Fukuhara, S., Saini, S., Majid, S., Deng, G., Shahryari, V., Chang, I., Tanaka, Y., Enokida, H., Nakagawa, M.*, et al.* (2014). Long non-coding RNA HOTAIR is targeted and regulated by miR-141 in human cancer cells. J Biol Chem *289*, 12550-12565.
Cho, H.S., Toyokawa, G., Daigo, Y., Hayami, S., Masuda, K., Ikawa, N., Yamane, Y., Maejima, K., Tsunoda, T., Field, H.I.*, et al.* (2012). The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. Int J Cancer *131*, E179-189.
Chou, D.M., Adamson, B., Dephoure, N.E., Tan, X., Nottke, A.C., Hurov, K.E., Gygi, S.P., Colaiacovo, M.P., and Elledge, S.J. (2010). A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. Proc Natl Acad Sci U S A *107*, 18475-18480.
Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D.C., Forsti, A., Vijayakrishnan, J., Migliorini, G., Dobbins, S.E., Holroyd, A.*, et al.* (2013). Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. Nat Genet *45*, 1221-1225.
Conway O'Brien, E., Prideaux, S., and Chevassut, T. (2014). The epigenetic landscape of acute myeloid leukemia. Adv Hematol *2014*, 103175.
Cuccato, G., Polynikis, A., Siciliano, V., Graziano, M., di Bernardo, M., and di Bernardo, D. (2011). Modeling RNA interference in mammalian cells. BMC Syst Biol *5*, 19.
Das, P.P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., De Los Angeles, A., O'Brien, K., Atsma, J.M., Fujiwara, Y., Nguyen, M.*, et al.* (2014). Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. Mol Cell *53*, 32-48.
Dawson, M.J., Elwood, N.J., Johnstone, R.W., and Trapani, J.A. (1998). The IFN-inducible nucleoprotein IFI 16 is expressed in cells of the monocyte lineage, but is rapidly and markedly down-regulated in other myeloid precursor populations. J Leukoc Biol *64*, 546-554.
Dawson, M.J., and Trapani, J.A. (1995). IFI 16 gene encodes a nuclear protein whose expression is induced by interferons in human myeloid leukaemia cell lines. J Cell Biochem *57*, 39-51.
de la Cruz, C.C., Kirmizis, A., Simon, M.D., Isono, K., Koseki, H., and Panning, B. (2007). The polycomb group protein SUZ12 regulates histone H3 lysine 9 methylation and HP1 alpha distribution. Chromosome Research *15*, 299-314.
de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M.*, et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell *7*, 663-676.
de Rooij, J.D., Hollink, I.H., Arentsen-Peters, S.T., van Galen, J.F., Berna Beverloo, H., Baruchel, A., Trka, J., Reinhardt, D., Sonneveld, E., Zimmermann, M.*, et al.* (2013). NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. Leukemia *27*, 2280-2288.
Degos, L., and Wang, Z.Y. (2001). All trans retinoic acid in acute promyelocytic leukemia. Oncogene *20*, 7140-7145.
Dimri, G.P., Martinez, J.L., Jacobs, J.J., Keblusek, P., Itahana, K., Van Lohuizen, M., Campisi, J., Wazer, D.E., and Band, V. (2002). The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. Cancer Res *62*, 4736-4745.
Dos Santos, G.A., Kats, L., and Pandolfi, P.P. (2013). Synergy against PML-RARa: targeting transcription, proteolysis, differentiation, and self-renewal in acute promyelocytic leukemia. J Exp Med *210*, 2793-2802.

Elderkin, S., Maertens, G.N., Endoh, M., Mallery, D.L., Morrice, N., Koseki, H., Peters, G., Brockdorff, N., and Hiom, K. (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. Molecular Cell *28*, 107-120.
Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A*., et al.* (2010a). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet *42*, 722-726.
Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A*., et al.* (2010b). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet *42*, 722-726.
Eskander, R.N., Ji, T., Huynh, B., Wardeh, R., Randall, L.M., and Hoang, B. (2013). Inhibition of enhancer of zeste homolog 2 (EZH2) expression is associated with decreased tumor cell proliferation, migration, and invasion in endometrial cancer cell lines. Int J Gynecol Cancer *23*, 997-1005.
Ewalt, M., Galili, N.G., Mumtaz, M., Churchill, M., Rivera, S., Borot, F., Raza, A., and Mukherjee, S. (2011). DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia. Blood Cancer Journal *1*, e9.
Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.
Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W*., et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife *1*, e00205.
Federico, A., Pallante, P., Bianco, M., Ferraro, A., Esposito, F., Monti, M., Cozzolino, M., Keller, S., Fedele, M., Leone, V*., et al.* (2009). Chromobox protein homologue 7 protein, with decreased expression in human carcinomas, positively regulates E-cadherin expression by interacting with the histone deacetylase 2 protein. Cancer Res *69*, 7079-7087.
Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stutzer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell *53*, 49-62.
Fischer, C., Mazzone, M., Jonckx, B., and Carmeliet, P. (2008). FLT1 and its ligands VEGFB and PlGF: drug targets for anti-angiogenic therapy? Nat Rev Cancer *8*, 942-956.
Fleckenstein, D.S., Uphoff, C.C., Drexler, H.G., and Quentmeier, H. (2002). Detection of p53 gene mutations by single strand conformational polymorphism (SSCP) in human acute myeloid leukemia-derived cell lines. Leuk Res *26*, 207-214.
Fu, L., Nara, N., and Tohda, S. (2007). Involvement of Notch signaling in myelodysplastic syndrome. Leukemia Research *31*, 1160-1161.
Fujisaki, S., Ninomiya, Y., Ishihara, H., Miyazaki, M., Kanno, R., Asahara, T., and Kanno, M. (2003). Dimerization of the Polycomb-group protein Mel-18 is regulated by PKC phosphorylation. Biochemical and Biophysical Research Communications *300*, 135-140.
Fujiwara, T., Saitoh, H., Inoue, A., Kobayashi, M., Okitsu, Y., Katsuoka, Y., Fukuhara, N., Onishi, Y., Ishizawa, K., Ichinohasama, R*., et al.* (2014). 3-Deazaneplanocin A (DZNep), an inhibitor of S-adenosylmethionine-dependent methyltransferase, promotes erythroid differentiation. J Biol Chem *289*, 8121-8134.
Gagnidze, K., Weil, Z.M., Faustino, L.C., Schaafsma, S.M., and Pfaff, D.W. (2013). Early histone modifications in the ventromedial hypothalamus and preoptic area following oestradiol administration. J Neuroendocrinol *25*, 939-955.
Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell *45*, 344-356.

Ghazal, P., Lubon, H., Fleckenstein, B., and Hennighausen, L. (1987). Binding of transcription factors and creation of a large nucleoprotein complex on the human cytomegalovirus enhancer. Proc Natl Acad Sci U S A *84*, 3658-3662.
Giagounidis, A.A., Germing, U., and Aul, C. (2006). Biological and prognostic significance of chromosome 5q deletions in myeloid malignancies. Clin Cancer Res *12*, 5-10.
Ginjala, V., Nacerddine, K., Kulkarni, A., Oza, J., Hill, S.J., Yao, M., Citterio, E., van Lohuizen, M., and Ganesan, S. (2011). BMI1 is recruited to DNA breaks and contributes to DNA damage-induced H2A ubiquitination and repair. Mol Cell Biol *31*, 1972-1982.
Gong, Y., Wang, X., Liu, J., Shi, L., Yin, B., Peng, X., Qiang, B., and Yuan, J. (2005). NSPc1, a mainly nuclear localized protein of novel PcG family members, has a transcription repression activity related to its PKC phosphorylation site at S183. FEBS Letters *579*, 115-121.
Gong, Y., Yue, J., Wu, X., Wang, X., Wen, J., Lu, L., Peng, X., Qiang, B., and Yuan, J. (2006). NSPc1 is a cell growth regulator that acts as a transcriptional repressor of p21Waf1/Cip1 via the RARE element. Nucleic Acids Res *34*, 6158-6169.
Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T., and Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. Mol Cell *54*, 485-499.
Gonzalez, M.E., Li, X., Toy, K., DuPrie, M., Ventura, A.C., Banerjee, M., Ljungman, M., Merajver, S.D., and Kleer, C.G. (2009). Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. Oncogene *28*, 843-853.
Gou, D., Jin, N., and Liu, L. (2003). Gene silencing in mammalian cells by PCR-based short hairpin RNA. FEBS Lett *548*, 113-118.
Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev *25*, 2210-2221.
Greenberg, P., Cox, C., LeBeau, M.M., Fenaux, P., Morel, P., Sanz, G., Sanz, M., Vallespi, T., Hamblin, T., Oscier, D*., et al.* (1997). International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood *89*, 2079-2088.
Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Sole, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F*., et al.* (2012). Revised international prognostic scoring system for myelodysplastic syndromes. Blood *120*, 2454-2465.
Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.
Haase, D., Germing, U., Schanz, J., Pfeilstocker, M., Nosslinger, T., Hildebrandt, B., Kundgen, A., Lubbert, M., Kunzmann, R., Giagounidis, A.A*., et al.* (2007). New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. Blood *110*, 4385-4395.
Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T*., et al.* (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241-247.
Hagarman, J.A., Motley, M.P., Kristjansdottir, K., and Soloway, P.D. (2013). Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. PLoS One *8*, e53880.
Hanabuchi, S., Ito, T., Park, W.R., Watanabe, N., Shaw, J.L., Roman, E., Arima, K., Wang, Y.H., Voo, K.S., Cao, W*., et al.* (2010). Thymic stromal lymphopoietin-activated plasmacytoid dendritic cells induce the generation of FOXP3+ regulatory T cells in human thymus. J Immunol *184*, 2999-3007.
Hatano, A., Matsumoto, M., Higashinakagawa, T., and Nakayama, K.I. (2010). Phosphorylation of the chromodomain changes the binding specificity of Cbx2 for methylated histone H3. Biochemical and Biophysical Research Communications *397*, 93-99.
He, J., Nguyen, A.T., and Zhang, Y. (2011). KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. Blood *117*, 3869-3880.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol *15*, 373-384.
Herbst, F., Ball, C.R., Tuorto, F., Nowrouzi, A., Wang, W., Zavidij, O., Dieter, S.M., Fessler, S., van der Hoeven, F., Kloz, U.*, et al.* (2012). Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo. Mol Ther *20*, 1014-1021.
Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nunez, J., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*, 649-662.
Huang, X., Ma, D., Dong, W., Li, P., Lu, T., He, N., Tian, T., Liu, N., Du, Y., and Ji, C. (2013). Gene expression profiling of the DNMT3A R882 mutation in acute leukemia. Oncol Lett *6*, 268-274.
Huang, Y., Leung, J.W., Lowery, M., Matsushita, N., Wang, Y., Shen, X., Huong, D., Takata, M., Chen, J., and Li, L. (2014). Modularized functions of the Fanconi anemia core complex. Cell Rep *7*, 1849-1857.
Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N.*, et al.* (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat Biotechnol *25*, 1315-1321.
Ismail, I.H., Andrin, C., McDonald, D., and Hendzel, M.J. (2010). BMI1-mediated histone ubiquitylation promotes DNA double-strand break repair. J Cell Biol *191*, 45-60.
Ismail, I.H., Gagne, J.P., Caron, M.C., McDonald, D., Xu, Z., Masson, J.Y., Poirier, G.G., and Hendzel, M.J. (2012). CBX4-mediated SUMO modification regulates BMI1 recruitment at sites of DNA damage. Nucleic Acids Res *40*, 5497-5510.
Isono, K., Fujimura, Y., Shinga, J., Yamaki, M., J, O.W., Takihara, Y., Murahashi, Y., Takada, Y., Mizutani-Koseki, Y., and Koseki, H. (2005). Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. Molecular and Cellular Biology *25*, 6694-6706.
Itahana, K., Zou, Y., Itahana, Y., Martinez, J.L., Beausejour, C., Jacobs, J.J., Van Lohuizen, M., Band, V., Campisi, J., and Dimri, G.P. (2003). Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. Mol Cell Biol *23*, 389-401.
Itzykson, R., and Fenaux, P. (2014). Epigenetics of myelodysplastic syndromes. Leukemia *28*, 497-506.
Iwama, A., Oguro, H., Negishi, M., Kato, Y., Morita, Y., Tsukui, H., Ema, H., Kamijo, T., Katoh-Fukui, Y., Koseki, H.*, et al.* (2004). Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. Immunity *21*, 843-851.
Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol *21*, 635-637.
Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA *12*, 1179-1187.
Jackson, A.L., and Linsley, P.S. (2004). Noise amidst the silence: off-target effects of siRNAs? Trends Genet *20*, 521-524.
Jankowska, A.M., Makishima, H., Tiu, R.V., Szpurka, H., Huang, Y., Traina, F., Visconte, V., Sugimoto, Y., Prince, C., O'Keefe, C.*, et al.* (2011). Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2 and DNMT3A. Blood.
Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63-66.
Jerez, A., Sugimoto, Y., Makishima, H., Verma, A., Jankowska, A.M., Przychodzen, B., Visconte, V., Tiu, R.V., O'Keefe, C.L., Mohamedali, A.M.*, et al.* (2012). Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. Blood *119*, 6109-6117.

Kajiume, T., Ninomiya, Y., Ishihara, H., Kanno, R., and Kanno, M. (2004). Polycomb group gene mel-18 modulates the self-renewal activity and cell cycle status of hematopoietic stem cells. Experimental Hematology *32*, 571-578.
Kamminga, L.M., Bystrykh, L.V., de Boer, A., Houwer, S., Douma, J., Weersing, E., Dontje, B., and de Haan, G. (2006). The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood *107*, 2170-2179.
Kao, J.M., McMillan, A., and Greenberg, P.L. (2008). International MDS risk analysis workshop (IMRAW)/IPSS reanalyzed: impact of cytopenias on clinical outcomes in myelodysplastic syndromes. Am J Hematol *83*, 765-770.
Kerppola, T.K. (2009). Polycomb group complexes--many combinations, many functions. Trends Cell Biol *19*, 692-704.
Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol *25*, 6857-6868.
Khan, S.N., Jankowska, A.M., Mahfouz, R., Dunbar, A.J., Sugimoto, Y., Hosono, N., Hu, Z., Cheriyath, V., Vatolin, S., Przychodzen, B*., et al.* (2013). Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. Leukemia *27*, 1301-1309.
Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. Nature Structural Biology *9*, 453-457.
Kim, C.A., Sawaya, M.R., Cascio, D., Kim, W., and Bowie, J.U. (2005). Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. Journal of Biological Chemistry *280*, 27769-27775.
Kim, J.Y., Kim, K.B., Eom, G.H., Choe, N., Kee, H.J., Son, H.J., Oh, S.T., Kim, D.W., Pak, J.H., Baek, H.J*., et al.* (2012). KDM3B is the H3K9 demethylase involved in transcriptional activation of lmo2 in leukemia. Mol Cell Biol *32*, 2917-2933.
Kim, S.J., Zhao, H., Hardikar, S., Singh, A.K., Goodell, M.A., and Chen, T. (2013a). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. Blood *122*, 4086-4089.
Kim, W., Bird, G.H., Neff, T., Guo, G., Kerenyi, M.A., Walensky, L.D., and Orkin, S.H. (2013b). Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. Nat Chem Biol *9*, 643-650.
Kirmizis, A., Bartley, S.M., and Farnham, P.J. (2003). Identification of the polycomb group protein SU(Z)12 as a potential molecular target for human cancer therapy. Mol Cancer Ther *2*, 113-121.
Klauke, K., Radulovic, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K*., et al.* (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat Cell Biol *15*, 353-362.
Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R*., et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.
Kojima, K., Konopleva, M., Samudio, I.J., Shikami, M., Cabreira-Hansen, M., McQueen, T., Ruvolo, V., Tsao, T., Zeng, Z., Vassilev, L.T*., et al.* (2005). MDM2 antagonists induce p53-dependent apoptosis in AML: implications for leukemia therapy. Blood *106*, 3150-3159.
Kolybaba, A., and Classen, A.K. (2014). Sensing cellular states--signaling to chromatin pathways targeting Polycomb and Trithorax group function. Cell Tissue Res *356*, 477-493.
Kondo, Y., Shen, L., Cheng, A.S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B*., et al.* (2008). Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. Nat Genet *40*, 741-750.
Koopman, G., Reutelingsperger, C.P., Kuijten, G.A., Keehnen, R.M., Pals, S.T., and van Oers, M.H. (1994). Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. Blood *84*, 1415-1420.
Kracmarova, A., Cermak, J., Brdicka, R., and Bruchova, H. (2008). High expression of ERCC1, FLT1, NME4 and PCNA associated with poor prognosis and advanced stages in myelodysplastic syndrome. Leuk Lymphoma *49*, 1297-1305.

Kruidenier, L., Chung, C.W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H.*, et al.* (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature *488*, 404-408.
Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Molecular Cell *14*, 183-193.
Law, J.C., Ritke, M.K., Yalowich, J.C., Leder, G.H., and Ferrell, R.E. (1993). Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. Leuk Res *17*, 1045-1050.
Lee, H.J., Shin, D.H., Kim, K.B., Shin, N., Park, W.Y., Lee, J.H., Choi, K.U., Kim, J.Y., Lee, C.H., and Sol, M.Y. (2014). Polycomb protein EZH2 expression in diffuse large B-cell lymphoma is associated with better prognosis in patients treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone. Leuk Lymphoma *55*, 2056-2063.
Lessard, J., Baban, S., and Sauvageau, G. (1998). Stage-specific expression of polycomb group genes in human bone marrow cells. Blood *91*, 1216-1224.
Lessard, J., Schumacher, A., Thorsteinsdottir, U., van Lohuizen, M., Magnuson, T., and Sauvageau, G. (1999a). Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation. Genes Dev *13*, 2691-2703.
Lessard, J., Schumacher, A., Thorsteinsdottir, U., van Lohuizen, M., Magnuson, T., and Sauvageau, G. (1999b). Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation. Genes Dev *13*, 2691-2703.
Leung, R.K., and Whittaker, P.A. (2005). RNA interference: from gene silencing to gene-specific therapeutics. Pharmacol Ther *107*, 222-239.
Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J.*, et al.* (2010a). DNMT3A mutations in acute myeloid leukemia. N Engl J Med *363*, 2424-2433.
Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J.*, et al.* (2010b). DNMT3A mutations in acute myeloid leukemia. N Engl J Med *363*, 2424-2433.
Li, B., Zhou, J., Liu, P., Hu, J., Jin, H., Shimono, Y., Takahashi, M., and Xu, G. (2007). Polycomb protein Cbx4 promotes SUMO modification of de novo DNA methyltransferase Dnmt3a. Biochemical Journal *405*, 369-378.
Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet *3*, 662-673.
Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev *24*, 368-380.
Li, H., Bitler, B.G., Vathipadiekal, V., Maradeo, M.E., Slifker, M., Creasy, C.L., Tummino, P.J., Cairns, P., Birrer, M.J., and Zhang, R. (2012). ALDH1A1 is a novel EZH2 target gene in epithelial ovarian cancer identified by genome-wide approaches. Cancer Prev Res (Phila) *5*, 484-491.
Li, K., Liu, C., Zhou, B., Bi, L., Huang, H., Lin, T., and Xu, K. (2013). Role of EZH2 in the Growth of Prostate Cancer Stem Cells Isolated from LNCaP Cells. Int J Mol Sci *14*, 11981-11993.
Lin, X., Ruan, X., Anderson, M.G., McDowell, J.A., Kroeger, P.E., Fesik, S.W., and Shen, Y. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. Nucleic Acids Res *33*, 4527-4535.
Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J.M., Herman, H., Pasini, D., Miro, X., Donohoe, M.E., Lee, J.T.*, et al.* (2008). Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. PLoS Genet *4*, e1000145.
Liu, J., Cao, L., Chen, J., Song, S., Lee, I.H., Quijano, C., Liu, H., Keyvanfar, K., Chen, H., Cao, L.Y.*, et al.* (2009). Bmi1 regulates mitochondrial function and the DNA damage response pathway. Nature *459*, 387-392.
Loeffler-Ragg, J., Germing, U., Sperr, W.R., Herrmann, H., Zwierzina, H., Valent, P., Ulmer, H., and Stauder, R. (2011). Serum CD44 levels predict survival in patients with low-risk myelodysplastic syndromes. Crit Rev Oncol Hematol *78*, 150-161.

Loser, P., Jennings, G.S., Strauss, M., and Sandig, V. (1998). Reactivation of the previously silenced cytomegalovirus major immediate-early promoter in the mouse liver: involvement of NFkappaB. J Virol *72*, 180-190.
Maciejewski, J.P., and Mufti, G.J. (2008). Whole genome scanning as a cytogenetic tool in hematologic malignancies. Blood *112*, 965-974.
Maciejewski, J.P., Tiu, R.V., and O'Keefe, C. (2009). Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. Br J Haematol *146*, 479-488.
Makishima, H., Jankowska, A.M., Tiu, R.V., Szpurka, H., Sugimoto, Y., Hu, Z., Saunthararajah, Y., Guinta, K., Keddache, M.A., Putnam, P.*, et al.* (2010). Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. Leukemia *24*, 1799-1804.
Marcucci, G., Metzeler, K.H., Schwind, S., Becker, H., Maharry, K., Mrozek, K., Radmacher, M.D., Kohlschmidt, J., Nicolet, D., Whitman, S.P.*, et al.* (2012). Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. J Clin Oncol *30*, 742-750.
Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.
Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008a). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell *32*, 503-518.
Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008b). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Molecular Cell *32*, 503-518.
Marschalek, R. (2011). Mechanisms of leukemogenesis by MLL fusion proteins. Br J Haematol *152*, 141-154.
Martin, C., Cao, R., and Zhang, Y. (2006). Substrate preferences of the EZH2 histone methyltransferase complex. Journal of Biological Chemistry *281*, 8365-8370.
Martinez, A.M., Schuettengruber, B., Sakr, S., Janic, A., Gonzalez, C., and Cavalli, G. (2009). Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling. Nature Genetics *41*, 1076-1082.
Meers, S. (2014). The Myelodysplastic Syndromes: The Era Of Understanding. Eur J Haematol.
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.
Metzeler, K.H., Maharry, K., Radmacher, M.D., Mrozek, K., Margeson, D., Becker, H., Curfman, J., Holland, K.B., Schwind, S., Whitman, S.P.*, et al.* (2011). TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol *29*, 1373-1381.
Meyer, C., Hofmann, J., Burmeister, T., Groger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E.*, et al.* (2013). The MLL recombinome of acute leukemias in 2013. Leukemia *27*, 2165-2176.
Mian, S.A., Smith, A.E., Kulasekararaj, A.G., Kizilors, A., Mohamedali, A.M., Lea, N.C., Mitsopoulos, K., Ford, K., Nasser, E., Seidl, T.*, et al.* (2013). Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica *98*, 1058-1066.
Mihara, K., Chowdhury, M., Nakaju, N., Hidani, S., Ihara, A., Hyodo, H., Yasunaga, S., Takihara, Y., and Kimura, A. (2006). Bmi-1 is useful as a novel molecular marker for predicting progression of myelodysplastic syndrome and patient prognosis. Blood *107*, 305-308.

Mistry, A.R., Pedersen, E.W., Solomon, E., and Grimwade, D. (2003). The molecular pathogenesis of acute promyelocytic leukaemia: implications for the clinical management of the disease. Blood Rev *17*, 71-97.
Miyazaki, M., Kawamoto, H., Kato, Y., Itoi, M., Miyazaki, K., Masuda, K., Tashiro, S., Ishihara, H., Igarashi, K., Amagai, T.*, et al.* (2005). Polycomb group gene mel-18 regulates early T progenitor expansion by maintaining the expression of Hes-1, a target of the Notch pathway. Journal of Immunology *174*, 2507-2516.
Mohamedali, A., Gaken, J., Twine, N.A., Ingram, W., Westwood, N., Lea, N.C., Hayden, J., Donaldson, N., Aul, C., Gattermann, N.*, et al.* (2007). Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes. Blood *110*, 3365-3373.
Mohamedali, A.M., Smith, A.E., Gaken, J., Lea, N.C., Mian, S.A., Westwood, N.B., Strupp, C., Gattermann, N., Germing, U., and Mufti, G.J. (2009). Novel TET2 mutations associated with UPD4q24 in myelodysplastic syndrome. J Clin Oncol *27*, 4002-4006.
Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005a). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr Biol *15*, 942-947.
Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005b). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr Biol *15*, 942-947.
Montgomery, N.D., Yee, D., Montgomery, S.A., and Magnuson, T. (2007). Molecular and functional mapping of EED motifs required for PRC2-dependent histone methylation. Journal of Molecular Biology *374*, 1145-1157.
Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F.*, et al.* (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet *42*, 181-185.
Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol Cell *45*, 255-262.
Mu, Z., Li, H., Fernandez, S.V., Alpaugh, K.R., Zhang, R., and Cristofanilli, M. (2013). EZH2 knockdown suppresses the growth and invasion of human inflammatory breast cancer cells. J Exp Clin Cancer Res *32*, 70.
Mufti, G.J., Bennett, J.M., Goasguen, J., Bain, B.J., Baumann, I., Brunning, R., Cazzola, M., Fenaux, P., Germing, U., Hellstrom-Lindberg, E.*, et al.* (2008). Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. Haematologica *93*, 1712-1717.
Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Muller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J *26*, 4078-4088.
Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T.*, et al.* (2010a). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nature Genetics *42*, 665-667.
Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T.*, et al.* (2010b). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat Genet *42*, 665-667.
Oguro, H., Iwama, A., Morita, Y., Kamijo, T., van Lohuizen, M., and Nakauchi, H. (2006). Differential impact of Ink4a and Arf on hematopoietic stem cells and their bone marrow microenvironment in Bmi1-deficient mice. Journal of Experimental Medicine *203*, 2247-2253.
Ohta, H., Sawada, A., Kim, J.Y., Tokimasa, S., Nishiguchi, S., Humphries, R.K., Hara, J., and Takihara, Y. (2002). Polycomb group gene rae28 is required for sustaining activity of hematopoietic stem cells. Journal of Experimental Medicine *195*, 759-770.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Molecular Cell *32*, 232-246.
Park, E., Kim, H., Kim, J.M., Primack, B., Vidal-Cardenas, S., Xu, Y., Price, B.D., Mills, A.A., and D'Andrea, A.D. (2013). FANCD2 activates transcription of TAp63 and suppresses tumorigenesis. Mol Cell *50*, 908-918.
Park, I.K., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature *423*, 302-305.
Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO Journal *23*, 4061-4071.
Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-310.
Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290-1302.
Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell *12*, 1577-1589.
Pietersen, A.M., Horlings, H.M., Hauptmann, M., Langerod, A., Ajouaou, A., Cornelissen-Steijger, P., Wessels, L.F., Jonkers, J., van de Vijver, M.J., and van Lohuizen, M. (2008). EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. Breast Cancer Res *10*, R109.
Pizzatti, L., Binato, R., Cofre, J., Gomes, B.E., Dobbin, J., Haussmann, M.E., D'Azambuja, D., Bouzas, L.F., and Abdelhay, E. (2010). SUZ12 is a candidate target of the non-canonical WNT pathway in the progression of chronic myeloid leukemia. Genes, Chromosomes and Cancer *49*, 107-118.
Qi, W., Chan, H., Teng, L., Li, L., Chuai, S., Zhang, R., Zeng, J., Li, M., Fan, H., Lin, Y.*, et al.* (2012). Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. Proc Natl Acad Sci U S A *109*, 21360-21365.
Raaphorst, F.M., van Kemenade, F.J., Fieret, E., Hamer, K.M., Satijn, D.P., Otte, A.P., and Meijer, C.J. (2000). Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers. Journal of Immunology *164*, 1-4.
Rao, Z.Y., Cai, M.Y., Yang, G.F., He, L.R., Mai, S.J., Hua, W.F., Liao, Y.J., Deng, H.X., Chen, Y.C., Guan, X.Y.*, et al.* (2010). EZH2 supports ovarian carcinoma cell invasion and/or metastasis via regulation of TGF-beta1 and is a predictor of outcome in ovarian carcinoma patients. Carcinogenesis *31*, 1576-1583.
Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H.*, et al.* (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol *14*, R25.
Reddington, J.P., Sproul, D., and Meehan, R.R. (2014). DNA methylation reprogramming in cancer: does it act by re-configuring the binding landscape of Polycomb repressive complexes? Bioessays *36*, 134-140.
Ribeiro, A.F., Pratcorona, M., Erpelinck-Verschueren, C., Rockova, V., Sanders, M., Abbas, S., Figueroa, M.E., Zeilemaker, A., Melnick, A., Lowenberg, B.*, et al.* (2012). Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. Blood *119*, 5824-5831.
Richie, E.R., Schumacher, A., Angel, J.M., Holloway, M., Rinchik, E.M., and Magnuson, T. (2002). The Polycomb-group gene eed regulates thymocyte differentiation and suppresses the development of carcinogen-induced T-cell lymphomas. Oncogene *21*, 299-306.
Richter, G.H., Plehm, S., Fasan, A., Rossler, S., Unland, R., Bennani-Baiti, I.M., Hotfilder, M., Lowel, D., von Luettichau, I., Mossbrugger, I.*, et al.* (2009). EZH2 is a mediator of EWS/FLI1 driven tumor

growth and metastasis blocking endothelial and neuro-ectodermal differentiation. Proc Natl Acad Sci U S A *106*, 5324-5329.
Rizo, A., Dontje, B., Vellenga, E., de Haan, G., and Schuringa, J.J. (2008). Long-term maintenance of human hematopoietic stem/progenitor cells by expression of BMI1. Blood *111*, 2621-2630.
Rizo, A., Olthof, S., Han, L., Vellenga, E., de Haan, G., and Schuringa, J.J. (2009). Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. Blood *114*, 1498-1505.
Rubbi, L., Titz, B., Brown, L., Galvan, E., Komisopoulou, E., Chen, S.S., Low, T., Tahmasian, M., Skaggs, B., Muschen, M.*, et al.* (2011). Global phosphoproteomics reveals crosstalk between Bcr-Abl and negative feedback mechanisms controlling Src signaling. Sci Signal *4*, ra18.
Russler-Germain, D.A., Spencer, D.H., Young, M.A., Lamprecht, T.L., Miller, C.A., Fulton, R., Meyer, M.R., Erdmann-Gilmore, P., Townsend, R.R., Wilson, R.K.*, et al.* (2014). The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. Cancer Cell *25*, 442-454.
Sambucetti, L.C., Cherrington, J.M., Wilkinson, G.W., and Mocarski, E.S. (1989). NF-kappa B activation of the cytomegalovirus enhancer is mediated by a viral transactivator and by T cell stimulation. EMBO J *8*, 4251-4258.
Schanz, J., Tuchler, H., Sole, F., Mallo, M., Luno, E., Cervera, J., Granada, I., Hildebrandt, B., Slovak, M.L., Ohyashiki, K.*, et al.* (2012). New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. J Clin Oncol *30*, 820-829.
Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.
Score, J., Hidalgo-Curtis, C., Jones, A.V., Winkelmann, N., Skinner, A., Ward, D., Zoi, K., Ernst, T., Stegelmann, F., Dohner, K.*, et al.* (2012). Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood *119*, 1208-1213.
Scott, C.L., Gil, J., Hernando, E., Teruya-Feldstein, J., Narita, M., Martinez, D., Visakorpi, T., Mu, D., Cordon-Cardo, C., Peters, G.*, et al.* (2007). Role of the chromobox protein CBX7 in lymphomagenesis. Proc Natl Acad Sci U S A *104*, 5389-5394.
Senthilkumar, R., and Mishra, R.K. (2009). Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics *10*, 549.
Seward, S., Semaan, A., Qazi, A.M., Gruzdyn, O.V., Chamala, S., Bryant, C.C., Kumar, S., Cameron, D., Sethi, S., Ali-Fehmi, R.*, et al.* (2013). EZH2 blockade by RNA interference inhibits growth of ovarian cancer by facilitating re-expression of p21(waf1/cip1) and by inhibiting mutant p53. Cancer Lett *336*, 53-60.
Shah, M.Y., and Licht, J.D. (2011). DNMT3A mutations in acute myeloid leukemia. Nat Genet *43*, 289-290.
Sharawat, S.K., Bakhshi, R., Vishnubhatla, S., and Bakhshi, S. (2014). High receptor tyrosine kinase (FLT3, KIT) transcript versus anti-apoptotic (BCL2) transcript ratio independently predicts inferior outcome in pediatric acute myeloid leukemia. Blood Cells Mol Dis.
Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303-1314.
Sher, F., Boddeke, E., Olah, M., and Copray, S. (2012). Dynamic changes in Ezh2 gene occupancy underlie its involvement in neural stem cell self-renewal and differentiation towards oligodendrocytes. PLoS One *7*, e40399.
Shiogama, S., Yoshiba, S., Soga, D., Motohashi, H., and Shintani, S. (2013). Aberrant expression of EZH2 is associated with pathological findings and P53 alteration. Anticancer Res *33*, 4309-4317.

Shivarov, V., Gueorguieva, R., Stoimenov, A., and Tiu, R. (2013). DNMT3A mutation is a poor prognosis biomarker in AML: results of a meta-analysis of 4500 AML patients. Leuk Res *37*, 1445-1450.
Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol Cell *49*, 808-824.
Sinclair, J.H., Baillie, J., Bryant, L.A., Taylor-Wiedeman, J.A., and Sissons, J.G. (1992). Repression of human cytomegalovirus major immediate early gene expression in a monocytic cell line. J Gen Virol *73 (Pt 2)*, 433-435.
Slape, C.I., Saw, J., Jowett, J.B., Aplan, P.D., Strasser, A., Jane, S.M., and Curtis, D.J. (2012). Inhibition of apoptosis by BCL2 prevents leukemic transformation of a murine myelodysplastic syndrome. Blood *120*, 2475-2483.
Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nat Cell Biol *5*, 834-839.
Smith, E.R., Lee, M.G., Winter, B., Droz, N.M., Eissenberg, J.C., Shiekhattar, R., and Shilatifard, A. (2008). Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Mol Cell Biol *28*, 1041-1046.
Smith, L.L., Yeung, J., Zeisig, B.B., Popov, N., Huijbers, I., Barnes, J., Wilson, A.J., Taskesen, E., Delwel, R., Gil, J*., et al.* (2011). Functional crosstalk between Bmi1 and MLL/Hoxa9 axis in establishment of normal hematopoietic and leukemic stem cells. Cell Stem Cell *8*, 649-662.
Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci U S A *107*, 20980-20985.
Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev *27*, 2663-2677.
Sroczynska, P., Cruickshank, V.A., Bukowski, J.P., Miyagi, S., Bagger, F.O., Walfridsson, J., Schuster, M.B., Porse, B., and Helin, K. (2014). shRNA screening identifies JMJD1C as being required for leukemia maintenance. Blood *123*, 1870-1882.
Steensma, D.P., McClure, R.F., Karp, J.E., Tefferi, A., Lasho, T.L., Powell, H.L., DeWald, G.W., and Kaufmann, S.H. (2006). JAK2 V617F is a rare finding in de novo acute myeloid leukemia, but STAT3 activation is common and remains unexplained. Leukemia *20*, 971-978.
Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol *9*, 1428-1435.
Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer *123*, 8-13.
Tadokoro, Y., Ema, H., Okano, M., Li, E., and Nakauchi, H. (2007). De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. J Exp Med *204*, 715-722.
Taghon, T., Stolz, F., De Smedt, M., Cnockaert, M., Verhasselt, B., Plum, J., and Leclercq, G. (2002). HOX-A10 regulates hematopoietic lineage commitment: evidence for a monocyte-specific transcription factor. Blood *99*, 1197-1204.
Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. Cancer Cell *20*, 563-575.
Tanaka, S., Miyagi, S., Sashida, G., Chiba, T., Yuan, J., Mochizuki-Kashio, M., Suzuki, Y., Sugano, S., Nakaseko, C., Yokote, K*., et al.* (2012). Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. Blood *120*, 1107-1117.
Tiacci, E., Spanhol-Rosseto, A., Martelli, M.P., Pasqualucci, L., Quentmeier, H., Grossmann, V., Drexler, H.G., and Falini, B. (2012). The NPM1 wild-type OCI-AML2 and the NPM1-mutated OCI-AML3 cell lines carry DNMT3A mutations. Leukemia *26*, 554-557.

Tie, R., Zhang, T., Fu, H., Wang, L., Wang, Y., He, Y., Wang, B., Zhu, N., Fu, S., Lai, X.*, et al.* (2014). Association between DNMT3A mutations and prognosis of adults with de novo acute myeloid leukemia: a systematic review and meta-analysis. PLoS One *9*, e93353.
Tiu, R.V., Gondek, L.P., O'Keefe, C.L., Elson, P., Huh, J., Mohamedali, A., Kulasekararaj, A., Advani, A.S., Paquette, R., List, A.F.*, et al.* (2011). Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. Blood *117*, 4552-4560.
Tokimasa, S., Ohta, H., Sawada, A., Matsuda, Y., Kim, J.Y., Nishiguchi, S., Hara, J., and Takihara, Y. (2001). Lack of the Polycomb-group gene rae28 causes maturation arrest at the early B-cell developmental stage. Experimental Hematology *29*, 93-103.
Tong, Q., He, S., Xie, F., Mochizuki, K., Liu, Y., Mochizuki, I., Meng, L., Sun, H., Zhang, Y., Guo, Y.*, et al.* (2014). Ezh2 regulates transcriptional and posttranslational expression of T-bet and promotes Th1 cell responses mediating aplastic anemia in mice. J Immunol *192*, 5012-5022.
Tonini, T., D'Andrilli, G., Fucito, A., Gaspa, L., and Bagella, L. (2008). Importance of Ezh2 polycomb protein in tumorigenesis process interfering with the pathway of growth suppressive key elements. Journal of Cellular Physiology *214*, 295-300.
Tonkin, E., Hagan, D.M., Li, W., and Strachan, T. (2002). Identification and characterisation of novel mammalian homologues of Drosophila polyhomeoticpermits new insights into relationships between members of the polyhomeotic family. Human Genetics *111*, 435-442.
Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P.*, et al.* (2011). L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. Mol Cell *42*, 438-450.
Tsai, M.C., Wang, J.K., and Chang, H.Y. (2010). Tumor suppression by the histone demethylase UTX. Cell Cycle *9*.
Ueda, K., Yoshimi, A., Kagoya, Y., Nishikawa, S., Marquez, V.E., Nakagawa, M., and Kurokawa, M. (2014). Inhibition of histone methyltransferase EZH2 depletes leukemia stem cell of mixed lineage leukemia fusion leukemia through upregulation of p16. Cancer Sci *105*, 512-519.
Ura, H., Usuda, M., Kinoshita, K., Sun, C., Mori, K., Akagi, T., Matsuda, T., Koide, H., and Yokota, T. (2008). STAT3 and Oct-3/4 control histone modification through induction of Eed in embryonic stem cells. Journal of Biological Chemistry *283*, 9713-9723.
Valent, P. (2012). Low blood counts: immune mediated, idiopathic, or myelodysplasia. Hematology Am Soc Hematol Educ Program *2012*, 485-491.
van den Boom, V., Rozenveld-Geugien, M., Bonardi, F., Malanga, D., van Gosliga, D., Heijink, A.M., Viglietto, G., Morrone, G., Fusetti, F., Vellenga, E.*, et al.* (2013). Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. Blood *121*, 2452-2461.
van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M.*, et al.* (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes Dev *8*, 757-769.
van der Vlag, J., and Otte, A.P. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nature Genetics *23*, 474-478.
van Haaften, G., Dalgliesh, G.L., Davies, H., Chen, L., Bignell, G., Greenman, C., Edkins, S., Hardy, C., O'Meara, S., Teague, J.*, et al.* (2009). Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet *41*, 521-523.
van Zutven, L.J., Onen, E., Velthuizen, S.C., van Drunen, E., von Bergh, A.R., van den Heuvel-Eibrink, M.M., Veronese, A., Mecucci, C., Negrini, M., de Greef, G.E.*, et al.* (2006). Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. Genes Chromosomes Cancer *45*, 437-446.
Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P.*, et al.* (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624-629.

Velichutina, I., Shaknovich, R., Geng, H., Johnson, N.A., Gascoyne, R.D., Melnick, A.M., and Elemento, O. (2010). EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. Blood *116*, 5247-5255.
Verena I. Gaidzik, M., Richard F. Schlenk, MD[2], Peter Paschka, MD[2*], Claus-Henning Köhne, MD[3*], Gerhard Held, MD[4*], Marianne Habdank[5*], Juliane Gohlke[2*], Hans-Günther Mergenthaler, MD[6*], Hans J. Salwender, MD[7*], Lars Bullinger[8], Hartmut Döhner, MD[2] and Konstanze Döhner, MD[2] (2010). *97 TET2 mutations in Acute Myeloid Leukemia (AML): results on 783 patients treated within the AML HD98A study of the AML Study Group (AMLSG)*. ASH abstract.
Vidal, M. (2009). Role of polycomb proteins Ring1A and Ring1B in the epigenetic regulation of gene expression. International Journal of Developmental Biology *53*, 355-370.
Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proceedings of the National Academy of Sciences of the United States of America *105*, 16572-16577.
Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M*., et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.
Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell *6*, 153-166.
Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847-851.
Wang, J., Ai, X., Gale, R.P., Xu, Z., Qin, T., Fang, L., Zhang, H., Pan, L., Hu, N., Zhang, Y*., et al.* (2013a). TET2, ASXL1 and EZH2 mutations in Chinese with myelodysplastic syndromes. Leuk Res *37*, 305-311.
Wang, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilangovan, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B*., et al.* (2010). Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure *18*, 966-975.
Wang, X., Dai, H., Wang, Q., Xu, Y., Wang, Y., Sun, A., Ruan, J., Chen, S., and Wu, D. (2013b). EZH2 mutations are related to low blast percentage in bone marrow and -7/del(7q) in de novo acute myeloid leukemia. PLoS One *8*, e61341.
Wong, C.K., Chen, Z., So, K.L., Li, D., and Li, P. (2007a). Polycomb group protein RING1B is a direct substrate of Caspases-3 and -9. Biochimica et Biophysica Acta *1773*, 844-852.
Wong, C.K., Chen, Z., So, K.L., Li, D., and Li, P. (2007b). Polycomb group protein RING1B is a direct substrate of Caspases-3 and -9. Biochim Biophys Acta *1773*, 844-852.
Wu, X., Gong, Y., Yue, J., Qiang, B., Yuan, J., and Peng, X. (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. Nucleic Acids Res *36*, 3590-3599.
Xiang, Y., Zhu, Z., Han, G., Lin, H., Xu, L., and Chen, C.D. (2007). JMJD3 is a histone H3K27 demethylase. Cell Res *17*, 850-857.
Xie, H., Xu, J., Hsu, J.H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S.H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell *14*, 68-80.
Xie, W., Wang, X., Du, W., Liu, W., Qin, X., and Huang, S. (2010). Detection of molecular targets on the surface of CD34+CD38- bone marrow cells in myelodysplastic syndromes. Cytometry A *77*, 840-848.
Xu, F., Li, X., Wu, L., Zhang, Q., Yang, R., Yang, Y., Zhang, Z., He, Q., and Chang, C. (2011a). Overexpression of the EZH2, RING1 and BMI1 genes is common in myelodysplastic syndromes: relation to adverse epigenetic alteration and poor prognostic scoring. Ann Hematol *90*, 643-653.
Xu, J., Wang, Y.Y., Dai, Y.J., Zhang, W., Zhang, W.N., Xiong, S.M., Gu, Z.H., Wang, K.K., Zeng, R., Chen, Z*., et al.* (2014). DNMT3A Arg882 mutation drives chronic myelomonocytic leukemia through

disturbing gene expression/DNA methylation in hematopoietic cells. Proc Natl Acad Sci U S A *111*, 2620-2625.
Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Xiao, M.T., Liu, L.X.*, et al.* (2011b). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell *19*, 17-30.
Yadav, A.K., Sahasrabuddhe, A.A., Dimri, M., Bommi, P.V., Sainger, R., and Dimri, G.P. (2010). Deletion analysis of BMI1 oncoprotein identifies its negative regulatory domain. Mol Cancer *9*, 158.
Yamashita, Y., Yuan, J., Suetake, I., Suzuki, H., Ishikawa, Y., Choi, Y.L., Ueno, T., Soda, M., Hamada, T., Haruta, H.*, et al.* (2010). Array-based genomic resequencing of human leukemia. Oncogene *29*, 3723-3731.
Yan, J., Ng, S.B., Tay, J.L., Lin, B., Koh, T.L., Tan, J., Selvarajan, V., Liu, S.C., Bi, C., Wang, S.*, et al.* (2013). EZH2 overexpression in natural killer/T-cell lymphoma confers growth advantage independently of histone methyltransferase activity. Blood *121*, 4512-4520.
Yan, X.J., Xu, J., Gu, Z.H., Pan, C.M., Lu, G., Shen, Y., Shi, J.Y., Zhu, Y.M., Tang, L., Zhang, X.W.*, et al.* (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet *43*, 309-315.
Yap, D.B., Chu, J., Berg, T., Schapira, M., Cheng, S.W., Moradian, A., Morin, R.D., Mungall, A.J., Meissner, B., Boyle, M.*, et al.* (2011). Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood *117*, 2451-2459.
Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Molecular Cell *38*, 662-674.
Yoshimi, A., Goyama, S., Watanabe-Okochi, N., Yoshiki, Y., Nannya, Y., Nitta, E., Arai, S., Sato, T., Shimabe, M., Nakagawa, M.*, et al.* (2011). Evi1 represses PTEN expression and activates PI3K/AKT/mTOR via interactions with polycomb proteins. Blood *117*, 3617-3628.
Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. Nat Rev Mol Cell Biol *9*, 47-59.
Young, M.D., Willson, T.A., Wakefield, M.J., Trounson, E., Hilton, D.J., Blewitt, M.E., Oshlack, A., and Majewski, I.J. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic Acids Res *39*, 7415-7427.
Yu, H., Simons, D.L., Segall, I., Carcamo-Cavazos, V., Schwartz, E.J., Yan, N., Zuckerman, N.S., Dirbas, F.M., Johnson, D.L., Holmes, S.P.*, et al.* (2012). PRC2/EED-EZH2 complex is up-regulated in breast cancer lymph node metastasis compared to primary tumor and correlates with tumor proliferation in situ. PLoS One *7*, e51239.
Yuan, J.B., Yang, L.Y., Tang, Z.Y., Zu, X.B., and Qi, L. (2012). Down-regulation of EZH2 by RNA interference inhibits proliferation and invasion of ACHN cells via the Wnt/beta- catenin pathway. Asian Pac J Cancer Prev *13*, 6197-6201.
Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.
Zee, B.M., Britton, L.M., Wolle, D., Haberman, D.M., and Garcia, B.A. (2012). Origins and formation of histone methylation across the human cell cycle. Mol Cell Biol *32*, 2503-2514.
Zeidler, M., Varambally, S., Cao, Q., Chinnaiyan, A.M., Ferguson, D.O., Merajver, S.D., and Kleer, C.G. (2005). The Polycomb group protein EZH2 impairs DNA repair in breast epithelial cells. Neoplasia *7*, 1011-1019.
Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A *100*, 9779-9784.
Zhang, J., Grindley, J.C., Yin, T., Jayasinghe, S., He, X.C., Ross, J.T., Haug, J.S., Rupp, D., Porter-Westpfahl, K.S., Wiedemann, L.M.*, et al.* (2006). PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. Nature *441*, 518-522.

Zhang, Y., Li, X., Chen, Z., and Bepler, G. (2014). Ubiquitination and degradation of ribonucleotide reductase M1 by the polycomb group proteins RNF2 and Bmi1 and cellular response to gemcitabine. PLoS One *9*, e91186.
Zhang, Y.B., Niu, H.T., Chang, J.W., Dong, G.L., and Ma, X.B. (2011). EZH2 silencing by RNA interference inhibits proliferation in bladder cancer cell lines. Eur J Cancer Care (Engl) *20*, 106-112.