TRANSCRIPTIONAL REGULATORY MECHANISMS IN STEM CELL MAINTENANCE AND DIFFERENTIATION

ANEESHA NATH

PhD THESIS

2020

SREE CHITRA TIRUNAL INSTITUTE

FOR

MEDICAL SCIENCES AND TECHNOLOGY

THIRUVANANTHAPURAM

TRANSCRIPTIONAL REGULATORY MECHANISMS IN STEM CELL MAINTENANCE AND DIFFERENTIATION

A THESIS PRESENTED BY

ANEESHA NATH

TO

SREE CHITRA TIRUNAL INSTITUTE

FOR

MEDICAL SCIENCES AND TECHNOLOGY

Thiruvananthapuram

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF

DOCTOR OF PHILOSOPHY

2020

DECLARATION

I, **Aneesha Nath,** hereby certify that I have personally carried out the work depicted in the thesis entitled **"Transcriptional regulatory mechanisms in Stem Cell maintenance and Differentiation",** except where due acknowledgement has been made in the text.

No part of thesis has been submitted for the award of any degree or diploma prior to this date.

Signature: Aneesha Nath

Name of the candidate**: Aneesha Nath**

Registration number**: 2014/PhD (CMCV)/20**

Date**: 25th September 2020**

CERTIFICATE

This is to certify that **Aneesha Nath** in the **Centre for Stem Cell Research of Christian Medical College** has fulfilled the requirements prescribe for the Ph.D. degree of the Sree Chitra Tirunal Institute for Medical Science and Technology, Thiruvananthapuram**.** The thesis entitled**, Transcriptional regulatory mechanisms in stem cell maintenance and differentiation"** was carried out under my direct supervision. No part of the thesis was submitted for the award of any degree or diploma prior this date.

Signature of the Guide **Date: 25th September 2020**

Dr.R.V.Shaji
Professor
Department of Haematology
Adjunct Scientist, Centre for Stem Cell Research

Shaji R.V. Ph.D.,
Professor, Dept. of Haematology,
Adjunct Scientist, Centre for Stem Cell Research,
(A Unit of inStem, Bengaluru)
Christian Medical College Campus,
Bagayam, Vellore - 632 002, Tamil Nadu. India.

Christian Medical College Campus,
Bagayam, Vellore -632002. INDIA.
Phone: +91 - 416 - 307 5101 / 5107
Fax: +91 - 416 - 307 5103.
E-mail: cscr@cmcvellore.ac.in

The thesis entitled

Transcriptional Regulatory Mechanisms in Stem Cell Maintenance and Differentiation

Submitted by

Aneesha Nath

for the degree of

Doctor of Philosophy

of

THE SREE CHITRA TIRUNAL INSTITUTE

FOR

MEDICAL SCIENCES AND TECHNOLOGY, TRIVANDRUM

Thiruvananthapuram

Is evaluated and approved by

Dr. R.V. SHAJI
Name and signature of the guide

PROF. DR. ERNEST DAVID
Name and signature of thesis examiner

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my mentor **Dr.R.V.Shaji**, Professor, Department of Haematology. Sir, the amount of dedication and efforts which you put into research stands undisputed and what is more remarkable is the patience and unconditional support which you have provided me during my entire tenure as your student. I whole heartedly thank you Sir for your guidance; valuable suggestions, constructive criticism and empathetic attitude towards me which has helped me evolve as a research scholar. I feel privileged to be your student Sir and you are one of the best guides a student can have which is why I can proudly proclaim I never hated my Phd days.

I am indebted to my DAC members **Dr. Poonkuzhali B, Dr. Sanjay Kumar, Dr. Sreenivasulu Kurukuti** and **Dr. Saravanabhavan Thangavel** for taking time out from their busy schedule and making themselves available for me to attend my presentations and DAC meetings. Their criticisms and suggestions motivated me to be on the right track and adhere to my schedule. Their insightful comments while I was writing my thesis helped me immensely in finalising my thesis. I am also thankful to the other faculty members for their valuable insights during my work presentations.

I extend my gratitude to **Dr. Alok Srivastava**, Prof and Head, Centre for Stem Cell Research, CMC for his valuable advice and critical suggestions during work presentations. I am thankful to him for providing me with the state of the art facilities to carry out my research work.

I am grateful to my seniors **Dr. Sumitha, Dr. Kannan, Dr. Musheer, Dr. Nancy and Dr. Thiyagaraj** for their help and suggestions during my PhD. I am extremely thankful to **Dr. Janakiram and Dhavapriya** for providing me training in the initial phase of my PhD with the various molecular biology and cell culture techniques which helped me carry out my experiments independently. I am thankful to my juniors especially **Abhirup, Smitha and Krittika** who helped me in carrying out several experiments.

I acknowledge **Saranya J, Abdul, Immani, and Dr Sandhya** for helping me execute my flow experiments smoothly. Their expertise and knowledge on flow cytometry helped me plan my experiments in a better way. I am also thankful to **Mr. Rajesh** for being available if any technical issues crop up. His timely intervention and prompt attitude helped in saving a lot of time and also reduced the tediousness of my work.

I am thankful to my friends and colleagues **Sowmya Ramesh, Franklin Herbert, Archana Kini, Karthikeyan, Elizabeth Pannakal, Smitha Ijee, Abisha Crystal, Thulaj Dattatreya Meherwade, Aishwarya Prasannan, Vignesh Rajendran,Vigneshwaran Venkatesan** for always being there whenever I needed any sort of help. I owe it to you **Sowmya, Smitha and Thulaj** for always being there for me during my difficult times and making my PhD days a little better. I also thank my friends **Sambit, Sushree, Madhusmita, Rajashree, Soumya, Sunit, Sourav** for always being there during my highs and lows.

I thank my Mom and Dad and my other family members who have always encouraged me to follow my dreams which motivated me to pursue my PhD.

I would also like to acknowledge the funding received from the **Department of Biotechnology** and **University Grants Commission**, India.

Last but not the least I thank **God** for providing me with courage and patience to overcome the tough days in my life.

Aneesha Nath

Table of contents

Contents	Page
Declaration by the student	i
Certificate by the guide	ii
Approval of the thesis	iii
Acknowledgement	iv
Table of contents	vii
List of figures	xi
List of tables	xiii
Synopsis	xiv
1.Introduction	1
1.1 Objectives of the study	3
2.Review of literature	6
2.1 Eukaryotic transcription and regulation	6
2.1.1 Transcription machinery in Eukaryotes	6
2.1.2 Regulation of transcription	10
2.1.2.1 Transcription factors	10
2.1.2.2 Transcriptional activators	11
2.1.2.3 Transcriptional repressors	11
2.1.2.4 Chromatin structure and transcription	12
2.2 Non coding RNAs and transcriptional regulation	13
2.2.1 Importance of studying miRNAs	15
2.3 Haematopoiesis	17
2.3.1Sites of Haematopoiesis	20
2.3.1.1 Yolk Sac	20
2.3.1.2 The Aorta-Gonad-Mesonephros (AGM) Region	21
2.3.1.3 The Liver	22
2.3.1.4 Bone Marrow	22
2.4 Regulation of Haematopoiesis	23
2.4.1.1 Transcription factors regulating haematopoiesis	23
2.4.1.1 SCL/TAL1	23
2.4.1.2 RUNX	24
2.4.1.3 GATA 1	25
2.4.1.4 GATA 2	26
2.5 Signalling pathways and Haematopoiesis	27
2.6 miRNAs and Haematopoiesis	29
2.7 Growth factors and Haematopoiesis	31
2.8 Erythropoiesis	32
2.9 Regulation of Erythropoiesis	35
2.9.1 Transcription factors in erythropoiesis	35
2.9.1.1 GATA Family	36
2.9.1.2 KLF1	37

2.9.1.3 NF-E2 40
2.9.2 Long non coding RNAs in erythropoiesis 41
2.9.3 Epigenetic factors in erythropoiesis 43
2.9.4 miRNAs in erythropoiesis 46
2.9.4.1 miRNA biogenesis 46
2.9.4.2 Canonical pathway of miRNA biogenesis 47
2.9.4.3 Non Canonical pathway of miRNA biogenesis 49
2.9.4.4 Regulation of miRNA biogenesis 51
2.10 Phosphorylation 53
2.11 Ubiquitylation and Sumoylation 54
2.12 Mode of action of miRNAs 54
2.13 Functional validation of miRNAs 58
2.14 miRNAs regulating erythroid differentiation and maturation 59
2.15 miRNAs regulating iron metabolism 62
2.16 miRNAs in haemoglobin synthesis 62
2.17 Model systems to study erythropoiesis 63
2.17.1 Immortalized cell lines 64
2.17.2 Embryonic stem cells 66
2.17.3 Animal models for erythropoiesis 67
2.18 Lentiviral Promoters in HSCs 67
2.19 Tools to study transcriptional regulation in erythropoiesis 68
3.Materials and Methods 70
3.1 Plasmid Extraction 70
3.2 Transformation 71
3.3 Cloning shRNAs in miRE shRNA backbone 71
3.4 Cloning shRNAs in UltramiR backbone using Gibson Assembly 72
3.5 Construction of UltramiR pGIPZ backbone 72
3.6 Cloning of shRNAs by Gibson cloning 73
3.7 Construction of miRNA overexpression vectors 75
3.8 Generation of lentiviral plasmids with different promoters 77
3.9 Transfection 78
3.10 Lentivirus Preparation 79
3.11 Transduction of adherent cells 80
3.12 Transduction of suspension cells 80
3.13 DNA sequencing 80
3.14 RNA extraction 82
3.15 cDNA synthesis 83
3.16 Real Time PCR 83
3.17 Isolation of peripheral blood CD 34+HSPCs 83
3.18 *Ex-vivo* erythropoiesis 84
3.18.1 Protocol 1 84
3.18.2 Protocol 2 85
3.19 Flow cytometry analysis of surface markers 85
3.20 Morphology analysis of erythroid cells 86

3.21 Haemoglobin analysis 86
3.22 Analysis of erythroid gene expression 86
3.23 Small RNA Sequencing 87
3.24 Analysis of transcription factor occupancy at the miRNA promoters 87
3.25 Analysis of erythroid transcriptome data 88
3.26 Analysis of miRNA targets 88
3.27 Pathway Analysis of miRNA target genes 88
3.28 Validation of miRNA expression by quantitative PCR 89
3.29 DNA Extraction 89
3.30 Construction of shRNA library 90
3.31 RNAi screening for signalling pathway genes 91
3.32 RNAi screening for HSC maintenance and differentiation genes 93
3.33 Culture and differentiation of HUDEP cells 94
3.34 Gene editing of miRNAs in HUDEP cells 95
3.34.1 Design of gRNAs 95
3.34.2 Cloning of gRNAs 95
3.35.3 Generation of Cas9 expressing HUDEP cells 96
3.35.4 T7 Endonuclease (T7E) assay 97
Results 106
4.miRNOME screening in human erythropoiesis 106
4.1 Introduction 106
4.2 A Two-Phase *ex vivo* Culture System for the Generation of Erythroid Cells 108
4.3 Small RNA sequencing of cultured erythroid cells. 109
4.4 Co-regulation of miRNAs and host genes 119
4.5 Multiple miRNA clusters are involved in regulating human erythropoiesis 120
4.6 Occupancy of erythroid-specific transcription factors near the upregulated miRNAs 123
4.7 Knocking out miRNAs using CRISPR-Cas9 126
4.8 Overexpression of upregulated and downregulated miRNAs did not change the kinetics of erythropoiesis 129
4.9 miRNA target analysis 131
4.10 Discussion 133
5 RNAi screen to delineate novel signalling regulators in HSC maintenance and erythropoiesis. 136
5.1 Introduction 136
5.2 Primary shRNA library screening to identify novel signalling regulators in erythropoiesis 137
5.3 Genes previously reported in haematopoiesis and erythropoiesis identified by RNAi screening in ex-vivo erythropoiesis 141
5.4 Novel genes identified by RNAi screening 144
5.5 Discussion 147
6. RNAi Screen to delineate novel regulators in HSC maintenance and erythropoiesis 150
6.1 Introduction 150
6.2 Selection of genes for the RNAi screening 154
6.3 Construction of an efficient lentiviral shRNA knockdown vector 159
6.4 Assessment of the activity of promoters during erythroid differentiation 161

6.5 Generation of a shRNA library targeting HSC maintenance genes 164
6.6 Validation of shRNA library in the ex vivo model of HSC differentiation 164
6.7 Bioinformatics to identify the genes that affect human erythropoiesis 167
6.8 Discussion 178
7.0 Summary and Conclusions 179
7.1 Hallmarks of the thesis 182
7.2 Future Directions 182
Bibliography 184

List of figures

Figure No	Caption	Page No
2.1.1	Transcription in Eukaryotes	8
2.2.1	Biological processes regulated by miRNAs	17
2.3	Different models of haematopoiesis	18
2.3.1	Anatomical sites of haematopoiesis in the foetal and adult system	21
2.6	miRNAs in the regulation of haematopoiesis	30
2.8	*In vivo* erythropoiesis	33
2.9	Regulation of erythropoiesis	35
2.9.4.2	Canonical miRNA biogenesis	48
2.9.4.3	Non canonical miRNA biogenesis	50
2.12	Mechanism of miRNA mediated downregulation	56
3.1	Generation of a lentiviral vector for shRNA cloning	74
3.2	Generation of a lentiviral shRNA vector using Gibson Assembly	75
3.3	Construction of miRNA expression vector	77
3.4	RNAi screening workflow	92
4.2.1	Schematic representation of the *ex-vivo* erythroid culture system used for the study	110
4.2.2	Flow cytometry analysis of the expression of transferrin receptor (CD 71) and glycophorin A (CD 235a) during *ex-vivo* erythroid differentiation	111
4.2.3 A	HPLC analysis showing the expression levels of different forms of haemoglobins in the cultured erythroid cells	112
4.2.3 B	Expression of HBB, HBA and HBG genes erythroid during differentiation	112
4.3.1	Heat map of miRNA expression	113
4.3.2	Global downregulation of miRNAs observed during the process of erythroid differentiation	116
4.3.3	Validation of small RNA Seq data by quantitative real-time PCR	116
4.4.1	Distribution of intragenic and intergenic miRNAs, which are differentially expressed in human erythropoiesis. Significant correlation between the expression of miRNAs and host genes suggesting shared transcriptional machinery	119
4.6.1	Occupancy of erythroid specific transcription factors within 5kb of selected miRNAs.	124
4.7.1	T7E analysis of the PCR products from miR-182 (lanes 1-3), miR-183 (lanes 4-6) and miR-96 (lanes 7-9) genomic regions	128

4.7.2	Flow cytometry analysis of CD71 and CD235a in the differentiated HUDEP cells in which miR-144; miR-451, miR-182, miR-183, miR-4732 and miR-96 were knocked out using CRISPR-Cas9 gene-editing method.	130
4.8.1	Expression levels of CD71 and CD235a in the cultured erythroid cells from different days of erythroid differentiation transduced with miRNA OE vectors	131
4.9.1	Pathways activated by the miRNAs in human erythropoiesis. (A) Bioplanet pathways (B) Wikipathways.	132
5.2.1	Pictorial map of the lentiviral vector used for the RNAi library for signalling pathway genes.	138
5.2.2 A	Schematic of ex-vivo erythropoiesis protocol.	138
5.2.2 B	Expression levels of CD71 and CD235a in the cultured erythroid cells from different days of erythroid differentiation.	138
5.2.3	Amplification of DNA with specific primers for NGS.	139
5.2.4	Graph showing enrichment of shRNAs.	140
5.2.5	Expression levels of the top hit genes identified by RNAi screen.	140
6.2.1	Predominant pathways in which the genes of RNAi are involved in.	156
6.2.2	Expression of the genes selected in the study in ex-vivo erythropoiesis	157
6.3.1 A	Representative GFP image depicting transfection efficiency of shRNA lentiviral plasmid in 293 T cells	159
6.3.1B	Representative GFP image showing transduction efficiency of shRNA lentiviral plasmid in HeLa cells	159
6.3.2 A	Knockdown efficiencies of different genes in pGIPZ-mirE-hCMV vector	160
6.3.2 B	Knockdown efficiencies of different genes in pZIP-UltramiR-hCMV and pZIP-UltramiR-hEF1α vectors.	160
6.4.1	Lentiviral construct with different promoters used in this study	161
6.4.2	Assessment of the activity of different lentiviral vector promoters during erythroid differentiation as estimated by quantitating the levels of GFP	163
6.5.1	Generation of a pooled shRNA library	165
6.6.1	Flow cytometry analysis to quantitate the percentage of GFP for estimation of transduction efficiency	166
6.6.2	Flow cytometry analysis of erythroid markers CD71 and CD235a of the cells on day 4,9,13 and 19, respectively.	166
6.6.3	Amplification of DNA with specific primers for NGS	167
6.7.1	The correlation plot of the samples analysed by NGS	168
6.7.2	Wikipathway analysis of the hits identified by RNAi screening.	174

List of Tables

Table No	Title	Page No
2.5	Signalling pathways involved in haematopoiesis	28
2.13	miRNAs play a role in several aspects of erythropoiesis	58
3.1	Cell culture media used in the study	98
3.2	Plasmids used in the study	99
3.3	Oligos used in the study	99
3.4	Flow cytometry antibodies used in this study	102
3.5	Sequence of shRNAs used for knocking down the expression of target genes.	102
3.6	shRNA specific barcode amplification PCR (Cellecta signalling Pathway shRNA library).	103
3.7	PCR to amplify and clone pooled shRNA oligos	104
3.8	shRNA specific amplification PCR	105
4.3.1	List of upregulated miRNAs with log2 FC≥+3	114
4.3.2	List of downregulated miRNAs with log2 FC≤-3	115
4.3.3	Previously reported miRNAs in erythropoiesis detected in our study	117
4.3.4	miRNAs with uniform expression throughout ex-vivo erythropoiesis	118
4.5.1	The top seven miRNA clusters differentially expressed in ex-vivo erythropoiesis	121
4.5.2	New miRNA clusters identified in erythropoiesis	122
5.1	The summary of the genes that showed fold change >3 and their functions in erythropoiesis.	149
6.2.1	Top 117 genes involved in HSC maintenance and differentiation identified by bioinformatics analysis	155
6.2.2	Gene ontology analysis of the genes selected for the RNAi screening in human erythropoiesis.	158
6.2.3	Major pathways in which the selected genes are involved in.	158
6.7.1	The genes with most depleted shRNAs in the RNAi screening of ex-vivo erythropoiesis.	169
6.7.3	Binding of transcription factors at the promoters of the genes with high depletion of shRNAs	175

PhD Synopsis

TRANSCRIPTIONAL REGULATORY MECHANISMS IN STEM CELL MAINTENANCE AND DIFFERENTIATION

Background of the study

Haematopoietic stem cells (HSCs) are one of the most versatile cells in the human body which are characterised by their ability to replenish themselves as well as simultaneously giving rise to mature blood lineages. With respect to haematopoietic differentiation, the process of erythropoiesis (formation of red blood cells) has been well studied. Erythropoiesis is a dynamic and multi-staged process in which mature enucleated erythrocytes are formed from multipotent haematopoietic stem cells. Each successive stage of erythroid differentiation is phenotypically distinct from its antecedent stage and this multifarious process of erythropoiesis is under tight regulation by an intertwined network of transcription factors, epigenetic modifiers, signalling pathways as well as long and short non coding RNAs. One of the hallmarks of erythropoiesis is apoptosis and extrusion of majority of the cellular organelles in the terminal stages of termination which leads to a shutdown of the transcriptional machinery during erythroid differentiation. Several studies carried out in different *ex- vivo* and *in- vivo* erythroid model systems have led to the identification of several factors involved in erythroid differentiation and the subsequent functional validation studies have unearthed their specific roles in erythropoiesis. These factors include genes involved in epigenetic regulation, signalling pathways, terminal erythroid maturation and iron metabolism. Even

though these studies have led to a better understanding of the complex process of erythropoiesis, the mechanisms of erythropoiesis is not completely understood. Several key regulators still need to be identified to better comprehend the mechanisms of erythropoiesis in normal individuals and in the patients with the diseases that affect erythropoiesis. This will help in understanding the pathogenesis of these diseases and in developing therapeutical approaches to treat these diseases.

Even though several studies have been carried out to emphasize the importance of miRNAs in erythropoiesis a comparative profiling of miRNAs between HSCs and differentiated erythroid cells using small RNA sequencing has not been carried out yet. In this study, small RNA analysis was used to identify candidate miRNAs playing crucial roles during the process of erythropoiesis. Similarly, a very few signalling pathways have been reported to play a role in erythropoiesis. Considering the complexity of the process many other signalling pathways might also play a crucial role. Therefore, in this study a high throughput RNAi screening approach was used to identify novel regulators in human erythropoiesis.

Aim of the study: Understanding the transcriptional regulatory mechanisms in human erythropoiesis.

Objectives of the study

1. miRNome screening in human erythropoiesis.

2. Identification of lentiviral vector promoters that show stable activity during the course of erythroid differentiation.

3. Establishment of efficient lentiviral knockdown tools for high throughput RNAi screen in *ex-vivo* erythropoiesis.

4. Identification of novel signalling pathways involved in erythroid differentiation by RNA interference (RNAi).

5. Identification of the genes involved in haematopoietic stem cell maintenance and erythroid differentiation.

Experimental methods used in the study

1. Generation of erythroid cells from $CD34^+$ haematopoietic stem and progenitor cells (HSPCs) using *ex- vivo* erythroid culture systems.

2. Validation of *ex- vivo* erythroid culture system using:

a) Real time PCRs to analyze the expression of globin genes.

b) Giemsa staining to assess the morphology of the erythroid cells.

c) Flow cytometry analysis to assess the expression of erythroid differentiation markers.

d) HPLC analysis to quantitate the different forms of haemoglobins.

3. Identification of lentiviral vector promoters suitable for performing experiments in *ex-vivo* erythropoiesis by transducing $CD34^+$ HSPCs with lentiviral vectors harbouring the promoters such as CBA, UbC, SFFV, mCMV, hCMV, mEF1α, hEF1α and MND.

4. Generation of lentiviral overexpression, knockdown and knockout vectors for

regulating the expression of target genes during erythroid differentiation.

5. RNAi screening of signalling pathway genes using a shRNA library and identification of enriched and depleted shRNAs during the process of erythroid differentiation by next generation sequencing.

6. Construction of a mini shRNA library targeting genes involved in HSC maintenance.

7. RNAi screen of HSC maintenance genes using a shRNA library.

8. Small RNA sequencing of erythroid cells collected at different stages of erythroid differentiation to delineate the miRNA profile during *ex- vivo* human erythropoiesis.

9. Functional validation of miRNAs identified from the small RNA sequencing data by knocking them out in HSCs by a CRISPR-Cas9 based gene editing method.

Major Findings

1. An efficient *ex- vivo* erythroid culture system was established in the laboratory which recapitulated the process of erythropoiesis *in- vivo*. The culture conditions used was suitable for prolonged maintenance of HSPCs and erythroid progenitors. In addition to this, the culture conditions were also amenable to carry out high throughput RNAi and small RNA sequencing studies. The cultured erythroid cells were fully characterized for their expression of erythroid specific markers, quantitation of different forms of haemoglobins and morphology.

2. A robust RNA polymerase II based lentiviral knockdown system was constructed

in house using UltramiR scaffold into which the shRNAs selected from Sherwood algorithm were cloned. With this combination we were able to achieve more than 70% knockdown efficiency of the target genes as validated in HeLa cells and by real time PCR.

3. In order to find the best promoter of lentiviral vectors which does not get silenced during the entire course of erythroid differentiation, lentiviral vectors with different promoters which drive the expression of GFP were evaluated. It was observed that MND, SFFV and hCMV promoters did not exhibit transgene silencing and they were the most suitable promoters to carry out knockdown and overexpression studies in *ex- vivo* erythropoiesis.

4. In order to study the complete miRNA profile during the process of *ex-vivo* human erythroid differentiation, cells were collected at different stages of erythropoiesis and subjected to small RNA sequencing. A phenomenon of global downregulation was observed as majority of the miRNAs were found to be downregulated. In addition to this, several new miRNAs such as miR-182 and miR-183 which were not previously studied in human erythropoiesis were identified in this study. This study also helped in the identification of several miRNA clusters which were previously not identified in the context of human erythropoiesis. Functional validation of the miRNAs was carried out by knocking out miRNAs using CRISPR-Cas9 in adult HSCs. This study was first of its kind where in gene editing was used to disrupt the function of miRNAs in adult HSCs and their effect was studied on erythroid differentiation.

5. For elucidating the role of various signalling pathways in the process of erythropoiesis, adult $CD34^+$ HSPCs were transduced with lentiviral vectors which

expressed shRNAs against 5000 human signalling pathways, and the transduced HSPCs were subjected to erythroid differentiation. Cells were collected before and after differentiation and were analyzed for the integrated shRNAs by next generation sequencing. Based on the specific shRNAs enriched or depleted in the differentiated cells specific signalling pathways involved in erythroid differentiation was identified.

6. Genes which were found to play a role in haematopoietic stem cell maintenance were selected based on meta- analysis of the previously published gene expression data sets. A custom-made lentiviral shRNA library was made with the shRNAs targeting these genes and it was used for transducing $CD34^+$ HSPCs. Cells were collected at various time points before and after differentiation and were analysed for the integrated shRNAs by next generation sequencing. Based on the specific shRNAs enriched or depleted in the differentiated cells specific genes involved in stem cell maintenance and erythroid differentiation were identified.

Significance and implications of the findings

Erythropoiesis is a dynamic process wherein cells traverse through various stages to form mature erythrocytes from HSCs. Even though several studies have delineated many factors which play significant roles in erythropoiesis, a lot still remains to be understood at the molecular level. In this study, the potential role of several miRNAs and signalling pathways associated with *in- vitro* human erythroid differentiation was elucidated. The findings from this study could deepen our understanding about the molecular mechanisms of erythropoiesis, and the knowledge gained can be extrapolated to understand ineffective erythropoiesis in several human red cell diseases. The signalling pathways identified in this study can help in understanding

disease pathogenesis better in case of haematological disorders such as polycythemia vera wherein the JAK-STAT signalling pathway is perturbed. In addition, several miRNAs identified in this study can serve as potent biomarkers in case of haematological malignancies.

1. INTRODUCTION

The haematopoietic system is one of the most versatile systems of the human body, and it is distinguished by the salient characteristics of self-renewal and its ability to differentiate into several lineages such as erythrocytes, platelets, leucocytes, B cells, and T cells (Kato and Igarashi, 2019). The regulation of the haematopoietic system is one of the most crucial physiological processes since it is susceptible to be subjected by various types of genotoxic insults (Thomas, Vadas and Lopez, 2004). Such a system must be endowed with several layers of checkpoints to avoid obliteration of the normal homeostatic process (Thomas, Vadas and Lopez, 2004). Several studies have been carried out to understand the key regulators playing a role in tweaking the balance between self-renewal and differentiation. These studies have led to the identification of transcription factors, epigenetic modifiers, growth factors, signalling pathways and non-coding RNAs. *In-vitro* erythroid differentiation occurs through several distinct stages, thus making it an excellent model system to study the dynamic transcriptional changes occurring during the process. Understanding the dynamics of erythroid differentiation could lead to the development of better approaches for the treatment of several haemoglobinopathies. The process of erythropoiesis is governed by an intricate network of transcription factors, epigenetic modifiers, growth factors, signalling pathways and non-coding RNAs (Hattangadi *et al.*, 2011). Even though several studies have underlined the importance of the above-mentioned factors in stem cell maintenance and differentiation, the molecular mechanisms defining the delicate balance between both remain to be understood. To comprehend the molecular mechanisms involved in stem cell maintenance and

differentiation additional factors involved in the process needs to be identified and their elaborate interplay needs to be dissected.

The role of several new factors in erythropoiesis can be identified by small RNA sequencing (Lessard *et al.*, 2018) as well as by knockdown studies using high - throughput RNAi screens (Galeev *et al.*, 2016). Small RNA sequencing has been particularly useful in identifying several new miRNAs involved in the process of erythropoiesis. Additionally, miRNAs at a specific stage of erythroid differentiation can also be identified using this method. Gene knockdown studies using RNAi screens have led to the identification of several novel factors in stem cell maintenance and differentiation. Since erythropoiesis is a multi-staged process wherein the cells at each stage are phenotypically different, the study design can have an impact on the outcome of such studies. The effect of gene knockdown is assessed based on the kinetics of erythropoiesis, which is evaluated based on the expression of markers CD71 and CD235a and the formation of different erythroid populations. One of the critical factors determining the efficacy of knockdown studies is the type of promoter used for the study since the performance of promoters varies with the target cell type. Therefore, it is essential to choose a promoter that does not undergo silencing during the course of *ex-vivo (in-vitro)* erythroid differentiation. Thus, the novel factors playing a role in erythropoiesis can be identified either by gene knockdown studies or by small RNA sequencing studies.

1.1 OBJECTIVES OF THE STUDY

The principal aim of this thesis is to identify the transcriptional regulatory mechanisms involved in haematopoietic stem cell maintenance and the differentiation of haematopoietic stem cells to the erythroid lineage. The study is described in chapters 4 to 6, with each chapter dealing with specific strategies to carry out the study.

CHAPTER 4: miRNOME SCREENING IN HUMAN ERYTHROPOIESIS

This chapter describes the miRNA screening strategies, which have been used in an *ex- vivo* erythroid culture system to identify the new miRNA players in the context of erythropoiesis. The strategies which have been used are as follows:

1. Establishment of a robust *ex vivo* erythroid culture system to study the miRNA profile in erythropoiesis.

2. To identify the differentially expressed miRNAs in HSCs and erythroid cells by small RNA sequencing.

3. To analyse the transcriptional regulation of miRNAs in erythroid cells by chromatin Immunoprecipitation (ChIP)-sequencing analysis.

4. To functionally validate miRNAs overexpressing them and by knocking them out using CRISPR-Cas 9.

CHAPTER 5: RNAi SCREEN TO DELINEATE NOVEL SIGNALLING PATHWAYS IN ERYTHROPOIESIS.

This chapter describes the robust methodology used to identify the candidate signalling genes in human erythropoiesis.

1. Primary shRNA library screening of 5000 signalling molecules in CD34+ haematopoietic stem and progenitor cells (HSPCs) to find the most enriched and depleted shRNAs.

2. Secondary screening of the most significantly enriched and depleted shRNAs by constructing a mini library with the candidate genes.

3. Validation of the role of the most significantly downregulated/ upregulated genes by knocking them down in CD 34+ HSPCs.

CHAPTER 6: RNAi SCREEN TO DELINEATE NOVEL REGULATORS IN HSC MAINTENANCE AND ERYTHROPOIESIS.

This chapter aims to identify novel regulators playing a role during haematopoietic stem cell maintenance and differentiation. The strategies are:

1. Meta-analysis to identify candidate genes in HSC maintenance.

2. Choosing the best algorithm and miR 30 based backbone for efficient shRNA knockdown.

3. Identification of promoters that show stable expression throughout erythroid differentiation.

4. Generation of shRNA library against HSC maintenance genes for RNAi screen.

5. Validation of the shRNA library in an *ex- vivo* model of HSC differentiation.

6. Identification of genes involved in HSC maintenance and erythropoiesis by NGS.

2. REVIEW OF LITERATURE

2.1 EUKARYOTIC TRANSCRIPTION AND ITS REGULATION

2.1.1 TRANSCRIPTION MACHINERY IN EUKARYOTES

Transcription is defined as the initial step of DNA based gene expression in which a specific segment of DNA is converted into messenger RNA (mRNA) by RNA polymerase. Transcription in eukaryotes is mediated by three different types of RNA polymerases, I, II, and III, which assist in the transcription of genes representing different classes of RNAs (Schramm and Hernandez, 2002). RNA polymerase I is found in the nucleolus and drives the synthesis of pre-rRNA, which is processed into 18S, 5.8S, and 28S rRNAs. RNA polymerase II (RNAPII) is responsible for transcribing all protein-coding mRNAs in addition to transcribing snoRNAs as well as micro RNAs (miRNAs). RNAPII recognizes the core promoter, which harbours sequence elements such as the TATA element (TBP binding site), BRE (TFIIB recognition element), Inr (initiator element), and DPE (downstream promoter element) (Hahn, 2005). Recognition of the core promoter by RNAPII is the most crucial step for transcription to begin. RNA polymerase III synthesizes tRNAs, 5S rRNA, and several other small RNAs such as the one that is involved in RNA splicing and the RNA component of the signal recognition particle, which directs the *de novo* proteins to the endoplasmic reticulum (Maraia, 2001).

Of these three eukaryotic polymerases, RNAPII has been widely studied and is of central importance to eukaryotic transcription. RNAPII derived from *Saccharomyces cerevisiae* has been studied extensively and is comprised of 12 components. RBP1 is

the largest subunit that exhibits a significant level of homology to the β' subunit of bacterial RNA polymerase. It is comprised of a carboxyl-terminal domain, which harbours numerous repeats of a heptad amino acid sequence –YSPTSPS– and is involved in several crucial functions. The RBP2 subunit shows several degrees of homology to the β subunit, whereas the RBP3 and the RBP11 subunit bear a striking resemblance to the α subunits.

The recruitment of several transcription factors is essential for RNAPII to form the active transcription complex. The process of transcription can be demarcated into three different phases: initiation, elongation, and termination (Gehring, Walker and Santangelo, 2016) **(Figure 2.1.1**). The initiation of transcription commences with the generation of a pre-initiation complex, which is formed by the assembly of RNAPII with the transcription factors such as TFIIB, TFIID, TFIIE, TFIIF, and TFIIH at the promoter DNA (Sainsbury, Bernecky and Cramer, 2015). Sequentially, in the first step, TATA-binding protein (TBP) interacts with the TATA box. TFIIA binds next in sequence, and conjunction with TFIIB, it helps in stabilization of the TBP DNA complex. The complex comprising of TFIIB, TFIIA, and TBP is then bound by TFIIF and Pol II (Sainsbury, Bernecky and Cramer, 2015). TFIIF facilitates targeting of Pol II to its promoters both by interacting with TFIIB and by diminishing the interaction of the polymerase with undefined sites on the DNA. The formation of the closed complex culminates when TFIIE and TFIIH bind in the end. TFIIH has several subunits and includes a DNA helicase activity that stimulates the uncoiling of DNA near the RNA start site, which leads to the generation of an open complex. In

addition to having helicase activity, TFIIH also possesses kinase activity, which is responsible for phosphorylating RNA polymerase II at several places in the CTD.

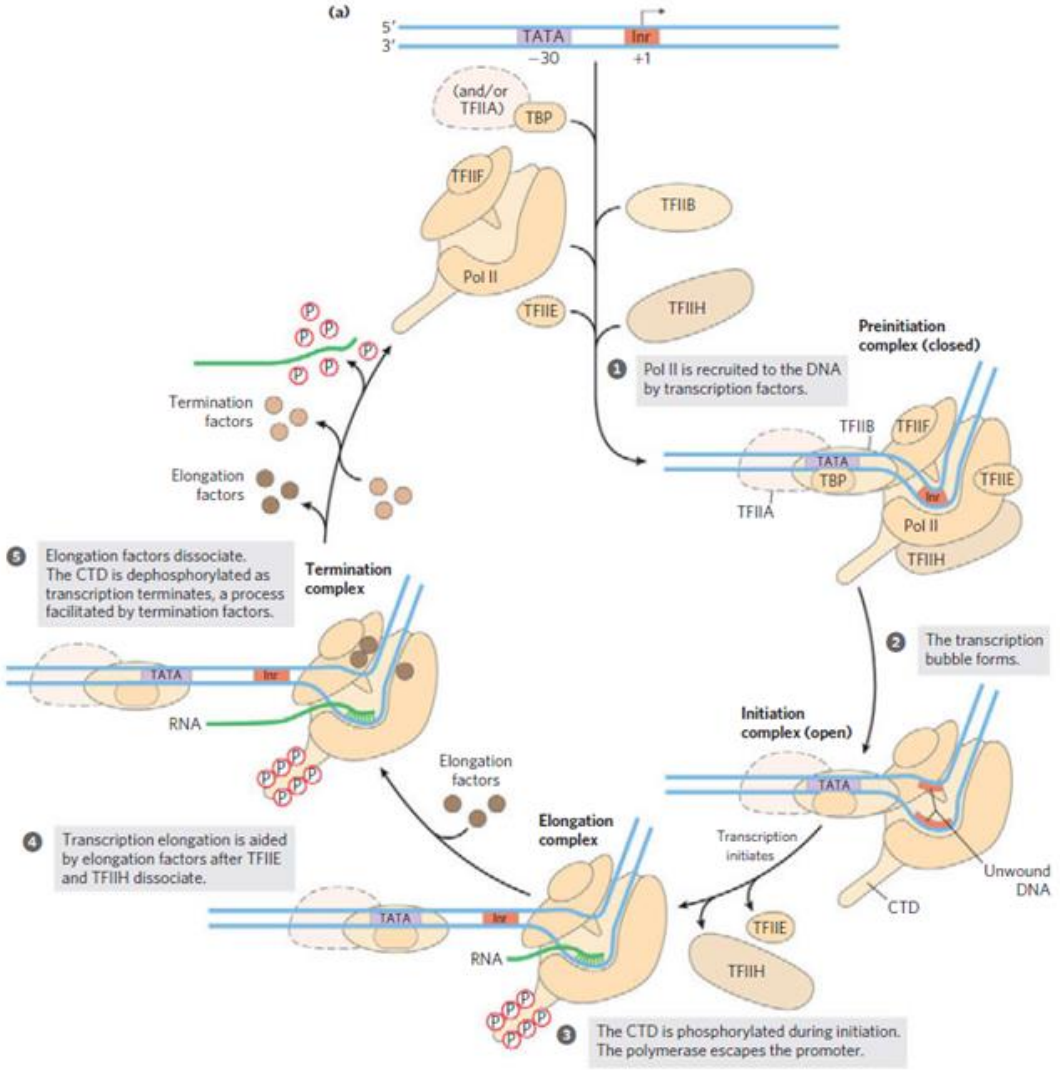


Figure 2.1.1: Transcription in eukaryotes. The process of transcription in eukaryotes can be demarcated into sequential phases such as assembly, initiation, elongation and termination. 1. Assembly: This stage results in the formation of a closed pre initiation complex in which RNAPII is recruited to the DNA by several factors such as TBP, TFIIA, TFIIB, TFIIE and TFIIF. 2. Initiation: The initiation phase of transcription is characterised by the formation of the initiation complex wherein several serine residues of the CTD domain are phosphorylated thus leading to a change in the conformation and initiation of transcription. 3&4 Elongation: The process of elongation begins with an elongation complex, which is formed by binding of several elongation factors to the phosphorylated residues of CTD domain. 5. Termination: The process of termination commences with the de-phosphorylation of CTD domain and the disjunction of the elongation factors which is mediated by multiple termination factors. *Adapted from Lehninger, 2009.*

This results in an alteration of the total conformation of the complex, which facilitates the initiation of transcription (Sainsbury, Bernecky and Cramer, 2015). When the first 60 to 120 nucleotides of RNA are synthesized, it results in the pausing of RNAPII, which is accompanied by the release of TFIIE and then TFIIH. This is followed by the entry of RNAPII into the productive elongation phase of transcription, which is mediated by positive transcription elongation factor b (P-TEFb) (Chen, Smith and Shilatifard, 2018). TFIIF remains associated with RNAPII throughout elongation. The activity of RNAPII is augmented by proteins called elongation factors that bind to the phosphorylated CTD and inhibit the cease of transcription and coordinate interactions between protein complexes involved in the post-transcriptional processing of mRNAs (Jonkers and Lis, 2015). Once the formation of the RNA transcript is complete, transcription is terminated, which is indicated by the disjunction of RNAPII from the DNA template. The most predominant pathway of transcription termination is the cleavage and polyadenylation factor-dependent pathway, which is mediated by the protein complexes, cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulatory factor (CstF) and Poly (A) polymerase. These termination complexes are recruited to the CTD repeats of RNAPII, which are phosphorylated at Ser-2 (Porrua and Libri, 2015). In addition to this, the CPSF complex associates with the transcribed poly (A) site on the pre-mRNA, which leads to the arrest of RNAPII and ultimately activates the cleavage and release of RNAPII at the gene terminus. The

event of termination also facilitates the recycling of RNAPII for further rounds of transcription (Shandilya and Roberts, 2012).

2.1.2 REGULATION OF TRANSCRIPTION

Three different types of proteins mediate the modulation of transcription initiation; specificity factors mediate promoter recognition by RNA polymerase and bring about a conformational change in RNA Polymerase which enables them to identify a given promoter or set of promoters; activators augment the association between RNA polymerase and the promoter and repressors obstruct the availability of RNA polymerase to the promoter (Aerts, 2012).

2.1.2.1 TRANSCRIPTION FACTORS

Transcription factors are essential proteins that decipher the information in the genome and express a certain subset of proteins and RNA molecules. These factors are crucial in determining cell fate during development and are responsible for maintaining a cell in a particular physiological state. It exhibits its mode of action by binding to specific DNA sequences designated as promoters or enhancers in the genome. With the cooperation of epigenetic factors as well as co-factors, it brings about the transcription of several target genes based on the cell type (Aerts, 2012). A typical transcription factor is comprised of the following domains such as the DNA binding domain which helps in recognizing sequences in DNA for specific binding, dimerization domain which assists in the formation of heterodimers and homodimers, activation domain which interacts with other transcription factors and regulators of transcription and the protein interaction domain which helps in interacting with histone methylases and other coactivators. Based on the type of DNA binding

domain which they have, transcription factors can adopt any one of the following confirmations such as the helix turn helix motif, zinc fingers, basic region leucine zippers, helix loop helix motif and β scaffold with minor groove contacts.

2.1.2.2 TRANSCRIPTIONAL ACTIVATORS*:* These proteins augment the transcription of a gene or set of genes. A prototypical activator has two crucial functions, i.e., to bind DNA and to activate transcription. The DNA binding domain of an activator is responsible for imparting specificity, whereas the activation domain is required to stimulate the process of transcription. The DNA binding domains form well-defined 3-dimensional structures and identify short specific DNA sequences by interacting with bases in the major groove of the DNA double helix. The activation domain, on the other hand, is less complex and is comprised of short protein sequences, which increases the rate of transcription (Ma, 2011).

2.1.2.3 TRANSCRIPTIONAL REPRESSORS: A repressor is defined as a DNA- or RNA-binding protein that impedes the expression of one or more genes by binding to the operator or associated silencers. A DNA-binding repressor inhibits the binding of RNA polymerase to the promoter, thus preventing transcription of the genes into messenger RNA. An RNA-binding repressor binds to the mRNA and prevents translation of the mRNA into protein. Mammalian transcriptional repressors can be classified as passive and active repressors (Cowell, 1994). Passive repressor proteins lack an innate repressing activity since they are devoid of a repression subunit. These proteins impede RNA synthesis by forming inactive heterodimers with transcriptional activators, thus rendering them incapable of interaction with DNA, or they exhibit their mode of action by binding to coactivators, which are essential for

the activation of transcriptional proteins. Examples of passive repressor proteins include Inducible cAMP early repressor (ICER) (Molina *et al.*, 1993) and Sp-1 like transcriptional repressors (Kaczynski *et al.*, 2001). Active mammalian transcriptional repressor proteins possess inherent repression activity, which targets the chromatin organization of the genome. Active repressors are usually activator independent and can exert their function over long distances. Active transcriptional repression is generally brought about by two different mechanisms, one which involves histone deacetylation and another, which requires histone methylation and heterochromatin formation (Thiel, Lietz and Hohl, 2004).

2.1.2.4 CHROMATIN STRUCTURE AND TRANSCRIPTION:

The DNA is tightly bound to histones, forming chromatin. The structural unit of chromatin consists of nucleosome, which consists of 146 base pairs of DNA wrapped around two molecules the histones. The chromatin is condensed by coiling into higher-order structures organized into large loops of DNA (McGinty and Tan, 2015). This packaging of eukaryotic DNA in chromatin is important for its availability as a template for transcription, thus, chromatin structure is a crucial aspect of gene transcription in eukaryotic cells. Both transcription activators and repressors regulate transcription in eukaryotes by inducing changes in the structure of chromatin. Actively transcribed genes are found in the decondensed fraction of chromatin that is more accessible to transcription factors than the rest of the genome. Decondensation of chromatin is the minimum criterion to make DNA available for transcription (McGinty and Tan, 2015). The compact winding of DNA around the nucleosome core particle is a major obstacle to transcription. This affects both the

ability of transcription factors to bind DNA and the ability of RNA polymerase to transcribe through a chromatin template. The inhibitory effect of nucleosomes is brought about by acetylation of histones and by the binding of two non-histone chromosomal proteins, HMG-14 and HMG-17 to nucleosomes of actively transcribed genes (Herrera *et al.*, 1999). In addition to this, nucleosome-remodelling factors facilitate the binding of transcription factors to chromatin by modifying the nucleosome structure. Nucleosome remodelling factors are protein complexes that promote the binding of transcription factors by bringing about changes in the nucleosome structure. The nucleosome remodelling factors catalyse the sliding of histone octamers along the DNA molecule, thereby relocating the nucleosomes to facilitate transcription factor binding (Becker and Workman, 2013).

2.2 NON CODING RNAS AND TRANSCRIPTIONAL REGULATION

The majority of the eukaryotic genome is transcribed into non-coding RNAs (ncRNAs), whereas only 1-2% of the transcripts code for proteins. ncRNAs can be categorized into infrastructural ncRNAs and regulatory ncRNAs. The class of regulatory ncRNAs includes the microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and long non-coding RNAs (lncRNAs). The recent addition to this list includes the promoter-associated RNAs (PARs) and enhancer RNAs (eRNAs).miRNAs are 22-24 nucleotides small single-stranded molecules, which are evolutionarily conserved and participate in gene regulation at the post-transcriptional level through mRNA degradation and translational inhibition. Recent reports have also suggested that miRNAs play a role in transcriptional gene silencing by

recruiting Polycomb group proteins, which subsequently form repressive heterochromatin marks in gene promoters. In addition to this, miRNAs have also been attributed to regulate gene expression by translational activation (Kaikkonen, Lam and Glass, 2011),(Wery, Kwapisz and Morillon, 2011). piRNAs are small ncRNAs of 24-31 nucleotides length and have the name because of their ability to form complexes with PIWI-clade of Argonaute proteins. Through a self-amplifying loop, which is known as the ping-pong cycle, piRNAs have been shown to play a role in suppressing transposon activity during the development of the germline. Moreover, the latest reports also suggest that piRNAs contribute to gene silencing in somatic cells through epigenetic regulation by recruiting heterochromatin factors such as the heterochromatin protein 1a (HP1a), to specific sites of the mammalian genome (Wery, Kwapisz and Morillon, 2011). Long non-coding RNAs (lncRNAs) are more than 200 nucleotides in length. They are defined by the hallmarks of localization in the nucleus, weak sequence conservation, and the presence of both Poly A+ and Poly A- transcripts (Kaikkonen, Lam and Glass, 2011). At the pre transcriptional level, lncRNAs can send appropriate signals at the chromatin level to modulate gene expression. LncRNAs have a major role in the process of methylation. For instance, it has been observed by Wang et al. that lncRNA Dum (developmental pluripotency-associated 2 (Dppa2) Upstream binding Muscle lncRNA) regulated the expression of Dppa2 by influencing DNA methylation. Methylation of Dppa2 promoter was stimulated by Dum through recruitment of Dnmt1, Dnmt3a, and Dnmt3b to its promoter site. This recruitment of the DNA methyltransferases led to silencing of Dppa2 expression in cis configuration, which further promoted myogenic differentiation. With respect to transcription, lncRNAs

can modulate the expression of genes by directly binding to TFs or RNA Polymerase II, or by inhibiting the binding of polymerase with the promoter. In addition to the above-mentioned functions, lncRNAs can also escort RNA polymerase II to bind to the promoter of specific genes and compete for enhancer with protein-coding gene promoter. Promoter associated RNAs (paRNAs) are usually 200-500 nucleotides long. They are transcribed from sequences that are located in the promoter regions of the gene. Current evidence has shown that paRNAs play an important role in regulating the transcription of genes in addition to making the chromatin architecture accessible for transcription and rescuing the negative supercoils during transcription (Yan and Ma, 2012). Enhancer RNAs (eRNAs) are defined as a subset of enhancers that are occupied by RNA polymerase II and are transcribed into eRNAs. eRNAs are determinants of active enhancers and play an important role in coordinating gene expression, which allows for the modulation of several genes through a single gene-binding site. Apart from the above-mentioned functions, eRNAs can also modify histone modifications to modulate transcriptional repression (Ding *et al.*, 2018).

2.2.1 IMPORTANCE OF STUDYING miRNAs:

Since the initial discovery of the miRNA *lin-4* in the early ’90s and the subsequent discovery of lin-14, which was found to be regulated by lin-4, the concept of gene regulation by miRNAs emerged as one of the areas which are being thoroughly studied (Wahid *et al.*, 2010). After the discovery of these miRNAs, several other miRNAs were identified based on several computational and experimental methods (Cai *et al.*, 2009) The advent of high-throughput sequencing technologies as well as computational and bioinformatics prediction methods has escalated research on

miRNAs which includes identification of its regulatory targets and decoding its plausible functions (Wahid *et al.*, 2010) miRNAs are ~22-nt long RNA molecules which are transcribed by RNA polymerase II endogenously and play a crucial role in fine-tuning the gene expression by either causing degradation of the target mRNAs or by repressing the translation of the target genes. The ability of the miRNAs to bind to the target gene with imperfect complementarity makes it possible to target multiple genes simultaneously which conforms to its involvement in regulating several biological processes such as cell proliferation, cell death, fat metabolism, neuronal patterning, hematopoietic differentiation and immunity (Wahid *et al.*, 2010), (Cai *et al.*, 2009) (**Figure 2.2.1**). Thus, it is important to study the role of miRNAs in different biological processes to understand the transcriptional regulatory network in specific processes. Furthermore, dysregulation of miRNAs has been attributed to causing several solid tumours, as well as haematological malignancies. Therefore, miRNA profiling in the case of diseased individuals might help us in delineating several biomarkers, which can serve as hallmarks for the disease. In addition to this, deregulation of miRNAs in diseased conditions can be harnessed as potential therapeutics by either miRNA replacement therapy using miRNA mimics or inhibition of miRNA function by antimiRs (Rupaimoole and Slack, 2017).

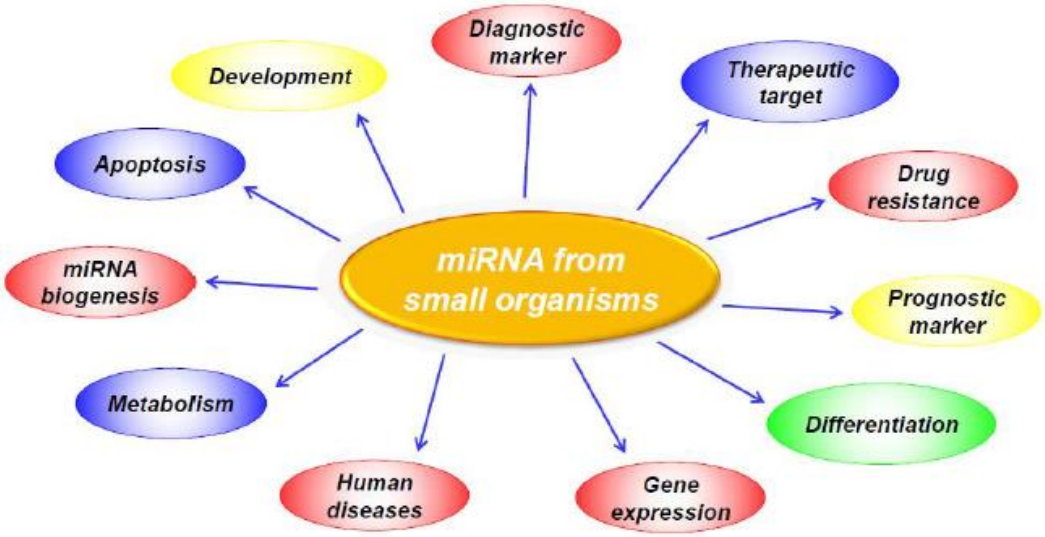


Figure 2.2.1: Biological processes regulated by miRNAs (Adapted from Chandra *et al*., 2017)

2.3 HAEMATOPOIESIS:

The haematopoietic system is one of the multifaceted systems in the human body, which is evolutionarily conserved across the vertebrate species, although some minute differences do exist. The haematopoietic stem cells (HSCs), which are positioned at the apex of the hierarchy, are multipotent cells that give rise to multiple lineages of blood cells by traversing through several stages of lineage constricted steps (Nimmo, May and Enver, 2015). Several models have been proposed to elucidate the process of hematopoietic differentiation (**Figure 2.3**). The earliest model, which is termed as the "classical model," was the result of high throughput research carried out around the year 2000. According to this model, HSCs give rise to multipotent progenitor (MPP), also known as short term HSC, which further leads to the formation of two lineage-restricted progenitors Common Myeloid Progenitors (CMP) and Common Lymphoid Progenitors (CLP).

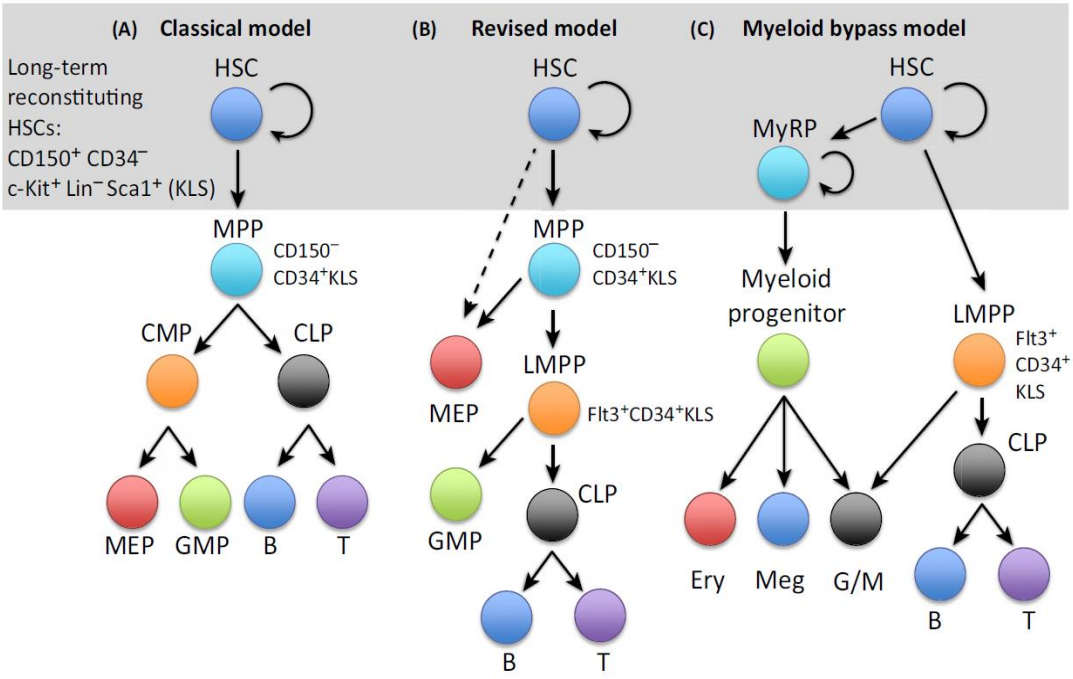


Figure 2.3: Different models of haematopoiesis. Classical Model: It states that the HSCs and the MPPs occupy the highest position in the hierarchy and they give rise to lymphoid and myeloid cells, which have lost the capacity to self-renew. Revised Model: In this model, LMPP has been identified which gives rise to GMP and CLP but not MEP. Myeloid Bypass model: In this model, the cell type MyRP gives rise to the myeloid progenitors whereas LMPP gives rise to lymphoid progenitors. HSC: Haematopoietic Stem Cell MPP: Multipotent Progenitor, CMP: Common Myeloid Progenitor CLP: Common Lymphoid Progenitor, MEP: Megakaryocyte/Erythrocyte Progenitor GMP: Granulocyte/Macrophage Progenitor B: B cell, T: T cell, LMPP: lymphomyeloid –restricted progenitors, MyRP: Myeloid Restricted Progenitor. Adapted from Nimmo *et al*., 2015

The CMPs and CLPs further bifurcate to form more differentiated cell types such as erythrocytes, megakaryocytes, eosinophils, B cells and T cells, respectively. The classical model was further improvised when a new cell type termed as lymphoid primed multipotent progenitor(LMPPs) was identified in the Flt3+ (FMS-like tyrosine kinase 3) portion of the MPP compartment. Single-cell gene expression analysis of LMPPs showed that these were the primordial cells expressing common lymphoid genes, which are associated with loss of MegE priming but the maintenance of myeloid priming. LMPPs have been shown to combine lympho-

myeloid differentiation capacity when cultured in multi-lineage supportive conditions. LMPPs have been shown to express one or more of the early lymphoid genes, such as Interleukin 7 receptor (Il7r), sterile immunoglobulin heavy chain complex (IgH), or recombination activating gene 1 (Rag1). In addition to these genes, they also expressed GM lineage-affiliated genes, thus suggesting that lympho-myeloid priming is functionally related to the competency of these progenitors. The sequential organization of lineage priming within HSPCs, with myeloid and MegE priming pioneering lymphoid priming, outlined the refined model of hematopoietic commitment. The most recent model of haematopoiesis is the myeloid bypass model, which appeared by multi-lineage clonal tracking of individual HSCs. As per this model, megakaryocyte and myeloid restricted progenitors are produced in the primordial HSC compartment, and they exhibit the properties of long-term self-renewal and reconstitution of specific myeloid lineages. This model is not in agreement with the previous two models since this model introduces the concept of segregation of self-renewal and multi-potentiality (Nimmo, May and Enver, 2015). Haematopoiesis occurs in two perspicuous waves in the vertebrate system: the primitive wave and the definitive wave. The formation of red blood cells in these two stages occurs at specific assigned sites, which changes with the course of development. The hierarchical stations of haematopoiesis include the yolk sac, the aorta gonad mesonephros region, the foetal liver, and, ultimately, the bone marrow. In addition to these locations, the placenta has also been recently described to play a role in the development of HSCs. The distinguishing characteristics of the HSCs in each site show a significant difference, thus indicating the existence of disparate niches that support the proliferation and differentiation of HSCs. To exemplify this,

the HSCs in the foetal stage are mostly in cycling conditions in contrast to the HSCs in the adult bone marrow which are usually in the quiescent state (Orkin and Zon, 2008). In mouse, distinct stages of embryonic and haematopoietic development can be differentiated by morphological transformations, which occurs within a span of 24 hours, whereas, in humans, the embryo development takes a longer time and the different stages are named as Carnegie stages (CS). Studying HSC development using Carnegie stages is more reliable as against gestation time, even though several levels of similarity do exist between mouse and human haematopoietic development. The first wave of human haematopoiesis emerges in the mammalian yolk sac and is designated as "primitive (Ivanov *et al.*, 2017).

2.3.1 SITES OF HAEMATOPOIESIS

Haematopoiesis occurs at several sites in the foetal and adult life, as shown in **Figure 2.3.1.**

2.3.1.1 YOLK SAC

Massive primitive nucleated erythrocytes represent the predominant population of the primary haematopoietic output from the yolk sac at CS 7-8 with the infrequent existence of primitive macrophages and megakaryocytes. Towards CS 10, the early primordial erythroblasts can be witnessed within the cardiac cavity, thus announcing the commencement of blood circulation, which is succeeded by the advent of the 1st CD45^{+} cells (PTPRC^{+}) cells (Ivanovs *et al.*, 2017).

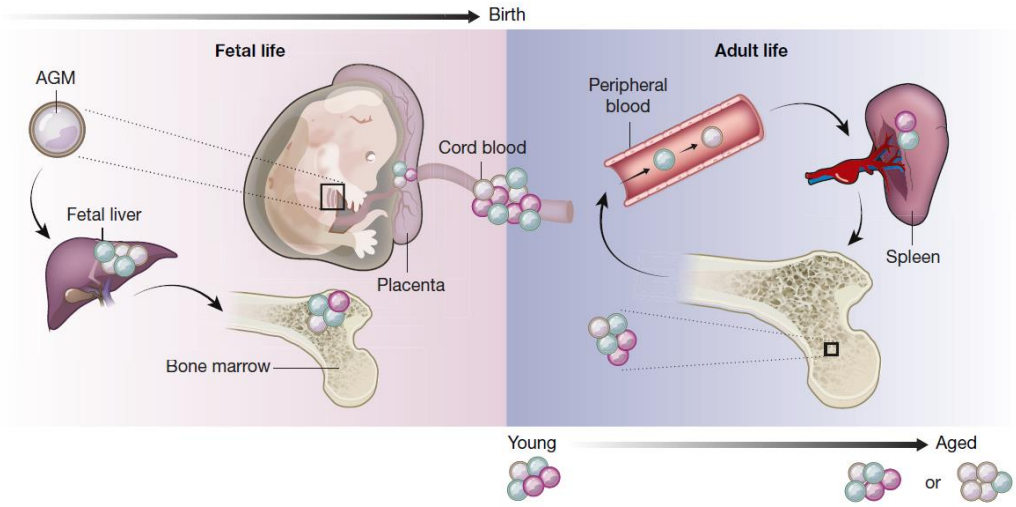


Figure 2.3.1: Anatomical sites of haematopoiesis in the foetal and adult system. The process of haematopoiesis in the embryo begins in the blood islands of the yolk sac specifically in the AGM region from where the primitive HSCs emerge. From the AGM it moves to the foetal liver and then to the bone marrow. In adults, haematopoiesis mostly occurs in the bone marrow of long bones. HSC: Haematopoietic Stem Cells AGM: Aorta Gonad Mesonephros. Adapted from Laurenti *et al*., 2018

2.3.1.2 THE AORTA-GONAD-MESONEPHROS (AGM) REGION

The emergence of the intra- aortic haematopoietic clusters (IAHCs) on the ventral wall of the human dorsal aorta in the AGM region indicates the origin of the intra embryonic, permanent adult haematopoietic wave in the vertebrate embryo (Ivanovs *et al.*, 2017). IAHCs emanate at CS 13 at the floor of the embryonic dorsal aorta and fade away by day 17. IAHCs make a sheet over the pre umbilical area on the dorsal aorta and invade the vitelline artery. Primordial mesoderm and the endothelial division in early post gastrulation embryos are characterized by KDR or FLK1 expression. The endothelial lining of the human dorsal aorta leads to an increase in the level of CD34, which later marks adult HSCs from CS9. By CS 13, the first

CD34+CD45+ cells make their appearance in the pre umbilical region of the dorsal aorta, and they become numerous by CS 15 (Oberlin *et al.*, 2002).

2.3.1.3 THE LIVER

The liver rudiment appears as a diverticulum from the embryonic gut initially at CS 10. From late CS 10, the liver rudiment comprises primitive yolk sac derived erythrocytes and CD45+ cells, mostly belonging to the monocytic or macrophage lineage (Ivanovs *et al.*, 2017). Towards CS 13, the liver is mainly comprised of enormous numbers of CD34+CD45+ cells, which displace primitive erythroblasts in the human embryo bloodstream when the liver gets colonized from CS13-14 (Tavian, Hallais, and Péault, 1999). The liver continues to be the most crucial anatomical site for haematopoietic differentiation and HSC proliferation until birth. The essential purpose of the red blood cells created from the primitive HSCs is to promote tissue oxygenation to abet the rapid growth of the embryo. The trademark of the primitive erythroid cells is to express embryonic globin proteins. This primitive phase is ephemeral and is immediately taken over by adult haematopoiesis which is termed as "definitive"(Orkin and Zon, 2008).

2.3.1.4 BONE MARROW

The main site of definitive haematopoiesis is the bone marrow, and the formation of the definitive bone marrow niche is intimately associated with the incursion of cartilaginous bone by blood vessels and bone ossification. The phenomenon of vascular attack expedites the colonization of the bone marrow with haematopoietic progenitors and HSCs. The development of the bone marrow is representative of the termination of the human embryonic period. The initial $CD34^{-}CD45^{+}$haematopoietic

cells, which colonize the cartilaginous bone, mostly consist of $CD68^+$ monocytes/macrophages. This is succeeded by the accumulation of $CD34^+CD45^+$progenitors and HSCs in the bone marrow (Ivanovs *et al.*, 2017).

2.4 REGULATION OF HAEMATOPOIESIS

The process of haematopoiesis is regulated by an intricate network of transcription factors, signalling pathways, miRNAs as well as growth factors.

2.4.1 TRANSCRIPTION FACTORS REGULATING HAEMATOPOIESIS

2.4.1.1 SCL/TAL1: The SCL (TAL1/TCL5) transcription factor belongs to the clan of basic helix- loop- helix transcription factors. SCL polypeptides do not self-associate to form homodimers but instead collaborate with E12, E47, and E2-2 and bind DNA in a sequence-specific manner. The SCL protein undergoes phosphorylation at serine 122, which is a substrate for MAP kinases such as ERKI, which is actively involved in the Ras signal transduction pathway (Eblen, 2018). In addition to this, cAMP-dependent protein kinase is involved in phosphorylating a second serine residue S172, which modifies binding of SCL-E12 heterodimers to some E box sequences (Porcher, Chagraoui and Kristiansen, 2017). These modifications suggest that the functional properties of SCL can be amended via the MAP Kinase pathways (Robb and Begley, no date)**.** The initial testimony to the involvement of SCL in the early stages of blood development was demonstrated by the in vivo extirpation of SCL activity in mouse models. SCL-/- embryos died at embryonic day 9.5, which was further marked by non-existence of yolk sac primordial erythropoiesis and myelopoiesis. In addition to this, not all adult definitive haematopoietic lineages were found in Scl-/- mouse chimeras. Ensuing

studies carried out in several experimental models suggested that SCL plays a crucial role in endothelial and blood development rather than from previously ingrained progenitors. Experiments carried out to differentiate Scl null cells in murine embryonic stem cell models, which mimicked the two waves of yolk sac haematopoiesis reinstated the fact that SCL was essential for the genesis of both blood and endothelial constituents of colonies obtained from FLK1+ mesodermal cells (Robb and Begley, no date). Furthermore, when SCL was conditionally deleted through the Tie2-Cre recombinase system, it was observed that haematopoietic commitment was not disturbed instead maturation of primitive and definitive megakaryocytes and erythroid cells was perturbed leading to the death of embryos at E13.5/E14.5. This study suggested that the role of SCL in blood fate commitment can be disengaged from its later role in blood cell maturation, and SCL is not essential for HSPC production. The requirement of SCL has also been observed for the development of vascular network after angioblast formation, which was evidenced by the perturbed angiogenic remodelling on knocking out of SCL. Thus, in a nutshell, SCL is indispensable for the designation of the 3 haematopoietic waves, maturation of blood lineages, and remodelling of the vascular network (Porcher, Chagraoui and Kristiansen, 2017).

2.4.1.2 RUNX: The RUNX group of transcription factors are comprised of Runx1, Runx2, and Runx3 and are evolutionarily conserved. These factors are responsible for the modulation of lineage-specific genes, cell identities, as well as for multiple other functions throughout development (de Bruijn and Dzierzak, 2017). All three RUNX factors play a critical role in haematopoiesis, and they regulate the expression

of target genes by binding to a heterodimeric non-DNA binding partner protein Cbfβ. During primitive haematopoiesis, RUNX1 is highly expressed in the mesoderm of the probable blood islands beginning from E7.5 of mouse development, and its expression continues till the blood island mesoderm differentiates into primitive erythrocytes. Although earlier studies have shown that primitive erythrocytes do not require Runx1 for their formation, recent studies have delineated the essentiality of Runx1 in the maturation of primitive erythrocytes (Yokomizo *et al.*, 2008). The absence of Runx1 leads to diminished levels of Ter119 as well as other transcription factors such as GATA1 and EKLF1.In the adult haematopoietic system, it has been shown that Runx1 plays a pivotal rule in its establishment and maintenance. For instance, Runx1 promotes haematopoietic specification, which is concomitant with the disappearance of the endothelial potential. It is also essential for the initial remodelling of the PU.1 transcription factor, which is one of the crucial downstream targets.Runx1also plays a role in modifying the epigenetic profile at the haematopoietic loci to initiate haematopoietic differentiation (de Bruijn and Dzierzak, 2017).

2.4.1.3 GATA1: Extensive studies carried out at the β-globin locus led to the identification of the transcription factor GATA-1 and its subsequent partners GATA 2-6. GATA 1, 2 & 3 are mostly expressed in haematopoietic cell types and play a role in regulating the development and functioning of multiple blood lineages. GATA 1 is expressed in more mature cell types such as eosinophils, dendritic cells, and erythrocytes. In contrast, the expression of GATA 2 is mostly restricted to haematopoietic stem and progenitor cells (HSPCs) as well as erythroid precursors

(Katsumura *et al.*, 2018). Studies carried out in GATA1 deleted mice have shown that it leads to malfunctioned erythroid cell development and embryonic day 10.5 to 11.5 lethality (Fujiwara *et al.*, 1996). Considering the crucial role the GATA family transcription factors play during haematopoiesis mutations mostly in GATA 1 and GATA 2 has been associated with several haematological disorders such as Acute Myeloid Leukemia (AML) and Myelodysplastic Syndrome (Katsumura and Bresnick, 2017).

***2.4.1.4 GATA2*:** The GATA 2 transcription factor is broadly distributed among haematopoietic stem cells and is predominantly expressed in early progenitors, megakaryocytes, and mast cell lineages (Vicente *et al.*, 2012). There have been many studies that have been carried out to delineate the role of GATA 2 in haematopoiesis. GATA 2 knock out mice exhibits acute deficiency in definitive haematopoiesis. Similar studies in GATA2 deleted mice have shown lethality at E10 occurring as a result of the collapse of haematopoiesis thus necessitating the importance of GATA 2 for HSPC establishment and function (Tsai *et al.*, 1994). Furthermore, GATA2 deficient embryos have been found to be anaemic and have significantly low numbers of primitive erythrocytes and haematopoietic progenitor cells (HPCs) (Tsai *et al.*, 1994). They usually show lethality at day 10-11 of gestation, thus emphasizing the indispensable role of GATA2 during the nascent stages of haematopoiesis. Haploinsufficiency of GATA2 results in aberrant homeostasis in adult mice (Tsai *et al.*, 1994).GATA 2 has two different roles during the ontogeny of HSCs, which include the formation and proliferation of HSC in the AGM region and the expansion and survival of HSCs in the adult bone marrow (Ling *et al.*, 2004). GATA 2 has also

been reported to bestow HSCs and HPCs with increased quiescence, which is a distinguishing property of HSCs (Vicente *et al.*, 2012). GATA-2+/− adult marrow demonstrates depletion in the number of functional HSCs, which is mechanistically associated with marked HSC apoptosis and cellular quiescence. Interestingly, the relative potential of HSCs to produce progenitors or to self-renew remains unaffected. In addition to this GATA-2+/−mice exhibit, a diminished population of granulocyte-macrophage progenitor (GMP) which performs poorly in colony-forming assays but the CMPs and CLPs remain unperturbed by loss of GATA2 (Rodrigues *et al.*, 2012). These findings suggest that GATA 2 is essential to sustain the inherent characteristics of HSCs, and a decrease in its levels is required for successful differentiation to ensue (Vicente *et al.*, 2012).

2.5 SIGNALLING PATHWAYS AND HAEMATOPOIESIS

Vertebrate haematopoietic development utilizes many of the important signalling pathways which are involved in ontogeny (Lento *et al.*, 2013). Some of the crucial pathways that play a vital role during haematopoiesis as well as HSC maintenance and self-renewal include the Wnt signalling pathway, Notch signalling pathway, MAPK pathway, and mTOR signalling pathway (**Table 2.5)**. The Wnt pathway has a multifarious impact on haematopoiesis which is attributed to the 19 different ligands and the 10 different receptors which the pathway harbours (Lento *et al.*, 2013). The proof of concept that Wnt signalling might be playing a role in haematopoietic ontogeny came from studies carried out by different groups. The presence of Wnt signalling in sites of primitive haematopoiesis (Cheng *et al.*, 2008), as well as the expression of Wnt ligands and frizzled receptors in the murine yolk sac, AGM and

Signalling Pathway	Function in Haematopoiesis	Reference
CXCL12/CXCR4	Regulates myelopoiesis and lymphopoiesis during embryonic development and plays a role in HSC maintenance.	Braun et al,2002 Tzeng et al Blood,2011
BMP	Haematopoietic lineage commitment and maintenance of HSC number and function.	Goldman et al, 2009 Durand et al,2007
c-mpl/TPO	Regulates megakaryopoiesis and involved in maintenance and self renewal of HSCs.	Broudy et al,1995 Bersenev et al,2008
Tie2/Ang-1	Maintenance of LT-HSCs	Gomei et al,2010
Hedgehog	Negative regulator of HSC quiescence	Trowbridge et al,2006
Notch	Proliferation, differentiation and cell fate decision	Lin et al,2011
Wnt	Maintains balance between quiescence and proliferation	Kinder et al,2010

Table 2.5: **Signalling pathways involved in haematopoiesis.** HSC-Haematopoietic Stem Cell, LT-Long Term (Adapted from Chotinantakul, 2012)

foetal liver, indicate a role for Wnt signalling in developmental haematopoiesis (Corrigan *et al.*,2008). Similarly in adult haematopoiesis, it has been shown that Wnt5a is highly expressed in HSCs and progenitors (Van Den Berg *et al.*, 1998), whereas Wnt 10b shows significant levels of expression in erythrocytes and immature B cells (Congdon *et al.*, 2008). Germline disruption of Wnt3a leads to substantial depletion in the numbers of haematopoietic stem and progenitor cells in the foetal liver, which further results in embryonic lethality at E12.5 (Luis *et al.*, 2011). The involvement of Notch signalling in haematopoiesis has been well studied, and all the four Notch receptors (1-4) are associated with haematopoietic cells. It has been shown that Notch 1 signalling influences stem cell renewal and differentiation by promoting the self-renewal of lymphoid and myeloid repopulating cells and simultaneously affecting lymphoid differentiation (Ohishi *et al.*, 2003). Several aspects of the MAPK signalling pathway have been found to exert their influence on haematopoiesis. For instance, the ERK pathway maintains the equilibrium between expansion, survival, and differentiation of haematopoietic progenitors.

On the contrary, the JNK and the p38MAPK pathway are indispensable for erythroid differentiation. In addition to this, the MAPK signalling pathway also plays a role in maintaining HSC quiescence and homeostasis (Geest and Coffer, 2009). mTOR belongs to the clan of PI3K related proteins, and it forms two major complexes: mTORC1 & mTORC2. The deletion of both these complexes in mouse models has highlighted the relevance of this signalling pathway in the context of haematopoiesis. Conditional knock out of negative regulators of mTORC1 such as PTEN and TSC1 led to an upsurge in the short term cycling HSCs and a simultaneous reduction in the long- term HSC self-renewal and quiescence. Similar observations were also made in mTOR conditional knock out mice in which BrdU labelling revealed accelerated cell cycling of HSCs, which led to impairment of quiescence and deficient HSC engraftment and repopulation upon transplantation into NSG mice (Malik, Sansom and Michie, 2018).

2.6 miRNAs AND HAEMATOPOIESIS:

miRNAs are 18-24 nucleotides long non-coding RNAs that function by causing translational repression and are involved in a wide repertoire of cellular functions, including regulation of haematopoiesis. Dysregulation of miRNAs has been associated with several haematological malignancies (Kotaki *et al.*, 2017). The role of miRNAs with perspective to haematopoiesis was delineated with the identification of three haematopoietic tissue-specific miRNAs miR142,miR181a, and miR223 in a murine model (Chen *et al.*, 2004). Since then, several miRNAs have been reported to play a role in various facets of haematopoiesis, such as in HSC maintenance, lineage commitment, and differentiation and globin gene regulation (**Figure 2.6**).

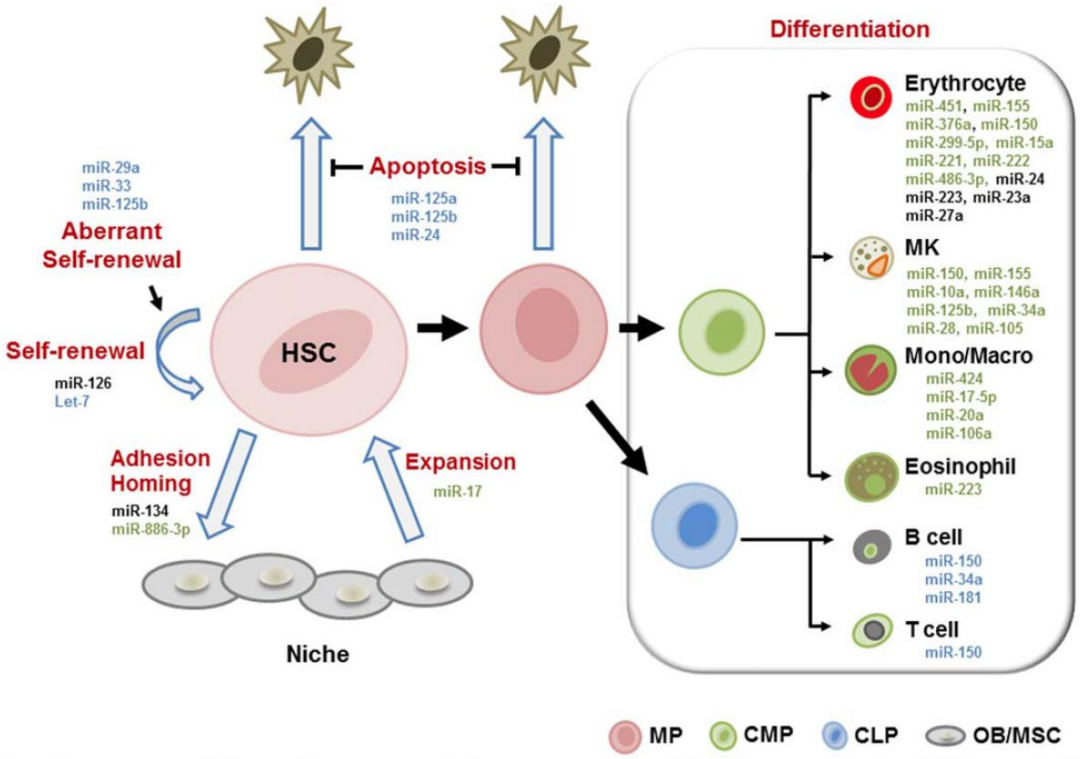


Figure 2.6: miRNAs in the regulation of haematopoiesis. A schematic representation of the miRNAs involved in the self-renewal and differentiation of haematopoietic stem cells. The role of these miRNAs has been functionally validated in murine (blue), human (green), both murine and human (black). HSC-Haematopoietic Stem Cell, MK-Megakaryocytes, MP- Multipotent haematopoietic progenitor, CMP: Common Myeloid Progenitor, CLP- Common Lymphoid Progenitor, OB-Osteoblast, MSC-Multipotent Stromal Cell. Adapted from Hong *et al*., 2015

To exemplify, miR 125a has been found to play a role in modulating the haematopoietic stem cell number (Guo *et al.*, 2010). Similarly, miR 125b and miR 126 have been found to regulate haematopoiesis by targeting Lin28A and attenuating the PI3/Akt signalling pathway, respectively (Chaudhuri *et al.*, 2012), (Lechman *et al.*, 2012). Apart from this, miRNAs are also involved in the differentiation processes. For instance, the miR144/451/4732 cluster has been found to play a crucial role in terminal erythroid differentiation (Dore *et al.*, 2008). On the contrary,

miR 24 has been found to negatively regulate erythropoiesis by targeting ALK4 (Wang *et al.*, 2019). miRNAs such as miR 223 (Fazi *et al.*, 2005) and miR 30c (Katzerke C *et al.*, 2013) play a significant role in modulating granulopoiesis. In addition to these functions, miRNA 96 (Azzouzi *et al.*, 2011) and 486 (Lulli *et al.*, 2013)also play a role in globin gene regulation by inhibiting the levels of γ globin during adult erythropoiesis.

2.7 GROWTH FACTORS AND HAEMATOPOIESIS: Several growth factors have been identified, which play a key role in stem cell maintenance, survival, and differentiation. Most of the growth factors have similar modes of action, which is responsible for the regenerative potential of the haematopoietic system (Thomas, Vadas, and Lopez, 2004). Stem cell factor (SCF), also known as the kit ligand, mast cell growth factor, or steel factor functions as a hematopoietic cytokine that triggers its biologic effect by binding to c-kit (the SCF receptor). It is mostly produced by the bone marrow stroma, and certain HSCs in both soluble and transmembrane forms and both forms are equally active. Its major activities include promoting haematopoiesis at several levels and influencing haematopoietic cell adhesive properties. The Flt3 ligand (FL) belongs to the clan of haematopoietic growth factors, which exhibit specificity for class III tyrosine kinase receptors. The receptor for FL is termed as FMS-like tyrosine kinase 3 (flt3), which is expressed significantly on primitive haematopoietic progenitors. FL promotes the growth of hematopoietic progenitors from the bone marrow, peripheral blood, and cord blood, and its effect is more profound when it functions in conjunction with other haematopoietic cytokines. FL, in combination with SCF and thrombopoietin, triggers

massive proliferation of cord blood CD34+ progenitors in stroma-free cultures. FL, together with myeloid growth factors such as granulocyte-macrophage and granulocyte colony-stimulating factor (GMCSF and G-CSF), or M-CSF, augments the number of myeloid colonies generated from the committed colony-forming units and more primitive long-term culture-initiating cells. FL also functions synergistically with the interleukins IL-7, IL-3, IL-11, IL12, and IL15 to promote B lymphopoiesis, T cell development, and the development of NK cells. In *in vivo* conditions, FL plays a crucial role in HSC expansion, mobilization as well as radioprotection (Wodnar-Filipowicz, 2003). Interleukin 3 (IL-3) is one of the earliest identified cytokines, which is synthesized by activated T and NK cells, keratinocytes, myeloid, stromal, neuronal, and microglia cells. IL-3 receptors are expressed mainly by hematopoietic cells. Several in vitro and in vivo studies have necessitated the role of IL3 in the durability, expansion, and differentiation of hematopoietic progenitors/stem cells.IL3 has been found to both elevate as well as alleviate the self-renewal potential of adult HSCs, and this phenomenon is attributed to the presence of other cytokines and culture conditions (Robin *et al.*, 2006).

2.8 ERYTHROPOIESIS:

Erythropoiesis is an intricate process (Shi *et al.*, 2014), which consists of multiple steps leading to the formation of mature erythrocytes from HSCs (**Figure 2.8**) (Zivot A *et al.*, 2018). The initial stages of erythroid differentiation consist of an engagement phase in which HSCs undergo differentiation to more dedicated erythroid progenitors from megakaryocyte-erythrocyte progenitor (MEP) (Zivot A *et al.*, 2018). The most

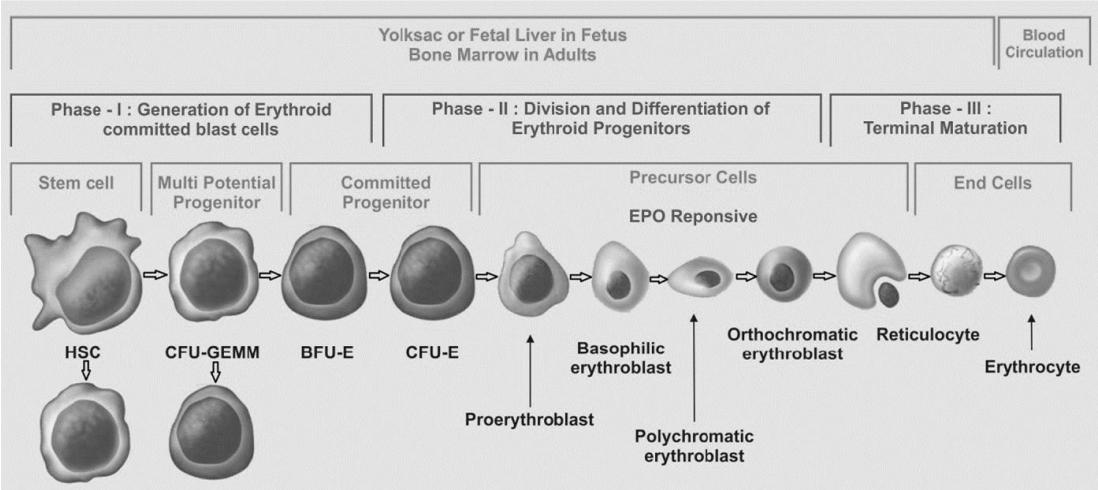


Figure 2.8: *In vivo* erythropoiesis. This figure illustrates the different stages of erythroid differentiation. In Phase I, BFU-E cells are formed from haematopoietic stem cells. In Phase II, BFU-E cells give rise to CFU- E and the subsequent erythroid progenitors. In Phase III, mature erythrocytes are formed through the process of enucleation. HSC: Haematopoietic Stem Cell, CFU-GEMM: Colony-forming unit – granulocyte, erythrocyte, monocyte and megakaryocyte, BFU-E: Burst forming unit erythroid CFU-E-Colony forming unit erythroid, EPO-erythropoietin (Adapted from Singh *et a*l., 2016)

aboriginal devoted erythroid progenitors recognized were the Burst forming Unit Erythroid (BFU-E), which exhibited gradual proliferation (Hattangadi *et al.*, 2011). The BFU-E consists of numerous haemoglobinised cells that make their appearance after 10 to 14 days in methylcellulose cultures. Their proliferation and survival are reliant on multiple growth factors such as SCF, thrombopoietin (TPO), interleukin 3 (IL3), IL11 and Flt3-ligand.The BFU-E cells further mature and differentiate into aggressively multiplying Colony-forming unit- erythroid (CFU-E) (Dzierzak and Philipsen, 2013). The CFU-E consists of minute colonies of 16-25 cells that make their appearance after 5-8 days in methylcellulose culture. They are present in huge numbers in the bone marrow as compared to BFU-E, but they are absent in the circulation (Dzierzak and Philipsen, 2013). The successive stage of erythroid maturation is dominated by the differentiation of nucleated precursors from

proerythroblasts to basophilic, polychromatophilic and orthochromatic erythroblasts (Zivot A *et al.*, 2018) which is accompanied by several considerable alterations such as a reduction in the cell diameter, chromatin condensation, haemoglobinisation, enucleation, and ejection of cellular organelles (Hattangadi *et al.*, 2011). The process of erythroid development culminates with the maturation of reticulocytes into erythrocytes, and dramatic changes occur during the stage wherein the cells attain their biconcave configuration through considerable membrane reassembly and enter into microcirculation until they are expelled by macrophages within the reticuloendothelial system. The phenomenon of terminal erythroid maturation occurs at distinct anatomical niches known as erythroblastic islands. These structures are exclusive to mammalian erythropoiesis and comprise of a centroidal macrophage encircled by several erythroid cells at different levels of maturation. The most predominant function of the central macrophage is to facilitate the anchoring of erythroblasts to themselves and anchoring of erythroblasts to the macrophages, which further provides the cell-to-cell interactions essential to promote erythroid differentiation and expansion. In addition to this, the macrophage also serves to destroy the expelled nucleus from the terminally differentiated erythroblasts and guides the transportation of iron to erythroid progenitors for heme production. Furthermore, they also play a role in regulating the rate at which erythropoiesis occurs by secreting cytokines such as Insulin growth factor I (IGF-I), which stimulates the process of erythroid maturation and proliferation (Zivot A *et al.*, 2018). To summarise, erythropoiesis occurs at a steady pace with a rapid turnover rate to cater to the surplus need for red blood cells (RBCs) within the human body. RBCs survive in the circulation for a span of 120 days during which time they are

monitored by macrophages within the liver and spleen. Macrophages within the spleen can recognize and expel undesirable, destroyed, and old RBCs on the culmination of their life cycle (Zivot A *et al.*, 2018).

2.9 REGULATION OF ERYTHROPOIESIS:

The fascinating process of erythropoiesis is coordinated by an intricate network of transcription factors, non-coding RNAs, as well as epigenetic modifiers (**Figure 2.9**).

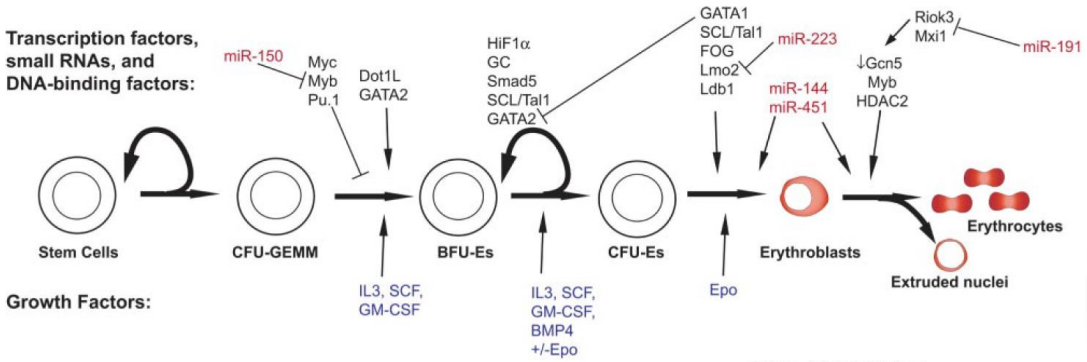


Figure 2.9: Regulators of erythropoiesis. The process of erythropoiesis is governed by a network of both extrinsic factors such as cytokines as well as intrinsic factors such as miRNAs and transcription factors. Adapted from Hattangadi *et al*., 2011

2.9.1 TRANSCRIPTION FACTORS IN ERYTHROPOIESIS:

Numerous transcription factors play a crucial role in modulating erythroid-specific genes by integrating with tissue-specific transcription factors to actuate and suppress transcription of genes during erythropoiesis. Furthermore, transcription factors function in collaboration with cofactors that modify chromatin architecture, usher transcription complexes, or modulate the transcription elongation rate of genes. The most important erythroid transcriptional regulators are GATA-1, GATA-2, KLF1 (EKLF, Erythroid Kruppel like Factor) and NF-E2 (Nuclear Factor, Erythroid 2.

2.9.1.1 GATA FAMILY

GATA-1 and GATA-2 belong to the GATA family of transcription factors, which are zinc finger proteins that play imperative roles during development and differentiation. They are the only two transcription factors of the GATA family, which are expressed in erythroid cells. GATA 1 regulates a diverse number of erythroid-specific genes such as α and β globin genes by interacting with coregulator friend of GATA-1 (FOG-1). It is considered the master transcription factor that modulates erythroid-specific genes. Various erythroid promoters are earmarked by the occupancy of GATA binding sites, which includes transcription factors such as Klf1, GATA-2, Tal1, and GATA1.GATA 2 is expressed significantly in HSCs, CMPs, and MEPs, but its expression is drastically reduced in cells primed for erythroid commitment.

On the contrary, GATA 1 portrays increased levels of expression as cells proceed towards differentiation, but their levels again diminish at the later stages of differentiation, such as in basophilic or orthochromatophilic stages. Numerous studies have implicated the positive regulation by GATA 2 of its own gene as well as GATA 1 in progenitor cells. The phenomenon of "GATA SWITCH "occurs in committed erythroblasts where GATA 2 is replaced by GATA 1, which is further accompanied by a simultaneous shift of the GATA factor function (Fan *et al.*, 2015). GATA 1 has also been shown to interact with the histone acetyltransferase CBP/p300. The occupancy of GATA 1 at different genomic locations has been delineated by genome-wide studies, which have further led to the revelation of three clusters of genes differentially monitored by GATA-1. The first class represented those genes, which are defunct in erythroid cells but showed significant levels of

expression in other hematopoietic cell lineages. The inactive genes were found to be associated with elevated levels of H3K27me3, a histone mark introduced by the Polycomb repressor complex (PRC). The second cluster comprised of those genes, which are expressed at increasing levels during erythropoiesis, and included erythroid-specific proteins involved in heme synthesis and the red blood cell membrane associated protein skeleton. The third cluster included those genes which were involved in the maintenance of housekeeping functions and whose levels declined during erythroid differentiation. These findings are indicative of the fact that GATA-1 plays a crucial role not only for the actuation of the genes which are erythroid-specific but also contributes towards a complete shutdown of the genes which are dispensable for the functioning of the specialized red blood cell (Fan *et al.*, 2015).

2.9.1.2 KLF1

Kruppel- like factor 1, also known as EKLF, is one of the primary transcription factors essential for the transcription of the genes pertaining to erythroid differentiation (Tallack *et al.*, 2012). It is distinguished by a preserved DNA binding domain that belongs to the group of C2H2 zinc fingers. KLF1 plays an essential role in erythropoiesis, and it is highly expressed in most of the hematopoietic sites throughout the process of vertebrate development. Its expression commences at embryonic day 7.5 (E7.5) in the yolk sac, after which it shifts to the foetal liver and ultimately ends in the adult bone marrow and the red pulp of the spleen. EKLF protein is found to be present in both primal and adult erythroid population. However, EKLF has not been detected in HSCs as well as multipotent myeloid

primogenitors. The expression of EKLF escalates gradually as cells mature towards the erythroid lineage and decline progressively in the MEPs, which move towards the megakaryocyte lineage. Mouse embryonic cell lines transduced with doxycycline-inducible EKLF have been used to deduce the effect of KLF1 on haematopoiesis as embryonic stem cells differentiate to form embryoid bodies. Overexpression of EKLF impedes the process of megakaryopoiesis, as evidenced by a decrease in the population of double-positive CD41/CD42b and CD41/CD42d cells. This is accompanied by a simultaneous increase in the levels of mature CD71/Ter119 cells. On morphological analysis of the cells, it was observed that the majority of the cells in the culture, induced with EKLF were basophilic and polychromatophilic erythroblasts. In contrast to the gain of function studies, loss of function studies carried out in primary EKLF-deficient mouse yolk sac or foetal liver cells has shown that there is an upsurge in the number of circulating platelets which are further accompanied by elevated levels of double-positive CD41/42band CD41/42d populations. In the erythroid compartment, EKLF deficiency resulted in abnormalities in the proteins expressed by both primitive and definitive erythrocytes. These results indicated that EKLF portrays a binary role in haematopoiesis by promoting the formation of the erythroid lineage while suppressing the formation of megakaryocytes. EKLF has also been found to target several genes during erythropoiesis. For instance, dematin, α haemoglobin stabilizing protein (AHSP) and KCNN4 are direct targets of EKLF1.Apart from these functions exhibited by EKLF1, it is also involved in some of the critical functions governing erythroid differentiation, such as in globin switching and terminal maturation of erythroid cells. With respect to globin switching, EKLF plays a crucial role in modulating the

switch between foetal and adult globin expression by direct actuation of β globin and indirect suppression of γ globin expression in adult erythroid primogenitors by controlling BCL11A. Similarly, in the context of erythroid maturation, the role of EKLF has been elucidated by carrying out studies in EKLF deficient mice. It is well known that EKLF, through activation of the E2f2 locus, is essential for the cell-cycle progression that antecedes terminal erythroid differentiation. A significant increase has been observed in the proportion of cells with G0/G1 (2N) DNA content in EKLF null E13.5 foetal liver definitive erythroid cells. This is accompanied by the simultaneous reduction in the number of cells containing S-phase (>2N) or G2/M (4N) DNA content. An identical decline in cells with S-phase DNA content has also been observed in EKLF-null cells during primordial (E10.5) erythropoiesis. The levels of E2F2 are drastically low in both the primitive and definitive erythroid compartment in embryos lacking EKLF. E2F2 is one of the principal components of the Rb-E2F complex, which is highly expressed during erythroid differentiation and is involved in the regulation of S-phase ingress. The absence of the E2F2 transcription factor in mice causes macrocytic anaemia, which is linked with defects in erythroid maturation. These findings suggested that the prominent role of E2F2 in erythropoiesis is to promote cell expansion by direct transactivation of subsequent cell cycle genes. With respect to terminal maturation, elevated levels of EKLF are associated with the progression of the cell cycle stage from expansion to differentiation. The increased levels of EKLF have been attributed to EKLF's direct activation of the cyclin-dependent kinase (Cdk) inhibitors p18 and p21, which contribute to cell cycle arrest. Taken together, EKLF plays a vital role in promoting the differentiation and maturation of red cells by directly modulating multiple

components of the cell cycle machinery. In the concluding phase of erythroid maturation, EKLF is responsible for activation of the gene ICAM4, which contributes towards the erythroblast/macrophage interactions within the erythroblastic island. Considering the crucial role played by EKLF during the process of erythroid differentiation, mutations in the EKLF gene have been linked to the prognosis and development of several haematological malignancies (Siatecka and Bieker, 2015).

2.9.1.3 NF-E2

NF-E2 is a heterodimer belonging to the class of the basic leucine zipper family of transcription factors. It has 2 different subunits the p45 subunit, which is tissue-restricted, and the 18kDa subunit, which is associated with the Maf protein family consisting of p18, MafF, MafG, and MafK (Andrews, 1998). The p45 subunit contains a cnc domain and has been found to be expressed during the process of erythropoiesis.NRF1 and NRF2 are two NF-E2 related factors whose expression had also been detected in erythroid cells (Fan *et al.*, 2015). The major function of NF-E2 is to regulate globin gene transcription by acting through locus control regions (LCRs) upstream of the alpha and beta-globin gene clusters (Andrews, 1998). Analysis of the complete protein profile of MafK during erythroid cell differentiation has shown that it interacts with Bach1 and corepressor complexes, which includes the NuRD/Mi2 complex and Brg1-containing chromatin remodelling complexes in undifferentiated cells. However, in cells undergoing differentiation, MafK collaborates with p45 or related factors and interacts with coactivating factors like CBP and the methyltransferase MLL (mixed-lineage leukaemia). Thus, the same

MARE sequences can play dual roles in recruiting repressor or activator protein complexes during the differentiation of erythroid cells. This has been wonderfully portrayed in the β-globin gene locus wherein the case of undifferentiated erythroid cells; BACH1 binds a tandem MARE sequence in locus control region HS2 which is responsible for keeping the LCR in an accessible but inactive configuration in progenitor cells. During differentiation, heme uncouples the BACH1 complex from the DNA, thus facilitating the interaction of MafK/NF-E2 heterodimers, which leads to the recruitment of different cofactors. NF-E2 is required for the transfer of RNA polymerase II from the LCR to the adult β-globin gene promoter at the β-globin gene locus. In addition to this it has been shown that NF-E2 functions in conjunction with USF2 (upstream stimulatory factor 2) which are essential for the recruitment of transcription complexes to the β-globin gene locus. Loss of function studies has shown that lack of p45, leads to an increased interaction of NRF2 with the β-globin gene locus (Andrews, 1998).

2.9.2 LONG NON-CODING RNAS IN ERYTHROPOIESIS

Long non-coding RNAs are described as a diverse group of ncRNAS longer than 200 nucleotides. They are devoid of protein-coding capacity and are positioned primarily in the nucleus and secondarily in the cytoplasm (Salviano-Silva *et al.*, 2018). Most of the long non-coding RNA is polyadenylated and is transcribed by RNA polymerase II (Salviano-Silva *et al.*, 2018). Based on their genomic location relative to the adjacent gene, lncRNAs have been assorted into several types such as intergenic (lincRNAs), bidirectional (divergent), sense, or overlapped (ilncRNAs), antisense (alncRNAs). small RNA (sRNA) hosts (shlncRNAs), enhancer derived (elncRNAs),

and pseudogene-derived (plncRNAs) (Kulczyńska and Siatecka, 2016). *X inactivation-specific transcript (Xist)* was one of the first lncRNA, which was studied. It was found to be positioned in the X inactivation centre region and was associated with X chromosome inactivation in mammalian females. The expression of *Xist* was also found to be repressed in the active X chromosome by another lncRNA, antisense to *Xist* promoter, called Tsix. Even though the function of most lncRNAs is anonymous, it has been seen that lncRNAs are involved in the modulation of several metabolic processes such as development, differentiation, division, survival, and death. In the nucleus, lncRNAs have been shown to engage in imperative chromatin remodelling processes regulating the nuclear architecture that affect gene expression (Salviano-Silva *et al.*, 2018). Several studies have been carried out to delineate the role of lncRNAs during erythropoiesis, and two strategies have been used extensively to interpret the role of lncRNA during erythroid differentiation. The first strategy globally evaluated the expression profile of lncRNAs during RBC ontogeny using state of the art technologies such as microarrays and next-generation sequencing, whereas the second strategy focussed on the functional evaluation of individual lncRNAs (Kulczyńska and Siatecka, 2016).In the last few years, many groups have carried out studies on long non-coding RNAs playing a role in red blood cell formation. For instance, Lodish and his group in the year 2014 dissected the lncRNA profile during erythroid development in vivo by carrying out RNA sequencing of the differentiated E14.5 mouse foetal liver erythroid cells. The RNA sequencing results led to the identification of 655 lncRNAs, of which 132 were novel and were specific to the erythroid lineage. When the expression profile of the top 100 lncRNAs was analyzed, it was observed that

many of the lncRNAs were targeted by key erythroid transcription factors such as GATA1, TAL1, or KLF1. Among these lncRNAs, many of them were found to be novel and showed specificity at every stage of development. When shRNAs were used to diminish the levels of some of the lncRNAs, it was noticed that they played critical roles in the progression stage from terminally differentiated erythroblasts to mature enucleated erythrocytes. In addition to the major findings discussed above in this study, this study also delineated the role of the alncRNA-EC7, which was found to be an enhancer RNA necessary for expression of the adjacent Band 3 (Cells *et al.*, 2014). Another study carried out by Shi et al. explored the role of lncRNAs in the regulation of heme biosynthesis. This study led to the elucidation of the role of the lncRNA Urothelial carcinoma-associated-1 (UCA1) with respect to heme production. UCA1 exhibits dynamic expression during human erythroid maturation, and its expression reaches a zenith in proerythroblasts. Knocking down UCA1 impedes the production of heme and interrupts with erythroid differentiation at the proerythroblast stage. Further analysis divulged that UCA1 associates with the RNA binding protein PTBP1 and behaves as an RNA scaffold to recruit PTBP1 to ALAS2 mRNA. The complex formed by PTBP1 and ALAS2 stabilizes aminolevulinic acid synthase 2 (ALAS2) mRNA, which is one of the key enzymes playing a role during heme biosynthesis (Liu *et al.*, 2018).

2.9.3 EPIGENETIC FACTORS IN ERYTHROPOIESIS

Epigenetic factors play a critical role in the modulation of gene expression, and the essentiality of these factors in gene regulation came into the picture around 10 years back when several scientists proposed that DNA methylation could contribute

towards a heritable modification of gene expression. Ensuing studies further portrayed that DNA methylation occurring specifically at the 5'position of cytosine in the dinucleotide CpG was involved in several biological processes such as in the development of the embryo, cellular differentiation, tumour formation as well as genomic imprinting (Ginder, Gnanapragasam and Mian, 2008). With respect to cellular differentiation, the role of epigenetic modifiers has been studied extensively during the process of erythropoiesis since differentiation of HSCs towards the erythroid lineage is defined by the actuation of erythroid-specific genes that is succeeded by progressive chromatin condensation and gene silencing. In addition to this, histone modifications are also involved in the transition of the switch from human foetal to adult globin expression. This holds significant relevance since the reactivation of foetal globin can alleviate the symptoms of several haematological disorders such as sickle cell anaemia and thalassemia (Broxmeyer, 2008). Towards this, Orkin's group in the year 2015 have demonstrated the role of the subunits EZH1 and EZH2 of the PRC2 complex during blood cell development (Manley *et al.*, 2017). They have shown that EZH1is transcriptionally activated by an erythroid-specific enhancer and a switch from GATA2 to GATA1 regulates the developmental EZH1/2 switch by interacting differentially with EZH1 enhancers.

Further analysis of the PRC2 complex using quantitative proteomics disclosed the presence of an EZH1-SUZ12 sub-complex, which was devoid of EED. EZH1, in conjunction with SUZ12, formed a non-canonical PRC2 complex that occupied active chromatin and positively regulated gene expression. The absence of EZH2 expression led to the relocation of EZH1 to EZH2 targets. Thus, the lineage- and developmental stage-specific regulation of the PRC2 subunit complex led to a shift

from canonical silencing to non-canonical functions during blood stem cell specification (Manley *et al.*, 2017). Another study carried out by Bresnick *et al.* has elucidated the role of the histone methyltransferase SetD8 in boosting erythroid cell survival and maturation (DeVilbiss *et al.*, 2015). According to this study, SetD8, which is a histone methyltransferase catalyses monomethylation of histone H4 at lysine 20 (H4K20me1) and serves as a context-dependentGATA-1 corepressor in erythroid cells. To functionally evaluate the role of SetD8 in erythropoiesis, loss of function studies using shRNAs were carried out in an in vitro primary murine erythroblast culture system. SetD8 in this system augmented erythroblast maturation and survival, but, interestingly, this did not coincide with an increase in the expression levels of the most widely accepted modulator of erythroblast survival Bcl-xL) (DeVilbiss *et al.*, 2015). SetD8 catalysed H4K20 monomethylation (DeVilbiss *et al.*, 2015) at a crucial Gata2 cis-element and confined occupancy by an enhancer of Gata2 transcription, Scl/TAL1, which led to inhibition of Gata2 transcription. An increase in the levels of GATA-2 in erythroid precursors resulted in a maturation arrest, which was synchronous to that caused by SetD8 knockdown. Surprisingly, a simultaneous decrease in the levels of GATA-2 did not lead to a rescue in the erythroid maturation, thus suggesting that SetD8 targets several genes and is a vital deciding factor of erythroid cell maturation (DeVilbiss *et al.*, 2015). Similar studies carried out by Yang in 2018 delineated the role of PHF2 histone demethylase in erythroid differentiation. Plant homeodomain finger 2 (PHF2) is a JmjC family histone demethylase that is involved in the demethylation of H3K9me2, a repressive gene marker. Investigation of PHF2 expression during erythroid differentiation of CD 34+ cells showed that the levels of PHF2 diminish during erythropoiesis and

knockdown of PHF2 augmented erythroid differentiation. The decline in the levels of PHF2 coincided with a decrease in the levels of p53, and further analysis revealed that the promoter of p53 is bound by PHF2, which demethylates H3K9me2 and contributes towards its regulation of expression. To summarise, PHF2 negatively modulates erythroid differentiation by fine-tuning p53 expression levels (Yang *et al.*, 2018).

2.9.4 miRNAs IN ERYTHROPOIESIS

2.9.4.1 miRNA BIOGENESIS

The discovery of miRNAs dates back to the early 90s (Treiber, Treiber and Meister, 2019) when it was first identified in *Caenorhabditis elegans*, but it took several years for it to be recognized as a major transcriptional regulator. Today miRNAs are found in most of the eukaryotes, including human beings, and it represents approximately 1-5% of the human genome (MacFarlane and R. Murphy, 2010). According to the most recent release of the miRNA database (miRBase) around 2588 mature miRNA sequences in humans have been reported (Ha and Kim, 2014). Mature miRNA is formed through a two-step division of primary miRNA (pri-miRNA), which assimilates into the effector complex RNA-induced silencing complex (RISC). The miRNA then acts as a guide by base-pairing with target mRNA to negatively regulate its expression (MacFarlane and R. Murphy, 2010). There are several mechanisms by which mature miRNAs are generated, as mentioned below.

2.9.4.2 CANONICAL PATHWAY OF miRNA BIOGENESIS

The canonical pathway of miRNA biogenesis is one of the most widely used pathways by which miRNAs are processed (**Figure 2.9.4.2**). According to this pathway, pri-miRNAs are transcribed from their host genes by RNA polymerase II/III and the length of the pri-miRNA is mostly over 1 kb. Conventionally a pri-miRNA is comprised of a stem of 33–35 bp, a terminal loop, and a single-stranded RNA segment at both the 5′ and 3′ ends. The earliest step of biogenesis leads to the formation of the single hairpin termed precursor miRNAs (pre- miRNAs) by a nuclear protein complex called microprocessor (Ha and Kim, 2014). The Microprocessor complex consists of the RNase III enzyme Drosha, the double-stranded RNA (dsRNA)-binding protein (dsRBP) DiGeorge critical region 8 (DGCR8) and several other factors such as the DEAD-box RNA helicases p68 (DDX5) and p72 (DDX17) (Treiber, Treiber and Meister, 2019). The N6-methyladenylated GGAC and other motifs within the pri-miRNA are identified by DGCR8, whereas Drosha is involved in cleaving the pri-miRNA duplex at the base of the typical hairpin structure of pri-miRNA. This leads to the generation of a 2 nt 3′ overhang on pre-miRNA. The pre-miRNAs are then transported to the cytoplasm by an exportin 5 (XPO5)/RanGTP complex and are acted upon by the RNase III- type enzyme Dicer (Broxmeyer, 2008). Dicer removes the terminal loop, which results in the formation of a mature miRNA duplex. Dicer functions in conjunction with the dsRBP trans- activation-responsive RNA- binding protein (TRBP), also known as TARBP2) to bring about the following modification (Treiber, Treiber and Meister, 2019). The process culminates with the formation of RNA- induced silencing complex (RISC) wherein the double-stranded miRNA produced by Dicer is loaded

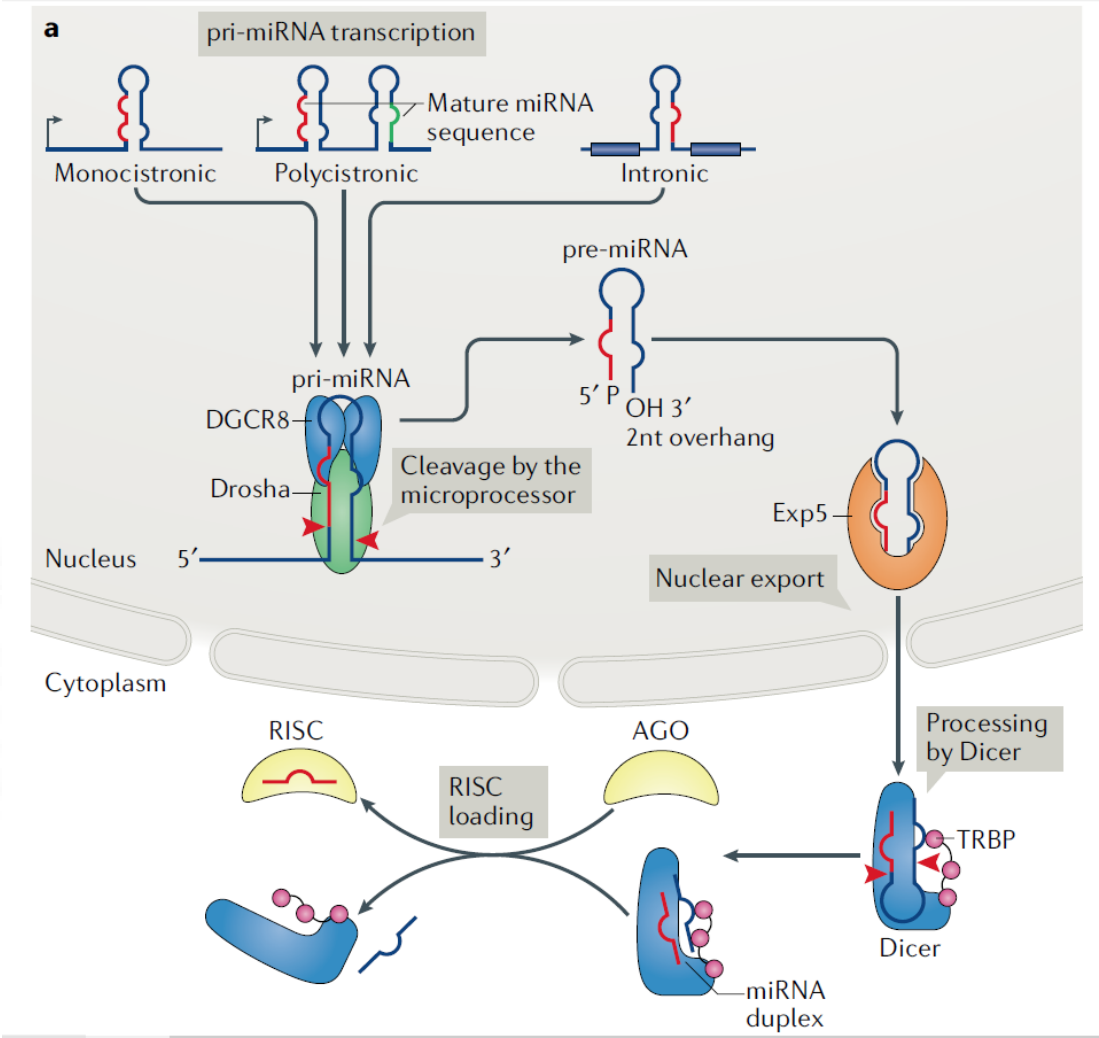


Figure 2.9.4.2: Canonical miRNA biogenesis. The primary miRNA (pri- miRNA) transcripts comprising of hairpins and 5′ and 3′ flanking sequences are generated by RNAPII. The pre-miRNA is then formed from the pri-miRNA by the microprocessor complex, which consists of Di-George Critical Region 8 (DGCR8), and Drosha. The pre-miRNA is then transported to the nucleus by Exportin 5 (Exp5) where it is acted upon by Dicer to generate a miRNA duplex intermediate. The RISC is then formed with the association of proteins such as trans- activation-responsive RNA- binding protein (TRBP) and Argonaute (AGO). One strand of the miRNA acts as a guide strand and the other strand is degraded. Adapted from Treiber *et al*., 2019

onto the Argonaute (AGO) family of proteins (AGO1-4 in humans) in an ATP-dependent manner (O'Brien *et al.*, 2018). The strand possessing diminished 5′ stability or 5′ uracil is primarily loaded into AGO and is designated as the guide strand. On the contrary, the strand that is more stable is designated as the passenger

strand and is separated from the guide strand, which is further degraded by the cellular machinery. The guide strand then binds to its target and causes translational repression (Treiber, Treiber and Meister, 2019).

2.9.4.3 NON CANONICAL PATHWAY OF miRNA BIOGENESIS

Non-canonical pathway of miRNA biogenesis involves several variations as against the conventional biogenesis method (**Figure 2.9.4.3)**. The earliest non-canonical pathway of miRNA biogenesis recognized was the miRtron pathway, which was defined by the union of intron splicing with dicing and is independent of the microprocessor complex (Gris, 2013). The miRtrons usually arise from introns, which, after undergoing splicing, behave as pre-miRNAs and thus bypass cleavage by the microprocessor. These are then transported to the cytoplasm for processing by Dicer. Before the pri miRNAs can acquire the pre miRNA configuration, the debranching enzyme opens the 2′–5′ linkage of the intron lariat. MiRtrons can have 3′ or 5′ flanking regions or extensions on both sides (tailed miRtrons), which are usually removed prior to export. Several features demarcate the canonical pre miRNAs from the miRtrons. The miRtron hairpins are comparatively longer, are frequently 3′ uridylated, and have numerous heterogeneous nucleotides at their ends (Treiber, Treiber and Meister, 2019). Another non-canonical pathway of miRNA processing involves bypassing the Dicer step, and this phenomenon of biogenesis is exhibited mostly by miR-451.miR 451 is one of the best-studied miRNAs in erythropoiesis and is evolutionarily conserved among vertebrates.

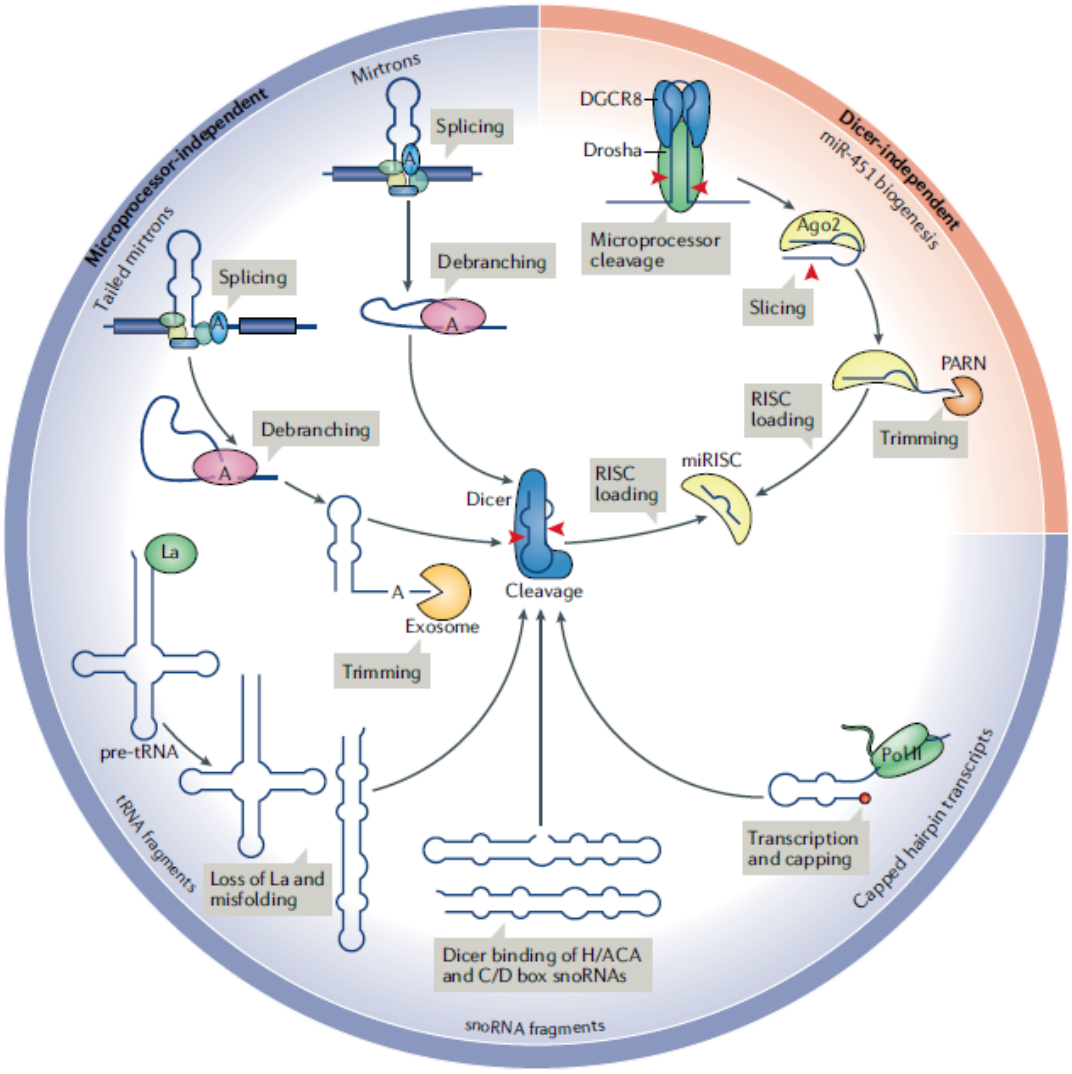


Figure 2.9.4.3: Non-Canonical miRNA biogenesis. Microprocessor independent pathways include miRtrons, tailed miRtrons and the miRNAs and the 7-methyl guanosine capped pre miRNAs. Both miRtrons and tailed miRtrons are formed by splicing which is followed by lariat debranching but the tailed miRtrons are subjected to further trimming by exosomes. Some tRNAs on loss of the RNA chaperone have a tendency to form a hairpin conformation, which is identical to the pre miRNA structure. The 7-methyl guanosine capped pre miRNAs evade the processing by Drosha and gain entry into the cytoplasm with the help of exportin. miR 451 is processed by the Dicer independent pathway since a short hairpin is generated which is not suitable to be processed by Dicer. Adapted from Treiber *et al*., 2019.

Cleavage of the pri-miR-451 by Drosha yields a short hairpin with a stem of~18 bp, which is short to be processed by Dicer. Thus, pre-miR-451 is directly loaded onto AGO2 and is sliced in the centre of its 3′ strand, which results in the formation of a 30-nucleotide-long intermediate species (AGO-cleaved pre-miR-451 (ac-pre-

miR-451)). The 3'end of ac-pre-miR-451 is trimmed by Poly (A)-specific ribonuclease (PARN), which leads to the production of the mature miR-451, which is ~23 nucleotides long(Ha and Kim, 2014). Most of the miRNAs which are produced because of the above-mentioned biogenesis methods usually have a two nucleotide-long 3′ overhang and are designated as the group I miRNAs. However, there are some miRNAs, especially of the let 7 family, which have a 1 nucleotide 3'overhang, and they are designated as group II miRNAs. They are usually extended by one nucleotide through monouridylation, which is facilitated by several terminal uridylyl transferases such as TUT2, TUT4, and TUT7, and this pathway is known as the TUTase dependent pathway (Ha and Kim, 2014).

2.9.4.4 REGULATION OF miRNA BIOGENESIS:

The formation of mature miRNAs is monitored stringently at several stages, which include transcriptional, post-transcriptional and post-translational mechanisms. Numerous post-transcriptional mechanisms have been reported which affect DROSHA and DICER processing, as well as miRNA modification and its yield. RBPs (RNA binding proteins) are one of the major mediators, which play a role in miRNA processing as well as loading of the miRNA duplex into the RISC complex. The Microprocessor-mediated processing of pri-miRNA in the nucleus and the DICER-mediated processing of a pre-miRNA in the cytoplasm is affected positively or negatively by the interaction of the RBP with the terminal loop (TL) or stem of miRNA (Michlewski and Cáceres, 2019). Several RBPs have been identified which play a role in regulating miRNA biogenesis. For instance, it has been observed that the proteins LIN28A and LIN28B play a key role in modulating the let-7 family of

miRNAs in embryonic stem cells (ESCs) and miRNA-9 in neurons of mouse cells. Interaction of LIN28 proteins with let-7 precursors leads to inhibition of their processing at both the nuclear and cytoplasmic levels. A change in the architecture of the let-7 precursor is brought about by binding of the cold shock domain (CSD) of LIN28 to the closed-loop of let-7. This change in the conformation promotes the binding of the CCHC zinc knuckles to a GGAG motif. The activity of the microprocessor in the nucleus is obstructed by binding of LIN28B to the TL of let-7 precursors. In the cytoplasm, LIN28A impedes the processing of the miRNA by Dicer by recruiting TUTase (either TUT4 or TUT7) that adds a short oligo (U) stretch to the 3'end of precursor miRNAs (Michlewski and Cáceres, 2019). Another RBP is the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1), which is specific for miR-18a and accelerates its processing by modifying the pri-miR-18a configuration, thus making it more available for Drosha. On the contrary, the interaction of hnRNPA1 with the conserved terminal loop of let-7a-1 attenuates its expression. The non-existence of hnRNPA1 facilitates the recognition of the let-7a loop by KH- type splicing regulatory protein (KSRP), also known as FUSE binding protein 2), which activates let-7a processing. Another splicing regulator that affects miRNA biogenesis is RNA- binding protein fox-1 homolog 3 (RBFOX3). RBFOX3 binds to a group of pri-miRNAs and positively or negatively modulates microprocessor recruitment in mouse P19 cells differentiated into neurons. Several other examples of RNA binding proteins are Y-box-binding protein 1(YB-1, also known as YBX1), TAR DNA-binding protein (TDP43) and RNA-binding protein Musashi homolog (Musashi 2) (Treiber, Treiber and Meister, 2019).In addition to the above-mentioned post-transcriptional mechanisms, many other post-translational

mechanisms also exist, which play a role in modulating the levels of miRNA biogenesis. Some of the post-translational mechanisms are mentioned below:

2.10 PHOSPHORYLATION

miRNA biogenesis is affected by MAPK signalling. Phosphorylation of DGCR8 results in accelerated microprocessor activity, thus leading to an increase in miRNA levels. The induction of several phosphorylations by MAPKs leads to the formation of a 'pro-growth' miRNA profile. Furthermore, TRBP also undergoes phosphorylation by MAPK ERK, which enhances the stability of the Dicer-TRBP complex and promotes miRNA production. TRBP is also phosphorylated by ribosomal protein S6 kinase (S6K), which is actuated by ERK and mTOR, thus assimilating input from various signalling pathways. It is also well documented that multiple stress-driven pathways stimulate the MAPK p38. MAPK- activated protein kinase 2 undergoes activation by p38, which phosphorylates the microprocessor component p68, thus escalating the biogenesis of several pri- miRNAs. DGCR8 is phosphorylated by the tyrosine-protein kinase ABL and promotes miRNA processing as a response towards DNA damage. The function of Drosha is also influenced by the glycogen synthase kinase 3β (GSK3β) pathway, which plays a role in glycogen metabolism but also endorses cell cycle progression, cell proliferation, and inflammation. Phosphorylation of Drosha at Ser300 and Ser302 is essential for its localization to the nucleus and acceleration in its activity. The biogenesis of miRNA is also regulated by oxygen levels. For instance, it has been observed that phosphorylation of Ago2 at Tyr 293 by epidermal growth factor receptor (EGFR) under hypoxic conditions causes perturbed interaction of Ago2 with Dicer during

RISC loading which leads to a decline of a certain group of miRNAs (Treiber, Treiber and Meister, 2019).

2.11 UBIQUITYLATION AND SUMOYLATION

Recently it has been reported that the communication between the mTOR and the miRNA biogenesis pathways influences miRNA biogenesis. mTOR perceives the energy status of the cell and modulates cellular metabolism in reaction to the environmental alterations. Levels of the p53 inhibitor E3 ubiquitin-ligase MDM2, which acts as an E3 ubiquitin ligase of Drosha, is elevated by mTOR, thus resulting in proteasome-mediated degradation and diminished miRNA processing. Sumoylation has also been shown to play a role in the modulation of miRNA biogenesis factors. Sumoylation of DGCR8 at a specific Lys disrupts its ubiquitination and degeneration. The phenomenon of sumoylation and stabilization of DGCR8 is augmented by ERK-mediated phosphorylation, thus providing a platform for establishing a link between the two modification pathways. TRBP also undergoes sumoylation, which leads to the acceleration of Ago2 binding and RISC loading (Treiber, Treiber and Meister, 2019).

2.12 MODE OF ACTION OF miRNAs:

miRNAs usually exhibit their mode of action by binding to the 3'UTR of target mRNAs and bring about translational repression. However, several studies have shown that miRNAs also bind to the 5'UTR as well as the coding regions (Valinezhad Orang, Safaralizadeh and Kazemzadeh-Bavili, 2014). MiRNAs bring about translational repression of target mRNAs (**Figure 2.12**) by several

mechanisms. Gene suppression by microRNAs at the level of mRNA translation is the most normally narrated mechanism, and it involves suppression of initiation and elongation, ribosome drop-off, and nascent polypeptide degradation. Gene repression by microRNAs at the level of mRNA stability includes microRNA-mediated mRNA decay, sequestration of target mRNAs in P-bodies, and target mRNA cleavage. Apart from this, miRNAs have also been found to act at the transcriptional level by facilitating chromatin remodelling (Morozova *et al.*, 2012). The repression of translational initiation can be brought about by several mechanisms. For instance, miRNA degeneration can be initiated by deadenylation from 3' end or decapping from 5' end by enzymes such as DCP1/2. Missing poly (A) tail and cap structure make the remaining RNA accessible to degradation by exonucleolytic Xrn1p enzyme. In addition to this, sequence-specific endonucleolytic mRNA cleavage by polysomal ribonuclease 1 (PMR1) may also co-occur. The Argonaute protein, which is recruited, binds to multiple initiation factors. Ago has been shown to compete with eIF4E, which is a eukaryotic translation initiation factor involved in recruiting ribosomes to the cap structure of mRNAs, for binding to cap structure. There are many other translation initiation factors which include PABP (the protein associated with poly (A) tail at 3'-end of mRNA, eIF4G which strongly interacts with eIF4E, the RNA helicase that unwinds mRNA secondary structure, eIF4A which is essential for the binding of mRNA to 40S ribosomal subunits, and eIF3 and eIFa which are associated with the ribosomal small subunits.

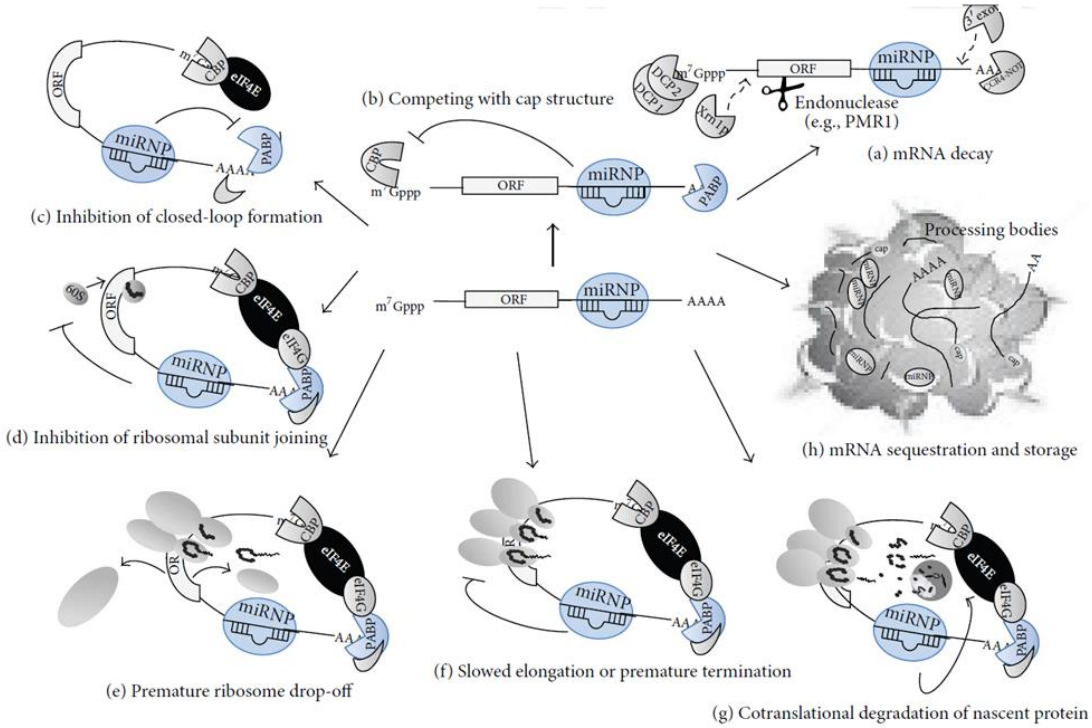


Figure 2.12: Mechanisms of miRNA mediated downregulation. The association between miRNPs and mRNAs leads to several mechanisms of downregulation a) mRNA decay is caused by recruitment of certain enzymes such as endonucleases. (b & c) Argonaute protein competes with cap binding proteins and prevents initiation of translation and formation of closed loop. (d & e) miRNPs impede the association of ribosome subunits and also stimulates its segregation f &g) miRNPs competes with the elongation factors and causes abortive termination and destruction of newly formed proteins. h) mRNAs marked for degradation are cloistered within P bodies. Adapted from Orang *et al*., 2014.

Another mechanism by which translation initiation can be obstructed is by the intrusion of Ago with the formation of closed-loop mRNA, which is attained by the association of cytoplasmic poly (A) binding protein with the cytoplasmic cap-binding protein. In the steps post-initiation, eIF6 is recruited by Argonaute, which hinders the union of the large ribosomal subunit with the miRNA-targeted mRNA (Valinezhad Orang, Safaralizadeh and Kazemzadeh-Bavili, 2014). Recent reports have also suggested the involvement of miRISC in translational repression through suppression of the elongation process. The inhibited mRNA is associated with polysomes, but when the initiation process is arrested with hippuristanol, they

become segregated in a miRNA-dependent manner. Thus, these findings demonstrated that miRISC assists initial ribosome detachment from the mRNAs. To support these findings, three different models have been delineated, which portray the mechanism by which miRISC suppresses the initiation mechanism. Initially, miRISCs were shown to compete with eIF4E for associating with the mRNA 5′ cap structure, which led to a collapse of the translation initiation process. However, some studies disproved this prototype and suggested that either GW182 or a downstream factor could be competing for eIF4E. The second model suggested that miRISC prohibited the circularization of the mRNA, which led to an impediment of translation. It has been seen that the C-C chemokine receptor 4-negative on TATA (CCR4–NOT) complex which is comprised of multiple proteins such as chemokine (C-motif) receptor 4 (CCR4), chromatin assembly factor 1 subunit(CAF1), and NOT1–NOT5 is involved in modulating gene expression and might play a role in hindering miRISC translation. The third model proposed that miRISC interferes with the assembly of the 60S ribosomal subunit with the 40S preinitiation complex. In this process, the 40S ribosomes associate with the targeted mRNA, but the 60S ribosomal subunit is incapable of joining the 40S subunit, thus leading to translation inhibition. Of late, it has also been reported that miRNA/RISC may facilitate translation inhibition through the aggregation of target mRNAs in processing bodies (P bodies). Translation machinery is usually absent in the P-bodies and, therefore, the mRNAs which are housed within the P bodies do not participate in the translation process (Wahid *et al.*, 2010).

2.13 FUNCTIONAL VALIDATION OF miRNAs

miRNAs are a group of small, 19–24 nucleotide noncoding RNAs that governs gene expression by degradation or translational suppression of their target mRNAs by binding to their 3'UTR with imperfect complementarity. This imperfect complementarity bestows the miRNAs with the property of binding to multiple targets simultaneously, thus leading to the involvement of miRNAs in coordinating various cellular functions such as apoptosis, haematopoiesis as well cellular differentiation and proliferation (Mohammdai-Asl *et al.*, 2015). With respect to cellular differentiation, miRNAs have been found to play vital roles in several aspects of erythropoiesis such as in erythroid differentiation and maturation, erythroid iron metabolism as well as haemoglobin synthesis as shown in (**Table 2.13)**.

miRNA	Expression	Target gene	Function	References
miR-15a, miR-16-1	Down	*c-Myb*	Differentiation toward the erythroid lineage from CFU-EM	Hattangadi et al., 2011; Azzouzi et al., 2012
miR-24	Bi-phasic	*hALK4*	Modulates negatively erythropoiesis	Wang et al., 2008; Havelange and Garzon, 2010
miR122	Down	*Hfe Hjv*	Encode activation of hepcidin expression	
miR126	Down	*PTPN9*	Growth and expansion of normal erythroid cell	Huang et al., 2011
miR-144/451	Up	*FOXO3 GATA-2 KLFD*	Regulates oxidative stress, positive regulation of terminal erythroid differentiation and in zebrafish hemoglobin synthesis (miR-144)	Fu et al., 2009; Pase et al., 2009; Havelange and Garzon, 2010
miR-155	Down	*PU-1/ETS-1 CEBPβ/SHIP1*	Inhibits erythropoiesis	Masaki et al., 2007; O'Connell et al., 2008; Havelange and Garzon, 2010
miR-191	Down	*Riok-3 Mxi-1*	Erythroblast chromatin condensation and enucleation	Zhang et al., 2011; Bianchi et al., 2012
miR-210	Up	*HIF Scf EPO*	Induced erythroid differentiation from erythroid progenitor	Kosaka et al., 2008; Byon and Papayannopoulou, 2012
miR-221/222	Down	*c-Kit*	Inhibits early erythroid proliferation Controls perinatal Hb switching	Felli et al., 2005; Bianchi et al., 2012
miR-223	Down	*LMO2*	Inhibits erythropoiesis	Felli et al., 2009; Havelange and Garzon, 2010
miR-320	Down	*CD71 (TFR-1)*	Central mediator of iron-loaded transferrin uptake	Chen et al., 2008; Lawrie, 2010
miR-376a	Down	*CDK-2 AGO*	Differentiation toward the committed erythroid/ megakaryocyte lineage from HSC	Wang et al., 2011
Let-7d	Down	*DMT-IRE*	Regulation Hemin-induced erythroid maturation and iron homeostasis	Andolfo et al., 2010; Listowski et al., 2013

Table 2.13 miRNAs play a role in several aspects of erythropoiesis.

2. 14 miRNAs REGULATING ERYTHROID DIFFERENTIATION AND MATURATION

Several miRNAs have been implicated to play a role during erythroid differentiation such as miR150, miR15a, miR126, miR221/222, miR 223, miR24, miR144/451, and miR191.The role of miR 150 has been studied exclusively in the context of megakaryopoiesis and erythropoiesis. In in vitro conditions, overexpression of miR 150 resulted in the proliferation of megakaryocytes as against erythrocytes (Lu *et al.*, 2008). Similar results were obtained in an in vivo murine bone marrow transplant model wherein ectopic expression of miR 150 in murine hematopoietic stem/progenitor cells using a retrovirus displayed a decline in erythroid cells and an upsurge in megakaryocytic cells (Lu *et al.*, 2008). Subsequent studies identified c-Myb to be the key target of miR-150, which was confirmed by both loss and gain of function studies of c-Myb. miR-15a was another miRNA that targeted c-Myb. miR-15a and c-Myb shared a reciprocal relation during the process of in vitro erythroid differentiation (Zhao *et al.*, 2009). The gain of function studies showed that miR-15a played a role both in impeding the process of erythroid and myeloid differentiation as well as in disrupting cell cycle progression. The regulation of miR-15a and c-Myb occurs through an autoinhibitory feedback loop wherein c-Myb activates the expression of miR-15a (Zhao *et al.*, 2009). The effects of miR-126 were found to be in similar lines with MiR-15. Studies carried out in zebrafish showed that knockdown of miR-126 resulted in the production of more erythrocytes and less of thrombocytes (Das C, Lucia MS, 2017). It also led to haemorrhage and loss of vascular integrity. Similar research carried out in human embryonic stem cells (hESCs) showed that the formation of the erythroid colony was hindered via

inhibition of the protein-tyrosine phosphatase, PTPN9.The miR221/222 cluster is one of the most well-known clusters in erythropoiesis, and its levels have been found to decrease with the progress of erythroid differentiation. Enforced expression of miR 221/222 in CD 34+ cells led to hindrance in the cell growth, increased erythroid maturation and accelerated apoptosis in the final stages of the culture (Felli *et al.*, 2005). C-Kit is the target of miR221/222, which is expressed in the initial stages of erythroid cells but not in the later stages. miR 223 is another miRNA whose expression levels diminish during erythropoiesis (Felli *et al.*, 2009). Ectopic expression of miR-223 impeded erythroid maturation as well as erythroid colony formation of UCB CD34+ cells (Warren *et al.*, 1994). The obstruction of maturation block was more evident towards the terminal stages of erythroid maturation with declining numbers of orthochromatophilic erythroblasts and the build-up of more premature erythroid cells. The transcriptional co-regulator LMO2 has been identified to be the major target of miR-223.LMO2 has been designated to be one of the crucial regulators of erythropoiesis since embryonic lethality is exhibited by LMO2 deficient mice (Warren *et al.*, 1994). Mice harbouring LMO2 mutations die at E10.5 due to the collapse of yolk sac erythropoiesis (Warren *et al.*, 1994). miR-24 is one of the other miRNAs which has significant levels of expression in CD 34+ cells, but its expression reduces during the course of erythroid differentiation. Reduction in the levels of this miRNA during differentiation leads to an increase in the levels of Alk4, which promotes erythroid maturation and haemoglobin synthesis by interacting with activin (Wang *et al.*, 2008). The best-characterized and the most highly expressed miRNAs in erythropoiesis is the miR144/451 cluster. The expression of this miRNA cluster is regulated by GATA1. In zebrafish, it has been demonstrated that disruption

of miR144/451 leads to anaemia and a reduction in the number of circulating RBCs (Dore *et al.,* 2008). GATA2 is one of the targets of miR144/451 cluster in zebrafish, and interestingly in mammalian haematopoiesis also downregulation of GATA 2 is indispensable for erythropoiesis to occur (Grass *et al.,* 2003). In vivo studies carried out in the murine bone marrow transplant model have indicated that disruption of miR144/451 leads to a decline in the number of mature erythrocytes with the deficiency occurring at the transition from the proerythroblast to the basophilic stage (Papapetrou, Korkola and Sadelain, 2010). Several groups have shown that a wide range of variability exists in the phenotypes exhibited by miR144/451 knock out mice. For instance, Donal O'Carroll's group has shown that mice deficient in miR144/451 exhibit erythroid hyperplasia which indicates inefficient erythroid differentiation and anaemia with the obstruction occurring at the basophilic stage (Rasmussen *et al.,* 2010). Mice harbouring miR 451 mutations display haemolytic anaemia and deferred restoration after being exposed to Phenyl Hydrazine. In addition to this, miR-144/451 null erythrocytes show significant accumulation of reactive oxygen species when they are exposed to hydrogen peroxide. Taken together, this is suggestive of the fact that the role of miR144/451 might be restricted more towards stress erythropoiesis than in normal erythropoiesis. This function of miR144/451 is supported by a study in which 14-3-3ζ has been identified as one of the important targets of miR 451. 14-3-3ζ is a phospho-serine/threonine binding protein that localizes FoxO3 to the cytoplasm and negatively regulates it (Yu *et al.,* 2010). FoxO3 is a master transcriptional regulator of anti-oxidant responses in several tissues, which also includes erythroid cells, and it has been seen that there is a reduced nuclear accumulation of FoxO3 in miR 144/451 null erythroid cells. In

addition to these miRNAs playing vital roles in erythroid maturation, miR 191 has been found to play a role in the process of enucleation during terminal differentiation (Zhang *et al.*, 2011). miR 191 was identified from RNA sequencing studies carried out in differentiating murine foetal liver erythroblasts and gain of function studies demonstrated that miR 191 impeded erythrocyte enucleation by targeting several genes such as Riok3 andMxi1 (Zhang *et al.,* 2011).

2.15 miRNAs REGULATING IRON METABOLISM

DMT1, Nramp2, SLC11A2 is one of the important metal transporters which plays a crucial role in absorbing and utilizing iron, and it is also responsible for iron discharge from the endosomes. It has four different isoforms, and the expression levels of DNMT1 have been found to increase during erythroid differentiation. The levels of DNMT1 are modulated by miRNA let-7d, and overexpression of let 7d in erythroid cell lines led to the impairment of hemin driven erythroid maturation, which interfered with iron homeostasis. This anomaly was attributed to the reduction in the levels of the ferritin heavy chain (FH1) and an increase in the levels of transferrin receptor C (TfRC) (John CH Byon., 2013).

2.16 miRNAs IN HAEMOGLOBIN SYNTHESIS

Haemoglobin synthesis is an extremely complicated process that requires the involvement of both cis and trans-acting factors. The primary evidence of miRNAs being involved in the regulation of haemoglobin synthesis came from studies carried out in zebrafish, in which it was demonstrated that an increase in the expression of miR 144 negatively correlated with the expression of embryonic α-globin during

zebrafish ontogenesis. Loss of function studies led to an escalation in the levels of embryonic α-globin without perturbing β-globin synthesis. The transcription factor KLFD was found to be the target of miR 144. In addition to this, it has also been observed that the existence of a negative feedback loop between miR-144 and KLFD attunes the levels of KLFD.miR-15a and miR-16-1which exist in a cluster have also been found to regulate the expression levels of γ globin by targeting c-myb. The expression of miR-15a and miR-16-1 in adult human hematopoietic progenitors ectopically led to the elevation of the γ-globin by almost 50%. Phenotypically, cells that overexpressed miR-15a and miR-16-1displayed diminished cell cycle progression only at the CFU-E/proerythroblast stage (John CH Byon., 2013).

2.17 MODEL SYSTEMS TO STUDY ERYTHROPOIESIS:

Erythropoiesis is envisioned as a process in which multipotent haematopoietic stem cells expand, differentiate, and culminate with the formation of mature erythrocytes. The most striking feature of erythropoiesis is illustrated by the fact that each mitosis gives rise to daughter cells, which are functionally and phenotypically distinct from their derivative parent cell. This salient feature highlights the importance of studying erythropoiesis in a stage-specific manner. Furthermore, it has also been observed that several haematological disorders are attributed to the disruptions in the gene expression pattern at specific developmental stages. Thus identifying the stage-specific defects and molecular pathways will help in comprehending the underlying mechanisms of defective erythropoiesis, which will pave the way for devising strategies to manage the clinical symptoms more effectively (Liu, Han, and An, 2015).In this aspect, several model systems have been developed to study

erythropoiesis at different developmental stages. In the earlier days, erythropoiesis was mostly studied in mouse chicken, goat, sheep, and rabbit, which led to a better understanding of the anatomic structure of the yolk sac blood islands, foetal liver, and bone marrow niches for erythropoiesis. The recent advancements in the field have led to the development of several cell lines, which can be driven to differentiate to the erythroid lineage. In addition to this, genetically modified animals have also come into play to study the process of erythropoiesis.

2.17.1 IMMORTALIZED CELL LINES: Cell lines are an important tool to study erythropoiesis because it is possible to manipulate them genetically. For instance, the Murine Erythroleukemia (MEL) cell lines are erythroid progenitor cells that are obtained from the spleens of mice infected with the Friend virus complex (I). These cells upon infection by the virus do not mature beyond the proerythroblast stage and can be maintained in culture over an indefinite period. These cells can differentiate towards the erythroid lineage and express adult haemoglobin when treated with compounds such as dimethyl sulfoxide and hexamethylene bisacetamide (Murray and Antoniou, 2003). Similarly, erythroid cell lines obtained from haematological malignancies have also been used to study erythropoiesis. The K562 cell line was procured from a patient with Chronic Myeloid Leukemia (Klein *et al.*, 1976). These cells undergo differentiation to the erythroid lineage and express embryo-foetal globin genes such as the Zita, epsilon, and gamma-globin gene when treated with hemin, anthracyclines, sodium butyrate, and hydroxyurea (Tsiftsoglou, Vizirianakis and Strouboulis, 2009). Human Erythroleukemia (HEL) is another cell line that was obtained from a patient initially diagnosed with Hodgkin's disease but who developed erythroleukemia later. This cell line has the ability to synthesize both Gγ

and Aγ chains of globin in considerable amounts when treated with hemin. In addition to these embryonic chains (ε, ζ) are also produced in very small quantities. However,β chains are undetectable (Martin and Papayannopoulou, 1982). Thus, these cell lines exhibit a globin expression profile, which is similar to embryonic or foetal cells. On the contrary, the human umbilical cord blood-derived erythroid progenitor (HUDEP-2) cell line which was developed by Kurita et al. by immortalizing umbilical cord blood CD34+ cells using lentiviral delivery of a tetracycline-inducible HPV16-E6/E7 transgene expresses β globin in a significant account and is identical to adult erythroid cells when cultured in erythroid differentiation media. Although these cell lines have proved to be an invaluable tool to carry out gene expression studies and identify genes that can lead to an increase in foetal haemoglobin, they harbour several limitations. For instance, the cell lines obtained from cancerous individuals pose the risk of having an abnormal genetic or chromosomal complement, and cannot be made to developmentally switch from expression of one globin to another. On the other hand, the immortalized cord blood cells are transformed cells rather than representative of native erythroid cells. In addition to these permanent cell lines, engineered cell lines such as the G1E cell line and its derivative G1E-ER4 was also developed to elucidate the transcriptional machinery governing the process of erythropoiesis (Tsiftsoglou, Vizirianakis and Strouboulis, 2009).G1E cells are a GATA null cell line derived from embryonic stem cells in which GATA1 is disrupted. These cells bear a similarity to proerythroblasts and do not undergo differentiation. These cells rely on the presence of Kit and erythropoietin when grown in culture. On the other hand, the G1E-ER4 cell line expresses a fusion product that contains GATA-1 with the estradiol receptor ligand-

binding domain. Treating the cells with either estradiol or tamoxifen leads to the functional activation of GATA-1, which induces erythroid differentiation up to a stage that is identical to the orthochromatic erythroblast, and it also expresses adult haemoglobin.

2.17.2 EMBRYONIC STEM CELLS: These cells are obtained from the inner cell mass of preimplantation blastocysts, and under in vitro conditions, they retain their ability to self-renew and differentiate into several different tissues. Mouse ES cells have been used for in vitro differentiation into several hematopoietic lineages, especially into erythroid cells. Two different approaches have been used to differentiate mouse embryonic stem cells into erythroid cells. The first approach involves the culture of the disaggregated embryoid bodies in the presence of growth factors such as SCF and Epo, which stimulates growth and differentiation of erythroid progenitors (Keller *et al.*, 1993). This approach leads to the production of both definitive and primitive erythroid colonies, which are distinguished by specific characteristic features such as the nucleus and expression of globin genes. The second approach developed by Nakano and his colleagues involved culturing the ES cells in the presence of the OP9 stromal cell line, which supports hematopoietic differentiation by providing a supportive hematopoietic niche. This approach also leads to the formation of both definitive and primitive erythroid colonies. Thus these approaches recapitulate murine erythropoiesis, which also occurs in both primitive and definitive phases, and it has served as an excellent tool to study transcription factors in erythropoiesis, especially GATA1 (Tsiftsoglou, Vizirianakis, and Strouboulis, 2009).

2.17.3 ANIMAL MODELS FOR ERYTHROPOIESIS: Mouse, zebrafish, and Xenopus are the gold standard models that are being used to study erythropoiesis. Leonard Zon and his colleagues initiated the use of zebrafish as a model to study erythropoiesis and globin gene regulation in the early ’90s. It serves as an ideal system because it has a shorter generation time than mice, which enables examination of larger no of progenies as compared to mice. Furthermore, it is also useful to carry out large genetic screens, which have led to a better understanding of erythropoiesis. The mice also serve to be an excellent system for studying erythropoiesis since mice have a short gestation period, and erythropoiesis in mice mirrors human erythropoiesis. The first transgenic mouse to study globin gene regulation was developed by Magram et al. in 1985 in which the cloned globin genes were present throughout erythroid development (Jeanne Magram, 1985). Subsequently, with the discovery of the β-globin locus control region, it was possible to create transgenic mice harbouring the entire β-globin gene locus. These mice were capable of expressing the human globin genes in a tissue- and developmental stage-specific manner, which was pivotal in studying the multiple aspects of globin gene regulation in vivo.

2.18 LENTIVIRAL PROMOTERS IN HSCs

An efficient lentiviral mammalian expression promoter, which does not undergo silencing during HSC maintenance and differentiation, is essential so that the shRNA or the transgene driven by the promoter is stably expressed throughout the process. Studies carried out in cord blood HSCs have shown the efficacy of EF1 and hCMV

promoters up to 6 weeks post-transduction (Varma *et al.*, 2011). Similarly, it has also been shown in several leukemic cell lines and cord blood HSCs that high level of multiple transgenes can be expressed by using lentiviral vectors with two independent internal promoters (Yu *et al.*, 2003).

2.19 TOOLS TO STUDY TRANSCRIPTIONAL REGULATION IN ERYTHROPOIESIS

The process of erythropoiesis is a multi-staged process that is comprised of a distinct population of cells at each stage. Therefore, it is essential to comprehend the different mechanisms, which govern this process. Several approaches, such as RNA sequencing and microarray, have been used to elucidate the transcriptome profile during the process of erythropoiesis. In addition to this library, screening approaches have also been used to decipher genes involved in various pathways during erythroid differentiation.CG et al. exploited the technique of microarray and looked at the gene expression profile during erythroid differentiation of murine erythroleukemia cells. This study led to the identification of several genes, such as H1f0, Bnip3, Mgl2, ST7L, and Cbll1, which could serve as potential markers for erythropoiesis (Heo *et al.*, 2005). Similar studies carried out by Robson et al. during erythroid differentiation of human primary differentiating erythroblasts led to the identification of several transcripts that were previously not described in the context of erythropoiesis.

Moreover, several transcription factors were also identified, which played a role during terminal erythropoiesis. The RNA sequencing technique led to the

identification of genes that exhibited different functions during different stages of erythropoiesis. In addition to this, it also helped in elucidating the differences between human and murine transcriptome (An *et al.*, 2014). Similar studies carried out during ex vivo differentiation of CD 34 positive cells to erythroid cells led to the identification of several novel intergenic and intronic transcripts as well as novel alternative transcript isoforms. This study also illustrated the existence of two transcriptional waves during terminal erythropoiesis. First, the intricacy of the transcript diversity was attributed to alternative splicing, and secondly, the heterogeneity of the splicing junctions reduced dramatically during erythroid differentiation. The most differentially expressed transcripts were found to play a role in haematopoiesis (Shi *et al.*, 2014). The latest technology, which has gained momentum to study the transcriptional profile of several biological processes, is the library screening approach in which shRNAs against several genes are simultaneously cloned into a lentiviral vector, which is then transduced into a particular cell type, and their effects are then studied. For instance, through this approach several genes such as PPARγ, cohesin, Msi2, Prox1and cytohesin 1 have been identified which play a role in stem cell homeostasis, stem cell renewal and differentiation and stem cell homing respectively (Hope *et al.*, 2010) (Galeev *et al.*, 2016) (Sertorio *et al.*, 2017) (Rak *et al.*, 2017).

3. MATERIALS AND METHODS

3.1 PLASMID EXTRACTION

Plasmid extraction was performed using a plasmid mini preparation kit (Favorgen) as per the manufacturer's protocol. The bacterial culture which was grown overnight (5 ml) was transferred to a 2 ml microcentrifuge tube and centrifuged at 12000g for 2 minutes. The supernatant was discarded, and 200 µl of resuspension buffer was added to the cell pellet and suspended well. To lyse the cells, 200 µl of lysis buffer was added, and it was mixed thoroughly by inverting the tubes several times. The samples were then incubated at room temperature for 5 minutes. The neutralization step was carried out by adding 300 µl of neutralization buffer to the tubes, and they were mixed thoroughly by inverting the tubes several times to neutralize the lysate. The tubes were then centrifuged at 12000g for 5 minutes, and the supernatant was transferred to the column. It was then centrifuged at 11000 g for 1 minute. The flow-through was discarded, and 400 µl of wash buffer was added to the column and centrifuged for 1 minute. The flow-through was discarded, and the washing step was repeated. The tubes were centrifuged for 3 minutes to dry the column. The column was then placed in a fresh 1.5 ml centrifuge tube, and 50 µl of elution buffer was added to the column. The tubes were then incubated for 3 minutes and centrifuged at 11000g for 1 minute to elute the plasmid DNA. The concentration of the plasmid was quantitated using Nanodrop, and they were then stored in -80^0C.

3.2 TRANSFORMATION

2 µl of ligated product or 300 ng of plasmid DNA was added to 50 µl of XL1-Blue competent cells. It was then incubated in ice for 30 minutes. Heat shock was given at 42^0C for 45 seconds, and it was incubated in ice for 2 minutes. 250 µl of SOC medium was then added to the tubes and kept at 37^0C at 800 rpm for 90 minutes. 50 µl from the mixture was plated into Luria Bertani Agar (LB-Agar) plates containing appropriate antibiotics. The plates were then incubated overnight at 37^0C for 14 -16 hours.

3.3 CLONING shRNAS IN "miR E" shRNA BACKBONE

The shRNAs were cloned into the "miR E" backbone (Fellman, 2013), as described below. The shRNA sequences were amplified from the plasmids into which the shRNAs were previously cloned using the primers having *Xho*I and *Mlu*I restriction sites. pGIPZ shRNA vector was digested with the enzymes, *Xho*I (New England Biolabs) and *Mlu*I-HF (New England Biolabs), for 4 hours at 37^0C, and then it was treated with antarctic phosphatase (New England Biolabs) for 1 hour at 37^0C. The amplified shRNA sequences were then cloned into the corresponding cloning sites of the shRNA expression vector using T4 DNA ligase (Takara) at a vector: insert ratio of 1:3, and the tubes were incubated overnight at 16^0C. 2 µl of ligated products were then transformed into XL1-Blue competent cells and plated onto LB-Agar plates containing ampicillin. The plates were then incubated at 37^0C for 16 hours, and 3 colonies from each plate were inoculated into LB- Broth and incubated in a bacterial shaker at 225 rpm at 37^0C for 16 hours. The plasmids were then extracted and were

then digested with *Xho*I and *Mlu*I-HF enzymes to confirm the presence of inserts. Subsequently, the plasmids were subjected to Sanger sequencing.

3.4. CLONING shRNAS IN "ULTRAmiR" BACKBONE USING GIBSON ASSEMBLY

UltramiR scaffold is another variant of the miR 30 scaffold, which was constructed by removal of the existing restriction sites from the previous miR 30 scaffolds (Knott, 2014). This modification helped in retaining the structure of endogenous miR 30, which further helped in enhancing the processivity of the shRNAs. The shRNAs were cloned into the UltramiR scaffold, as described below.

3.5 CONSTRUCTION OF ULTRAmiR pGIPZ BACKBONE

A lentiviral vector containing the UltramiR scaffold was generated, as shown in **Figure 3.1** The UltramiR scaffold was amplified with primers UltramiR-F and UltramiR-R **(Table-3.3),** which contained *Hpa*I and *Mlu*I restriction sites, respectively. They were then subjected to restriction digestion with enzymes *Hpa*I (New England Biolabs) and *Mlu*I-HF (New England Biolabs) for 4 hours at 37^0C. It was then cloned into the corresponding cloning sites of the pGIPZ vector at a vector: insert ratio of 1:3 using T4 DNA Ligase (Takara). The tubes were then incubated overnight at 16^0C. 2 µl of ligated products were then transformed into XL1 competent cells, and the transformed product was plated onto LB-Agar ampicillin plates. The plates were incubated at 37^0C for 16 hours, and 2 colonies from each plate were inoculated into LB- Broth and incubated in a bacterial shaker at 225 rpm for 16 hours. The plasmids were then extracted using a plasmid extraction kit, and

they were then digested with *Hpa*I and *Mlu*I-HF enzymes to confirm for the presence of inserts. After confirmation for the presence of inserts, the plasmids were subjected to Sanger sequencing.

3.6 CLONING OF shRNAS BY GIBSON CLONING

shRNA sequences were cloned into the lentiviral vector, which was generated in section 3.5 using the Gibson assembly, as shown in **Figure 3.2**. The shRNA sequences were amplified using the primers having *Hpa*I restriction site (UltramiR HpaI-F and UltramiR HpaI-R) **(Table-3.3)**. The UltramiR vector was digested with *Hpa*I enzyme (New England Biolabs) at 37^0C for 4 hours, and both the vector and inserts were purified using a PCR product purification kit. The vector and inserts were then ligated at 1:2 ratio using NEBuilder Hi-Fi DNA Assembly Master Mix (New England Biolabs) at 50^0C in thermocycler for 1 hour. After the ligation was complete, the transformation was done on LB Agar plates containing ampicillin, and 5 colonies were inoculated in ampicillin containing LB-Broth. The tubes were incubated for 16 hours, and plasmids were extracted using a plasmid extraction kit. The plasmids were digested with *Hpa*I (New England Biolabs) and *Mlu*I-HF (New England Biolabs) enzymes to confirm the presence of inserts, and they were then subjected to Sanger sequencing.

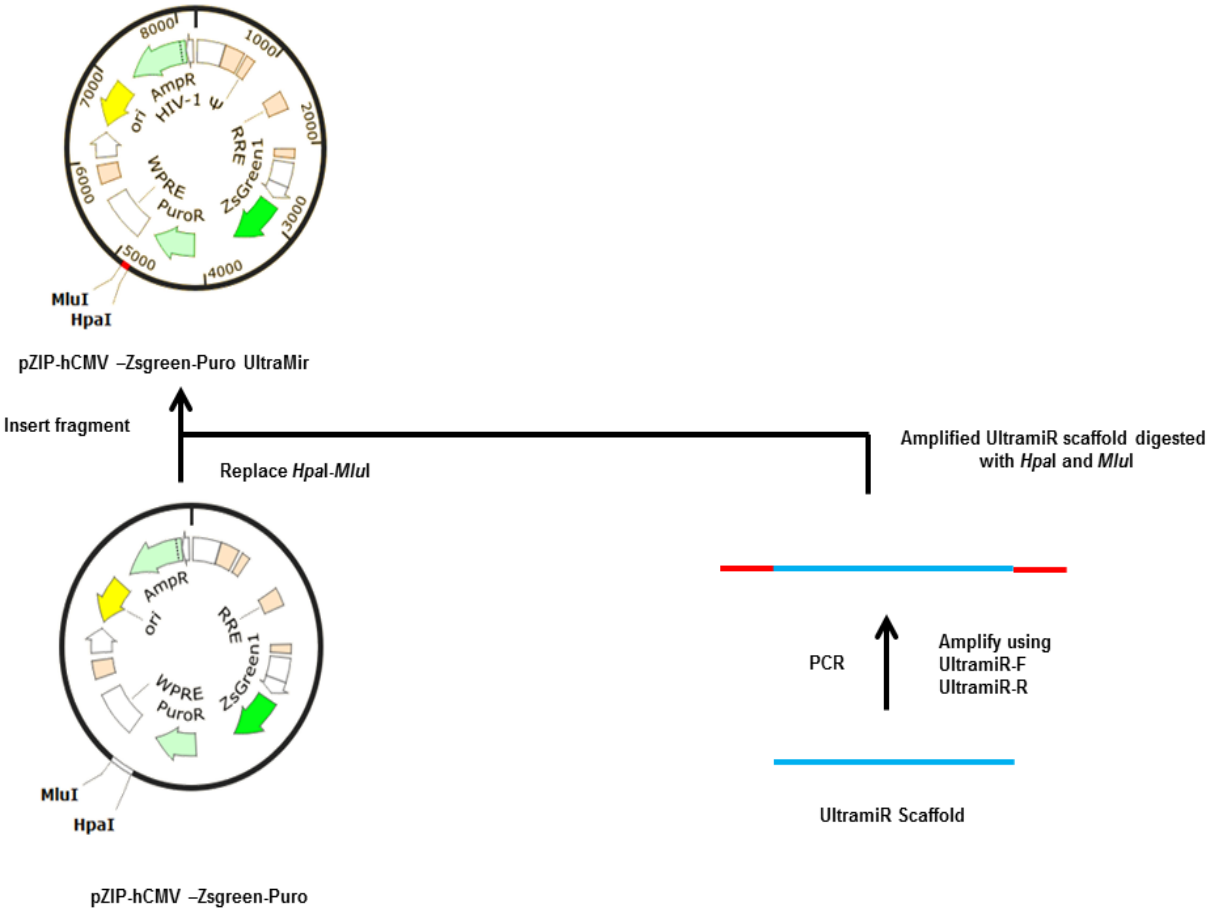


Figure 3.1. Generation of a lentiviral vector for shRNA cloning. An oligo which contains the scaffold sequence was synthesized and amplified using specific primers. The amplified product is cloned into the pZIP shRNA expression vector by Gibson cloning.

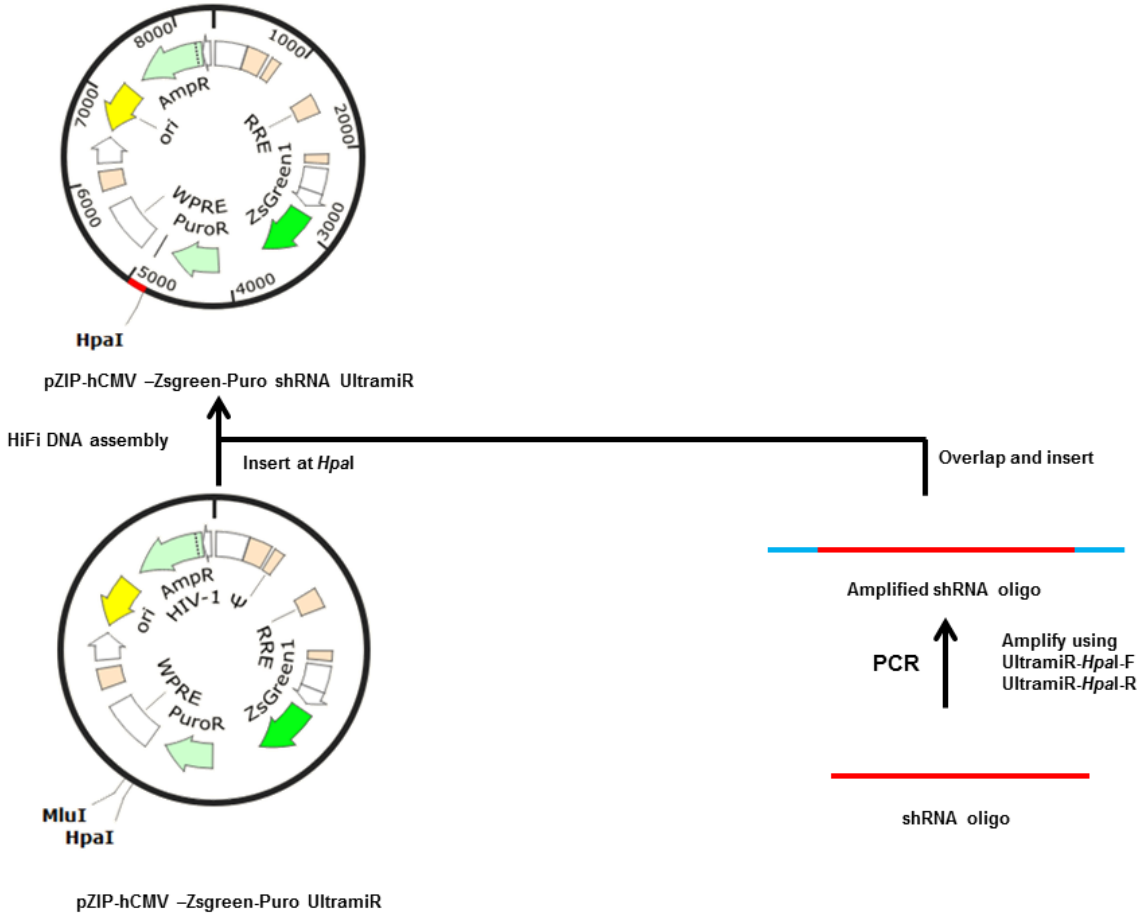


Figure 3.2: Generation of lentiviral shRNA vectors using Gibson Assembly. Oligos with the shRNA sequences were synthesized, amplified using specific primers and cloned into the pZIP plasmid by Gibson cloning.

3.7. CONSTRUCTION OF miRNA OVEREXPRESSION VECTORS

A lentiviral vector for overexpressing miRNAs was generated, as shown in **Figure 3.3**. For overexpression of miRNAs, miR 144, miR 451, miR 4732, miR 182, miR 1910, miR 145, miR3690, miR1271 and miR 758 were selected. The selected miRNAs were the top upregulated and downregulated miRNAs obtained from the small RNA sequencing data performed in the human erythroid cells. The pre-miR sequences of the above-mentioned miRNAs were amplified from human genomic

DNA using the primers flanking 150-200 bp on either side of the pre-miRNA sequence, and the primers contained *Not*I and *Mlu*I restriction sites **(Table 3.3)**. pZIP-SFFV and pZIP-hCMV lentiviral vectors containing the promoters SFFV and hCMV respectively were digested with *Not*I and *Mlu*I restriction enzymes. They were then cloned into the corresponding cloning sites of the expression plasmids using T4 DNA ligase (Takara) at a ratio of 1:3, and the reaction mixes were incubated overnight at 16^0C. The ligated products were then transformed into XL1 competent cells and plated onto LB-Agar plates containing ampicillin. The plates were incubated at 37^0C for 16 hours, and 2 colonies from each plate were inoculated into LB-Broth and incubated in a bacterial shaker at 225 rpm for 16 hours. The plasmids were then extracted using a plasmid extraction kit, and they were then digested with *Not*I and *Mlu*I enzymes to confirm for the presence of inserts. After confirmation for the presence of inserts, the plasmids were subjected to Sanger sequencing.

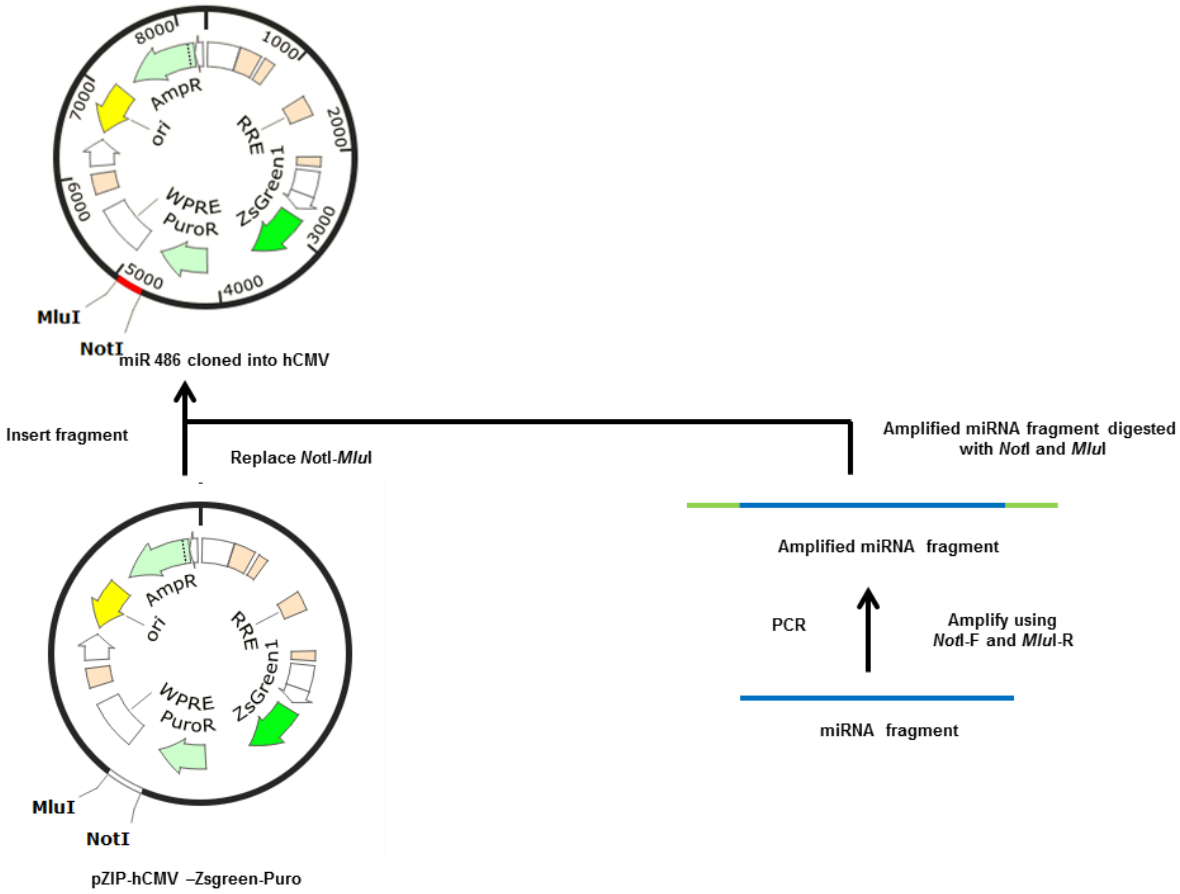


Figure 3.3: Construction of miRNA overexpression vector. The miRNA gene and the flanking sequences were amplified from genomic DNA and amplified using specific primers containing NotI and MluI restriction digestion enzyme sites. The digested products were cloned into pZIP lentiviral vector digested with the same enzymes.

3.8. GENERATION OF LENTIVIRAL PLASMIDS WITH DIFFERENT PROMOTERS

UBC, CBA, and MND promoter sequences were amplified using primers containing *Cla*I and *Age*I restriction sites **(Table 3.3)** from the plasmids pINDUCER20, AAV-CBA-RFP, and pTRIP-MND-GFP, respectively. They were then cloned into the corresponding cloning sites of the pZIP-hCMV-ZsGreen-Puro vector (obtained from

Transomic) in which the hCMV promoter was replaced with the above-mentioned promoters after digestion of the vector with *Cla*I and *Age*I. The amplified inserts and the vector were ligated in a ratio of 2:1 using T4 DNA ligase (Takara), and the reaction mixes were incubated overnight at 16^0C. The ligated products were then transformed into XL1 competent cells and plated onto Luria Agar ampicillin plates. The plates were incubated at 37^0C for 16 hours, and 3 colonies from each plate were inoculated into LB-Broth and incubated in a bacterial shaker at 225 rpm for 16 hours. The plasmids were then extracted using a plasmid miniprep extraction kit, and they were then digested with *Cla*I and *Age*I to confirm for the presence of inserts. After confirmation for the presence of inserts, the plasmids were subjected to Sanger sequencing.

3.9. TRANSFECTION

293T cells were revived and seeded in D10 (DMEM supplemented with 10% FBS (Gibco, Thermo Fisher), 1% of Pen/Strep (Gibco, Thermo Fisher) and 1X L-Glutamine (Hyclone, GE Life Sciences) medium in a 10 cm plate. On attaining 80% confluency, they were split into 2 X 12 well plates or 2 X 6 well plates. The transfection mix was prepared for each of the plasmids, and it contained 250 µl Opti-MEM (Gibco, Thermo Fisher) medium, 1 µg pMD2.G (envelope plasmid), 1 µg psPAX2 (packaging plasmid) ,1 µg lentiviral transfer plasmid and 2 µl of Trans IT-LT transfection reagent (Mirus). The contents of the tube were tapped well and incubated at room temperature for 20 minutes. The transfection mix was then added to the plate containing 293T cells. The medium was changed after 24 hours, and imaging was done after 48 hours to check for GFP.

3.10. LENTIVIRUS PREPARATION

Viral supernatant was collected at 48 hours, 60 hours, and 72 hours after the transfection of 293T cells, and it was concentrated either by ultracentrifugation or by using Lenti-X concentrator. For the concentration of viruses by the ultracentrifugation method, 24 ml of the viral supernatant was filtered through a 0.45 µm filter. The viral supernatant was then transferred to ultracentrifuge tubes by using a 10 ml syringe. The volume of the viral supernatant was then equalized in all the tubes by adding plain cold DMEM. The tubes were then sealed with sealant and kept for ultracentrifugation at 20000 rpm for 2 hours at 4^0C. After completion of centrifugation, the pellet was suspended in an appropriate volume of the desired medium to get a concentration of 100X. It was then incubated in ice for 1 hour and aliquoted into cryovials and stored at -80^0C. For the concentration of viruses using the Lenti-X concentrator, the viral supernatant was collected at the above mentioned time points and was centrifuged at 1000 rpm for 5 minutes at 4^0C. The supernatant was then transferred to fresh tubes, and the Lenti-X concentrator (Takara) was added in the ratio of 1:3. It was then mixed well and incubated at 4^0C for 4 hours. It was then centrifuged at 1500g for 45 minutes at 4^0C. The supernatant was then discarded, and the pellet was tapped gently to dislodge it. An appropriate volume of the desired cold medium was added to the pellet to get 50X concentration, and it was aliquoted in cryovials and stored at -80^0C.

3.11. TRANSDUCTION OF ADHERENT CELLS

HeLa cells were revived and seeded in D10 medium in a 10 cm plate. On attaining 80% confluency, they were split into 2 X 12 well plates, and polybrene (Sigma-Aldrich) containing D10 medium was added to the cells. The cells were then transduced with 100 and 500 µl of shRNA lentiviruses. The medium was changed 24 hours after transduction, and the cells were observed under the microscope after 48 hours to check for the expression of GFP. The cells were split at regular intervals into 1:4 ratios to avoid making them confluent. After 96 hours of transduction, the cells were treated with puromycin having a concentration of 1µg/ml to select the GFP positive cells. Puromycin selection was completed after five days, and the cells were collected for RNA in Tri Reagent (Sigma- Aldrich).

3.12. TRANSDUCTION OF SUSPENSION CELLS

Approximately 6.4 X 10^7 $CD34^+$ cells were revived and seeded in the appropriate medium. The cells were then transferred to 6 X 6 well plates and transduced with viruses. 74 ml of medium was prepared into which 5.6 µl of polybrene and 300 µl of 50X concentrated virus were added. It was then added to the cells in 6 well plates. The plates were then centrifuged at 2250 rpm for 1 and a half hours at 37^0 C. Medium was then changed after 24 hours.

3.13. DNA SEQUENCING

100-300ng of plasmid DNA was added to the reaction master mix consisting of sequencing buffer, Big Dye Terminator v 3.1 reaction mix (Applied Biosystems),

sequencing primer, betaine and water. The sequencing reaction was then carried out in a thermocycler with the reaction conditions as mentioned below

96^0 C- 6 minutes

96^0 C- 20 seconds
65^0 C-30 seconds
60^0 C-4 minutes
} 25 cycles

15^0 C- Final hold

10 µl of PCR product was mixed with 10 µl of post clean up beads (Magbio Genomics) and 40 µl of 85% ethanol in a 1.5 ml tube. The tubes were then incubated in a magnetic stand for 5 minutes. The clear solution was discarded without disturbing the pellet. 200 µl of 85% ethanol was added to the tubes and incubated for 2 minutes in the magnetic stand. The ethanol was then discarded, and the above step was repeated. The ethanol was removed completely, and the tubes were left to dry for 10 minutes. The tubes were then removed from the magnetic field, and 40 µl of injection solution was added to the tubes. It was mixed well and incubated in a magnetic field for 5 minutes. The clear solution was then loaded on to the sequencing plate. Sequencing was performed on an ABI capillary sequencer.

3.14. RNA EXTRACTION

The cell pellets were suspended in Tri-Reagent (Sigma-Aldrich) and stored at -80^{0}C before RNA extraction. The samples in the Tri-Reagent were thawed on ice. They were then vortexed gently for approximately 3 seconds and given a momentary spin. 100 µl of chloroform was then added to the samples and mixed well. It was then incubated in ice for 10 minutes with occasional mixing every 2 minutes. The samples were then centrifuged at 13000 rpm for 30 minutes at 4^{0}C. The aqueous phase was then transferred to pre-chilled fresh tubes. An equal volume of isopropanol was added to the aqueous phase and mixed gently by inverting the tubes for 1 minute. The tubes were then incubated in ice for 10 minutes. It was then centrifuged at 13000 rpm for 30 minutes at 4^{0}C. The position of the pellet was checked, and the supernatant was discarded without disturbing the pellet. The RNA pellet was washed with freshly prepared 75% ethanol by centrifuging at 13000 rpm for 15 minutes at 4^{0}C. The RNA pellet is then checked, and the supernatant was discarded. The tubes were then kept aside without disturbing them for 10 seconds, which allowed the residual 75% ethanol attached to the sidewalls of the tubes to settle down. The residual 75% ethanol was then removed with a 200 µl tip without disturbing the pellet, and the tubes were left for drying in a thermal mixer with caps open. The appropriate volume of RNase free water (30-50 µl) was then added to the pellet and incubated at 55^{0}C for 30 minutes at 350 rpm on a thermal mixer. RNA was then quantitated using Nanodrop with RNase free water as blank.

3.15. cDNA SYNTHESIS

500 ng of RNA was used for cDNA synthesis using High Capacity cDNA Reverse Transcription Kit (Life Technologies). A master mix comprising of the buffer, dNTPs, random primers, reverse transcriptase enzyme, RNase inhibitor, and water is made, which is added to the RNA sample. cDNA synthesis is then carried out in a thermocycler with the following reaction conditions, as mentioned below.

25^0C-10 minutes

37^0C-120 minutes

85^0C-5 minutes

The cDNA products are then stored at -20^0C.

3.16. REAL-TIME PCR

RNA was extracted from cell lines, $CD34^+$ cells, and differentiated erythroid cells using Tri-reagent (Sigma-Aldrich). 1 µg of total RNA was used for reverse transcription reaction using High Capacity cDNA Reverse Transcription Kit (Life Technologies) using the protocol mentioned in section 15. Quantitative RT-PCR was set up with SYBR Premix Ex TaqII using specific primers **(Table 3.3)** and analyzed with QuantStudio 12K Flex (Life Technologies) real-time PCR system.

3.17. ISOLATION OF PERIPHERAL BLOOD $CD34^+$ HSPCS

$CD34^+$ positive haematopoietic stem and progenitor cells (HSPCs) were isolated from healthy donors after GCSF administration. Informed consent was obtained from

the donors as approved by the institutional review board. Peripheral blood mononuclear cells (PBMNCs) were isolated by density gradient centrifugation with Ficoll-Paque PLUS (GE Health Sciences). The cells were washed with PBS, and the contaminating red blood cells were removed by washing with RBC lysis solution (130.8mM NH_4Cl and 0.9mM NH_4HCO_3). $CD34^+$ HSPCs were enriched from PBMNCs through direct immunomagnetic selection using the Easy Sep human $CD34^+$ selection kit (Stem Cell Technologies Inc) that yielded ≥95% $CD34^+$ cells.

3.18. EX-VIVO ERYTHROPOIESIS

3.18.1 PROTOCOL 1

For the *ex-vivo* culture of erythroid cells, a modified 2 phase erythroid culture system was used. In phase I of the culture (day 0 to day 5), approximately 1 X 10^7 purified $CD34^+$ cells were cultured in serum-free StemPro-34 complete medium supplemented with antibiotics (100U/ml penicillin, 1mg/ml streptomycin), 2mM L-glutamine, stem cell factor (SCF) (100ng/ml), Fms related tyrosine kinase 3 -Ligand (FLT3-L) (100ng/ml), Interleukin-6 (IL-6) (20ng/ml) and Interleukin-3 (IL-3) (20ng/ml). After Phase I, the cells were transferred to the Phase II of the culture (day 5 to day 14) in StemPro-34 complete medium containing $FeSO_4$ (900ng/ml) (Sigma-Aldrich), $Fe(NO_3)_3$ (90ng/ml) (Sigma-Aldrich), β-mercaptoethanol (10µM), 2 U/mL erythropoietin (Epo), 20 ng/mL SCF and 2ng/ml IL-3. The cell count was maintained at less than 0.5 X 10^6 /ml, and the medium was changed every alternate day. All the cell culture reagents were purchased from Life Technologies, and all the cytokines were from Immunotools.

3.18.2 PROTOCOL 2

In this protocol, a modified 2 phase erythroid culture system was used in which the cells could be maintained for a prolonged period in the differentiation medium without proceeding to differentiation completely. In phase I of the culture (day 0 to day 9), approximately 9.6 X 10^6 purified $CD34^+$ cells were cultured in serum-free Stem Pro HSC Expansion Medium (Prototype) supplemented with antibiotics (100 U/ml penicillin,1 mg/ml streptomycin) 2mM L-glutamine and cytokines such as stem cell factor (SCF) (100 ng/ml), Fms related tyrosine kinase 3-ligand (Flt3-L) (100ng/ml),Interleukin-6 (IL-6) (20ng/ml), Interleukin-3 (IL-3) (50ng/ml) and Thrombopoietin (TPO) (100 ng/ml).After culturing the cells in Phase I, the cells were transferred to Phase II of the culture (day 10-day 29) in Stem Span SFEM II medium containing stem cell factor (SCF) (50ng/ml), Insulin growth factor (IGF1) (40 ng/ml), Interleukin-3 (IL-3) (10 ng/ml) and Erythropoietin (Epo) (3U/ml). The cell density was maintained at 0.5 X 10^6 cells/ml, and the medium was changed every alternate day. All the cell culture reagents were purchased from Life Technologies, and all the cytokines were from Immunotools.

3.19. FLOW CYTOMETRY ANALYSIS OF SURFACE MARKERS

Approximately, 1 X 10^5 of cultured erythroid cells were collected, washed with phosphate-buffered saline (PBS), and incubated in PBS containing antibodies for 20 minutes at room temperature. The cells were rewashed with PBS and resuspended in 500 µl PBS, and approximately 10,000 cells were analyzed on the BD FACS

CELESTA flow cytometer using Cell Quest software (BD Biosciences). The antibodies, CD34-APC, CD71-FITC, CD 71-APC, CD235a-PE, and CD235a-BV421, were obtained from BD Biosciences-Pharmingen.

3.20. MORPHOLOGY ANALYSIS OF ERYTHROID CELLS

Approximately, 1 X 10^5 cells were collected on different days of differentiation and centrifuged onto glass slides for 5 minutes at 500 rpm using Cytospin 4 centrifuge (Thermo Scientific). The cells were fixed in methanol for 3 minutes at room temperature and stained with modified Giemsa stain (Sigma-Aldrich) as per the manufacturer's recommendations. The stained slides were examined under a bright-field microscope (Leica DM 1000, Leica) at 100X oil immersion objective, and the images were captured using a Leica DFC 290 camera.

3.21. HAEMOGLOBIN ANALYSIS

Approximately, 5 X 10^6 cells were collected in the late stages of erythroid differentiation (between days 12 to 14) and washed with PBS and stored at -80°C. The frozen cells were thawed on ice, lysed with haemolysis solution, and centrifuged at 13,000rpm for 1 minute at room temperature. The haemolysates were analyzed for HbF, HbA and HbA_2 levels using cation exchange chromatography (Bio-Rad, VARIANT).

3.22. ANALYSIS OF ERYTHROID GENE EXPRESSION

5 X 10^6 cells were collected on different days of *ex-vivo* erythropoiesis, and RNA was isolated using Tri-Reagent (Sigma-Aldrich). 1 μg of RNA was used for cDNA

synthesis using High Capacity cDNA Reverse Transcription Kit (Life Technologies). The expression of erythroid-specific genes *HBB, HBG, HBA, EPOR* and *KLF1* were analyzed using SYBR Premix Ex Taq II (Takara).

3.23. SMALL RNA SEQUENCING

Approximately 3 X 10^6 cells were collected from biological triplicates on days 3 and 5 in phase I and days 9, 11 and 14 in phase II, and RNA was isolated using Tri-Reagent (Sigma-Aldrich). The purity and integrity of the total RNA were checked using Bio-Analyzer (Agilent Technologies). Small RNA library was created using TruSeq RNA sample preparation kit v2 ((Illumina) according to the manufacturer's protocols. Libraries were sequenced using TruSeq SBS Kit v3-HS on HiSeq1000 platform (Illumina) according to the manufacturer's recommendations. For bioinformatics analysis of the small RNA expression, a web-tool Oasis 2.0 (https://oasis.dzne.de/) was used. Using this tool, trimming of adapter sequences, alignment to the known human small RNA transcript sequences and to the human genome sequence and differential expression using DE-Seq2 algorithm were performed

3.24. ANALYSIS OF TRANSCRIPTION FACTOR OCCUPANCY AT THE miRNA PROMOTERS

ChIP-Seq data for *GATA1, TAL1 and KLF1* published by Huang et al. was uploaded using Cistrome Data Browser (http://dbtoolkit.cistrome.org) and UCSC genome browser, to identify their genome-wide occupancy in human erythroid cells.

3.25. ANALYSIS OF ERYTHROID TRANSCRIPTOME DATA

The expression of the predicted target mRNAs in HSPCs and erythroid cells was analysed using previously published erythroid transcriptome data through the online portal to access the data.
(http://guanlab.ccmb.med.umich.edu/data/Shi_L_Developmental/index.php).

3.26. *ANALYSIS OF MIRNA TARGETS*

Predicted target genes for upregulated miRNAs were obtained from the online database, miRWalk2), which provides the targets from 12 different miRNA prediction programs (miRwalk, miRDB, PITA, MicroT4, miRMap, RNA22, miRanda, miRNAMap, RNAhybrid, miRBridge, PICTAR2 and Targetscan) with p-value <0.05. Targets that were present in at least 8 databases were then selected. Validated targets for upregulated miRNAs were obtained from recent miRTarBase2016 (http://mirtarbase.mbc.nctu.edu.tw/). The predicted and validated targets were then pooled and then screened for their expression in erythroid cells (Shi et al. 2014 and Li et al. 2014) and those which showed significant downregulation (≥ 4 fold) in human erythropoiesis were considered to generate the pathways.

3.27. PATHWAY ANALYSIS OF miRNA TARGETS

The gene ontology (GO) analysis of the predicted and validated targets was performed using Enrichr (https://maayanlab.cloud/Enrichr/). The significant pathways obtained from Wikipathways 2019 and Bioplanet pathways 2019 were considered.

3.28. VALIDATION OF MIRNA EXPRESSION BY QUANTITATIVE PCR

Total RNA was converted to cDNA using the miRNA first-strand cDNA synthesis kit (OriGene) as per the manufacturer's protocol. Briefly, 1 µg of total RNA was polyadenylated by polyA polymerase and then reverse transcribed to cDNA using tagged oligo-dT primers. The cDNAs were used as templates to quantitate miRNAs using forward primers **(Table 3.3)** that bind to the mature miRNA sequence of the known miRNAs from miRBase or predicted miRNAs from miRdeep and a common reverse primer that binds to the tag sequence of the oligo-dT primer used for cDNA synthesis. Human U6 snRNA and let7 primers were used for normalization.

3.29. DNA EXTRACTION

DNA was extracted from adherent and suspension cells as per the protocol of the Gentra DNA Extraction Kit (Qiagen). The adherent and suspension cells were pelleted down in 15 ml tubes by centrifuging at 1000 rpm for 5 minutes. The pellet was then suspended in 1 ml of cell lysis buffer to make a homogenous solution. The tubes were then incubated at 37^0C in a water bath for 2 to 3 hours.333 µl of protein precipitation solution was then added to the tubes, and it was vortexed for 3 secs and then centrifuged at 16000g for 1 minute. The supernatant was then transferred to a fresh 15 ml tube, and 1 ml of isopropanol was added to the tube, and it was mixed gently. Visible DNA thread was then transferred to 1.5 ml microfuge tubes containing 1 ml of 75% ethanol. The tubes were then centrifuged at 16000g for 1 minute. The ethanol was then removed, and the tubes were kept for drying in a

thermomixer at 55^0C. DNA hydration solution was then added to the DNA pellet, and it was incubated at 55^0C in a thermomixer at 800 rpm for 30 minutes. This incubation was followed by a second incubation at 37^0C at 800 rpm for 30 minutes. The concentration of DNA was then measured in different layers to ensure the homogeneity of the dissolved DNA.

3.30. CONSTRUCTION OF shRNA LIBRARY

A pool of oligos containing shRNA sequences for 100 genes that were identified by meta-analysis was synthesized from a commercial source (Custom Array). shRNA sequences were designed using the Sherwood (Knott *et al*. 2014) and Splash (http://splashrna.mskcc.org) algorithms. 113 ng of pooled shRNA oligos were amplified in 4 tubes for 12 cycles with the primers used for cloning of shRNAs in lentiviral vectors (**Section 3.6**) using KOD polymerase (Sigma-Aldrich). The amplified PCR products were then pooled, and 10 µl was loaded on the gel to check for the correct size of the amplified product. Secondary PCR was then carried out for 12 cycles in 10 tubes using 2 µl of the primary PCR product as the template. 10 µl of the secondary PCR product was then loaded on the gel to confirm for the correct size of the amplified product. The secondary PCR products were then subjected to PCR purification and stored in -80^0C. The pZIP-MND-ZsGreen-Puro UltramiR vector was then digested with *Hpa*I. The purified shRNA fragment (66 ng) and the vector backbone (100ng) were ligated in a 20 µl ligation reaction using NEBuilder Hi-Fi DNA Assembly Master Mix (New England Biolabs). A 2 µl ligation reaction was used to transform 50 µl of Mach 1 competent cells (Thermo Fisher) in multiple tubes. 100 µl of the transformation reaction was plated on one 10 cm LB-Amp (100

μg/ml) plate to estimate the total number of colonies, and the rest of the transformation reaction was plated on two 15 cm LB-Agar plates containing ampicillin (100 μg/ml) plates and grown overnight at 37°C. At least 100X coverage was needed to maintain the representation of the library (i.e., colony number = 100 X the number of shRNAs in the library). The next day, lawns formed on the two 15 cm plates were scraped and cultured in 200 ml LB-Broth medium containing ampicillin (100 μg/ml) and grown at 37°C for 3 hours. The bacteria were collected, and the pool of plasmids containing the shRNAs was extracted by the Midiprep kit (Macherey-Nagel).

3.31. RNAi SCREENING FOR SIGNALLING PATHWAY GENES

The RNAi experiments were carried out following the conventional RNAi screening workflow, as shown in **Figure 3.4**. Lentiviruses were prepared using the Cellecta shRNA library module 2. 6.4 X 10^7 $CD34^+$ cells were transduced with 300 μl of 100X concentrated virus according to the protocol mentioned in section 12. The transduction efficiency was kept low (approximately 30%) to achieve a single copy integration of shRNAs. The cells were maintained in the Phase I HSPC expansion medium for 5 days, and the transduction efficiency was estimated using FACS analysis. After 5 days, the cells were then grown in an erythroid differentiation medium for 9 days. Cells were collected on day 5 before differentiation (Sample 1) and then on day 9 after differentiation (Sample 2). DNA was extracted from the samples using the Gentra DNA extraction kit (Qiagen). The shRNA specific barcodes were amplified from the genomic DNA by two rounds of PCRs using the primers described by the manufacturer of the shRNA library **(Table 3.3)**. The 1st

round of amplified products from each sample were pooled, and 2 µl was used for setting up the 2nd round PCR in 5 X 50 µl reactions. The second-round PCR was performed for 14 cycles. Different indexing primers were used for the amplification of the 2nd round PCR products from the two DNA samples. Next-generation sequencing (NGS) was carried out on HiSeq 2500, and the GexSeq primer **(Table 3.3)** was used for sequencing purposes. The barcode deconvolution and enumeration were carried out by using the software available on the Cellecta Decipher Project website (www.decipherproject.net/software). A T-test was performed to identify differentially expressed shRNAs.

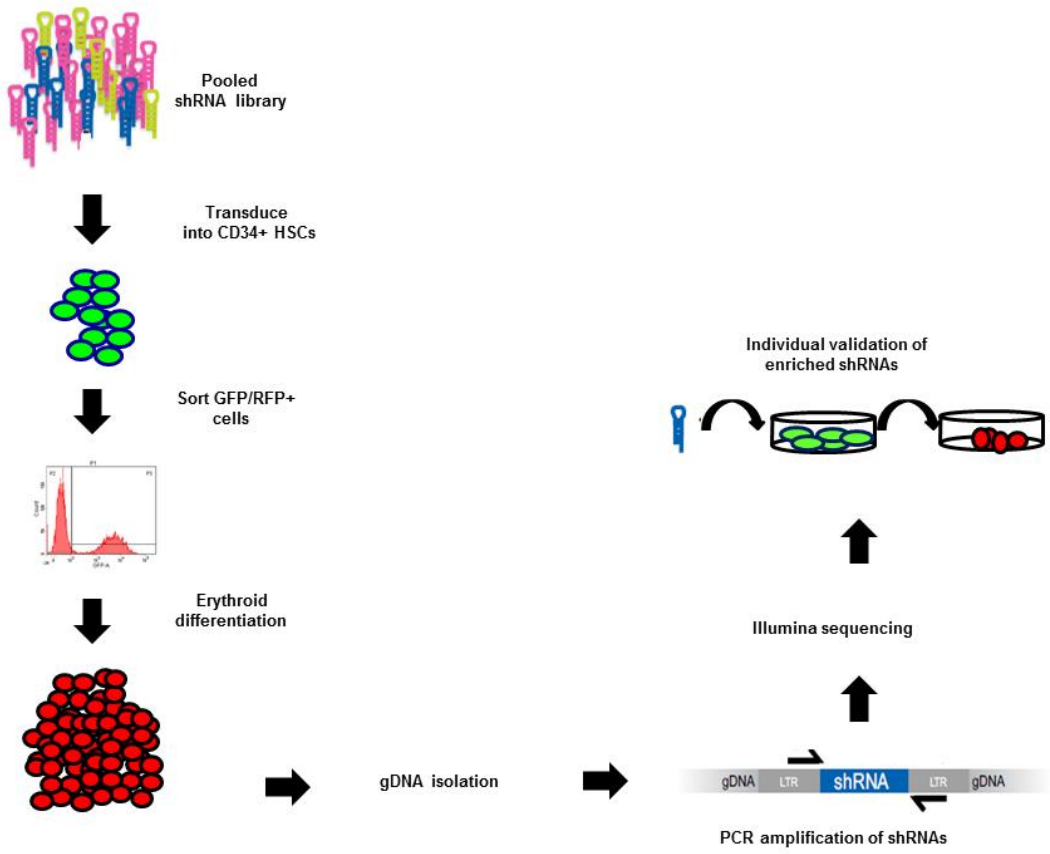


Figure 3.4 RNAi Screening workflow

3.32. RNAi SCREENING FOR HSC MAINTENANCE AND DIFFERENTIATION GENES

Lentiviruses were prepared using the lentiviral shRNA plasmid pool generated from section 3. 1. 6 X 10^6 $CD34^+$ cells were transduced with 60 µl of 100 X concentrated virus according to the protocol mentioned in section 3.12. The transduction efficiency was kept low (approximately 30%) to achieve a single copy integration of shRNAs. The cells were maintained in the phase I expansion medium for 9 days, and the transduction efficiency was estimated using FACS analysis. After 9 days, the cells were then grown in differentiation medium for 19 days. Cells were collected on day 8 before differentiation and days 4, 9, 13, and 19 after differentiation in duplicates. DNA was extracted from the samples using the Gentra DNA extraction kit (Qiagen). The shRNA specific regions were amplified from the genomic DNA in the first round of PCR, whereas the second round of PCR was done with various indexing primers to make them suitable for sequencing. In the first round of PCR for each sample, 850 ng of DNA was used in each tube, and the reaction was set up in 10 tubes of 50 µl reactions. The amplified products from each sample were pooled, and 10 µl was loaded on the gel to check for the size of the product. They were then PCR purified using a PCR purification kit. 500 ng of the primary PCR product was taken, and secondary PCR was carried out in 10 tubes for each sample. Amplified products from the second round of PCR were pooled, and 10 µl was loaded on the gel to check for the size of the product. 6 tubes were then gel purified, and 4 tubes were subjected to PCR purification. The first round of PCR was done for 25 cycles, whereas the

second round of PCR was performed for 12 cycles after the initial optimization of the number of cycles to obtain the equal representation of the shRNAs from all the DNA samples. To facilitate multiplex sequencing, indexing primers (**Table 3.3**) were used for amplification of each sample in the 2nd round PCR, and the high throughput sequencing was carried out on HiSeq 2500 using shRNA loop sequencing primer (**Table 3.3**).

3.33. CULTURE AND DIFFERENTIATION OF HUDEP CELLS:

Immortalized erythroid progenitor cells, HUDEP(Kurita *et al.*, 2013), were cultured in the medium containing StemSpan SFEM-II (Stem Cell Technologies) supplemented with 3 U/ml erythropoietin (Epo) (Zydac), 50ng/ml stem cell factor (SCF) (ImmunoTools GmbH), 20 ng/ml interleukin-3 (IL-3) (ImmunoTools GmbH)), 40 ng/ml insulin growth factor-1 (IGF-1) (ImmunoTools GmbH), 1μM dexamethasone (Sigma-Aldrich), 1μg/ml doxycycline (Sigma-Aldrich), 2mM L-glutamine (Hyclone), 100 U/mL penicillin, and 100μg/mL streptomycin (Thermo Fisher Scientific). To induce erythroid differentiation, a previously described protocol was used with minor modifications(Hawksworth *et al.*, 2018). Briefly, in phase I of the differentiation, iEPCs (immortalized erythroid progenitor cells) were seeded at a density of 2-3 X 10^5 cells per ml in erythroid differentiation medium I (EDM I) for 2 days. EDM I consisted of Iscove's Modified Dulbecco's Medium (IMDM) containing Glutamax (ThermoFisher Scientific), 3% human AB serum (MP Biomedicals), 2% Foetal Bovine Serum (Gibco), 200 mg/ml holotransferrin (Sigma-Aldrich), 3 U/ml heparin (Sigma-Aldrich), 10 μg/ml insulin (Sigma-Aldrich), 3 U/ml Epo (Zydus), 10 ng/ml SCF (Immunotools GmbH), 1ng/ml IL-3 (Immunotools GmbH), 1 mg/ml doxycycline, 100 U/mL penicillin and 100 μg/mL streptomycin (Thermo Fisher Scientific). On day-2, the cells were reseeded at a density of 3-4X10^5 cells/ml in EDM-I and cultured for 2 days. On day-4, cells were counted and were seeded at a density of 5-6X10^5 cells/ml in EDM-II (DDM I without

doxycycline) for 2 days. On day-6, the cells were counted and seeded at a density of $1X10^6$ cells/ml in EDM-III (EDM II with 500µg/ml holotransferrin) for 2 days. On day-8, the cells were reseeded at $1X10^6$ cells/ml in EDM-IV (EDM III without SCF & IL3) for another 4 days till the end of differentiation.

3.34. GENE EDITING OF miRNAS IN HUDEP CELLS

3.34.1. DESIGN OF gRNAS: The genomic location of each primary miRNA stem-loop was downloaded from miRBase v21 (Kozomara and Griffiths-Jones 2014). The stem-loop sequence plus 20 bp on either side was then obtained from the UCSC genome browser. The sgRNA sequences for miR-182, miR-183 and miR-96, miR-144, miR-451 and miR-4732 were designed by MIT CRISPR design programme (http://crispr.mit.edu/). For miRNA knock out experiments using single gRNAs (sgRNAs), the gRNAs were designed adjacent to the Drosha and Dicer sites. For dual gRNA knock out experiments, pairs of gRNAs flanking the miRNA genomic sequences were designed in such a way that at least one gRNA cleaves Drosha or Dicer sites.

3.34.2. CLONING OF gRNAS:

For the generation of lentiviral vectors containing paired gRNAs, MULE (Albers *et al.*, 2015) plasmids were used. One gRNAs of the pair was cloned in D9 entry plasmid containing attL1 and attL4 sites, and the other gRNA was cloned in D12 entry plasmid containing attR4 and attR3sites, and the final vector was generated by LR recombination of the D9 and D12 containing gRNAs with H8 destination vector,

which contains attR1 and attR2 sites. The plasmids, D9, D12, and H11, were gifts from Ian Few (Addgene # 1000000060).

The complementary oligos, Forward: 5' ACCG NNNNNNNNNNNNNNNNNNNN 3' and Reverse: 5' AAAC NNNNNNNNNNNNNNNNNNNN 3' were synthesized commercially, and 1 µl Fwd oligo (100uM) and 1 µl Rev oligo (100uM) were mixed with 1 µl 10x T4 ligation buffer (NEB), 0.5 µl T4 PNK (NEB) and 6.5 µl H2O to make up a final volume of 10 µl. The reaction mix was incubated at 37°C for 30 minutes. Subsequently, the oligos were annealed by heating at 95°C for 5 minutes and then ramping down to 25°C at 5°C/min. 1 µg of D9 and D12 plasmids were digested with BfuAI (NEB) enzyme and purified by gel extraction. The ligation was carried out by mixing 50 ng of digested vector, 1 µl 10x T4 ligation buffer, 1 µl T4 ligase, and 1 µl annealed oligos in a 10 µl reaction volume at 16°C overnight. The transformation of the ligated products was carried out in Stbl3 competent cells. Plasmids extracted were sequenced to confirm the inserts.

3.34.3. GENERATION OF CAS9 EXPRESSING HUDEP CELLS:

HUDEP cells were transduced with pLentiCas9-T2A-BFP (Pulido-Quetglas *et al.*, 2017) lentiviruses using the lentiviral transduction protocol described above. The transduced cells were selected with 1ug/ml of blasticidin. The Cas9-HUDEP cells were transduced with the lentiviruses generated with the vectors containing dual gRNAs targeting the miRNAs. After 5 days, GFP^+ cells were flow-sorted, and they were cultured for a total of 10 days from the day of transduction.

3.34.4 T7 ENDONUCLEASE (T7E) ASSAY

Genomic DNA was extracted from transduced cells using the Gentra DNA extraction kit (Qiagen). DNA fragments spanning target sites of CRISPR/cas9 were amplified by Emerald AmpGT PCR Mastermix (Takara Clonetech), and PCR products were purified by PCR Purification Kit (Machery Nagel). T7E assay is a known method to detect mismatched DNA (Guschin *et al.,* 2010). In brief, total 200 ng of purified PCR products were denatured and reannealed in 1× NEBuffer 2 (NEB) in 20 µl volume in a thermocycler with following steps (95 °C, 5 min; 95-85 °C at − 2 °C/s; 85-25 °C at − 0.1 °C/s; hold at 4 °C) (Niu *et al.,* 2014). Then 10 U of T7EN1 enzymes (NEB, M0302L) were added to hybridised PCR products, and the reaction mix was incubated at 37 °C for 15 minutes. The digested PCR products were run on a 2% agarose gel. The presence of in-dels was detected by the presence of the additional two bands in the gels.

Table 3.1: Cell culture media used in this study

Application	Media	Components	Concentration	Volume
Culture of HEK293T cells and HeLa cells	D10 medium	DMEM FBS 200mM Glutamine 100X Penicillin Streptomycin	- 10% 2mM 1X	100ml 10ml 1ml 1ml
Culture of CD 34^+ cells	Stem Pro-34	Stem Pro-34 200mM Glutamine 100X Penicillin Streptomycin 100 µg/ml SCF 50 µg/ml Flt3-L 10 µg/ml IL-3 20 µg/ml IL-6 Nutrient Supplement 40X	- 2mM 1X 100ng/ml 100ng/ml 20ng/ml 20ng/ml 1X	10 ml 100 µl 100 µl 10 µl 20 µl 20 µl 10 µl 260 µl
Culture of CD 34^+ cells	Stem Pro HSC medium (prototype)	Stem Pro medium 200mM Glutamine 100X Penicillin Streptomycin 100 ng/ µl SCF 50 ng/ µl Flt3-L 10 ng/ µl IL-3 20 ng/ µl IL-6 100 ng/ µl TPO Nutrient Supplement 50X	- 2mM 1X 100ng/ml 100ng/ml 50ng/ml 20ng/ml 100ng/ml 1X	10ml 100µl 100µl 10µl 20µl 50µl 10µl 10µl 200µl
Erythroid Differentiation	Stem Pro-34	Stem Pro-34 200mM Glutamine 100X Penicillin Streptomycin Ferrous Sulphate 18 mg/ml Ferric Nitrate 1.8g/ml β-mercaptoethanol 55mM 50 ng/ µl SCF 10 ng/ µl IL-3 2U/ µl Epo Nutrient Supplement 40X	- 2mM 1X 900 ng/ml 90ng/ml 10µM 20ng/ml 2ng/ml 2U/ml 1X	10ml 100µl 100µl 50µl 50µl 1.8µl 4µl 2µl 10µl 260µl
Differentiation of CD 34^+ cells to erythroid cells	SFEM-II	Stem Span SFEM-II 200mM Glutamine 100X Penicillin Streptomycin 100 ng/ µl SCF 10 ng/ µl IL-3 2U/ µl EPO 100 ng/ µl IgF Dexamethasone	- 2mM 1X 50 ng/ml 10 ng/ml 3U/ml 40 ng/ml 1µM	10 ml 100µl 100µl 5µl 10µl 15µl 4µl 1.2µl
Differentiation of HUDEP cells to erythroid cells	IMDM Glutamax	IMDM Glutamax 100X Penicillin Streptomycin AB serum Insulin Sodium Heparin Holotransferrin Mifepristone Doxycycline	- 1X 5% 10µg/ml 2U/ml 500µg/ml 1µM 1 µg/ml	10ml 100µl 490µl 9.17µl 10µl 490µl 10µl 10µl

Table 3.2: Plasmids used in this study

Plasmids	Source	Reference
pGIPZ	Dharmacon	
pZIP-hEF1α	Transomics TLN005,Promoter Selection Kit	
pZIP-mEF1α		
pZIP-hCMV		
pZIP-mCMV		
pZIP-SFFV		
pZIP-UBC	Constructed in house	
pZIP-CBA		
pZIP-MND		
pINDUCER20	Gift from Stephen Elledge,HHMI	Meerbrey,KL et al.,2011
AAV-CBA-RFP	Gift from Dr.Sanjay Kumar,CSCR	-
pTRIP-MND-GFP	Gift from Dr.Poonkuzhali B,CMC	-
pZIP-UltramiR	Constructed in house	Knott et al.,2014
DECIPHER Pooled shRNA Library-Human Module 1-signalling pathways	Addgene 28285	

Table 3.3: Oligos used in this study

Oligos used for pZIP-ultramiR-shRNA vector construction by Gibson assembly	
Primer Name	**Sequence**
Ultramir Scaffold	GTTAACTGAATACCTTGCTATCTCTTTGATACAATTTTTACAAAGCTGAATTAAAATGGTATAAATTAAATCACTTTACGCGT
Ultramir-F	CTTCTTCAGGTTAACTGAATACCTTGCTATCTC
Ultramir-R	GTCCAGACGCGTAAAGTGATTTTAATTTATAC
Ultramir-Hpa-I-F	CTGGGATTACTTCTTCAGGTTAACCCAACAGAAGGCTAAAGAAGGTATATTGCTGTTGACAGTGAGCG
Ultramir-Hpa-I-R	AGAGATAGCAAGGTATTCAGTTTTAGTAAACAAGATAATTGCTCCTAAAGTAGCCCCTTGAAGTCCGAGGCAGTAGGC
Primers used for sequencing shRNAs in the pZIP-ultramiR-shRNA vector	
pZIP-Seq-For	GGTGCCCGAAGGACCG
Primers for cloning different promoters into pZIP shRNA lentiviral vector	
UBC-FP	CAATCTATCGATGGCCTCCGCGCCGGGTTTTG
UBC-RP	CACGACCGGTGTCTAACAAAAAAGCCAAAAACG
MND FP	CGATATCGATCGCGTCATCGATCGATTAGTC
MND RP	CGATACCGGTCAGATCGCGCCGAGGGG
CLAI F	CTACGGATCGATACGCGTGGTACCTCTGGTCG
AGEI R	ATTAATACCGGTCTCCCGCCCGCCGCG
Real time PCR primers for analysing the expression of haemoglobin genes	
HBB F	CAGGCTGCTGGTGGTCTAC
HBB R	GCCATGAGCCTTCACCTTAG
HBG F	GCAAAGAATTCACCCCTGAG
HBG R	AGCCTATCCTTGAAAGCTCTG
HBA F	TCAAGCTCCTAAGCCACTGC
HBA R	AGAAGCCAGGAACTTGTCCA

Primers used for construction of miRNA overexpression vectors	
miR 144 FP	CTAGGCGGCCGCAATGGATCTGCAGCATCTCTCCTGTCCTC
miR 144 RP	CTAGACGCGTCTTTTCAAGCCATGCTTCCT
miR 451a FP	CTAGGCGGCCGCAATGGATCTGCCTTGTTTGAGCTGGAGTC
miR 451a RP	CTAGACGCGTGAGCCTGACAAGGAGGACAG
miR 4732 FP	CTAGGCGGCCGCAATGGATCTGCAGCATCTCTCCTGTCCTC
miR 4732 RP	CTAGACGCGTATTACGCCATCTCTGGCTTG
miR 7155 FP	CTAGGCGGCCGCAATGGATCTCCGAGCTGGAGATGCTGT
miR 7155 RP	CTAGACGCGTCCTCAGGCCTCAAGCAGA
miR 375 FP	CTAGGCGGCCGCAATGGATCTACCCGTACGGTTGAGATGG
miR 375 RP	CTAGACGCGTAGACCAGGACCAGGAGATCA
miR 182 FP	CTAGGCGGCCGCAATGGATCTGAGAACAGCAGGTCCAGCAT
miR 182 RP	CTAGACGCGTGCCTGGACCATCCTAACTGT
miR 1910 FP	CTAGGCGGCCGCAATGGATCTGATGTGGGGTGTGTGGAACT
miR 1910 RP	CTAGACGCGTCCCAGCACCTCAGACTCG
miR 145 FP	CTAGGCGGCCGCAATGGATCTGGCTGGATGCAGAAGAGAAC
miR 145 RP	CTAGACGCGTGCCTTCTTCTTGAACCCTCA
miR 3690 FP	CTAGGCGGCCGCAATGGATCTGGTGTCCCTAGTCCACATCC
miR 3690 RP	CTAGACGCGTTGGAGGTGGTGCAGTCTT
miR 1271 FP	CTAGGCGGCCGCAATGGATCTGCCATGTCCTTCCCAGTTAG
miR 1271 RP	CTAGACGCGTTCACCTTTCTTTTCCTCTCTGC
miR 758 FP	CTAGGCGGCCGCAATGGATCTCAGCCTTCTTCTCAGCACACT
miR 758 RP	CTAGACGCGTCAGCCACATTTGGCATACAT
Real time PCR primers for analysing the expression of miRNA genes	
miR-145	GTCCAGTTTTCCCAGGA
miR-143	GCAGTGCTGCATCTCTG
miR-20b	TGCTCATAGTGCAGGTAG
miR-18b	AGGTGCATCTAGTGCAG
miR-222	CTCAGTAGCCAGTGTAG
miR-221	ACCTGGCATACAATGTAG
miR-144	GGATATCATCATATACTGTA
miR-451	ACCGTTACCATTACTGAG
miR-4732	TGTAGAGCAGGGAGCAGGAAG
miR-182	GGCAATGGTAGAACTCAC
miR-183	ATGGCACTGGTAGAATTC
miR-96	TTTGGCACTAGCACATTTT
miR-rev	GAACATGTCTGCGTATCTC
Real Time PCR primers for calculating knockdown efficiency	
PRDM14-F	GAGCCTTCAGGTCACAGAGC
PRDM14-R	GTCCACACAGGGGGTGTACT
PRDM1-F	AACGGCCTTTCAAATGTCAG
PRDM1-R	TCTTGAGATTGCTGGTGCTG
MECOM-F	ACAGCAGTGTGAAGCCCTTT
MECOM-R	ATTTGGGTTCTGCAATCAGC
SETD7-F	GGGAGGCCTCACTTTGAACT
SETD7-R	GATCTGGAAGAAGAGCATTGG
PRDM12-F	CTCGGCATCTTCTCCAAGAC
PRDM12-R	CGTGCCATCCTCATTGAAC
KDM1B-F	CCCGGGTACTCGGTGATAAT
KDM1B-R	GGGAGGAACGTGACCAAAA
KDM6B-F	TATCGCCTCTGAGGTGGAAG
KDM6B-R	GGGGAAAGGTAGGACTCTCG
KDM2B-F	CATGCAGCAGAAAAGCAAAA
KDM2B-R	TCCGACAAGTCCTCGTTCTC
DNMT3A-F	AGACCCCTGGAACTGCTACA
DNMT3A-R	GGGACAGGTGGGTAAACCTT
TET1-F	GGTGAATGCACTTACTGCAAGA
TET1-R	TTTGGGCTTCTTTTCCCTCT
JMJD1C-F	AGGCTCAACCTTACGGGATT

Real Time PCR primers for calculating knockdown efficiency	
JMJD1C-R	TGTTAGGCATAGTCCGTCCA
JARID2-F	TCCAAAGTGTGCTGTGGGTA
JARID2-R	TGCTTCTGCTTGTGCAATCT
KDM1A-F	CGGCATCTATAAGAGGATAAAACC
KDM1A-R	CCTGGCTTCCAAAAGTGTGA
KDM2A-F	ACTGCATAACCAACCGTTCC
KDM2A-R	TTCTCGATCCACTGCTTCCT
KDM4A-F	GGGCAGTCTGGCCTCTTTAC
KDM4A-R	CGCTCGAGCTCTTCAAACTC
TET2-F	TGGATACACCTGTCAAGACTCAA
TET2-R	GGACCTGCTCCTAGATGGGTA
BMI1-F	ATGCAGCTCATCCTTCTGCT
BMI1-R	TCTGGTCAAAGAATTCAATGGATA
GRHL1-F	CTGAGCCAGATCACAGCAAA
GRHL1-R	GCACTCTGTTTTCTCCAGCA
GRHL2-F	AGTTGGGGCTGAGGAGTACA
GRHL2-R	CCTGCTTCTGACGGAGAGAT
GRHL3-F	GTTCCGGAGGAAGGTCAAGT
GRHL3-R	CAGTCTCTGGCCGAAGGTAG
SUV39H2-F	TCGATACGGCAATGTGTCTC
SUV39H2-R	AAAGTCAGCTCTTCTCCAGCA
EZH2F	TTCATGCAACACCCAACACT
EZH2R	ACGTTTTGGTGGGGTCTTTA
DOT1LF	CCACCAACTGCAAACATCAC
DOT1LR	ATTCGCTCCCTCCACTCTTC
EHMT2F	GGGACCTTCATCTGCGAGTA
EHMT2R	GGTCACACAGGTGGTTGATG
SUV39H1F	CCTGCCCTCGGTATCTCTAA
SUV39H1R	CTGCTTGAGGATACGCACAC

Primers used for high throughput sequencing of shRNA barcodes (Ref: Cellecta Decipher RNAi screen manual)	
FwdHTS	TTCTCTGGCAAGCAAAAGACGGCATA
RevHTS	TGCCATTTGTCTCGAGGTCGAGAA
FwdGex	CAAGCAGAAGACGGCATACGAGA
IND1AS	AATGATACGGCGACCACCGAGAGGTTCAGAGACGTACAGTCCGAAA
IND2AS	AATGATACGGCGACCACCGAGAGGTTCAGATTATTACAGTCCGAAA
GexSeqN	ACAGTCCGAAACCCCAAACGCACGAA

Oligos used for high throughput sequencing of shRNAs (Ref: Transomic pZIP lentiviral pooled shRNA screening library manual)	
Primary PCR Forward	CAGAATCGTTGCCTGCACATCT TGGAAAC
Primary PCR Reverse	CGTATCCACATAGCGTAAAAGGAGCAAC
Secondary PCR Forward	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACA CGACGCTCTTCCGATCTTAGTGAAGCCACAGATGTA
Secondary PCR Reverse Index 1	AAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCTCGTATCCACATAGCGTAAAAGG
Secondary PCR Reverse Index 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCGTATCCACATAGCGTAAAAGG
Secondary PCR Reverse Index 3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCGTATCCACATAGCGTAAAAGG
Secondary PCR Reverse Index 4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCGTATCCACATAGCGTAAAAGG
Secondary PCR Reverse Index 5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCGTATCCACATAGCGTAAAAGG
shRNA Loop Sequencing Primer	ACGACGCTCTTCCGATCTTAGTGAAGCCACAGATGTA

Table 3.4: Flow cytometry antibodies used in this study

Purpose	Antibody-Flurochrome	Catalogue No	Source
Flow cytometry			
Quantitation of % of CD 34 cells	CD 34-APC	555824	BD Pharmingen,San Diego,CA
	CD 34-PE	343506	Biolegend
Assessment of erythroid differentiation of HSCs	CD 71-FITC	555536	BD Pharmingen,San Diego,CA
	CD 71-APC	341028	
	CD 235a-PE	555570	
	CD 235a-APC	551336	
	CD 235a-BV421	562938	

Table 3.5: Sequences of shRNAs used for knocking down the expression of target genes

Gene	shRNA Sequences
hSETD7	TGCTGTTGACAGTGAGCGCTGGAGAGATGATAGAAGGCAATAGTGAAGCCACAGATGTATTGC CTTCTATCATCTCTCCATTGCCTACTGCCTCGGA
hPRDM12	TGCTGTTGACAGTGAGCGAATCAAGTGTGCACGTAACGAATAGTGAAGCCACAGATGTATTCG TTACGTGCACACTTGATGTGCCTACTGCCTCGGA
hKDM1B	TGCTGTTGACAGTGAGCGCCTGGAGATGAAGTGCAGGTTATAGTGAAGCCACAGATGTATAAC CTGCACTTCATCTCCAGATGCCTACTGCCTCGGA
hKDM6B	TGCTGTTGACAGTGAGCGACAGCAGGAATGCCAAGGTGAATAGTGAAGCCACAGATGTATTCA CCTTGGCATTCCTGCTGGTGCCTACTGCCTCGGA
hKDM2B	TGCTGTTGACAGTGAGCGACCCGATTGACCGCCAGCGATATAGTGAAGCCACAGATGTATATC GCTGGCGGTCAATCGGGCTGCCTACTGCCTCGGA
hDNMT3A	TGCTGTTGACAGTGAGCGATCCAGATGTTCTTCGCTAATATAGTGAAGCCACAGATGTATATTA GCGAAGAACATCTGGAGTGCCTACTGCCTCGGA
hTET1	TGCTGTTGACAGTGAGCGCCAGGAAGGAAGATGTAAACAATAGTGAAGCCACAGATGTATTGT TTACATCTTCCTTCCTGATGCCTACTGCCTCGGA
hJMJD1C	TGCTGTTGACAGTGAGCGCAACCTCCAAGATAAATGTAAATAGTGAAGCCACAGATGTATTTA CATTTATCTTGGAGGTTTTGCCTACTGCCTCGGA
hJARID2	TGCTGTTGACAGTGAGCGCCTCCATGGAGAAGTTACTCTATAGTGAAGCCACAGATGTATAGA GTAACTTCTCCATGGAGATGCCTACTGCCTCGGA
hKDM1A	TGCTGTTGACAGTGAGCGACAGCAGCTCGACAGTTACAAATAGTGAAGCCACAGATGTATTTG TAACTGTCGAGCTGCTGCTGCCTACTGCCTCGGA
hKDM2A	TGCTGTTGACAGTGAGCGCCAGCATGGATTTGGAGTTAAATAGTGAAGCCACAGATGTATTTA ACTCCAAATCCATGCTGATGCCTACTGCCTCGGA
hKDM4A	TGCTGTTGACAGTGAGCGCCTGGAAAAATCTTACATTCAATAGTGAAGCCACAGATGTATTGA ATGTAAGATTTTTCCAGTTGCCTACTGCCTCGGA
hTET2	TGCTGTTGACAGTGAGCGAATGGTGAAAATCAGTATTCAATAGTGAAGCCACAGATGTATTGA ATACTGATTTTCACCATGTGCCTACTGCCTCGGA
hBMI1	TGCTGTTGACAGTGAGCGAATGAAGATAAGAGAATTATAATAGTGAAGCCACAGATGTATTAT AATTCTCTTATCTTCATCTGCCTACTGCCTCGGA
hGRHL1	TGCTGTTGACAGTGAGCGCCAGAGTGCAAGTACTGAAAAATAGTGAAGCCACAGATGTATTTT TCAGTACTTGCACTCTGTTGCCTACTGCCTCGGA
hGRHL2	TGCTGTTGACAGTGAGCGCCAGCGAGACCGGAGACAACAATAGTGAAGCCACAGATGTATTGT TGTCTCCGGTCTCGCTGATGCCTACTGCCTCGGA
hGRHL3	TGCTGTTGACAGTGAGCGCCAGGGGCAATGAGACGACCTATAGTGAAGCCACAGATGTATAG GTCGTCTCATTGCCCCTGATGCCTACTGCCTCGGA
mGRHL1	TGCTGTTGACAGTGAGCGCTCGACACATACAGCTATAACATAGTGAAGCCACAGATGTATGTT ATAGCTGTATGTGTCGATTGCCTACTGCCTCGGA
mGRHL2	TGCTGTTGACAGTGAGCGAAGACAAGAGACTTCTGTCTGATAGTGAAGCCACAGATGTATCAG ACAGAAGTCTCTTGTCTCTGCCTACTGCCTCGGA
mGRHL3	TGCTGTTGACAGTGAGCGCCCTCATGAAATTTTTGACAGATAGTGAAGCCACAGATGTATCTGT CAAAAATTTCATGAGGTTGCCTACTGCCTCGGA
hSUV420H2	TGCTGTTGACAGTGAGCGCCTGGACTATGAGTCTGATGAATAGTGAAGCCACAGATGTATTCA TCAGACTCATAGTCCAGATGCCTACTGCCTCGGA
hSUV39H2	TGCTGTTGACAGTGAGCGCGCCCGTTACTGCTTCAGCAATTAGTGAAGCCACAGATGTAATTG CTGAAGCAGTAACGGGCATGCCTACTGCCTCGGA

Gene	shRNA Sequences
Scrambled	TGCTGTTGACAGTGAGCGTCTCGCTTGGGCGAGAGTAAGATAGTGAAGCCACAGATGTATCTT ACTCTCGCCCAAGCGAGATGCCTACTGCCTCGGA
hDOT1L	TGCTGTTGACAGTGAGCACGCGAGTTCAGGAAGTGGATATAGTGAAGCCACAGATGTATATCC ACTTCCTGAACTCGCGGTGCCTACTGCCTCGGA
hEZH2	TGCTGTTGACAGTGAGCAGACCACAGTGTTACCAGCATATAGTGAAGCCACAGATGTATATGC TGGTAACACTGTGGTCCTGCCTACTGCCTCGGA
hEHMT2	TGCTGTTGACAGTGAGCCGAGGATGATTCTTACCTCTTATAGTGAAGCCACAGATGTATAAGA GGTAAGAATCATCCTCTTGCCTACTGCCTCGGA
hSUV39H1	TGCTGTTGACAGTGAGCCCCTGGTGAAATGGCGTGGATATAGTGAAGCCACAGATGTATATCC ACGCCATTTCACCAGGTTGCCTACTGCCTCGGA

The sequences highlighted in red indicate the sense strand of the shRNA; the sequences highlighted in purple indicate the stem loop and the sequences highlighted in green indicate the antisense strand.

Table 3.6: shRNA specific barcode amplification PCR (Cellecta signalling pathway shRNA library).

I. First round of PCR

Reaction Components	50 µl reaction
Water	6 µl
10X buffer	5 µl
dNTP(10mM)	4 µl
Forward Primer (10 µM)	2 µl
Reverse Primer (10 µM)	2 µl
Titanium Taq Polymerase	1 µl
DNA Template	30 µl

Reaction Conditions

94^0 C-3 minutes

94^0 C-30 seconds
65^0 C-10 seconds
72^0C- 20 seconds
} 16 cycles

68^0 C-2 minutes

II. Second round of PCR

Reaction Components	50 µl reaction
Water	26 µl
10X buffer	5 µl
dNTP(10mM)	4 µl
Forward Primer (10 µM)	2 µl
Reverse Primer (10 µM)	2 µl
Titanium Taq Polymerase	1 µl
DNA Template	10 µl

Reaction Conditions
94^0 C-3 minutes
94^0 C-30 seconds
65^0 C-10 seconds } 14 cycles
72^0C- 10 seconds
68^0 C-2 minutes

Table 3.7: PCR to amplify and clone pooled shRNA oligos

Primary PCR

Reaction Components	50 µl reaction
Water	28 µl
10X buffer	5 µl
dNTP(10mM)	5 µl
25 mM $MgSO_4$	3 µl
Forward Primer (10 µM)	1.5 µl
Reverse Primer (10 µM)	1.5 µl
KOD DNA Polymerase	1.0 µl
DNA Template	5 µl

Reaction Conditions
95^0 C-2 minutes (Hold)
95^0 C-20 seconds
60^0 C-10 seconds } 12 cycles
70^0C- 10 seconds

Secondary PCR

Reaction Components	50 µl reaction
Water	31 µl
10X buffer	5 µl
dNTP(10mM)	5 µl
25 mM MgSO4	3 µl
Forward Primer (10 µM)	1.5 µl
Reverse Primer (10 µM)	1.5 µl
KOD DNA Polymerase	1.0 µl
DNA Template	2 µl

Reaction Conditions

95^0 C-2 minutes (Hold)
95^0 C-20 seconds
60^0 C-10 seconds } 12 cycles
70^0C- 10 seconds

Table 3.8: shRNA specific amplification PCR

Primary PCR

Reaction Components	50 µl reaction
Water	24 µl
10X buffer	5 µl
dNTP(10mM)	5 µl
25 mM MgSO4	4 µl
Betaine	5 µl
Forward Primer (10 µM)	1.7 µl
Reverse Primer (10 µM)	1.7 µl
KOD DNA Polymerase	1.5 µl
DNA Template	2 µl

Reaction Conditions

98^0 C-5 minutes
95^0 C-30 seconds
57^0 C-30 seconds } 25 cycles
72^0C- 30 seconds
72^0C- 5 minutes

Secondary PCR

Reaction Components	50 µl reaction
Water	13 µl
10X buffer	5 µl
dNTP(10mM)	5 µl
25 mM MgSO4	4 µl
Betaine	2 µl
Forward Primer (10 µM)	7.5 µl
Reverse Primer (10 µM)	7.5µl
KOD DNA Polymerase	1.5 µl
DNA Template	5.6 µl

Reaction Conditions

98^0 C-5 minutes
94^0 C-30 seconds
52^0 C-30 seconds } 12 cycles
72^0C- 30 seconds
72^0C- 5 minutes

4. miRNOME SCREENING IN HUMAN ERYTHROPOIESIS

4.1 INTRODUCTION

Erythropoiesis is a unique and dynamic cellular differentiation process in which haematopoietic stem cells and progenitors (HSPCs) undergo different stages of maturation towards erythrocytes or red blood cells (RBCs). During differentiation, each stage of erythroid maturation is distinct from the other in their morphological, cellular and molecular properties. In the path of maturation, HSPCs differentiate to form blast forming units-erythroid (BFU-e), which divide rapidly and differentiate into colony-forming unit-erythroid (CFU-e). CFU-e cannot self-renew, and on each successive division, it leads to the formation of erythroid precursors such as basophilic erythroblasts (Baso-E), polychromatophilic erythroblasts (Poly-E), orthochromatophilic erythroblasts (Ortho-E), reticulocytes (Retics) and finally to RBCs. From BFU-e stage, the erythroid cells show a marked decrease in size, nuclear condensation, haemoglobinization and expulsion of cell organelles and, by reticulocyte stage, they become enucleated and are released into the bloodstream (Mello *et al.*, 2019)(Hattangadi *et al.*, 2011) (An *et al.*, 2014). An intricate network of transcription factors, epigenetic modifiers, signalling proteins, growth factors, and, more recently, the long non-coding RNAs and short non-coding RNAs, especially the miRNAs orchestrate the complex process of erythropoiesis. Several studies have been carried out which underlines the importance of miRNAs in several stages of erythropoiesis such as in terminal differentiation (Dore *et al.*, 2008), enucleation (Zhang *et al.*, 2011) (Rouzbeh *et al.*, 2015), regulation of oxidant stress in erythroid cells (Yu *et al.*, 2010),(Sangokoya, Telen and Chi, 2010),(Zhang *et al.*, 2018)

erythrocyte survival (Rivkin *et al.*, 2017), globin regulation (Sankaran *et al.*, 2011), (Azzouzi *et al.*, 2011), (Lulli *et al.*, 2013), (Li *et al.*, 2018) and iron metabolism (Andolfo *et al.*, 2010).

Most of the previous studies carried out in the identification of miRNAs relied on microarrays; however, with the advancement of the state of the art technology such as RNA sequencing, it is now possible to delineate the vast repertoire of known and novel miRNAs involved in any cellular process. The technique of RNA sequencing has led to the identification of miR-4732 as a bonafide erythroid miRNA.miR-4732 is present in a cluster along with miR144/451, but its role as a regulator of erythroid cell proliferation by targeting SMAD4 pathway was recently delineated in this study (Doss *et al.*, 2015). Similarly, RNA sequencing has also made possible the comparative analysis of the small RNA profile in fetal and human erythroblasts, which led to the identification of let7 group of miRNAs and 14q32 miRNAs as key players in the adult and fetal erythroid compartment (Lessard *et al.*, 2018). Therefore, given the predominant role that miRNAs play at different stages of erythropoiesis, it is essential to extensively elucidate the miRNA profile during the process of erythroid differentiation. In this study, complete miRNA profiling was carried out by collecting cells at different stages of erythropoiesis and subjecting them to small RNA sequencing. A robust ex vivo erythroid culture system which mimicked in vivo erythropoiesis was used to carry out the study.

In comparison to all other studies reported in the literature, where they have used red blood cells and cells from fetal and adult erythroid compartments, this study has utilized cell types from different developmental stages of adult erythropoiesis. In

addition to this, the most significantly upregulated miRNAs were functionally validated by overexpressing and knocking them out using *Crispr-Cas9.* It has been shown previously that it is possible to design single guide RNAs, which can specifically disrupt the biogenesis sites and cause a significant decline in the expression of miRNAs (Chang *et al.*, 2016). Therefore this was used as a strategy to effectively knock out miRNAs.

The strategies which were used to carry out the study are stated below:

1. Establishment of a robust ex vivo erythroid culture system to study the miRNA profile in erythropoiesis.

2. To identify the differentially expressed miRNAs in HSCs and erythroid cells by small RNA sequencing.

3. To analyze the transcriptional regulation of miRNAs in erythroid cells by ChIP-sequencing analysis.

4. To functionally validate miRNAs overexpressing them and by knocking them out using *Crispr-Cas9*.

RESULTS

4.2 A TWO-PHASE EX-VIVO CULTURE SYSTEM FOR THE GENERATION OF ERYTHROID CELLS

Purified human CD 34 + haematopoietic stem progenitor cells were differentiated *ex vivo* (**Figure 4.2.1**). Briefly, CD 34+ cells isolated from mobilized peripheral blood

were expanded for five days and then driven towards erythroid differentiation. After culturing for four days in the differentiation medium, it was observed that the expression of CD71 (transferrin receptor) had reached a peak of 94 %. In contrast, CD 235a (glycophorin A) exhibited heterogeneous levels of expression (**Figure 4.2.2**). This suggested that the cells were in the earlier stages of differentiation. By day 7 in culture, the CD 71 levels had attained its maximum expression level of 100%, and a steady increase in the levels of CD235a was observed even though its expression pattern was still heterogeneous. Most of the cells in this stage were proerythroblasts. By day 10 of differentiation, almost 95 % of the cells were double positive for CD 71 and CD 235a. By day 14, most of the cells had undergone terminal erythroid maturation, with 14 % of cells expressing CD 235a alone and representing the population of enucleated cells. Quantitative PCR analysis showed a progressive increase in the expression of erythroid-specific genes, α-globin, γ-globin and β-globin in the cells collected on different days of differentiation (**Figure 4.2.3**). Haemoglobin analysis on day 12 showed that HbA (84%) was the predominant form of haemoglobin, which was expressed in these cells (**Figure 4.2.3**). This data thus suggested that erythroid cells could be successfully generated in *in-vitro* conditions, and it mimicked in vivo erythropoiesis during the 14-day culture period. This makes it an excellent system to carry out high throughput transcriptome studies.

4.3 SMALL RNA SEQUENCING OF CULTURED ERYTHROID CELLS

Cells were harvested on days 3 and 5 before differentiation and days 9, 11, and 14 after differentiation. Small RNA sequencing was performed on three

biological replicates at each stage using a sequencing format on an Illumina Hi-Seq 1000 platform. To scrutinize the alterations in the expression of the miRNAs at the global level, a principal component analysis was performed. All biological triplicates of a particular time point clustered together, which authenticated the reproducibility of the data. The expression of each miRNA at each time point (in triplicates) is presented on the heat map. Bioinformatics analysis showed that 358 miRNAs were differentially expressed during erythroid differentiation (**Figure 4.3.1)**. The fact most of the miRNAs were downregulated during the process of erythroid differentiation (**Figure 4.3.2**) suggests that the transcriptional changes during erythroid differentiation cause global downregulation of miRNAs. This observation was concordant with the small RNA expression observed in mouse erythroid cells (Zhang *et al.*, 2011). Sixteen differentially expressed miRNAs identified from the small RNA sequencing data were validated by real-time PCR (**Figure 4.3.3)**.

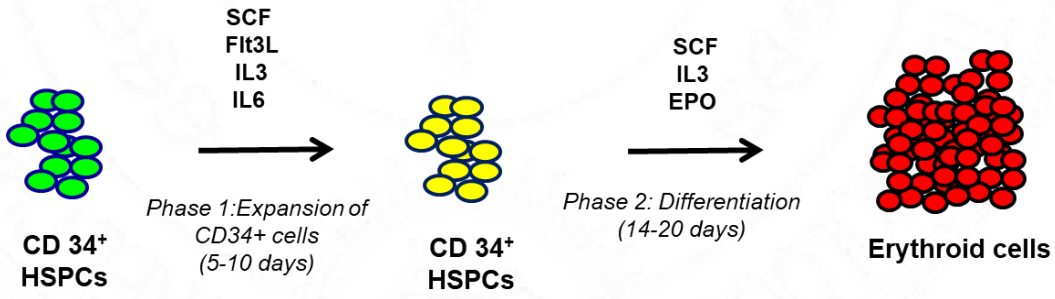


Figure 4.2.1: Schematic representation of the *ex vivo* erythroid culture system used for the study

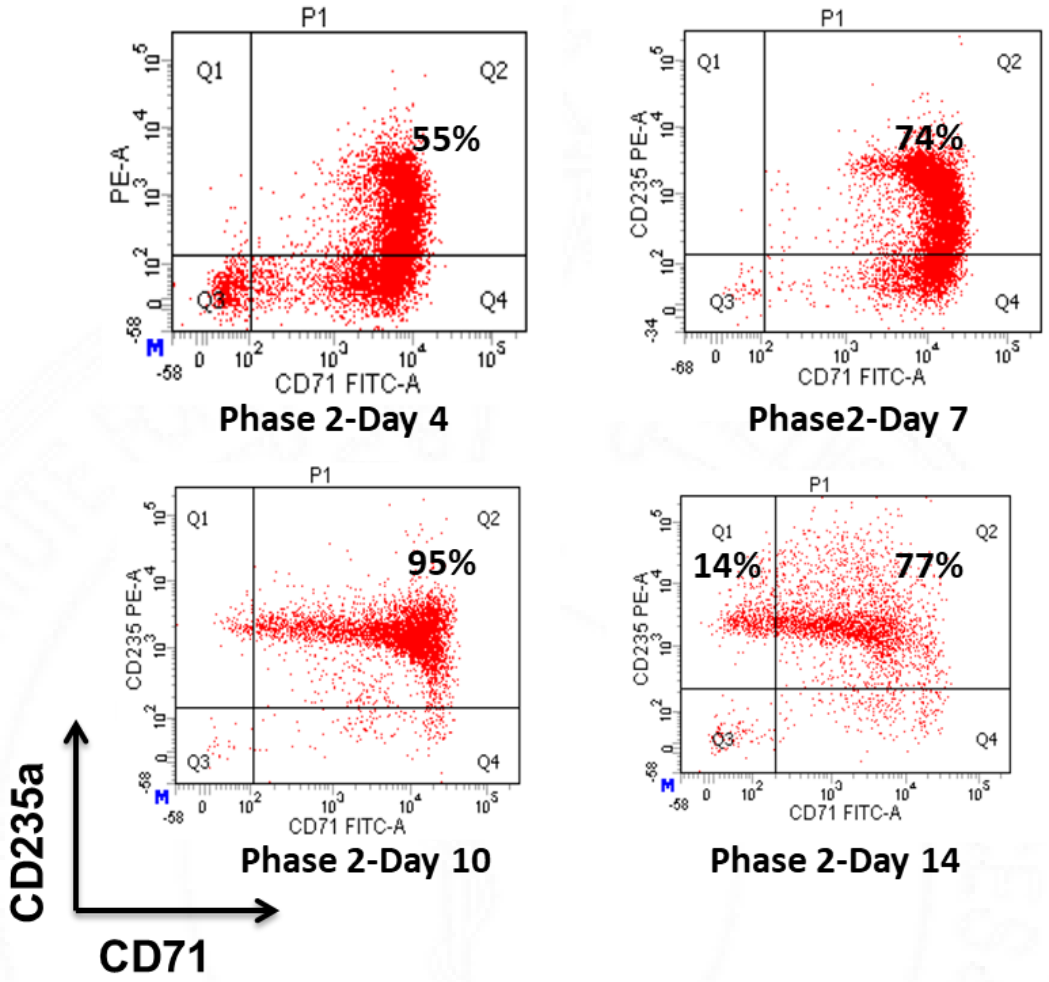


Figure 4.2.2: Flow cytometry analysis of the expression of transferrin receptor (CD 71) and glycophorin A (CD 235a) during *ex-vivo* erythroid differentiation. Cells double positive for CD71 and CD 235a increased from 55% to 95% between days 4 and 10. Cells positive for CD235a alone were present on day 14.

Among the differentially expressed miRNAs, 54 miRNAs showed upregulation, and 61 miRNAs showed downregulation, with a fold change $\geq$+3 and $\leq$-3, respectively (**Tables 4.3.1 and 4.3.2**). A good correlation was observed for the expression of 36 miRNAs, which were previously reported in erythropoiesis (**Tables**

4.3.3), suggesting the authenticity of the miRNAs profile analysis performed in this study. Our study identified 79 miRNAs which were not identified earlier in erythropoiesis. We observed 42 highly expressed miRNAs at all stages of ex-vivo erythropoiesis, without significant difference in their expression levels (**Table 4.3.4**), suggesting that a large number of miRNAs which are expressed abundantly in erythropoiesis have a stable expression. These miRNAs belong to the class of those required for the maintenance of the general housekeeping pathways or a cellular process that is essential throughout erythropoiesis.

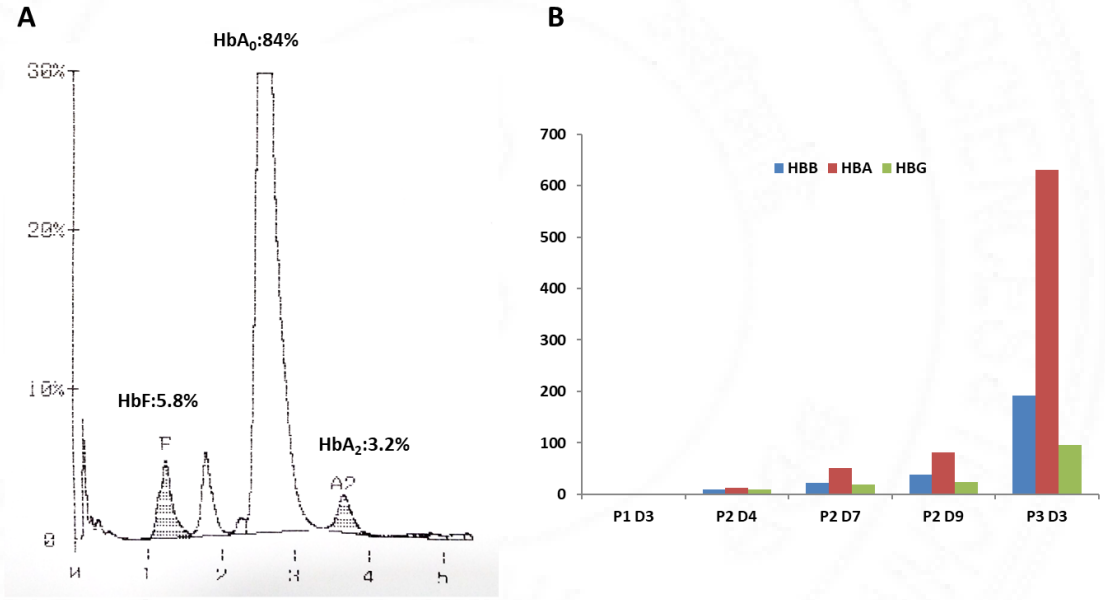


Figure 4.2.3: (A) HPLC analysis showing the expression levels of different forms of haemoglobins in the cultured erythroid cells (B) Expression of HBB, HBA and HBG genes erythroid during differentiation. P1, P2 and P3 represent phases 1, 2 and 3 of the erythroid culture, and D represents the days of the culture.

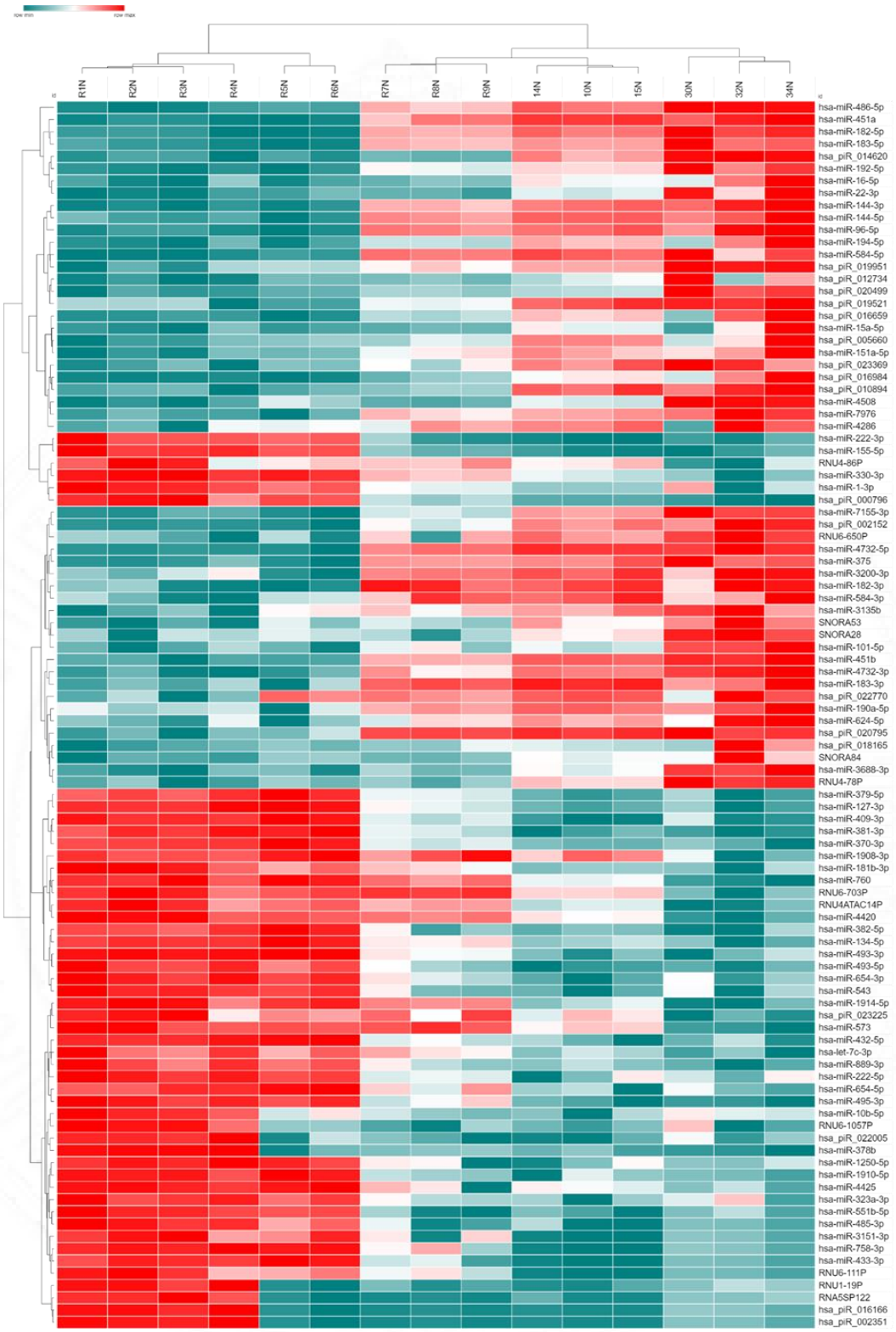


Figure 4.3.1: Heat map of miRNA expression from day 5 of the culture till day 14 of erythroid differentiation. Only miRNAs with and expression of ≥ 100 RPM for at least one time point were included.

Table 4.3.1 List of upregulated miRNAs with log2 FC≥+3

miRNA	log2FC	miRNA	log2FC
hsa-miR-616-5p	3.02	hsa-miR-548ag	3.94
hsa-miR-195-5p	3.02	hsa-miR-192-5p	3.94
hsa-miR-4497	3.01	hsa-miR-4448	3.97
hsa-miR-548bb-5p	3.00	hsa-miR-624-5p	3.98
hsa-miR-34a-5p	3.00	hsa-miR-194-5p	4.02
hsa-miR-548az-5p	3.00	hsa-miR-548h-5p	4.07
hsa-miR-501-5p	3.01	hsa-miR-215-5p	4.12
hsa-miR-22-3p	3.02	hsa-miR-548an	4.20
hsa-miR-210-3p	3.06	hsa-miR-548ak	4.28
hsa-miR-625-5p	3.09	hsa-miR-4306	4.45
hsa-miR-16-5p	3.14	hsa-miR-548bb-3p	4.47
hsa-miR-548aj-3p	3.18	hsa-miR-548ad-5p	4.55
hsa-miR-6513-5p	3.20	hsa-miR-548ae-5p	4.55
hsa-miR-548n	3.26	hsa-miR-486-3p	4.55
hsa-miR-6073	3.30	hsa-miR-548t-5p	4.57
hsa-miR-3688-3p	3.40	hsa-miR-144-5p	4.71
hsa-miR-5695	3.42	hsa-miR-548c-3p	4.84
hsa-miR-548aT-5p	3.43	hsa-miR-548d-3p	4.87
hsa-miR-548ay-5p	3.48	hsa-miR-486-5p	4.87
hsa-miR-548ar-5p	3.66	hsa-miR-548m	4.95
hsa-miR-12136	3.66	hsa-miR-4732-5p	5.09
hsa-miR-584-5p	3.81	hsa-miR-96-5p	6.12
hsa-miR-3663-5p	3.82	hsa-miR-183-5p	6.96
hsa-miR-3163	3.84	hsa-miR-182-5p	7.20
hsa-miR-4531	3.86	hsa-miR-144-3p	7.76
hsa-miR-15a-5p	3.86	hsa-miR-451a	10.64

Table 4.3.2 List of downregulated miRNAs with log2 FC≤-3

miRNA	log2FoldChange
hsa-miR-181c-3p	-3.00
hsa-miR-6890-3p	-3.00
hsa-miR-145-3p	-3.00
hsa-miR-6792-5p	-3.03
hsa-miR-338-5p	-3.03
hsa-miR-598-5p	-3.05
hsa-miR-191-3p	-3.06
hsa-miR-92a-1-5p	-3.08
hsa-let-7e-3p	-3.09
hsa-miR-4494	-3.10
hsa-miR-6503-3p	-3.12
hsa-miR-1273h-3p	-3.13
hsa-miR-494-5p	-3.16
hsa-miR-330-3p	-3.21
hsa-let-7a-2-3p	-3.26
hsa-miR-7974	-3.26
hsa-miR-6804-3p	-3.30
hsa-miR-766-5p	-3.31
hsa-miR-411-5p	-3.35
hsa-miR-124-5p	-3.35
hsa-miR-181a-3p	-3.36
hsa-miR-149-5p	-3.40
hsa-miR-1257	-3.42
hsa-miR-4515	-3.44
hsa-miR-335-3p	-3.50
hsa-miR-2115-3p	-3.52
hsa-miR-130b-5p	-3.62
hsa-miR-125b-1-3p	-3.64
hsa-miR-1271-5p	-3.64
hsa-miR-98-3p	-3.67
hsa-miR-487b-3p	-3.70

miRNA	log2FoldChange
hsa-miR-222-3p	-3.74
hsa-miR-4472	-3.76
hsa-miR-181c-5p	-3.77
hsa-miR-181b-2-3p	-3.82
hsa-miR-379-5p	-3.87
hsa-miR-1-3p	-3.91
hsa-miR-143-3p	-3.91
hsa-miR-382-5p	-3.93
hsa-miR-99b-3p	-4.07
hsa-miR-181b-3p	-4.14
hsa-miR-323b-3p	-4.17
hsa-miR-558	-4.18
hsa-miR-155-5p	-4.22
hsa-miR-99a-3p	-4.29
hsa-miR-206	-4.38
hsa-miR-196a-5p	-4.39
hsa-miR-196b-5p	-4.42
hsa-let-7c-3p	-4.44
hsa-miR-654-3p	-4.53
hsa-miR-222-5p	-4.53
hsa-miR-432-5p	-4.56
hsa-miR-300	-4.61
hsa-miR-381-3p	-4.63
hsa-miR-409-3p	-4.65
hsa-miR-323a-3p	-4.66
hsa-miR-493-3p	-5.05
hsa-miR-493-5p	-5.44
hsa-miR-543	-5.51
hsa-miR-134-5p	-6.12
hsa-miR-370-3p	-6.14

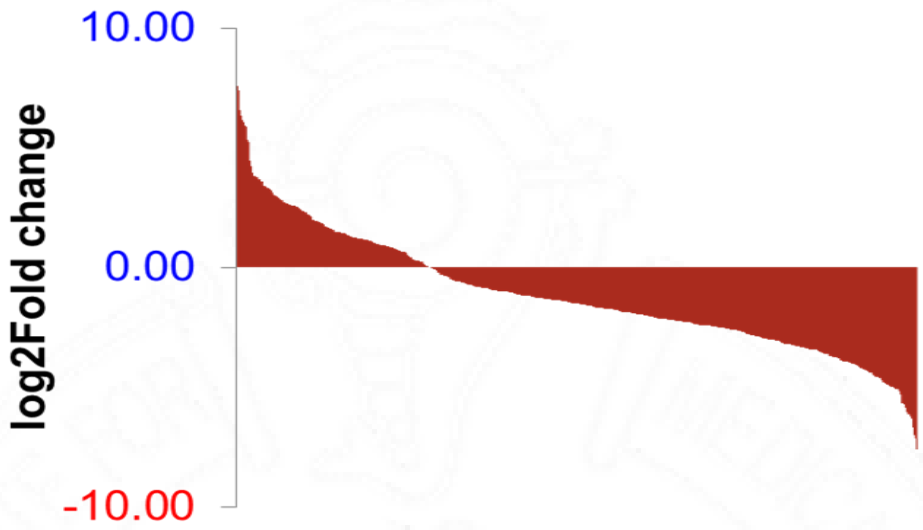


Figure 4.3.2: Global downregulation of miRNAs observed during the process of erythroid differentiation.

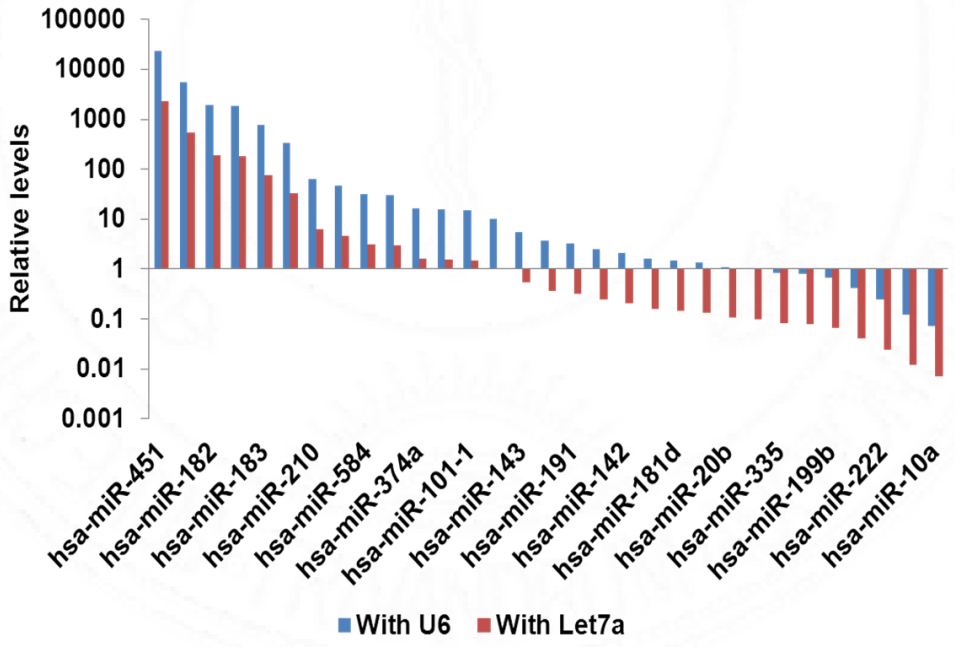


Figure 4.3.3: Validation of small RNA Seq data by quantitative real-time PCR. The relative expression of the miRNAs was estimated by normalizing with U6 and let 7a levels.

Table 4.3.3: Previously reported miRNAs in erythropoiesis detected in our study.

miRNA	Role in Erythropoiesis
miR 23a~27a~24-2	Supress KLF3 and SP1
miR-126	Terminal differentation in primitive erythropoiesis
miR-144	Terminal Maturation, Oxidative Stress
miR-4732	Promotes erythropoiesis
miR-145	KD showed impaired erythroid & megakaryocyte differentiation
miR-150	Lineage decision between megakaryopoiesis and erythropoiesis
miR-155	Inhibits Megakaryopoiesis
miR-155	Favors erythropoiesis
miR-15a	Cell cycle (Maintains G0/G1 Phase)
miR-15a	Overexpression blocks Differentiation, HbF regulation
miR-16-1	Overexpression blocks Differentiation, HbF regulation
miR-16-2	abnormal expansion of erythroid cells in PV patients
miR-191	Involved in erythroblat enucleation
miR-210	It's upregulation corelated with γ globin expression
miR-221	Inhibits Terminal Maturation
miR-222	Inhibits Terminal Maturation
miR-223	Inhibits erythroid differentiation
miR-24	Inhibits erythroid differentiation
miR-376a	Promotes erythroid differentiation
miR-451	Terminal Maturation, Oxidative Stress
miR-486	Positively regulates HbF,regulates proliferation and erythroid differentiation
miR-96	Regulates γ Globin gene expression
miR-let-7d	Iron metabolism in erythroid cells, Regulation of globin gene expression
miR-125a	Inhibits erythroid differentiation
miR-199b	Promotes erythroid differentiation
miR-485	Regulates Cellular Iron homeostasis
miR-99a/125b-2	Inhibits erythroid differentiation and promotes megakaryopoiesis
miR-181a/b	Promotes foetal erythroid differentiation
miR-146b	Promotes erythroid differentiation
mir 320a	Inhibits erythroid differentiation
mir 218	Inhibits erythroid differentiation and alters iron metabolism
mir 200a-3p	Inhibits erythroid differentiation
mir 142	Maintains erythrocyte survival,cytoskeleton and function
mir 9	Inhibits erythropoiesis
miR-326	Regulates HbF synthesis

Table 4.3.4 miRNAs with uniform expression throughout ex-vivo erythropoiesis. (TP-1 and TP5 designate the first and the last time points of erythroid culture. Normalized read counts from triplicates of each time point are shown)

miRNA	Read Counts					
	TP-1	TP-1	TP-1	TP-5	TP-5	TP-5
hsa-miR-378c	12902.87	8083.418	8270.806	9831.968	5753.689	8164.334
hsa-miR-28-3p	13397.36	8316.446	8561.483	9911.592	5819.74	8411.211
hsa-miR-92a-3p	17637.67	13436.78	13665.98	19910.77	25997.88	53975.65
hsa-miR-1307-3p	20029.06	16099.16	19063.24	24758.28	38172.98	37664.35
hsa-miR-378e	28439.41	35000.24	35020.38	6815.812	29916.54	26771.52
hsa-miR-18a-5p	33815.88	26745.19	32507.06	6602.42	32452.92	106075.9
hsa-miR-378b	33820.01	32587.63	32683.54	15338.76	29251.22	27488
hsa-miR-18b-5p	35424.19	37581.68	41501.44	6540.313	13502.13	22483.25
hsa-miR-363-3p	51418.82	32298.09	45964.38	41910.87	13027.76	40102.17
hsa-miR-514a-3p	52900.62	34021.38	37286.62	18479.13	6851.345	24886.34
hsa-miR-625-3p	68648.01	44445.54	40407.25	32897.44	61406.31	47309.19
hsa-miR-24-3p	68770.39	63896.4	84431.36	93586.83	30193.96	58472.3
hsa-miR-92b-3p	73853.32	71070.73	74471.51	62570.11	20944.34	29745.38
hsa-miR-191-5p	78819.65	53864.34	58039.94	47361.93	18980.81	21158.26
hsa-miR-30e-3p	89766.91	67902.52	67113.22	85949.3	134375.2	200156
hsa-miR-30a-3p	89818.18	68115.32	69156.27	82880.59	205799.8	158666.2
hsa-miR-106a-5p	96267.14	91497.65	104875.3	86581.51	25396.21	42318.79
hsa-miR-17-5p	103989.5	99001.99	113317.4	84391.85	24682.86	42184.4
hsa-miR-423-3p	109905.1	93823.04	113184.5	54244.63	97855.95	260002.4
hsa-miR-148a-3p	109922.5	71037.94	70045.95	83815.38	10647.51	12281.98
hsa-miR-20b-5p	111659	95399.82	115042.8	54429.35	98012.07	261401.4
hsa-miR-27a-3p	111663.9	95399.12	115054.2	54429.35	98012.07	261404.4
hsa-let-7d-5p	112141	72698.43	71716.31	84598.88	10723.17	12454.12
hsa-miR-103a-3p	147120.4	87155.93	101678.9	13375.23	51580.24	222252
hsa-miR-103b	165139.2	146636.7	167029.4	56867.44	107595.5	189871
hsa-miR-107	173744.7	104133.5	122661.6	15289.39	54957.28	257412.1
hsa-let-7i-5p	177251.6	106832.1	125968.1	15670	58835.1	265914.6
hsa-miR-98-5p	284636.3	335697	323103.3	131217.1	172949.3	135932.3
hsa-let-7b-5p	395300.6	239615.5	289274.7	224940.9	80100.09	215682.9
hsa-let-7g-5p	424260.9	202367.3	244969.3	602747.1	208524.7	366968.4
hsa-let-7a-5p	447753	228593.4	281638.2	664094.2	270987.8	425119.1
hsa-let-7f-5p	490658.1	231650.7	281212.6	633294	241748.6	378074.2
hsa-let-7c-5p	729869.8	691635.3	683626.2	1144233	311548.3	317661.6
hsa-let-7e-5p	930871.5	672076.3	848838.8	1246292	889974.7	1129814
hsa-miR-93-5p	1227079	557582.1	667055.5	2250745	502666.6	1031763
hsa-miR-106b-3p	1306414	607562.4	729096.4	2326393	572445.8	1109822
hsa-miR-21-5p	1823012	882534	1079877	3084066	878702.8	1753745
hsa-miR-25-3p	1894796	924089.7	1130976	3141882	979471	1821272
hsa-miR-26b-5p	2030688	1127827	1388693	3370085	1346175	2103939

4.4 CO-REGULATION OF MIRNAS AND HOST GENES

The genomic location of miRNAs is the deciding factor that dictates the mode of regulation of the miRNAs, and based on their genomic location, miRNAs can be demarcated as either intragenic or intergenic. We found that the majority of the miRNAs were intragenic in location (**Figure 4.4.1**), and there was a significant correlation between the expression of the miRNAs and their host genes. This hinted at the existence of a co-transcriptional regulatory mechanism that transcribes the miRNAs and their host genes, and it also suggested that miRNAs use the transcriptional start sites of their host genes to initiate their transcription**.**

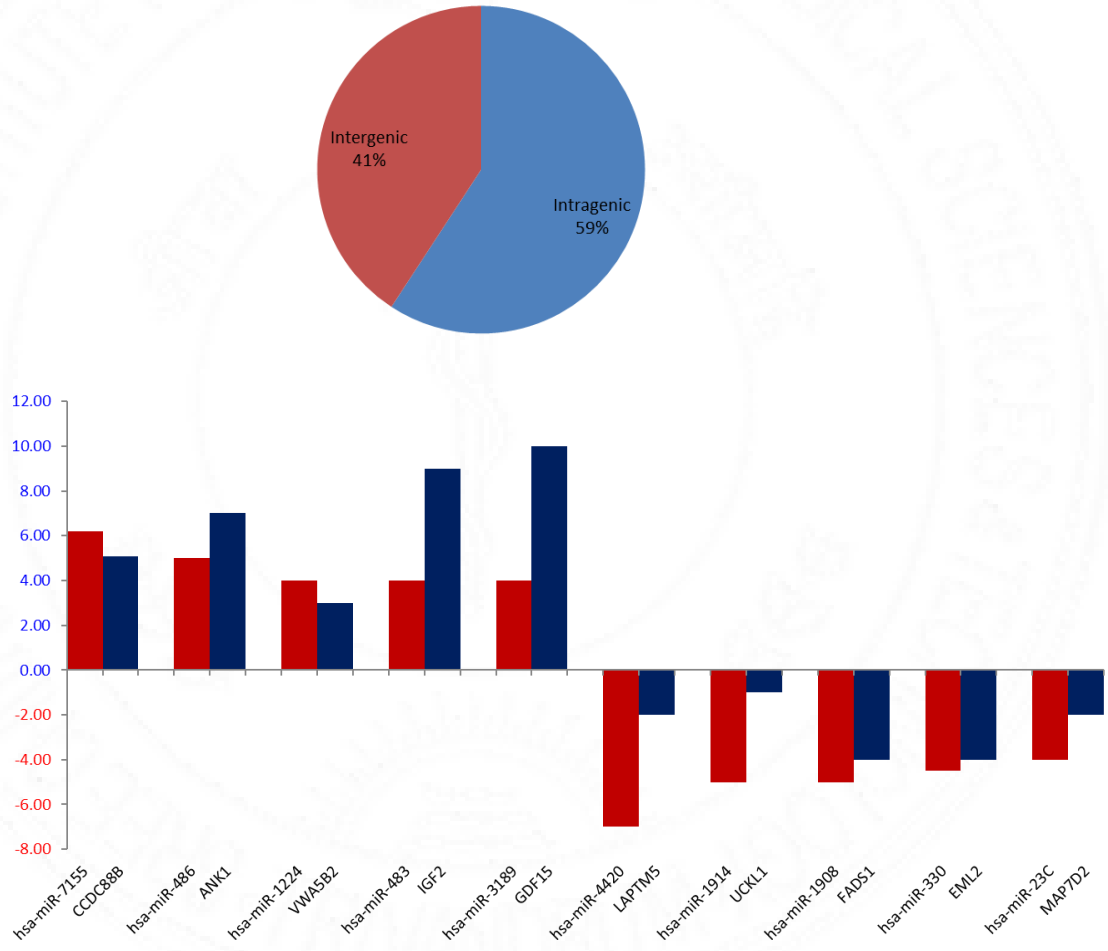


Figure 4.4.1: Distribution of intragenic and intergenic miRNAs, which are differentially expressed in human erythropoiesis. Significant correlation between the expression of miRNAs and host genes suggesting shared transcriptional machinery.

4.5 MULTIPLE miRNA CLUSTERS ARE INVOLVED IN REGULATING HUMAN ERYTHROPOIESIS

miRNAs can either exist as single entities in the human genome or can occur in a cluster. miRNAs occurring within a span of 50 kb within the same gene are designated as a cluster. About 15% of human miRNAs occur in clusters, and their expression patterns suggest that they are coexpressed by the same transcriptional regulatory mechanisms (Megraw *et al.*, 2007), (Lee *et al.*, 2004), (Yu *et al.*, 2006). miRNAs in a cluster may be functionally related by co-regulating or co-ordinately regulating biological processes (Lim *et al.*, 2003),(Xu and Wong, 2008). The systematic analysis of clustered miRNAs is necessary to uncover their potential functional correlation and co-ordination for cell fate determination and, therefore, we analyzed the expression of miRNA clusters in erythroid cells and undifferentiated HSPCs.

The top seven differentially regulated miRNA clusters identified in this study are shown in **Table 4.5.1.** miR-144/ 451, the most studied cluster in erythropoiesis, was upregulated as reported earlier (Sangokoya, Telen and Chi, 2010). They are involved in terminal differentiation of erythroid cells, and knockout mice showed defective erythropoiesis(Yu *et al.*, 2010)**.** The additional functions of these miRNAs include regulation of oxidative stress in the erythroid cells by inhibiting specific target mRNAs and regulating the expression of erythroid transcription factors (Patrick *et al.*, 2010). We identified a new member in this cluster, miR-4732, located ~0.4kb away from miR-144 and miR-451, and it was upregulated by ~600 fold in the cultured erythroid cells. A recent study proposed that miR-4732 is involved in cell

proliferation during erythroid differentiation by repressing the SMAD2/4 dependent TGF-β signalling pathway (Doss *et al.*, 2015). The miR-23a~27a~24-2 cluster has been reported to be regulated by GATA1 and, by a positive feedback loop with transcription factors KLF3 and SP1, it promotes β-like globin gene expression in erythroid cells (Ma *et al.*, 2013). The miR-15a/16-1 cluster is upregulated in the terminal stages of adult erythropoiesis, and it targets transcription factor MYB (Zhao *et al.*, 2008) (Sankaran *et al.*, 2011). In patients with trisomy 13, upregulation of miR-15a and miR-16-1 results in the downregulation of MYB and elevation of foetal γ-globin expression. The miR-221/222 cluster is downregulated during erythropoiesis has been found to negatively regulate erythroid proliferation by targeting C-KIT (Felli *et al.*, 2005). We found upregulation and downregulation of all these miRNA clusters, as reported previously.

Table 4.5.1 The top seven miRNA clusters differentially expressed in ex-vivo erythropoiesis

miRNAs in cluster	Expression
miR- 221 and 222	Down
miR-301b and 130b	Down
miR-143 and 145	Down
miR-181c and 181d	Down
miR-451,144 and4732	Up
miR-6777 and 33b	Up
miR-182,96 and 183	Up

We have also identified several miRNA clusters which were either upregulated or downregulated, but were not reported earlier in human erythropoiesis (**Table 4.5.2)**. These include the miR-183/96/182 and the miR 192/194-2 clusters, which showed very high upregulation in the erythroid cells compared to the undifferentiated HSPCs. The miR-183 cluster expression usually is highly specific to

the sensory organs and is necessary for sensory development (Jacobs *et al.*, 2007) (Dvornyk, Vinogradova and Nevo, 2003). It also plays a significant role in the regulation of several cellular pathways such as apoptosis, circadian rhythm, metabolism and immune regulation. The miR192/194-2 cluster is mostly expressed in the liver and intestine and is essential for hepatocyte function as well as intestinal epithelial cell differentiation (Hino *et al.*, 2008). Dysregulation of the miR-183/96/182 and miR 192/194-2 clusters has been associated with various cancers, and this has been attributed to the p53 binding site in their promoters, which gets disrupted when p53 is lost in most of the malignancies (Song *et al.*, 2008). The other clusters which were identified in this study were the miR 143/145 and miR125a clusters. The miR17/92 cluster plays an important role in the cell cycle, proliferation, and apoptosis, and its widespread role as a potential oncomir is attributed to the transcriptional regulation of the cluster by C-Myc and N-Myc. The miR 125a cluster has been recognized as one of the critical regulators of primitive haematopoietic stem cells in mice, and it has been observed that the majority of this function is carried out by miR 125a. Similarly, it has also been reported in another study that miR 125a controls the haematopoietic stem cell number.

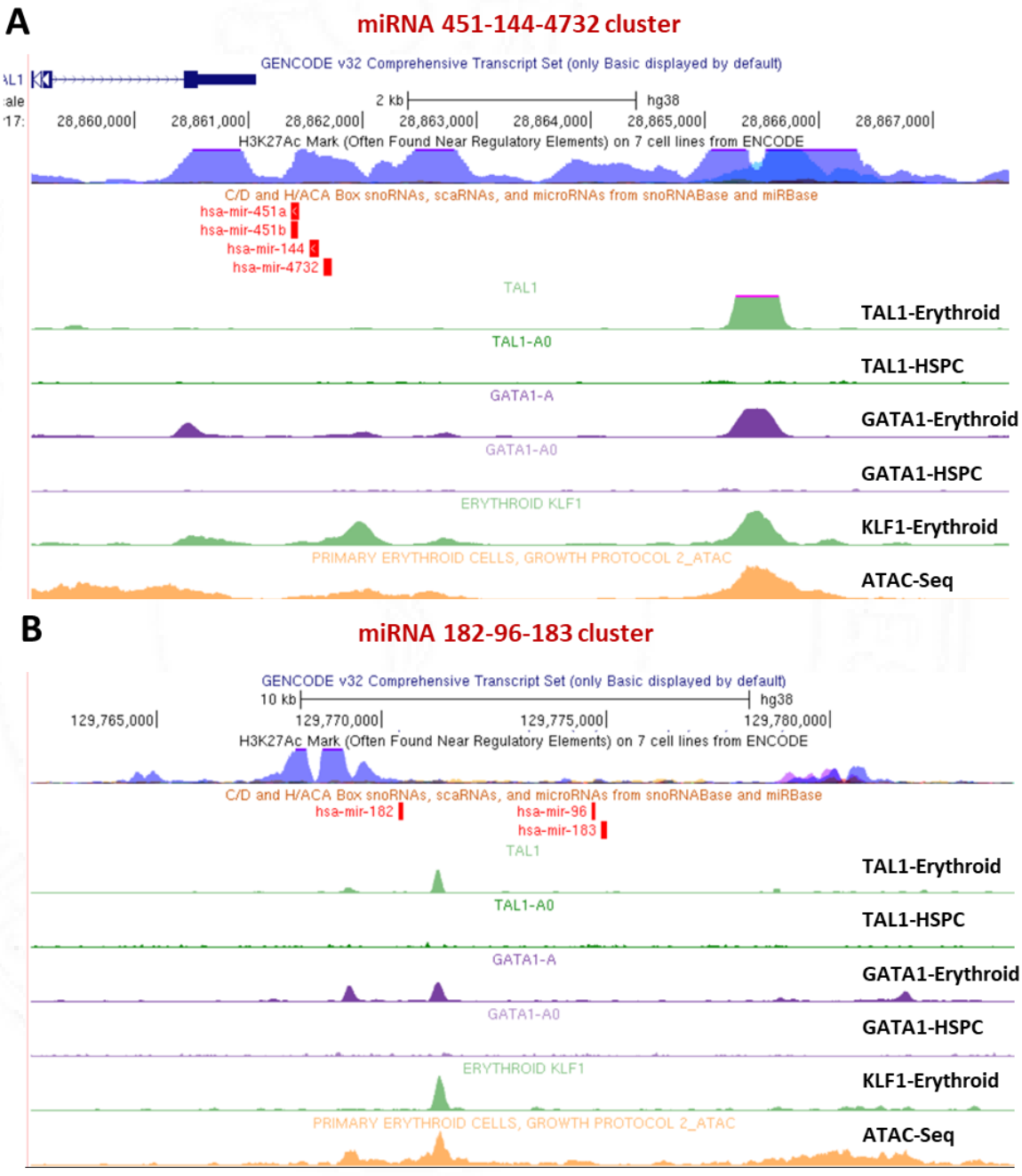
Table 4.5.2. New miRNA clusters identified in erythropoiesis (miRNAs with log2FC>2 and <-2) were chosen as the partners of miRNA clusters)

miRNA clusters	Regulation
miR 125a-let7e-99b	Down
miR 183-182-96	Up
miR 483 cluster(8 miRNAs)	Down
miR-130b-301	Down
miR 143-145	Down
miR 192-194	Up
miR-181c-181d	Down
miR 143-145	Down

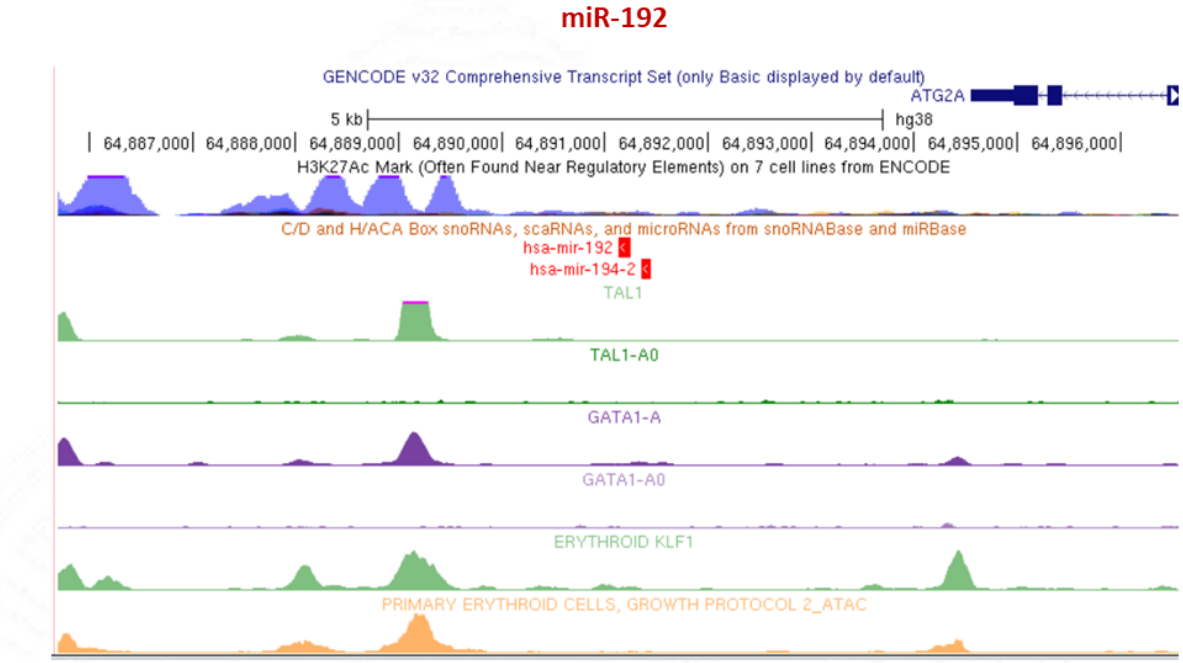
4.6 OCCUPANCY OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTORS NEAR THE UPREGULATED MIRNAS:

It has been reported earlier that the cell-specific transcription factors bring about the cell-specific expression of miRNAs. Analyzing the occupancy of erythroid-specific transcription factors near the genomic loci of the upregulated miRNAs can help in identifying bonafide erythroid-specific miRNAs during erythropoiesis. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) helps in assessing chromatin accessibility and in the identification of promoters and enhancers. We analyzed the publicly available ChiP-Seq and ATAC-Seq data from the undifferentiated CD34+ cells and differentiated proerythroblasts for the erythroid transcription factors (Huang *et al.*, 2016). We analyzed GATA-1, EKLF (KLF1) and TAL1 at the regions <5kb downstream or upstream of 12 upregulated miRNAs in the *ex-vivo* erythropoiesis. We observed almost all the upregulated miRNA genes have binding sites of these transcription factors with high levels of chromatin accessibility, either upstream or downstream of their transcription start sites (**Figure 4.6.1**). We compared the ChIP-Seq results in the erythroid cells and the differentiated erythroid progenitors to confirm that the ChIP peaks obtained are specific in the miRNA regions in the erythroid cells. We did not observe the occupancy of the erythroid transcription factors near the genomic regions of the miRNA genes that are downregulated during erythropoiesis. The upregulated miRNA clusters also showed significant occupancy of erythroid-specific transcription factors. The new miRNA cluster identified to be upregulated in human erythropoiesis, miR-182-96-183 cluster, had significant binding of the erythroid transcription factors, while undifferentiated

HSPCs. This data confirms that the specific cell-specific transcription factors mediate the expression of this cluster.



.C



D

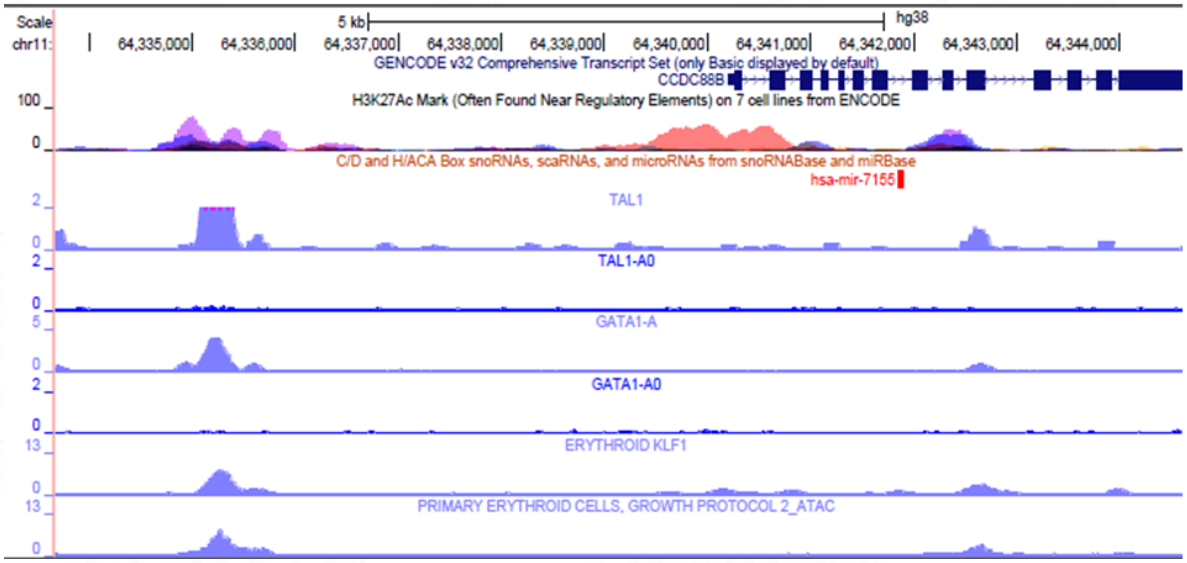


Figure 4.6.1: Occupancy of erythroid-specific transcription factors within 5kb of selected miRNAs.

4.7 KNOCKING OUT miRNAS USING CRISPR-Cas9:

Antisense technology using miRNA inhibitors that are synthetic oligonucleotides designed to specifically bind to and inhibit endogenous miRNAs has been widely used to functionally validate miRNAs. However, this method is not robust due to the off-target effects on miRNAs sharing similar seed sequences. miRNA sponges, which contain complementary sites to 'soak' miRNAs, also suffer from the specificity issue. Therefore, gene knockout (KO) would be an ideal choice for the deletion and study of an individual miRNA locus. Although the homologous recombination-based methodology has been conventionally used in animal models, it is not widely accessible, owing to the complicated procedure and limitations in the choice of organism.

More recently, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system has been used as a tool for targeted genome editing. Cas9 nuclease is directed by a single-guide RNA (sgRNA) consisting of a ~100nt guide sequence, which includes a scaffold sequence necessary for Cas9 binding and a 20-nt guide sequence which is complementary to the desired target site. Stable binding of the Cas9-sgRNA complex to a target sequence results in the creation of a DNA double-strand break (DSB) 3 bp upstream of the PAM sequence. This DSB can be repaired by the cell's normal DNA repair mechanisms, including homology-directed repair and non-homologous end joining (NHEJ), which results in insertions or deletions (in-dels) at the cleavage site. CRISPR-Cas9 system has also been used on non-coding genomic elements, including regulatory regions and non-coding RNAs. The CRISPR system can be employed to inhibit microRNA expression by destroying

the loop region, the Dicer processing site, or the Drosha processing site in a specific microRNA gene. However, this method has several drawbacks. It is challenging to find photospacer adjacent motif (PAM) sequences for gRNA recognition near the biogenesis sites or the stem-loop regions of the interested miRNAs due to a limited design space (60-90 bp) to find PAM sequences for gRNA recognition, without significant off-target effects of gene editing. Also, Cas9-induced small in-dels within the stem-loop of the microRNA may not be disruptive enough to knock out the function of the microRNA effectively.

To validate the function of miRNAs, we used gene editing of the target miRNAs by CRISPR-Cas9. Our initial experiments in CD34+ HSPCs showed that efficient transduction of these cells with the lentiviral vectors expressing Cas9 and a fluorescence marker is challenging as we obtained very low expression of Cas9 and the coexpressed fluorescence markers in multiple attempts. Therefore, we used HUDEP (Kurita *et al.*, 2013), a human erythroid cell line, for our experiments to evaluate the roles of miRNAs in erythropoiesis. These cell lines can be differentiated to late stages of erythropoiesis robustly and have been extensively used for the analysis of the genes involved in human erythropoiesis(Velez *et al.*, 2018)(Ch, 2016). For successful knock out of miRNA genes, we generated HUDEP cells expressing Cas9 using a Cas9- lentiviral vector. We generated lentiviral vectors with paired gRNAs that target miR-144, miR-451, miR-182, miR-183 and miR-96. The gRNAs were designed to target the miRNA biogenesis sites as it has been shown earlier that targeting the miRNA biogenesis sites effectively reduces their expression. As the gRNA expressing vector also expressed GFP, GFP+ cells were flow-sorted,

and they were cultured for a total of 10 days. The T7-endonuclease assay showed efficient in-del formation at the targeted miRNA regions (**Figure 4.7.1)**.

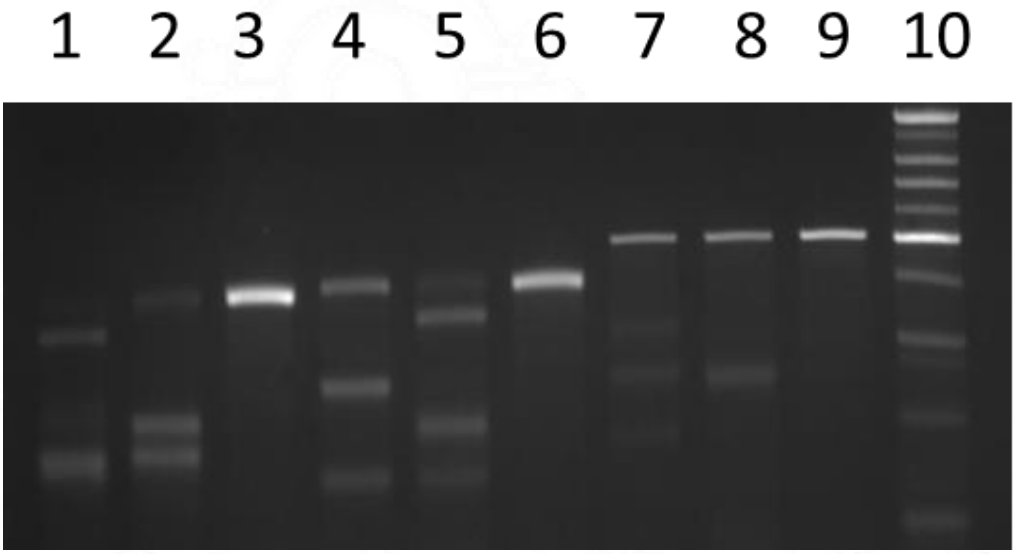


Figure 4.7.1. T7E analysis of the PCR products from miR-182 (lanes 1-3), miR-183 (lanes 4-6) and miR-96 (lanes 7-9) genomic regions. Lanes 1, 2, 4, 5, 7 and 9 contain the PCR products after T7E treatment and 3, 6 and 9 contain the PCR products without T7E treatment.

The edited HUDEP cells were differentiated to the cells at the later stages of erythropoiesis, using a protocol described recently. The cells at different stages of erythropoiesis were collected, and the effect of the knock out of the target miRNAs in the late-stage differentiation of the erythroid cells was assessed by performing flow cytometry for CD71 and CD235a expression. During late-stage differentiation, the erythroid cells show reduced expression of CD71, and, therefore, flow cytometry analysis of this property can be used for evaluating the differentiation of erythroid cells. We observed that the HUDEP cells with the gene-edited for the miRNAs, miR-144 and miR-451 showed reduced differentiation while gene editing of miR-4732, miR-182 and miR-183 did not affect differentiation (**Figure 4.7.2**), suggesting that

knocking out of these miRNAs impairs erythroid differentiation. miR 144 and miR-451 have been shown to have critical cellular functions, and knocking out of these miRNAs to understand its effect has not been studied earlier.

This is the first study that evaluated miRNAs by functionally disrupting them using Crispr-Cas9. Mir-182-183-96 cluster, which has been identified to be upregulated during erythropoiesis in this study, has not been reported earlier in erythropoiesis. Using gene editing, we could confirm that this miRNA cluster is not involved in erythroid differentiation as the HUDEP cells in which these miRNAs were knocked out did not show any effect in their differentiation. Altogether, these experiments showed that gene editing of miRNAs is very efficient using CRISPR-Cas9, and it can be used for evaluating the roles of miRNAs in erythropoiesis.

4.8 OVEREXPRESSION OF UPREGULATED AND DOWNREGULATED miRNAS DID NOT CHANGE THE KINETICS OF ERYTHROPOIESIS

To gain insights into the functions of the upregulated and the downregulated miRNAs, a miRNA overexpression system was established using the pZIP-hCMV-Zsgreen-Puro lentiviral vector system. Transduction of $CD34^+$ cells with miRNA overexpression vector did not result in a significant increase in the miRNA levels since these miRNAs had high levels of endogenous expression. In addition to this, when the transduced $CD34^+$ cells were differentiated towards the erythroid lineage, no significant difference was observed in the kinetics of erythropoiesis **(Figure 4.8)** thus suggesting that the overexpression system was not robust enough to carry out

functional studies in CD34^{+} cells. This also indicated the importance of a particular cell type in carrying out specific functional studies.

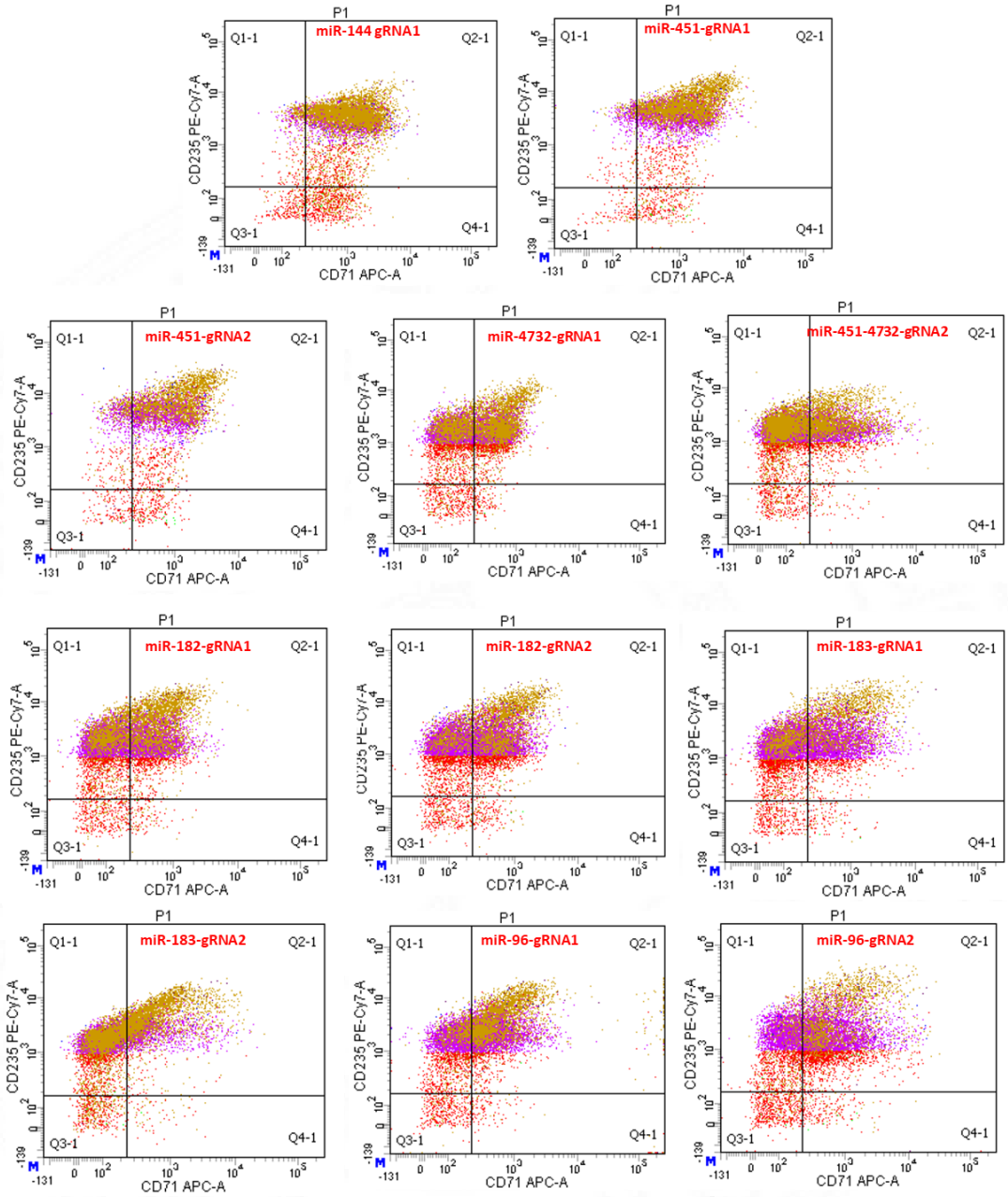


Figure 4.7.2 Flow cytometry analysis of CD71 and CD235a in the differentiated HUDEP cells in which miR-144; miR-451, miR-182, miR-183, miR-4732 and miR-96 were knocked out using CRISPR-Cas9 gene-editing method.

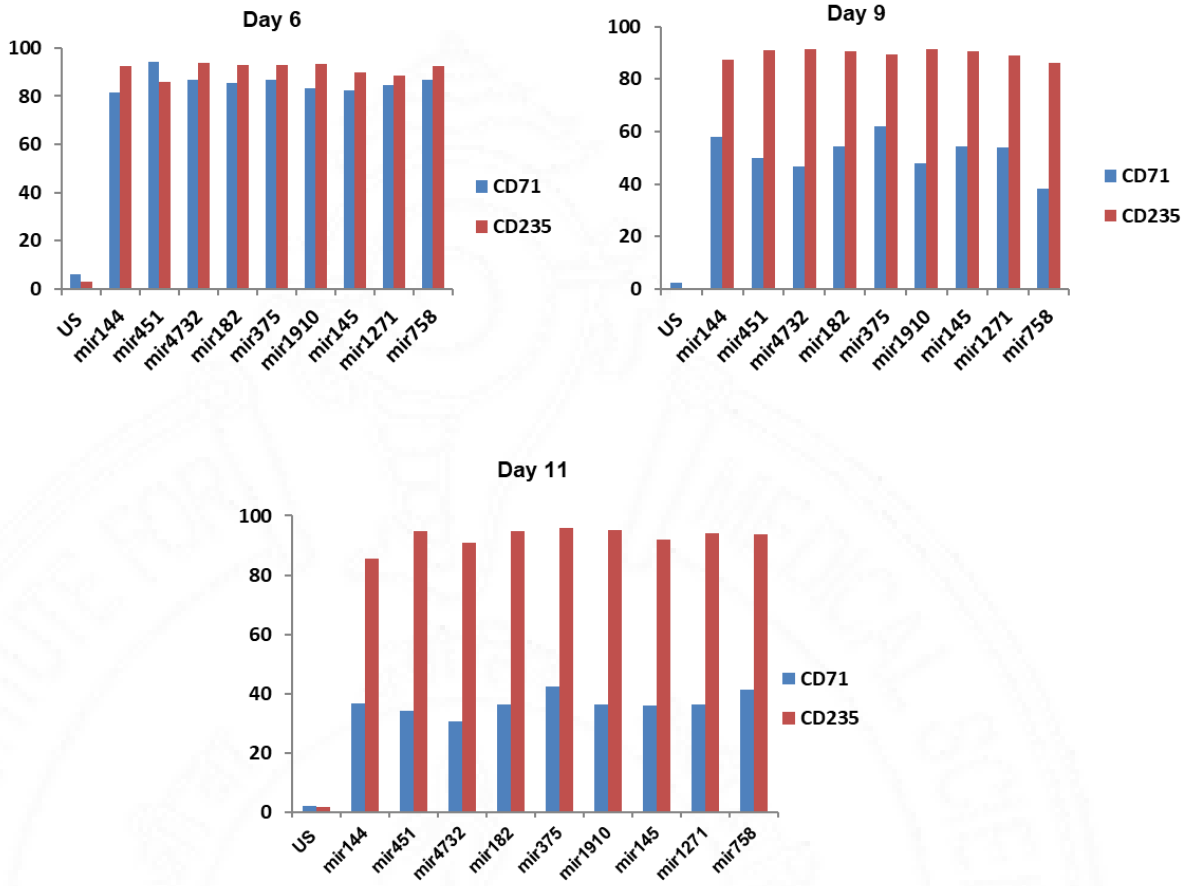


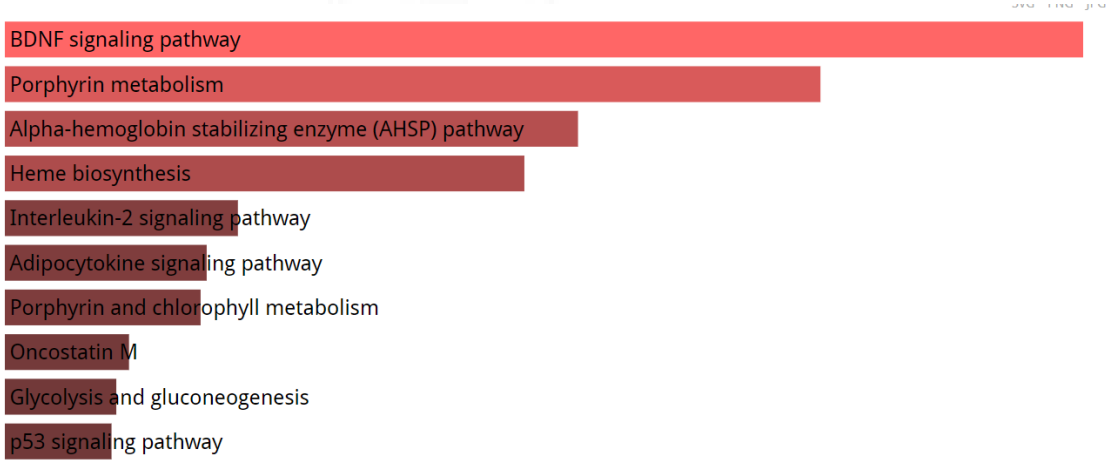
Figure 4.8.1 Expression levels of CD71 and CD235a in the cultured erythroid cells from different days of erythroid differentiation, days 6, 9 and 11. $CD34^{+}$ cells were transduced with lentiviral vectors to express miR-144, 451, 4732, 182, 375, 1910,145, 1271 and 758 and differentiated to erythroid cells.

4.9 MIRNA TARGET ANALYSIS

We analyzed the predicted targets of the miRNAs, which showed significant difference in their expression during erythropoiesis. We first analyzed the targets of the miRNAs that were downregulated. We found that 540 target genes that were upregulated during erythropoiesis. Pathway analysis was performed using Enrichr (https://maayanlab.cloud/Enrichr/). The Bioplanet pathway results showed that the downregulated miRNAs activate the genes in the BDNF and porphyrin metabolism pathways predominantly (**Figure 4.9.1 A**). Wikipathway results showed that heme

biosynthesis, unfolded protein response and NRF2 pathways are upregulated (**Figure 4.9.1 B**).

A



B

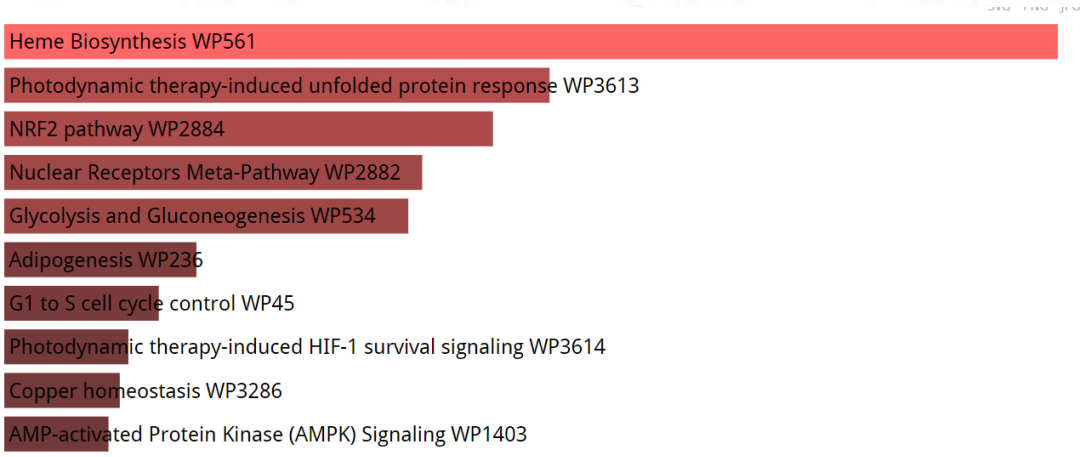


Figure 4.9.1 Pathways activated by the miRNAs in human erythropoiesis. Results obtained from (A) Bioplanet pathways and (B) Wikipathways are shown.

Brain-derived neurotrophic factor (BDNF) and its receptor, TrkB, are reported to be important for development of adult brain (Kowiański *et al.*). This signalling pathway is critical for neuronal survival and morphogenesis. Binding of BDNF to TrkB activates MAPK/ERK and phosphoinositide 3-kinase (PI3K) pathways. This pathway has not been reported earlier in haematopoiesis or erythropoiesis. Through the miRNAs and their

target genes we identified a new pathway involved in human erythropoiesis. Porphyrins are precursors of haem and they are an essential constituent of haemoglobins in the red cells. It is known that the porphyrin metabolism is upregulated during erythropoiesis. This analysis shows that miRNAs play an important role in porphyrin metabolism in the red cells. Wikipathway analysis also confirmed that the haem biosynthesis pathway is regulated by the differentially regulated miRNAs in erythropoiesis.

4.10 DISCUSSION

Identification of the differentially regulated miRNAs during a cellular process provides insights into the transcriptional regulatory mechanisms that control the expression of various proteins required for cell fate. Small RNA sequencing analysis helps in the comprehensive analysis of the known and novel miRNAs in different cells. Ex-vivo erythropoiesis of HSPCs has been shown to mimic in-vivo erythropoiesis with progressive erythroid differentiation with the characteristic surface marker expression and morphology changes. Using this cellular model, we identified several miRNAs that are involved in human erythropoiesis. This is the first comprehensive screening of miRNAs during human *ex-vivo* erythropoiesis, and the findings from this study would be of great significance in understanding various regulatory transcriptional mechanisms mediated by miRNAs during erythropoiesis.

This study also identified many miRNA-clusters that are coregulated during erythroid differentiation, and several miRNAs were found to be intragenic, and their expression correlated with the expression of the host genes. We performed further

analysis to study the occupancy of the erythroid-specific transcription factors, GATA1, KLF1 and TAL1. We found that these transcription factors regulate the expression of 90% of the upregulated miRNAs.

Robust, specific and stable strategies for miRNA silencing are essential for studying the functional role of miRNAs in human diseases. The major loss-of-function technologies in miRNA studies include miRNA specific antisense inhibitors, miRNA sponges, and genetic knockout. However, it is known that the degradation of miRNAs by antisense inhibitors rarely occurs, given the short length. Therefore, even though the miRNA loses the function, the mature miRNA sequences can still be detected by qRT-PCR, resulting in low or no correlations between the functional phenotype resulting from antisense inhibitors and the quantification of miRNA expression. This pitfall may cause difficulty in the accurate measurement of antisense efficiency and potential toxicity. Besides, antisense inhibitors are limited to only short term studies because of irreproducibility after transfection. miRNA sponges have tandem binding sites that can compete with specific targets dampening miRNA repressive function, which has shown the long term stability and improved efficiency in miRNA loss-of-function studies. However, the miRNA sponge is vulnerable to Ago2-mediated endonucleolytic cleavage that limits its application in the selected models only. In addition, the expression efficiency of miRNA sponge constructs and the total number of binding sites contained in the sponge will decide the silencing efficiency. Genetic knockout of miRNA is the most reliable technique to study loss-of-function of miRNAs with high efficiency and specificity for in vivo analysis; however, the manipulation procedure is complicated, and the process is

time-consuming. Therefore, to establish a robust and convenient technique in the study of miRNA, loss-of-function is in need. There are very few studies carried out so far to evaluate the differentially expressed miRNAs functionally. Using the strategy that we adopted, it was observed that miRNAs could be effectively disrupted for assessing their role in cellular processes. We could confirm that the previously reported miRNAs, miR-144 and miR-451, are required for erythropoiesis by editing these miRNAs.

5. RNAi SCREEN TO DELINEATE NOVEL SIGNALLING REGULATORS IN HSC MAINTENANCE AND ERYTHROPOIESIS

5. 1. INTRODUCTION

The process of erythropoiesis initiates in the bone marrow from haematopoietic stem cells (HSCs) after which it traverses through several distinct stages to form mature reticulocytes (An *et al.*, 2014). This complex process of erythropoiesis is tightly regulated by transcription factors, epigenetic modifiers, non-coding RNAs, as well as signalling proteins (Hattangadi *et al.*, 2011). Several studies carried out previously have highlighted the role of several signalling pathways which include Epo dependent signalling pathways such as PI3-AKT (Jafari *et al.*, 2019), MAP kinase, STAT5, protein kinase C (Ingley, 2012), FGF2R3 (Van Vuren *et al.*, 2019) and IGF-1 signalling (Kadri *et al.*, 2015) pathways as well as the Epo independent signalling pathways such as the TGF-β signalling pathway (Zermati *et al.*, 2000).

It is possible that several other signalling pathways might also be involved in the differentiation of HSCs into the erythroid lineage, and they can be identified by comprehensive screening methods. Previous studies have used cell lines, murine models and cord blood $CD34^+$ cells (Ingley, 2012). With the emergence of high throughput technologies such as the RNAi screens, it is possible to identify novel regulators in any biological process. The technology of the RNAi screen has led to the identification of cohesion genes as one of the key players in stem cell renewal and differentiation (Galeev *et al.*, 2016). Furthermore, PPARγ was also identified as a crucial regulator of haematopoietic stem cell homeostasis (Sertorio *et al.*, 2017).

Therefore, given the strength of this technology in deciphering novel regulators, this can be used to identify novel signalling regulators in HSC maintenance and differentiation. This library was then used in a robust *ex - vivo* erythropoiesis system, which mimics *in - vivo* erythropoiesis, to identify the novel signalling regulators in erythropoiesis.

The following experiments were performed to understand the signalling pathways involved in the maintenance of HSPC and erythropoiesis. $CD34^+$ haematopoietic stem and progenitor cells (HSPCs) were transduced with a shRNA library that targets 5000 signalling pathway genes, and the cells were differentiated by *ex-vivo* erythropoiesis. Bioinformatics analysis was performed to identify the enriched and depleted shRNAs.

5.2 PRIMARY shRNA LIBRARY SCREENING TO IDENTIFY NOVEL SIGNALLING REGULATORS IN ERYTHROPOIESIS

To delineate the novel signalling regulators in HSC differentiation, a pooled lentiviral shRNA screen was performed in adult human $CD34^+$ cells. A lentiviral library consisting of 27,500 shRNAs targeting 5000 human signalling genes were used which was purchased from Addgene as pooled lentiviral plasmids. In this library, the shRNAs were cloned in a lentiviral plasmid which uses a RNA Polymerase III promoter (U6) (**Figure 5.2.1**). A total of 6.4 X 10^7 $CD34^+$ cells were transduced with the library to obtain a high complexity and representation of the library. Transduction efficiency of ~30% was used to avoid multiple shRNA integrations within a single cell. Following transduction, the cells were grown for 18

days in a serum-free medium which was supplemented with growth factors to facilitate *ex-vivo* erythropoiesis (**Figure 5.2.2**).

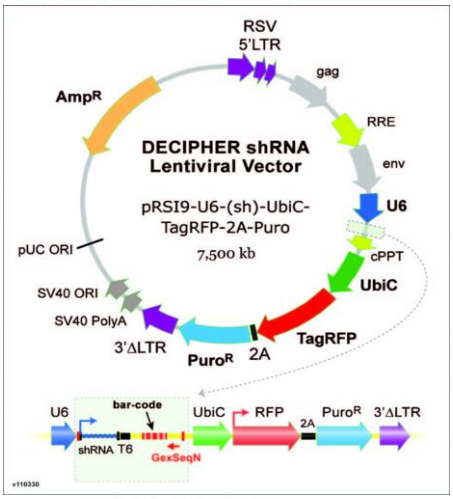


Figure 5.2.1. Pictorial map of the lentiviral vector used for the RNAi library for signalling pathway genes.

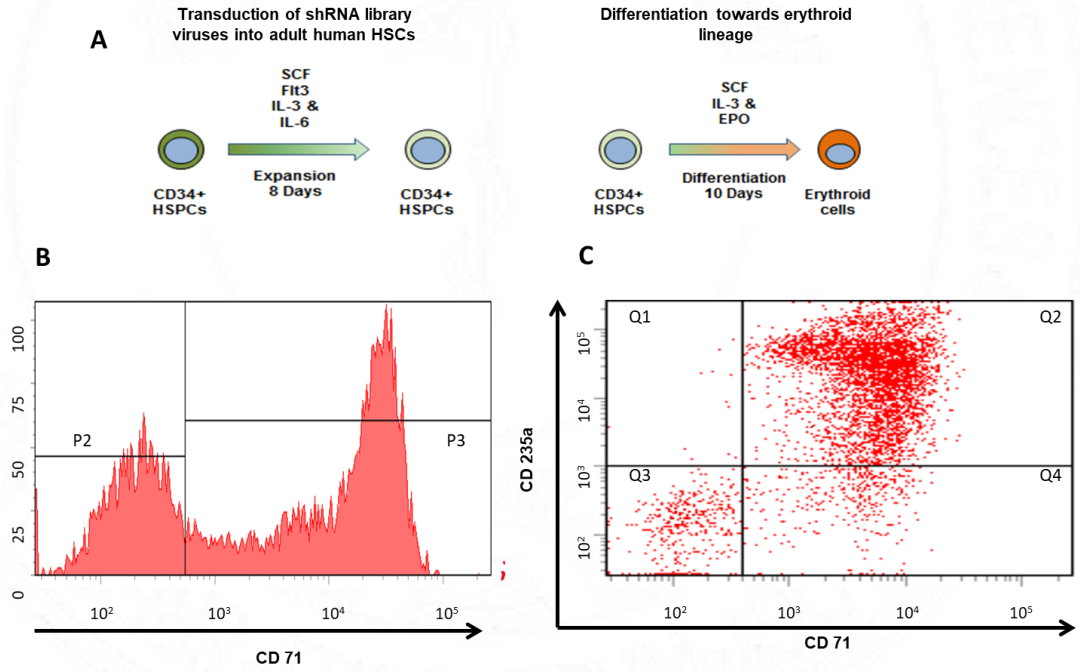


Figure 5.2.2: (**A**) Schematic of ex-vivo erythropoiesis protocol (**B**) Expression of CD71 (early erythropoiesis marker) on day 7 of the culture and (**C**) expression of CD71 and CD235a (late stage marker) on day 9.

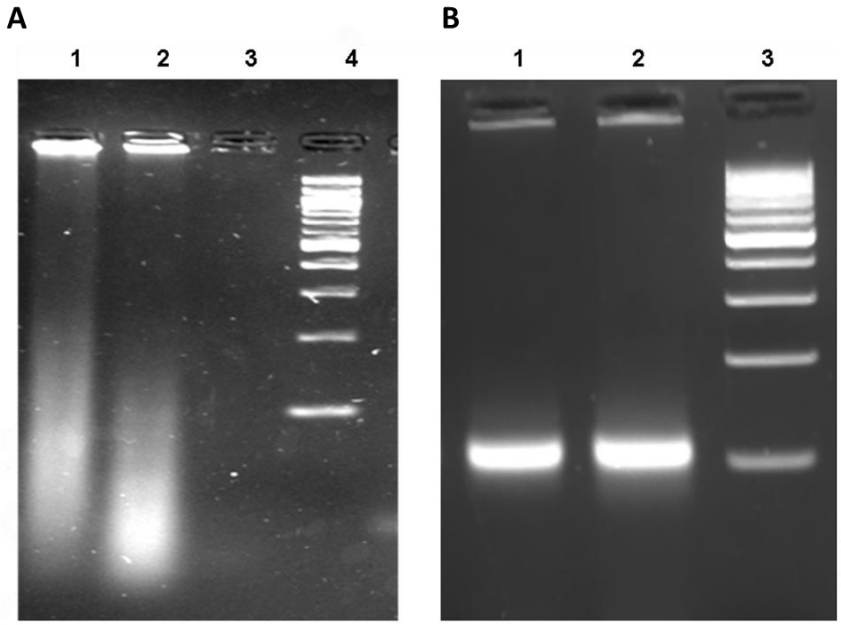


Figure 5.2.3: Amplification of DNA with specific primers for NGS. (A) Lane 1: from the day 5 of phase 1 of the culture, Lane 2: from the day 9 of phase 2 of the culture, Lane 3: Negative control and Lane 4:100bp (B) Amplification of first round PCR product with specific primers. Lane 1: obtained from the 1st round PCR product of the DNA of the cells from day 5 of phase 1 of the culture, Lane 2: obtained from 1st round PCR product of the DNA of the cells from day 9 of the phase 2 of the culture. Lane 3:100bp molecular weight marker.

Cells were then collected for DNA before and after differentiation, and they were amplified in two rounds of PCRs to enrich the shRNA barcodes (**Figure 5.2.3**). The PCR products were sequenced using Illumina Next Seq, and the bioinformatics analysis was carried out to identify the enriched and depleted shRNAs. The barcode reads from the DNA extracted from the cells on day 9 in phase II of the culture were normalized with those obtained from the cells on day 4 in Phase I of the culture. Bioinformatics analysis results showed that there were 25 shRNAs that were enriched (Fold Change FC>3) and 1 shRNA that was depleted (**Figure 5.2.4**).

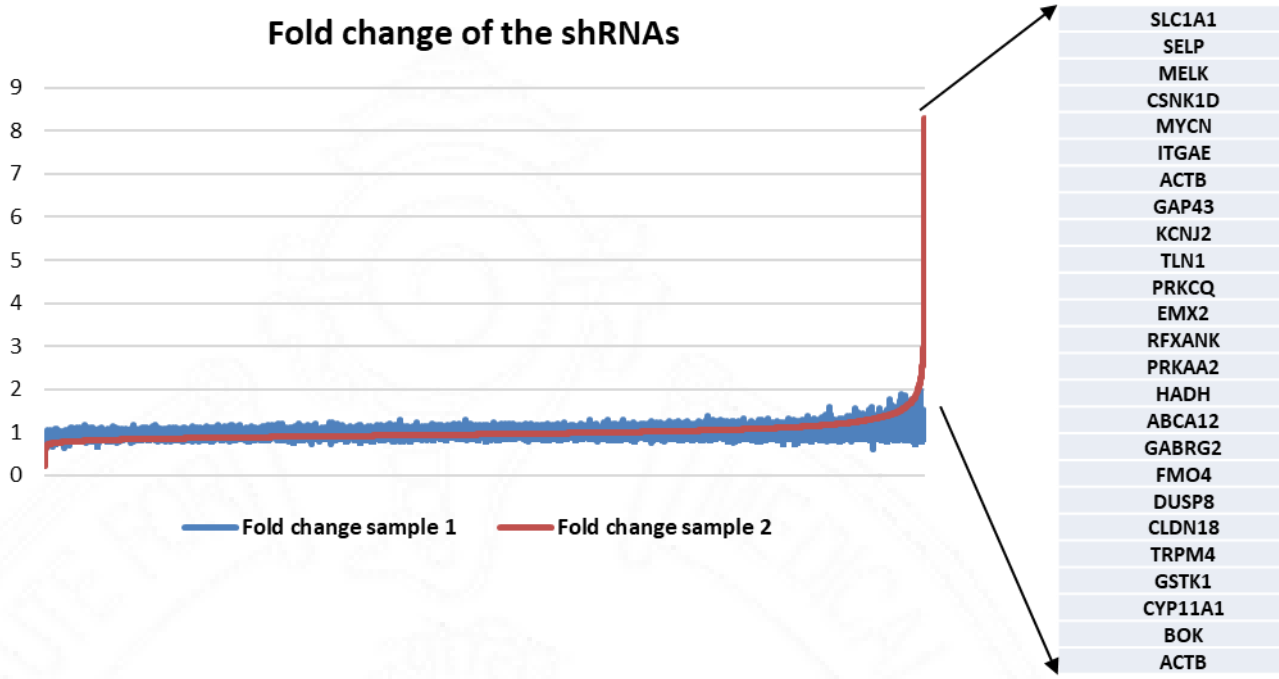


Figure 5.2.4: Graph showing the enrichment of the shRNAs in time point 2 (Phase 2 Day 9) as compared to time point 1 (Phase 1 Day 4) and top 25 genes as identified by the RNAi screen.

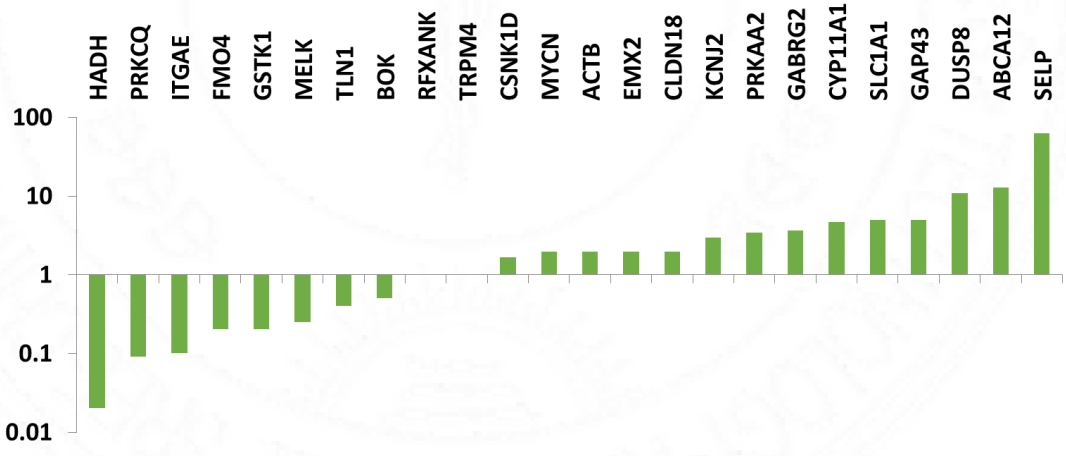


Figure 5.2.5: Expression levels in ex-vivo erythropoiesis of the top hit genes identified by the RNAi screen

5.3 GENES PREVIOUSLY REPORTED IN HAEMATOPOIESIS AND ERYTHROPOIESIS IDENTIFIED BY RNAi SCREENING IN EX-VIVO ERYTHROPOIESIS

From the genes obtained after bioinformatics analysis of the NGS data, top 25 genes were selected for further analysis. The expression of these genes was analysed in the erythroid cells from published datasets, and it was found that almost all the genes were expressed in the erythroid cells (**Figure 5.2.5**). These genes were then evaluated for their role both in haematopoiesis/erythropoiesis and out of the 25 genes, 11 genes were found to have a role in mouse/human erythropoiesis as listed below:

MELK: MELK is a serine/threonine kinase which is significantly expressed in the thymus and bone thus indicating a role for MELK in haematopoiesis. Functional studies carried out using morpholinos to disrupt the function of MELK like gene in zebrafish models led to deferred circulation of blood which was accompanied with the diminished levels of haematopoietic transcription factors such as SCL,GATA1 and LMO2 (Saito *et al.*, 2005). On the contrary, overexpression of the MELK like gene was instrumental in rescuing the anaemia like phenotype which was caused due to abrogation of the MELK like gene. These findings were indicative of the fact that the expression of MELK gene is critical for the fine tuning of several transcription factors which played a key role in haematopoiesis (Saito *et al.*, 2005).

CSNK1D: CSNK1D plays a crucial role in phosphorylating HIF-2α at Ser383 and Thr528 and destruction of these phosphorylation sites leads to a reduction in the

levels of target genes of HIF-2 which further causes reduced production of erythropoietin especially in cancer lines (Volker K Haase, 2014). In addition to this, inhibition of phosphorylation caused significant mislocalisation of HIF-2α (Pangou *et al.*, 2016). Thus these findings suggested that CSNK1D augments erythropoietin secretion in hypoxic environment by stimulating the nuclear accumulation of HIF-2α and this phenomenon might play a key role in increasing the levels of erythropoietin during stress erythropoiesis (Volker K Haase, 2014).

MYCN: MYCN is significantly expressed in the long term haematopoietic stem cells (LT-HSCs) and progenitors and its levels decline during the process of cellular differentiation (Dolores Delgado and León, 2010). Loss of function studies carried out in mouse embryos demonstrate multiple abnormalities with defective haematopoiesis and angiogenesis being the most dominant ones and embryonic lethality is observed beyond embryonic day 10.5. Such phenotype has also been witnessed in other vertebrate models. In addition to this, MYCN also impedes monocytic, granulocytic and erythroid differentiation. These roles exhibited by MYCN suggest that it plays a significant role in the self-renewal and differentiation of HSCs.

ACTB: Actins are the most extensively conserved proteins and are involved in various stages of development. Recent studies have led to the identification of actin localization in the nucleus thus suggesting its role in the regulation of transcription. Functional studies carried out in heterozygous beta actin knock out mouse to delineate the role of beta actin during development showed that there was a drastic decline in the number of primitive erythroid cells as well as blood islands which

serve as the main hub for primitive erythropoiesis. These findings were concordant with the diminished levels of GATA-2 which is an important transcription factor involved in erythropoiesis. Thus these results indicated that beta actin played a pivotal role in regulating the process of erythropoiesis by tweaking the levels of GATA 2 in mouse embryo (Tondeleir *et al.*, 2013).

SELP: SELP also known as P-Selectin belongs to a family of tri transmembrane calcium dependent lectins. It is mostly expressed by the megakaryocytes, platelets, and is involved in rolling of leukocytes which expresses its transmembrane ligand PSGL-1. P-selectin in conjugation with E-selectin plays a crucial role in the recruitment of HSPCs to the bone marrow. Similarly, in association with megakaryocytes, P-selectin also plays an important role in the attachment of HSCs in the bone marrow (Levesque and Winkler, 2016).

PRKAA2/PRKAA1: Protein Kinase AMP-activated catalytic subunit α-2 (PRKAA2) is an important gene that plays a crucial role in regulating cellular energy metabolism as well as cell growth and proliferation. PRKAA2 has been found to have two roles in the context of haematopoiesis. PRKAA2 is highly expressed in HSCs and is Foxo3a dependent. Thus, HSCs may be particularly poised to activate AMPK in an environment of reduced cellular energy, hence inactivating mTOR and rapidly driving the induction of autophagy. In support of this hypothesis, genetic mouse models that are either deficient in FOXO family members or display constitutive mTOR activation both result in HSC depletion and loss of function. AMP-activated protein kinase α1 (PRKAA1) regulates erythrocyte homeostasis by mediating the selective degradation of mitochondria by mitophagy. Disruption of the function of PRKAA1 leads to a decline in the phosphorylation levels of ULK1, which further restrains autophagy-dependent mitochondrial clearance and causes

accumulation of mitochondria. This leads to the extirpation of red blood cells via assimilation of excessive ROS.PRKAA1 knockout mice exhibit splenomegaly and anaemia (Zhu *et al.*, 2014).

***Glutathione S Transferase Kappa 1 (GSTK1)*:** The role of GSTK1 has not been studied widely in the case of haematopoiesis. However, it has been shown that GSTK1 is expressed mostly on LT-HSCs and it renders LT-HSC resistant to stress which is essential for long-lived cells such as stem cells (Camilla Forsberg *et al.*, 2005).

BOK: BCL2 Ovarian Killer (BOK) belongs to the BCL2 family and behaves as anti or pro-apoptotic regulators that are involved in a wide variety of cellular processes. It is mostly an apoptotic regulator that functions through different apoptotic signalling pathways. It is mostly involved in regulating stress erythropoiesis (Kang *et al.*, 2017).

5.4 NOVEL GENES IDENTIFIED BY RNAi SCREENING

In addition to the genes which were found to play a role in erythropoiesis/haematopoiesis, several other genes such as SLC1A1, GAP43, TLN1, DUSP8, and TRPM4 were also identified whose roles were not delineated in the context of erythropoiesis/haematopoiesis.

SLC1A1: The SLC1A1 gene codes for the excitatory amino acid transporter 3 (EAAT3) protein which belongs to a class of excitatory amino acid transporter family and is significantly expressed in the hippocampus. It is involved in three major biological functions such as in the absorption of synaptic glutamate and

obstructing the glutamate spillover from active synapses to extrasynaptic regions, making glutamate available for synthesis of gamma-aminobutyric acid (GABA) synthesis and neuronal absorption of cysteine, which is one of the critical substrates for the synthesis of intracellular glutathione (GSH) in neurons (Afshari, Yao and Middleton, 2017). Glutathione is one of the essential antioxidants which play a crucial role in the metabolism of reactive oxygen species (ROS), thus protecting cells from oxidative stress. Both reduced and oxidised forms of glutathione work in conjugation with other redox compounds to modulate the levels of cellular redox compounds (Lushchak, 2012). With respect to erythroid cells, it has been observed that GSH exercises a pivotal role in the slowing down of the ageing process in erythrocytes. Its role becomes more profound when either normal or diseased erythrocytes come in contact with aromatic catalysts. It helps in the sequestration of the reactive oxygen species which are produced in vast amounts and protects cells from permanent damage by producing methaemoglobin which acts as a buffer against the damage (Datta and Kontomichalou, 1965). Recent reports have also shown that Diamond Blackfan anaemia patients show elevated levels of reduced glutathione which can serve as a novel biomarker for the diagnosis of DBA (Utsugisawa *et al.*, 2016). These findings suggest that SLC1A1 might also play a role in erythropoiesis, considering its role in the synthesis of glutathione.

GAP43: This protein, also known as neuromodulin, mostly plays a role in the development and generation of neurons and also contributes to memory formation and learning. In addition to this, it associates with heterotrimeric G proteins and is involved in G protein-coupled receptor transduction and is also one of the primary

substrates for protein kinase C (Denny, 2006). Protein Kinase C has also been recognised to play an essential role in the formation of platelets, especially in the secretion of granules and thrombosis by providing upstream signals to its downstream partner PKD2. Mice in which key phosphorylation sites of PKD2 have been mutated exhibit defects in the assemblage of platelets and formation of thrombus. These findings suggest that GAP43 might play an indirect role in the process of megakaryopoiesis (Konopatskaya *et al.*, 2011).

TLN1: Talin1 is one of the isoforms of a family of proteins which are characterised by the presence of an N-terminal head domain and a long flexible rod domain that can be separated from each other by calpain 2 cleavage. Talin1 predominantly functions to perpetuate migration of cells and cell spreading. Gene knock-out studies in mice have further substantiated the role of Talin1 in embryonic development and platelet formation. Furthermore, it has also been shown that the interaction of Talin1 and integrin is indispensable for the process of angiogenesis, thus emphasising its role on the process of haematopoiesis(Monkley *et al.*, 2011).

DUSP8: Dual-specificity phosphatases (DUSPs) belong to a class of protein tyrosine phosphatases that are involved primarily in dephosphorylating the residues of phosphor-serine/threonine and phosphor-tyrosine on mitogen-activated protein kinases (MAPKs). DUSP8 particularly plays a significant role in the signal transduction of MAPK and mediates cell oxidative stress response as well as cell apoptosis. DUSP8 is involved in the dephosphorylation of JNK, p38 MAPK and ERKs which are the terminal kinases of the MAPK pathway. This dephosphorylation impedes continuous phosphorylation and activation, which can otherwise lead to

severe abnormalities(Ding *et al.*, 2019).MAPK is one of the key signalling pathways playing a role in the survival, proliferation and differentiation of erythroid cells and hence it's regulation by DUSP8 might also be essential to curbing the development of haemoglobinopathies (Hodges *et al.*, 2007).

TRPM4: TRPM4 (Transient Receptor Potential member 4) belongs to the group of proteins which are unbiased cation channels and are activated by calcium and penetrable only to monovalent ions such as sodium and potassium. They are ubiquitously expressed in the majority of human tissues and play a role in several physiological processes such as in T cell activation and allergic reactions (Cho *et al.*, 2015). Furthermore, it has also been shown that TRPM4 augments cell proliferation by up-regulating the β-catenin signalling pathway (Armisén *et al.*, 2011). This function of TRPM4 might indirectly contribute to vertebrate haematopoiesis since several studies have shown the importance of the canonical Wnt signalling pathway in primitive haematopoiesis (Luis *et al.*, 2011) as well as in the development and maintenance of haematopoietic stem cells (Luis *et al.*, 2012).

The summary on the genes that showed fold change >3 is shown in the **Table 5.1**.

5.5 DISCUSSION

Erythropoiesis is a process in which haematopoietic stem cells give rise to mature erythrocytes by traversing through several stages with each stage being distinct from the other. The different stages of erythropoiesis can be reiterated in vitro which is characterised by changes in cell size and morphology thus making it an efficient system to carry out high throughput transcriptome studies(Shi *et al.*, 2014). Several

high throughput transcriptome studies carried out using ex vivo erythroid culture systems has led to the identification of several novel coding and non-coding transcripts (Shi *et al.*, 2014, Ding *et al.*, 2016, Doss *et al.*, 2015). In addition to this, it has also served as an essential tool to carry out a comparative analysis between mouse and human erythropoiesis (An *et al.*, 2014) as well as in decoding the transcriptional regulatory networks in foetal and adult erythroblasts (Lessard *et al.*, 2018). Thus these findings indicate the effectiveness of the system in carrying out transcriptional studies. In our study, we used an RNAi screening approach in a well-established *ex vivo* erythroid culture system to delineate the signalling pathways involved in the process of erythroid differentiation. Such an approach has also been used previously in the mouse as well as in human $CD34^{+}$ haematopoietic stem cells to identify novel genes involved in stem cell maintenance and renewal respectively (Kinkel *et al.*, 2015, Galeev *et al.*, 2016, Hope *et al.*, 2010). However, no such studies have been carried out earlier in an erythroid differentiation model to delineate the different signalling pathways involved in erythropoiesis. We found several genes which were already reported to have a role in erythropoiesis/haematopoiesis as well as genes whose roles were not identified in the context of erythropoiesis/haematopoiesis. These findings validated the RNAi screening approach and the erythroid culture system, which was used for the study. Such an approach stands better as well as cost-effective against mouse models since a vast repertoire of genes including housekeeping genes and genes for survival can be analysed simultaneously while avoiding the tedious maintenance and analysis of animals

Table 5.1. The summary of the genes that showed fold change >3 and their functions in erythropoiesis.

Gene	Fold change sample in the RNAi screen	Expression levels (fold change)	Role in erythropoiesis/haematopoiesis	References
MYCN	4.44	2	Maintains balance between self-renewal and differentiation of HSCs	Delgado *et al*,2010
CSNK1D	5.04	1.7	HIF-2α phosphorylation by CK1δ promotes erythropoietin secretion in liver cancer cells under hypoxia	Liakos *et al*,2016
MELK	5.13	0.25	Affects haematopoiesis in zebra fish (elaborate)	Saito *et al*,2005
BOK	3.12	0.5	Promotes erythropoiesis in a mouse model of myelodysplastic syndrome.	SG *et al*,2019
ACTB	4.12	2	Modulation of erythropoiesis during development by fine tuning GATA2 expression levels	Tondeleir *et al*,2013
PRKAA2	3.43	3.5	Required for autophagy dependent mitochondrial clearance during erythrocyte maturation	Zhu *et al*,2014; Landes *et al*,2013
SLC1A1	8.30	5	Unknown	
SELP	5.39	64		
ITGAE	4.26	0.1		
GAP43	3.95	5		
KCNJ2	3.88	3		
TLN1	3.83	0.4		
PRKCQ	3.63	0.09		
EMX2	3.62	2		
RFXANK	3.48	1		
HADH	3.39	0.02		
ABCA12	3.38	13		
GABRG2	3.35	3.7		
FMO4	3.33	0.2		
DUSP8	3.31	11		
CLDN18	3.30	2		
TRPM4	3.27	1		
CYP11A1	3.13	4.7		

6 RNAi SCREEN TO DELINEATE NOVEL REGULATORS IN HSC MAINTENANCE AND ERYTHROPOIESIS

6.1 INTRODUCTION

Haematopoiesis is a process in which a specific population of multipotent stem cells designated as haematopoietic stem cells simultaneously self-renew and differentiate into multiple blood lineages at a rapid pace for the entire life span of a human being. Under normal homeostatic conditions, an adequate amount of blood cells must be generated without the stem cell population being exhausted. Although the mechanisms governing the balance between stem cell renewal and differentiation have been extensively studied, the underlying molecular cues influencing the maintenance of haematopoietic stem cells and its differentiation remain to be understood. One of the major challenges faced in this aspect is to dissect the molecular mechanisms of the connate ability of HSCs to self-renew and perpetuate their undifferentiated state. This has dampened the efforts to expand HSCs *ex- vivo* for therapeutic applications. Over the years, reverse genetic approaches have been used to elucidate the functional role of candidate genes. The advent of RNA interference (RNAi) has paved a highly effective way to delineate the functional aspects of hundreds of genes involved in both normal and malignant haematopoiesis using shRNA libraries (Galeev *et al.*, 2016).

It is possible to obtain the list of genes involved in a particular biological process by combining datasets from multiple cell types using high throughput computational approaches, which can be further screened simultaneously to delineate their role in a

specific cellular process. The mechanism of RNAi exploits the use of siRNAs as well as shRNAs to reduce the expression levels of target genes, and it is one of the promising methods as against gene knockouts since it provides a platform to analyze the essential genes *in- vivo* in model organisms without posing any risks towards their survival. Although short interfering (siRNAs) facilitate rapid gene knockdown, their transient nature makes it unsuitable for the long term *in- vivo* applications. However, this shortcoming has been overcome by stem-loop short hairpin RNAs (shRNAs), which provide an uninterrupted source of RNAi trigger when expressed from appropriate vectors. Nevertheless, shRNAs also exhibit several limitations such as imprecise processing, off-target effects and toxicities, which are mostly mediated through the saturation of the endogenous RNA polymerase II (RNAPII) machinery. Therefore, to establish an efficient shRNA expression vector that can give high levels of a knockdown, the architecture of shRNAs, as well as the vectors expressing them, has undergone several alterations over the years. shRNAs are short sequences of RNA that make tight hairpin turns and assist in silencing the target genes. Previously, they were mostly expressed by vectors harbouring RNA polymerase III (RNAPIII) promoters. shRNAs driven by RNAPIII promoters provided high levels of knockdown of target genes, but it also caused severe toxicities in most of the cell lines in which it was evaluated (Grimm *et al.,* 2006). To alleviate the toxic effects of shRNAs, which are driven by RNAPIII promoters, the concept of shRNA^{miR} evolved wherein the shRNA sequences are embedded within flanking regions, which are usually derived from endogenous miRNA precursors (Brendel *et al.*, 2016). The shRNA^{mirs} can be expressed from RNAPII promoters, thus enabling tissue-specific RNAi studies. Furthermore, they are less prone to cause toxicities and off-target

defects since this configuration enables the shRNAs to be processed like the endogenous miRNAs, thus causing a reduction in the toxicity level. Several miRNA scaffolds such as those of miR 155, miR 223 and miR 30 have been used by several groups (Fowler *et al.*, 2015) (Amendola *et al.*, 2009) (Fellmann *et al.*, 2013) to enhance the processivity of shRNAs, and it has led to promising results. However, miR 30 has been mostly used as a miRNA for the construction of scaffolds to increase knockdown levels owing to the ease by which the structure of miR 30 can be modified.

Towards this, several variants of miR 30 have been created to increase the knockdown efficiencies. In the first variant of the miR-30 scaffold, the shRNA sequence was flanked by 125 bases of the 5′ and 3′ of the pre–miR-30 sequence (Chang *et al.*, 2013). In addition to this, the *Xho*I restriction site was artificially introduced at the 5’ end, and the *EcoR*I site was introduced at the 3’ end. Although these modifications heled in the cloning of shRNAs in the vectors the introduction of the *EcoR*I site at the 3’end led to the disruption of a conserved CNNC motif, which led to reduced levels of biogenesis, this shortcoming was overcome by shifting the *EcoR*I site more towards the 3’ end as a result of which it did not interfere with the conserved sequence, thus leading to increased biogenesis of the shRNAs. This variant of miR 30 was termed as miR E (Fellmann *et al.*, 2013). The next variant of miR 30 was termed as the ultramiR scaffold, which was created by removal of the restriction sites from the previous miR 30 scaffolds (Knott *et al.*, 2014). This modification thus helped in retaining the original structure of endogenous miR 30,

which further helped in enhancing the processivity of the shRNAs and efficient knockdown of the target genes.

To carry out high throughput shRNA studies in primary cells, several parameters need to be considered to achieve adequate knockdown levels, such as the structure of the scaffold, algorithms used to design the shRNAs and the promoters driving the shRNAs. Several algorithms have been used to design shRNAs such as DSIR(Filhol *et al.*, 2012), ShRNAPred (Singh *et al.*, 2012), Sherwood (Knott *et al.*, 2014) and Splash (Pelossof et al., 2017). But most of the recent RNAi studies employ either Sherwood or Splash algorithms to design shRNAs since they use a sensor-based approach for the design, which leads to efficient high levels of knockdown. The promoter driving the expression of shRNAs is a critical determinant of knockdown levels since the efficacy of promoters varies with the cell types, and it must be ensured that the promoter used to express shRNAs should not get silenced by DNA methylation during the entire duration of the study.

This chapter describes the establishment of strategies for high throughput RNAi screen studies in HSCs.

1. Meta-analysis to identify the candidate genes involved in HSC maintenance and differentiation.
2. Choosing the most efficient shRNA design algorithm and miR30 based backbone for gene knockdown.
3. Identification of promoters that show stable expression throughout ex-vivo erythroid differentiation.

4. Generation of an shRNA library for RNAi screen of the genes involved in HSC maintenance.

5. Identification of positive and negative regulators of HSC maintenance and differentiation by RNAi.

6.2 SELECTION OF GENES FOR THE RNAi SCREENING

The Gene-expression databases (GEO) consisting of 677 and 569 microarray experimental data sets related to HSCs in mouse and human, respectively, were used for identifying the genes for the RNAi experiment. First, for each microarray dataset, the ranked list of differentially expressed genes between HSCs and the differentiated cells (DCs) were identified. The common top-ranked genes of human and mouse HSCs were considered to be the potential candidate genes responsible for HSC maintenance. In the second step, we used gene-ontology terms to identify the list of genes involved in signal transduction, transcription and epigenetic modifications. The presence of known self-renewal regulators (such as MPL and c-KIT) in top of the ranked list were used as a verifiable indicator of the progress. The top genes that were identified after the meta-analysis are shown in **Table 6.2.1**. The gene ontology analysis of the genes and the pathways that they are involved in are shown the **Table 6.2.2,** Table **6.2.3** and **Figure 6.2.1**. The most significant pathway was found to be Wnt Signalling pathway (6 genes), followed by VEGA-VEGFR2 pathway (10 genes), PI3K-Akt-MTOR pathway (11 genes), focal adhesion pathway (9 genes) and Hedgehog Signalling pathway (5 genes).

We analyzed the expression of the 117 genes that we selected for this study in ex-vivo erythropoiesis. For this, we used previously published RNA-Sequencing data (GSE119315). The bioinformatics analysis showed that 53 genes had a significant difference in gene expression. Out of these, 14 genes were upregulated with log2FC>2 and 18 were downregulated with log2FC<-2 and 21 genes had a lower difference in gene expression (log2FC<2 or >-2) (**Figure 6.2.2 A, B** and **C**).

Table 6.2.1 Top 117 genes involved in HSC maintenance and differentiation identified by bioinformatics analysis

ACTB	*CTNNB1*	*PPBP*	*GCNT2*	*ITGA6*
ADCY9	*CTTN*	*PRDM5*	*GFI1B*	*ITGAE*
ADRB3	*CXXC5*	*PRDX2*	*GHR*	*ITSN1*
ANXA4	*CYP27B1*	*PRKCQ*	*GLIS2*	*KCNJ2*
APLP2	*CYP7B1*	*PRKG1*	*GNG11*	*KLF4*
ARHGEF12	*DST*	*PTTG1IP*	*GPR56*	*KLF9*
ARRB1	*DTX3*	*RFX5*	*GRB10*	*LTBP2*
BEX1	*EGR1*	*RPS6*	*H1F0*	*LZTS2*
BNIP2	*ELTD1*	*RRAS*	*HBEGF*	*MAFF*
C4A	*ENG*	*SELP*	*HIST1H1C*	*MAGED1*
CA2	*EPAS1*	*TRIM6*	*HMGN3*	*MELK*
CAPN5	*F11R*	*TSC22D1*	*HOXA5*	*MFNG*
CASP12	*F2R*	*TSPAN13*	*IER3*	*MMP2*
CAV2	*FAM110C*	*TSPAN3*	*IFIH1*	*MYCN*
CBFA2T3	*FAM132A*	*TSPAN4*	*IFITM1*	*NDN*
CBX2	*FBXO9*	*VEGFC*	*IFITM3*	*NEDD4*
CBX6	*FRAT1*	*VLDLR*	*IL15*	*NFIX*
NR1D2	*SOCS5*	*VNN1*	*CCND1*	*NKX2-3*
NR4A1	*SPNS2*	*ZFAND5*	*CSNK1D*	*SLC1A1*
NRIP1	*TCEAL1*	*ZNF467*	*CSNK1E*	*SMO*
PDGFC	*TEK*	*IRF6*	*FSTL1*	*SMYD2*
PGLYRP2	*TGM2*	*ITGA2B*	*GAB1*	
PLXNC1	*TLN1*	*NFKBIL1*	*GAP43*	
POR	*TMBIM1*	*NGFRAP1*	*IL18*	

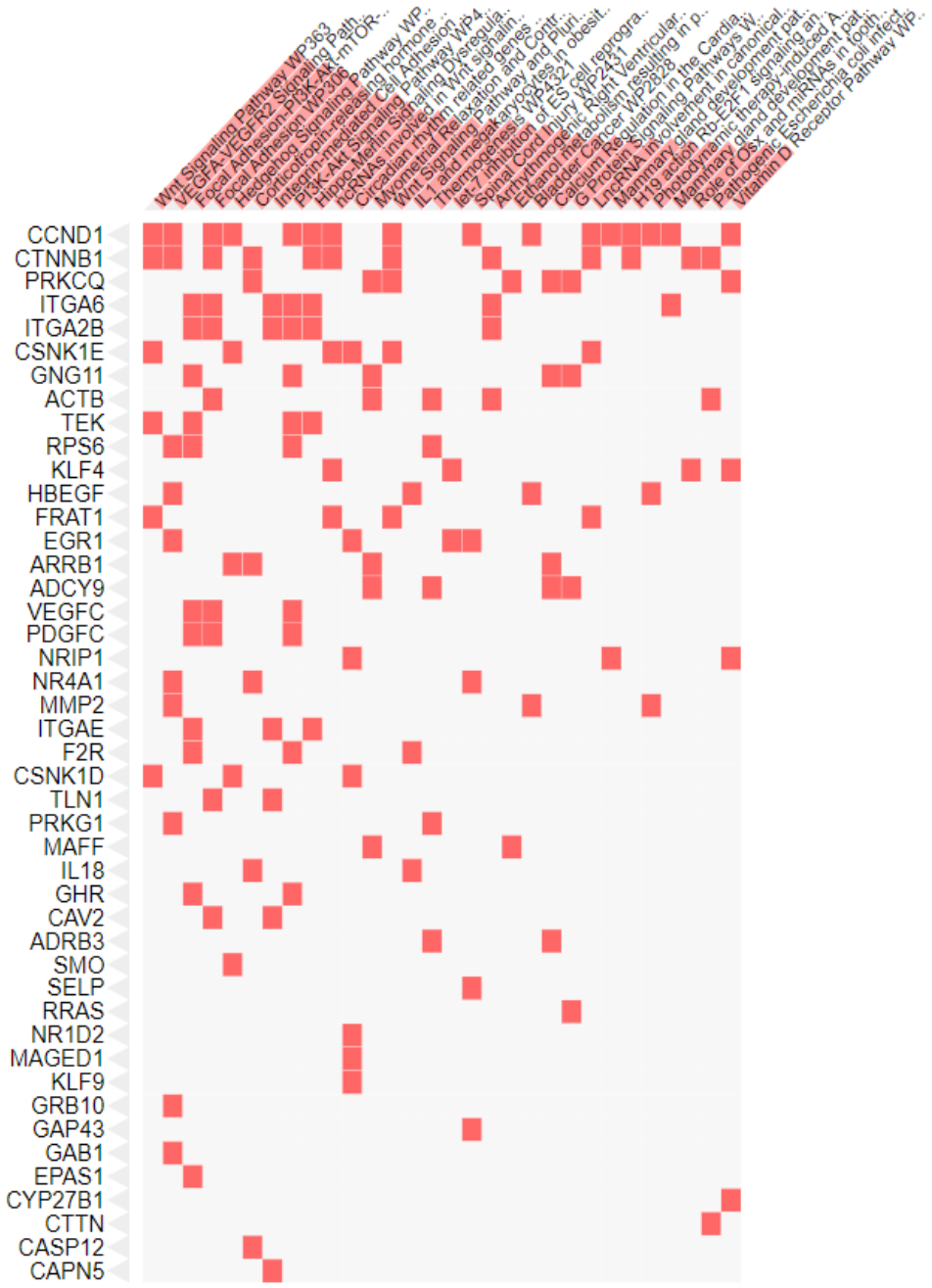


Figure 6.2.1 The predominant pathways in which the genes selected for the RNAi are involved in.

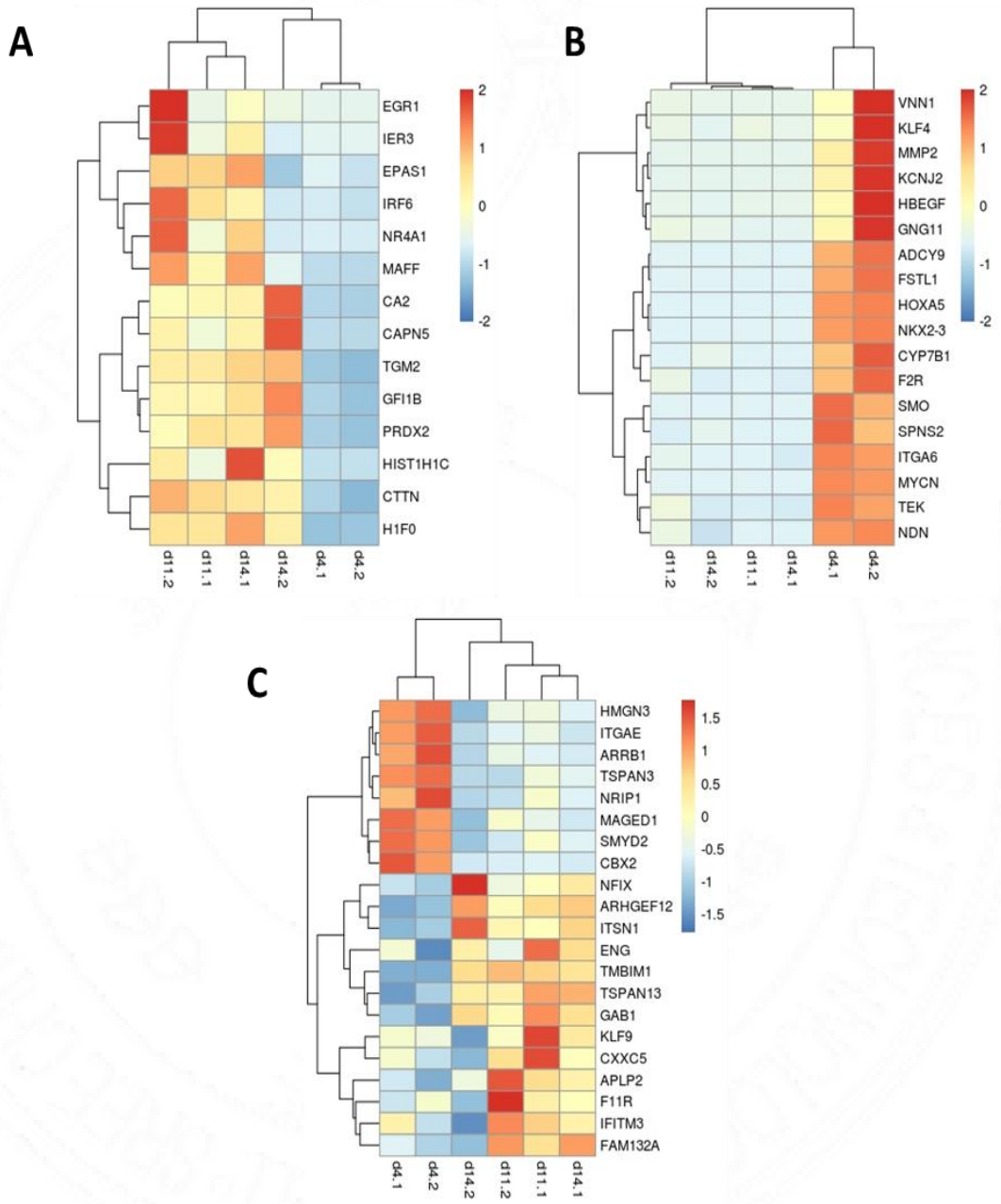


Figure 6.2.2 Expression of the genes selected in the study in ex-vivo erythropoiesis. The different time points in the culture (days) are shown as d4, d11 and d14. The data from the replicates were used for analysis. (A) Upregulated genes (B) Downregulated genes and (C) Genes without significant difference in the expression.

Table 6.2.2 Gene ontology analysis of the genes selected for the RNAi screening in human erythropoiesis.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding (GO:0001228)	0.00004479	0.05156	5.42	54.24
2	transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding (GO:0000982)	0.0002440	0.1404	4.88	40.63
3	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding (GO:0001077)	0.0005777	0.2216	5.86	43.70
4	glucocorticoid receptor binding (GO:0035259)	0.001189	0.3422	37.99	255.82
5	core promoter proximal region sequence-specific DNA binding (GO:0000987)	0.001256	0.2892	4.30	28.75
6	sequence-specific DNA binding (GO:0043565)	0.002230	0.4279	3.47	21.19
7	kinase binding (GO:0019900)	0.003203	0.5267	3.27	18.79
8	nuclear hormone receptor binding (GO:0035257)	0.004316	0.6210	9.16	49.87
9	RNA polymerase II core promoter proximal region sequence-specific DNA binding (GO:0000978)	0.004449	0.5690	3.91	21.20
10	transforming growth factor beta binding (GO:0050431)	0.005437	0.6258	17.99	93.83

Table 6.2.3 Major pathways in which the selected genes are involved in.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Wnt Signaling Pathway WP363	5.757e-7	0.0002717	19.72	283.38
2	VEGFA-VEGFR2 Signaling Pathway WP3888	0.000001280	0.0003021	7.24	98.28
3	Focal Adhesion-PI3K-Akt-mTOR-signaling pathway WP3932	0.000001681	0.0002644	6.21	82.51
4	Focal Adhesion WP306	0.000002503	0.0002953	7.77	100.22
5	Hedgehog Signaling Pathway WP4249	0.000005690	0.0005371	19.43	234.59
6	Corticotropin-releasing hormone signaling pathway WP2355	0.00001774	0.001396	11.03	120.65
7	Integrin-mediated Cell Adhesion WP185	0.00002840	0.001915	10.15	106.31
8	PI3K-Akt Signaling Pathway WP4172	0.00003152	0.001860	5.03	52.11
9	Hippo-Merlin Signaling Dysregulation WP4541	0.00007481	0.003923	8.55	81.20
10	ncRNAs involved in Wnt signaling in hepatocellular carcinoma WP4336	0.0001501	0.007084	9.94	87.50

6.3 CONSTRUCTION OF AN EFFICIENT LENTIVIRAL shRNA KNOCKDOWN VECTOR

Recent advances in the field of RNAi have led to the establishment of various synthetic shRNA^{mir} scaffolds, which contribute to higher knockdown efficiencies. In addition to this, the algorithms used to design the shRNAs also play a significant role in determining the knockdown efficiencies. Towards this, the endogenous miR-30 sequence has been engineered to give rise to miR-E and UltramiR scaffolds. These scaffolds cause increased processivity of shRNAs, which further leads to higher knockdown efficiencies. Furthermore, the sensor-based algorithm, such as Sherwood has made possible the generation of effective single-copy shRNAs that facilitate knockdown studies in cultured cells.

The knockdown efficiencies of shRNAs designed by different algorithms were assessed by cloning them into different vector scaffolds such as miR-E, pZIP-hEF1α and pZIP-hCMV and transfecting them in 293T cells for lentivirus generation (**Figure 6.3.1 A** and **B**). The shRNA sequences were obtained from Sherwood (Sherwood.cshl.edu) and from (Fellmann *et al.*, 2013). For estimation of the knockdown efficiency, RNA was extracted from the transduced HeLa cells after puromycin selection, and real-time PCR was performed to estimate the expression levels of target genes (**Figure 6.3.2 A** and **B)**.

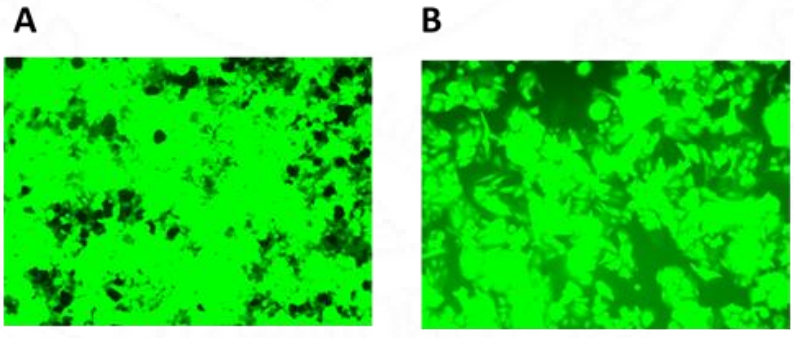


Figure 6.3.1 (**A**) Representative GFP image depicting transfection efficiency of shRNA lentiviral plasmid in 293 T cells. (**B)** Representative GFP image showing transduction efficiency of shRNA lentiviral plasmid in HeLa cells.

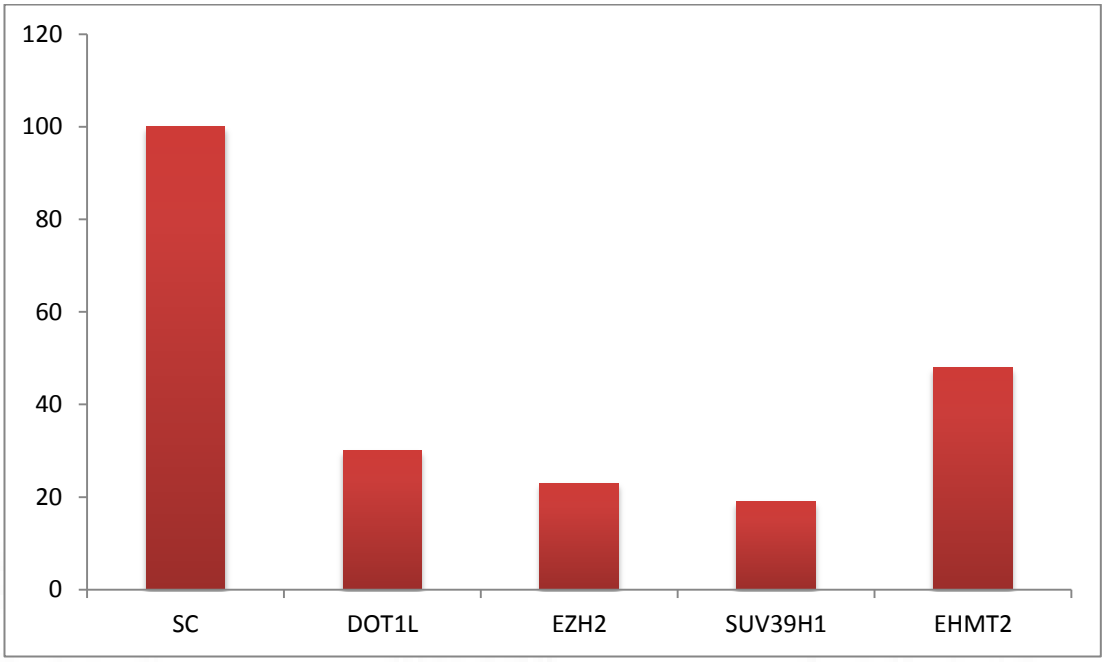


Figure 6.3.2 (**A**): Knockdown efficiencies of different genes in pGIPZ-mirE-hCMV vector. Data are represented as the percentage of residual mRNA expression. HeLa cell lines transduced with the empty vectors served as control.

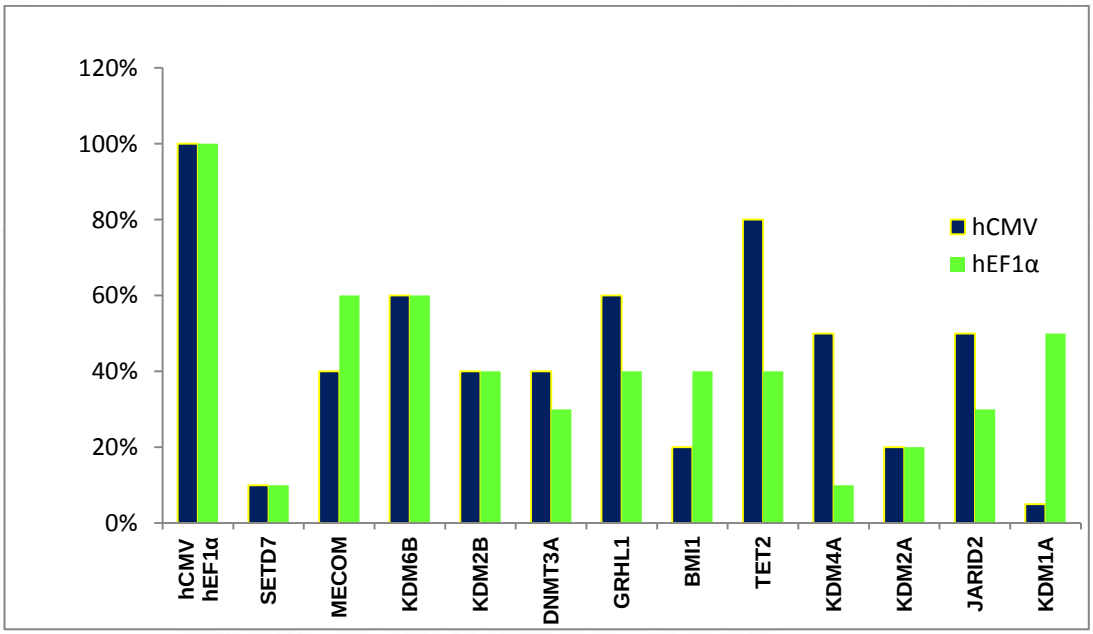


Figure 6.3.2 (B) Knockdown efficiencies of different genes in pZIP-UltramiR-hCMV and pZIP-UltramiR-hEF1α vectors. Data are represented as the percentage of residual mRNA expression. HeLa cell lines transduced with the empty vectors served as control.

6.4 ASSESSMENT OF THE ACTIVITY OF PROMOTERS DURING ERYTHROID DIFFERENTIATION

A promoter that remains active throughout erythroid differentiation is essential to ensure consistent knockdown of genes. Systematic studies to identify the most efficient promoters during the process of *ex-vivo* erythropoiesis have not been carried out so far. In this thesis, lentiviral vectors with eight different promoters (**Figure 6.4.1)**, spleen focus forming virus (SFFV), human cytomegalo virus (hCMV), mouse cytomegalovirus (mCMV), human elongation factor 1 alpha (hEF1α), mouse elongation factor 1 alpha (mEF1α), chicken β actin (CBA), ubiquitin C (UbC)) and MND were analyzed for their stability and strength of expression during *ex-vivo* erythroid differentiation.

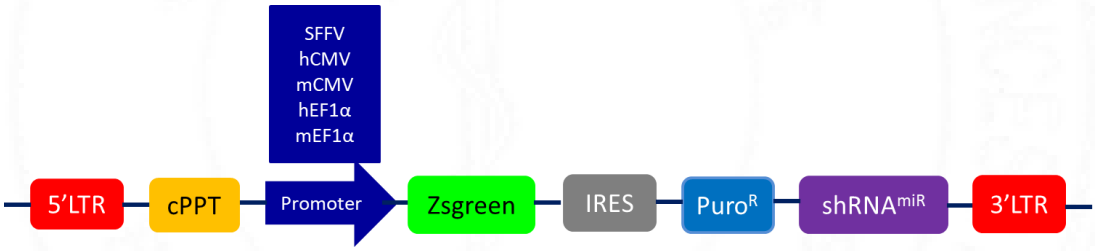


Figure 6.4.1: Lentiviral construct with different promoters used in this study. SFFV: spleen focus forming virus promoter, hCMV: human cytomegalovirus promoter, mCMV: mouse Cytomegalo Virus promoter, hEF1α: human Elongation Factor 1 alpha promoter, mEF1α: mouse elongation factor 1 alpha promoter, LTR: long terminal repeat, cPPT-central polypurine tract and IRES-internal ribosome entry site.

Lentiviral shRNA plasmids that expressed GFP under the control of these promoters (SFFV, CBA, hCMV, mCMV, mEF1α, hEF1α, UbC and MND) were used for the generation of lentiviruses using the protocol mentioned in the materials and methods. $CD34^+$ cells were transduced with lentiviruses, and the GFP^+ cells were flow-sorted.

The sorted cells were then seeded for erythroid differentiation, and the promoter activity was estimated at regular intervals based on the % of GFP^+ cells and the mean fluorescence intensity (MFI) of GFP expression (**Figure 6.4.2**).

It was observed that there was a significant difference in the transcriptional activity of the promoters during the process of erythroid differentiation. Promoters such as SFFV, mEF1α and MND showed a consistent expression of GFP throughout erythroid differentiation. In contrast, the promoters such as hEF1α and UBC showed a decline in the levels of GFP during erythroid differentiation. The hCMV promoter showed heterogeneous levels of GFP expression throughout erythroid differentiation, and it displayed an interesting pattern wherein the expression levels of GFP decreased during the intermediate stage of erythroid differentiation but again picked up during the later stages of erythropoiesis. Promoters such as mCMV and CBA showed complete silencing in the intermediate stage of erythropoiesis. With respect to the MFI, most of the promoters showed a sharp decrease in the MFI levels except hCMV and MND promoters, which did not show a significant reduction it the MFI of the GFP expression.

Thus this data suggested that MND and SFFV promoters showed consistent expression kinetics during erythroid differentiation, whereas other promoters were less efficient. The shRNA vector with MND promoter was selected for the subsequent experiments due to the consistently high expression of GFP from this promoter. Previous studies have shown that MND has a strong promoter activity in haematopoietic cells, and it has been used for gene therapy for Wiscott Aldrich Syndrome.

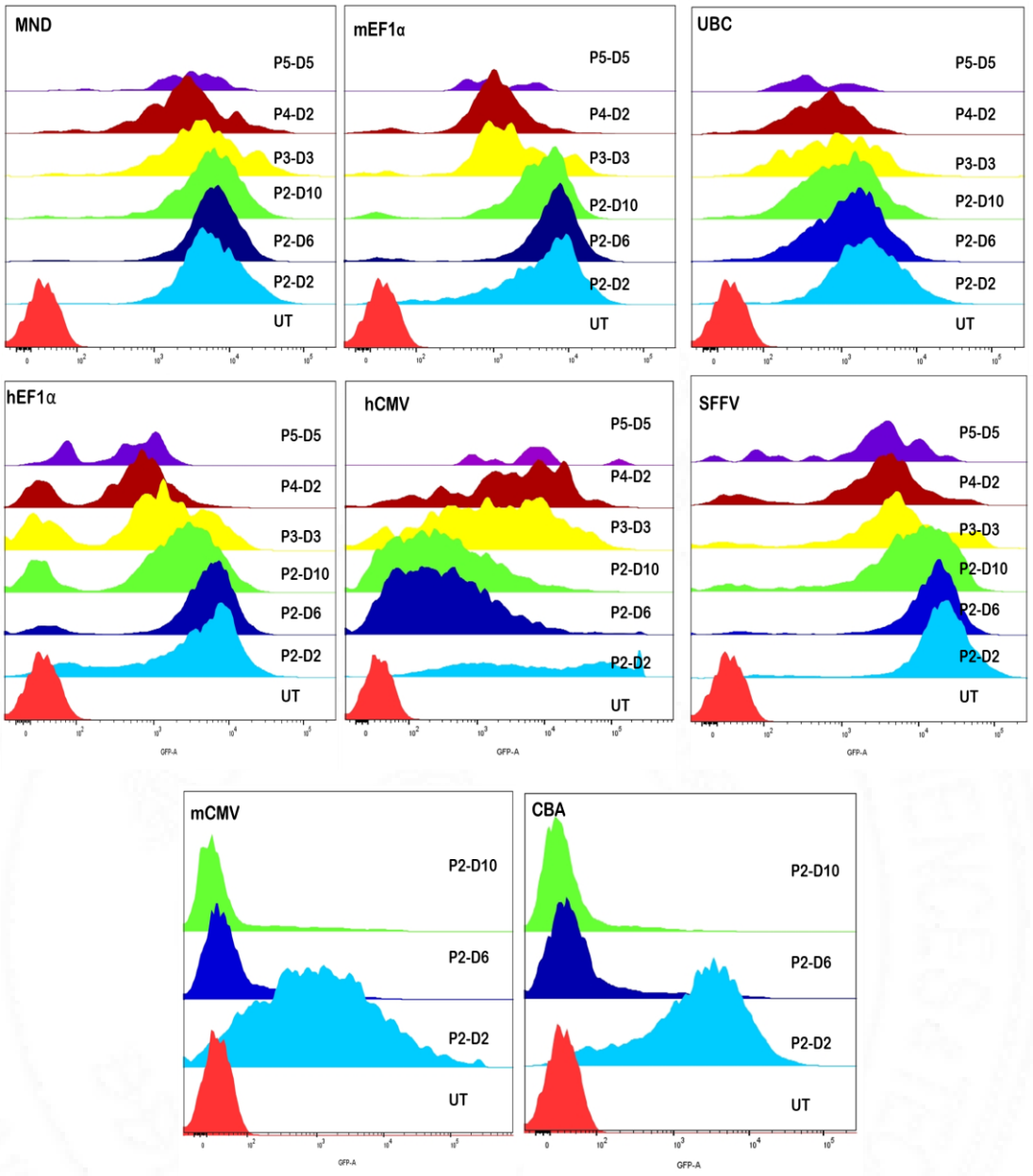


Figure 6.4.2: Assessment of the activity of different lentiviral vector promoters during erythroid differentiation as estimated by quantitating the levels of GFP. P2-D2, P2-D6, P2-D10, P3-D3, P4-D2 and P5-D5 indicate different phases (P) of erythroid culture and different days (D) of differentiation.

6.5 GENERATION OF A shRNA LIBRARY TARGETING HSC MAINTENANCE GENES

For the generation of an shRNA library against HSC maintenance genes, shRNA sequences were synthesized as a pool of oligos. As the synthesized shRNA oligos had a very low quantity, we amplified them by two rounds of PCRs **(Figure 6.5.1 A** and **B)**. To obtain a sufficient amount of shRNAs, a PCR with 25 cycles was performed using the primers used for cloning. The PCR products containing double-stranded shRNAs were purified, quantitated, and subjected to the second round of PCR using the same set of primers for 14 cycles. By doing two rounds of PCRs, we obtained a sufficient quantity of shRNAs for Gibson assembly cloning into the pZIP-UltramiR-MND vector. For getting more than 1000 times representation of the shRNAs after cloning, the ligations and transformations were carried out in 4 and 11 tubes, respectively, and a bacterial lawn made to obtain equal growth of the transformed colonies. The extracted plasmids were digested with restriction enzymes, which confirmed that the cloning was successful and no abnormal pattern was observed, suggesting that there was no recombination during the cloning process (**Figure 6.5.1C**). The region containing the shRNA sequences were amplified (**Figure 6.5.1 D** and **E**) and subjected to NGS to ensure the equal representation of the shRNAs. A bell-shaped curve indicates a good representation of shRNAs in the plasmid library (**Figure 6.5.1 F**).

6.6 VALIDATION OF shRNA LIBRARY IN THE EX-VIVO MODEL OF HSC DIFFERENTIATION

The custom made shRNA library was used for transducing the $CD34^+$ HSPCs, and they were differentiated to erythroid cells to identify the genes involved in HSC maintenance and erythroid differentiation. To obtain a single copy of the integration of lentiviral shRNA vectors after transduction, we maintained a transduction efficiency of ~30% (**Figure 6.6.1**), which was measured by the flow cytometric analysis of ZsGreen expressed from the lentiviral vector. To obtain the representation of all the shRNAs in the transduced cells, we transduced at the complexity of 1000.

The transduced cells were cultured in the Phase 1 medium for five days, and the GFP^+ cells were flow-sorted. The GFP^+ cells were cultured for five more days in the phase 1 medium, and DNA was collected from 2 X 10^6 numbers of cells. The remaining cells were cultured in the phase II medium, and DNA was extracted from 4 different time points in phase II of the culture to represent the cells from the different stages of erythropoiesis, as estimated by flow cytometric analysis of CD71 and CD235a (**Figure 6.6.2**). DNA from the five different time points was extracted, and the PCRs were carried as described in the materials and methods to enrich the shRNA sequences (**Figure 6.6.3**).

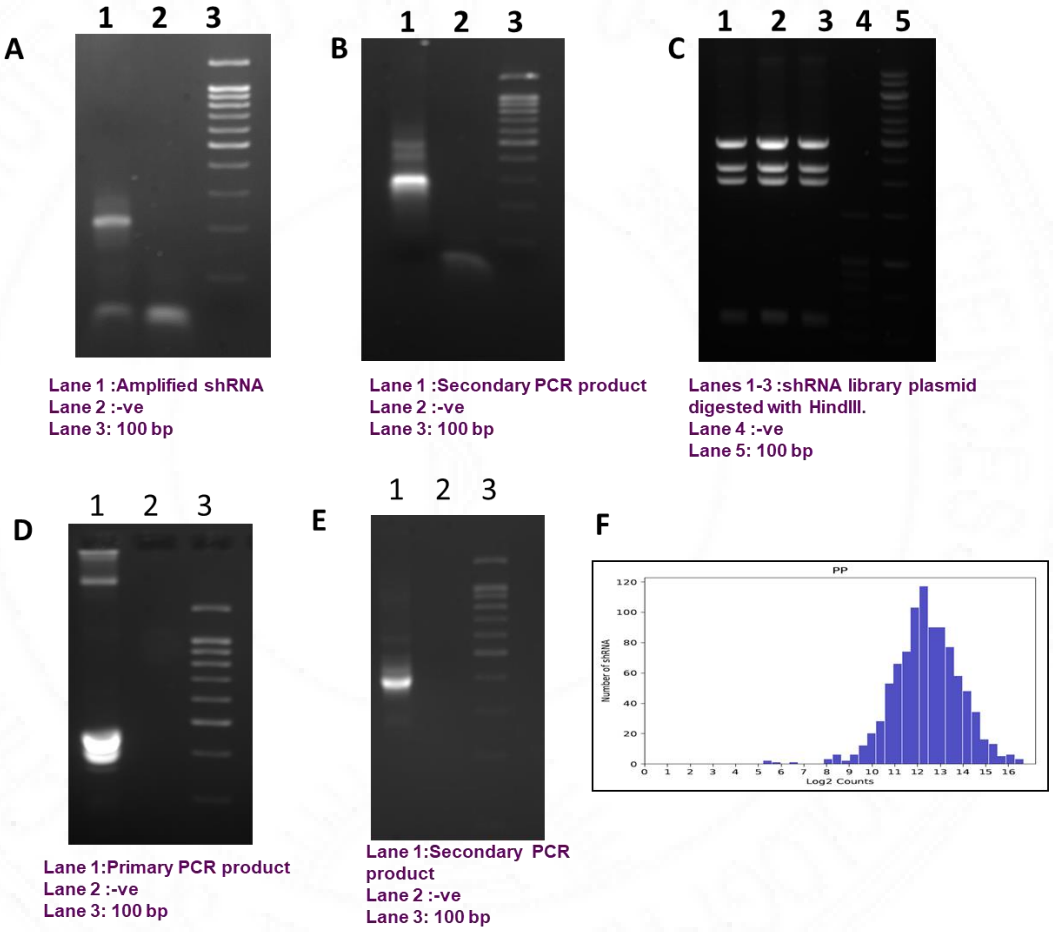


Figure 6.5.1: Generation of a pooled shRNA library. (A) and (B) Amplification of the shRNA oligos by the first and second round PCRs. (C) Restriction digestion of the shRNA library pooled plasmids (D) and (E) Amplification of the shRNA containing region from the plasmid by the first round and the second round PCRs (F) Histogram showing the representation of the shRNA sequences after NGS.

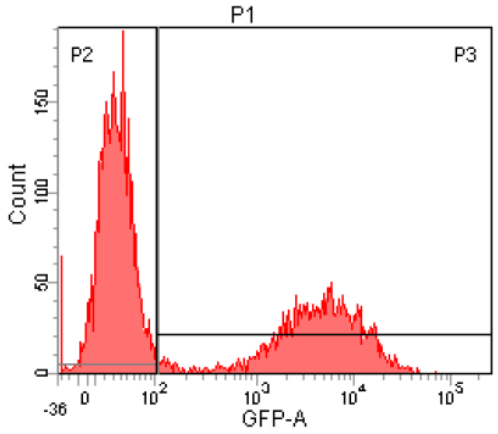


Figure 6.6.1 Flow cytometry analysis to quantitate the percentage of GFP for estimation of transduction efficiency.

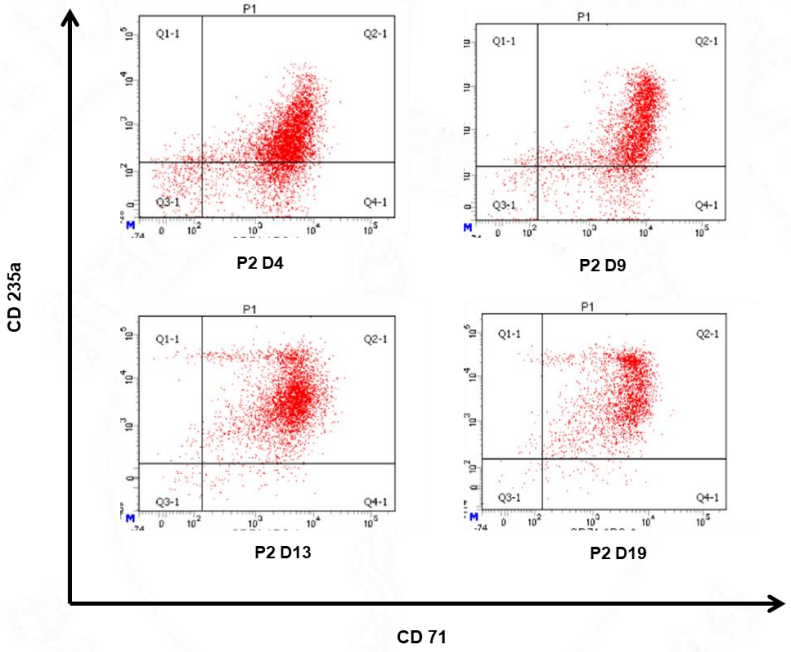


Figure 6.6.2: Flow cytometry analysis of erythroid markers CD71 and CD235a of the cells on day 4,9,13 and 19, respectively.

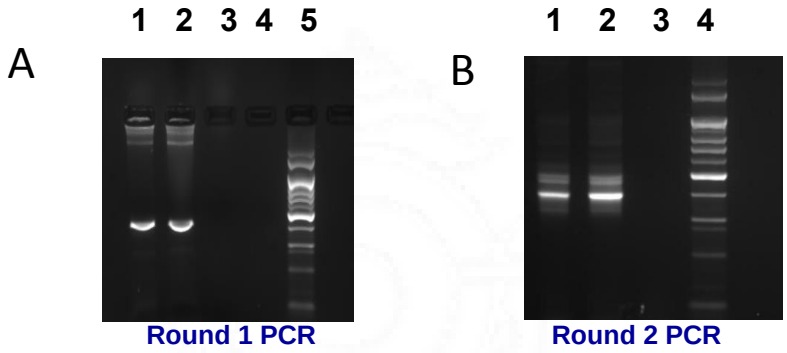


Figure 6.6.3 Amplification of DNA with specific primers for NGS. (A) Lane 1 & 2: DNA samples from Day 13 of phase 2 of the culture, Lane 3 & 4: Negative controls and Lane 5:100bp (B) Second round PCR Lane 1 & 2: 1st round PCR product of the DNA of the cells from day 13 of phase 2 of the culture, Lane 3: negative control Lane 4:100bp molecular weight marker.

6.7 BIOINFORMATICS TO IDENTIFY THE GENES THAT AFFECT HUMAN ERYTHROPOIESIS

The fastq files were analysed using CRISPRCloud2 (Jeong *et al.***).** The analysis results provided the list of genes and the list of shRNAs that are enriched and depleted. We compared the enrichment and depletion of shRNAs and the genes on days 4 (TP2), 9 (TP3), 13 (TP4) and 19 (TP5) in phase II of the culture compared to day 8 (TP1) of phase I of the culture. The correlation plot of the results from the samples is shown in **Figure 6.7.1**.

We did not find enrichment of shRNAs against any genes in this RNAi screening. This is may be because of the selection of the shRNAs that are important for HSC maintenance and differentiation in the RNAi screening.. The list of genes for which significantly high depletion was observed is shown in **Table 6.7.1**.

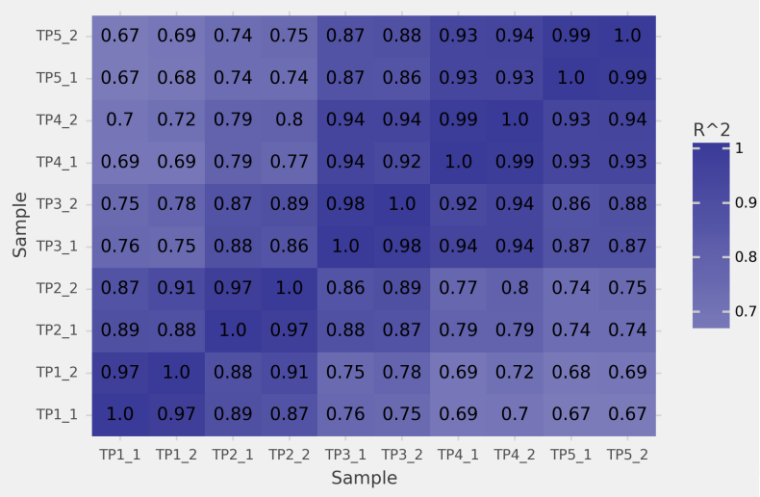


Figure 6.7.1 The correlation plot of the samples analysed by NGS. TP1, TP2, TP3, TP4 and TP5 represent the different time points of the DNA collection for the experiment. 1 and 2 are replicates of the samples.

Out of the 24 genes with maximum depletion of shRNAs, the function of 10 genes in human erythropoiesis has been described earlier. The details are given below.

SMO: SMO is a G protein-coupled receptor for hedgehog ligands. Gene expression analysis showed that SMO has higher expression in early stages of erythropoiesis and lower expression in the late stages. We found that SMO shRNAs are significantly depleted in ex-vivo erythropoiesis. These data suggest that hedgehog signalling may be preferentially activated in erythroblasts undergoing self-renewal cell divisions. Deletion of Smo in mice inhibits the development of erythroid progenitors in the spleen (Paulson *et al.*, 2020).

Table 6.7.1 The genes with most depleted shRNAs in the RNAi screening of ex-vivo erythropoiesis. Fold change between the last time point TP5 and the first time point in two experiments and the average is shown in the table.

Gene	Experiment 1	Experiment 2	Average Log2FC
	log2FC	log2FC	
SMO	-5.63	-4.76	-5.19
TSPAN4	-4.08	-5.35	-4.72
IFITM3	-5.27	-4.09	-4.68
PRDX2	-4.73	-4.52	-4.63
CSNK1E	-5.11	-4.14	-4.63
MAFF	-5.29	-3.86	-4.58
NFIX	-4.17	-4.98	-4.57
CXXC5	-4.51	-4.62	-4.56
KLF9	-4.71	-4.21	-4.46
POR	-4.91	-4.00	-4.45
C4A	-3.91	-4.67	-4.29
NFKBIL1	-3.00	-5.41	-4.21
MAGED1	-4.88	-3.44	-4.16
RPS6	-5.03	-3.21	-4.12
GCNT2	-3.70	-4.50	-4.10
CSNK1D	-3.79	-4.38	-4.08
HIST1H1C	-4.91	-3.24	-4.07
RRAS	-3.23	-4.90	-4.07
CBX2	-4.80	-3.28	-4.04
GLIS2	-3.41	-4.60	-4.00
FAM132A	-3.46	-4.53	-4.00
ENG	-4.07	-3.86	-3.96
CBX6	-4.05	-3.83	-3.94
ARRB1	-4.94	-2.93	-3.94

IFITM3: IFITM3 (Interferon (IFN)-induced transmembrane protein 4) belongs to the class of proteins including IFITM1, IFITM2 and IFITM3, which are important gatekeepers of pathogen infection in vertebrates. Knockdown of IFITM3 increased the susceptibility of cell lines to influenza A virus (IAV) while its IFITM3 could inhibit infection by other strains of IAV. IFITM3 is the most active against IAV. IFITMs are active against many other viruses, but they are not active against DNA

viruses. *ifitm3*$^{-/-}$ mice are found to be more susceptible to infection by IAV. IFITM3 is important for the antiviral activity of human stem cells, and a recent study has shown that downregulation of this gene increases the transduction efficiency of lentiviruses (Petrillo *et al*., 2018). Inhibition of mTOR pathway down-regulates IFITM3 through a lysosomal degradation pathway in hematopoietic and nonhematopoietic cells (Shi et al., 2018). We found that the shRNAs of IFITM3 significantly depleted, suggesting that this gene is important for human erythropoiesis. Further studies are required to understand the fundamental role of IFITM3 in erythropoiesis.

PRDX2: Peroxiredoxin II (PRDX2) is highly expressed during terminal erythropoiesis. It is is one of the most abundant proteins in erythroid cells. It is an antioxidant enzyme, which reduces the reactive oxygen species (ROS), such as hydrogen peroxide and alkyl hydroperoxides, which are present in high levels in erythroid cells (Romanello *et al.*, 2018) *PRDX2*$^{-/-}$ mice were found to have hemolytic anaemia with evidence of oxidative damage of the erythrocyte proteins resulting to decreased red blood cell (RBC) survival. Recently, a family with a *PRDX2* variant was found with a phenotype of dominantly-inherited congenital dyserythropoietic anaemia. Our RNAi screen also suggests that RPDX2 is important for human erythropoiesis.

CSNK1E: CSNK1E gene is a serine/threonine-protein kinase and a member of the casein kinase I protein family. The members of this family are involved in the control of several cytoplasmic and nuclear processes, including DNA repair and replication.

CSNK1E protein is a cytoplasmic protein and can phosphorylate a variety of proteins. The shRNAs for *CSNK1E* were found to be significantly depleted during ex-vivo erythropoiesis. ChiP-Sequencing analysis of published data sets using METAP2 (https://amp.pharm.mssm.edu/archs4/gene/METAP2) showed that CSNK1E is a target of the major erythroid transcription factor KLF1, confirming that CSNK1E has an essential role in erythropoiesis. Further studies in mouse and functional studies in human erythroid cells will help in understanding its role in human erythropoiesis.

NFIX: Chaand *et al.* showed increased expression of *NFIX* and chromatin accessibility at the *NFIX* promoter in the adult erythroid cells relative to cord blood erythroid cells. Knockdown of NFIX in primary erythroblasts derived from human CD34+ BM cells caused induction of gamma-globin mRNA, foetal haemoglobin and F-cells and led to a delay in erythroid differentiation as measured by flow cytometric analysis of CD36 and GYPA expression without affecting enucleation. These results suggested that NFIX is a major regulator of human erythropoiesis.

***KLF9*:** Krüppel-like factors (KLFs) are involved in diverse physiological processes, such as the differentiation and development of red blood cells. It has been noted that the expression of KLF1 and KLF9 was significantly higher in the late stages of erythropoiesis than in hematopoietic stem cells. Overexpression of KLF1 and KLP9 in K562 cells revealed that their co-expression significantly promotes erythroid differentiation and enhances β-globin gene expression. KLF1 and KLF9 synergistically regulate erythroid differentiation through the PI3K-Akt and FoxO signalling pathways (Ren *et al*., 2018).

TSPAN4: It is also known as CD151, and belongs to the tetraspanin family consisting of cell membrane proteins involved in the regulation of cell development, activation, growth and motility. TSPAN4 expresses MER2 blood group antigen and is expressed in erythrocytes. Those individuals with mutations in CD151 had beta thalassaemia phenotype, suggesting that CD151 has a role in erythropoiesis. There was no study so far to confirm the role of Cd151 in human erythropoiesis. Our RNAi experiments confirmed that CD151 is important for normal erythropoiesis.

RPS16: Haploinsufficiency of RPS16 is associated with Diamond Blackfan anaemia and 5q- myelodysplastic syndrome, two diseases characterized by dysregulated erythropoiesis.

ARRB1 (Arrestin beta 1): It is a member of arrestin/beta-arrestin protein family, which participates in agonist-mediated desensitization of G-protein-coupled receptors and it causes dampening of cellular responses to stimuli such as hormones, neurotransmitters, or sensory signals. It is a cytosolic protein and acts as a cofactor in the beta-adrenergic receptor kinase (BARK) mediated desensitization of β-adrenergic receptors. In addition to the central system, it is expressed at high levels in white blood cells. The BARK/β-arrestin system is proposed to play a major role in the regulation of receptor-mediated immune functions. In definitive erythroid cells, ARRB1 is recruited to active sites of transcription near the globin genes. However, its role in human erythroid differentiation was not identified earlier.

HIST1H1C: This protein has both transcription repression and activation function. It has been proven that it is involved in multiple functions in the cell cycle, DNA_damage induced apoptosis and fibrillary stabilization. It can interact with RNA

Polymerase II to initiate transcription. This protein binds to Ldb1 complex and carries out transcription and nuclear processing in erythroid cells.

Pathway analysis of the genes with the significant difference in shRNA depletion with log2FC<-3.5 identified the pathways that are required for human erythropoiesis. Wikipathways identified the Hedgehog signalling pathway as the most important pathway required for human erythropoiesis (**Figure 6.7.2**). SMO, CSNK1E, CSNK1D and ARRB1 were the genes in the Hedgehog pathway whose shRNAs were found to be depleted in the late stages of erythropoiesis. Although this pathway has been described in stress erythropoiesis, its role in normal haematopoiesis and erythropoiesis has not been studied earlier. Our study clearly showed this pathway has significant importance in erythroid differentiation. In addition to the Hedgehog pathway, there are genes involved in other molecular mechanisms, including transcriptional activation of specific genes required for erythropoiesis.

We analyzed the transcription regulation of the genes with high levels of shRNA depletion in our RNAi experiment. For this, we performed transcription factor enrichment analysis using published data sets for the haematopoietic transcription factors, GATA2, GATA1, RUNX1 and NFE2L2, at the promoter regions of these genes (**Figure 6.7.3**). We found that among the top hits with depleted shRNAs, there 8 genes regulated by GATA2, 7 by GATA1, 9 by RUNX1 and 7 by NFE2L2. This data suggest that the top hits that were identified by the RNAi screening have important roles in HSC maintenance and differentiation.

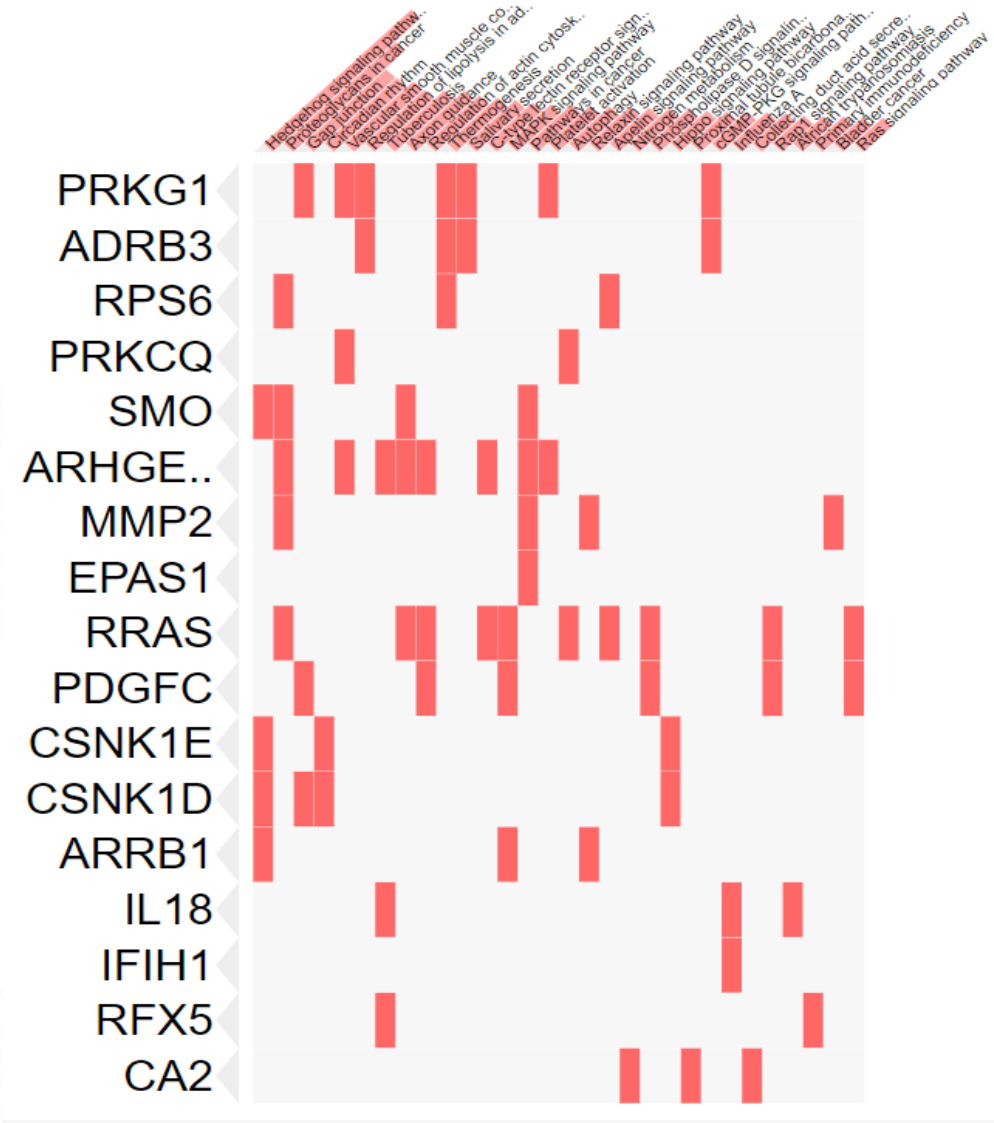
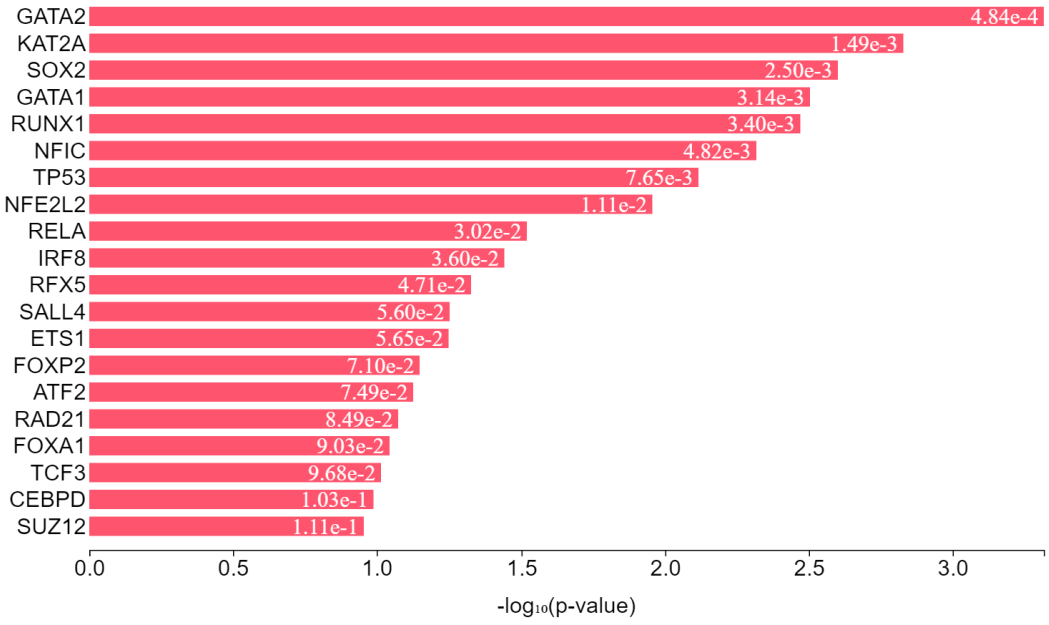


Figure 6.7.2 Wikipathway analysis showing the predominant pathways in which hits by RNAi screen are involved in.



Rank	Transcription Factor	Hypergeometric p-value	Enriched Targets
1	GATA2	0.0004838	8 targets
2	KAT2A	0.001488	3 targets
3	SOX2	0.002504	7 targets
4	GATA1	0.003135	7 targets
5	RUNX1	0.003404	9 targets
6	NFIC	0.004818	4 targets
7	TP53	0.007649	4 targets
8	NFE2L2	0.01110	7 targets
9	RELA	0.03017	4 targets
10	IRF8	0.03604	2 targets

Table **6.7.3** Binding of transcription factors at the promoters of the genes with high depletion of shRNAs. The data was generated from published data sets using X2KWeb (https://amp.pharm.mssm.edu/X2K/).

6.8 DISCUSSION

The process of human erythropoiesis serves as a paradigm of physiologic cellular differentiation. Genetic and cell biological experiments have provided several important insights into erythroid differentiation in normal individuals and patients with red cell diseases. To understand the molecular processes involved in normal and pathological erythropoiesis, currently, the cells generated by *in-vitro* erythroid differentiation of primary HSPCs obtained from patients are being used for this purpose. The CD34+ HSPCs from normal donors and the patients can be expanded and differentiated to erythroid cells, and the cells from each stage of erythroid differentiation can be isolated and used for comprehensive transcriptomic and proteomic profiling to understand the regulation of erythropoiesis. This ex-vivo erythroid differentiation protocol, with the constant improvements to increase the number of erythroid cells, has been extensively used to understand the chromatin occupancy of transcription factors. However, there has been no study so far to understand the application of this cellular model for human erythropoiesis for RNAi screen to identify the regulators of erythroid differentiation. We tested the applicability of the ex-vivo erythropoiesis system for RNAi experiments. Although RNAi has been performed by transducing Sca1+ mouse HSCs to study the role of genes involved in erythropoiesis, use of human CD34+ cells and ex-vivo erythropoiesis is more significant as the target genes in haematopoiesis may have different functions in humans and mice.

Gene silencing through sequence-specific targeting of mRNAs by RNAi has enabled genome-wide functional screens in cultured cells and *in vivo* in animal models. These

screens helped in the identification of new cellular pathways involved in biological processes. In the past 5 years, considerable progress has been made in this area of research. For the long term knockdown of the target genes in the cultured cells and mice models, stably integrating lentiviruses are recommended. The choice of the vectors and the promoters to express the shRNAs is important for efficient RNAi experiment. We systematically analysed the promoters and identified the promoter that provided robust expression of ZsGreen that is bicistronically expressed with the shRNAs. It was interesting to note that human EF1α and mouse CMV promoters that have been extensively used in haematopoietic cells showed low expression or silencing of ZsGreen during the erythroid culture. We found that SFFV and MND promoters have the highest expression without significant transgene silencing in the erythroid culture. The MND promoter is constitutively active in the hematopoietic system, and it resists transcriptional silencing (Halene *et al*., 1998). This promoter has been used in successful gene therapy clinical trials. Our data showed that the combination of the lentiviral promoter, design of shRNAs and the selection of the backbone to express the mir30 based shRNAs determine the efficiency of knockdown. Through simple Gibson assembly, it is possible to generate shRNA libraries to knock down thousands of target genes. This is the first study to use lentiviral RNAi screen in ex-vivo erythropoiesis. Using this approach, we found essential pathways involved in human erythropoiesis. Some of these pathways reported earlier using cell lines, and our study confirmed their roles in erythropoiesis. We also identified several genes that were not reported earlier in human erythropoiesis. It is important to perform further experiments to understand the role of the other genes with significant fold depletion of shRNAs. Ex-vivo erythropoiesis

using the shRNAs against each gene followed by transcriptome analysis will help in understanding the mechanism through which these genes affect human erythropoiesis. Further, mouse experiments can be performed for in-vivo validation of these data. Sca1+ mouse HSCs transduced with lentiviral shRNAs have been used for studying haematopoiesis and erythropoiesis earlier, and this model will help us in our future studies to understand the role of individual genes in erythropoiesis.

7 SUMMARY AND CONCLUSION

Genetic and cell biological experiments have provided several important insights into erythroid differentiation in normal individuals and patients with red cell diseases. To understand the molecular processes involved in normal and pathological erythropoiesis, currently, the cells generated by *in-vitro* erythroid differentiation of primary HSPCs obtained from patients are being used for this purpose. Ex-vivo erythropoiesis of HSPCs has been shown to mimic in-vivo erythropoiesis with progressive erythroid differentiation with the characteristic surface marker expression and morphology changes. We have established robust protocols for ex-vivo erythropoiesis and we used this cellular model for studying erythropoiesis in this thesis. The different stages of erythropoiesis can be reiterated in vitro which is characterised by changes in cell size and morphology thus making it an efficient system to carry out high throughput transcriptome studies. Several high throughput transcriptome studies carried out using *ex- vivo* erythroid culture systems has led to the identification of several novel coding and non-coding transcripts.

Many previous studies have attempted to study the transcriptional regulation of erythropoiesis by analyzing gene expression and transcription factors. However, a systematic study on miRNAs has been lacking. Identification of the differentially regulated miRNAs during a cellular process provides insights into the transcriptional regulatory mechanisms that control the expression of various proteins required for cell fate. The results of the small RNA sequencing analysis performed in HSPCs and the differentiated erythroid cells would be of great significance in understanding various regulatory transcriptional mechanisms mediated by miRNAs during erythropoiesis. This study also identified many miRNA-clusters that are coregulated

during erythroid differentiation, and several miRNAs were found to be intragenic, and their expression correlated with the expression of the host genes. We performed further analysis to study the occupancy of the erythroid-specific transcription factors, GATA1, KLF1 and TAL1. We found that these transcription factors regulate the expression of 90% of the upregulated miRNAs. We also performed gene editing for evaluating the roles of the most upregulated miRNAs identified in ex-vivo erythropoiesis. Gene editing has several advantages compared to other methods which were used for downregulation miRNA expression in cultured cells. It includes specificity and robustness of this method. There are very few studies carried out so far to evaluate the differentially expressed miRNAs functionally. Using the strategy that we adopted, it was observed that miRNAs could be effectively disrupted for assessing their role in cellular processes. We could confirm that the previously reported miRNAs, miR-144 and miR-451, are required for erythropoiesis by editing these miRNAs.

Ex-vivo erythropoiesis generates sufficient number of erythroid cells and the progressive erythropoiesis helps in studying the transcriptional regulation in each stage. We used an RNAi screening using lentiviral shRNA vectors to delineate the signalling pathways involved in erythropoiesis. Such an approach has also been used previously in the mouse as well as in human $CD34^+$ haematopoietic stem cells to identify novel genes involved in stem cell maintenance and renewal respectively (Kinkel *et al.*, 2015, Galeev *et al.*, 2016, Hope *et al.*, 2010). However, no such studies have been carried out earlier in an erythroid differentiation model to delineate the different signalling pathways involved in erythropoiesis. We found several genes

which were already reported to have a role in erythropoiesis/haematopoiesis as well as genes whose roles were not identified in the context of erythropoiesis/haematopoiesis. These findings validated the RNAi screening approach and the erythroid culture system, which was used for the study. Such an approach stands better as well as cost-effective against mouse models since a vast repertoire of genes including housekeeping genes and genes for survival can be analysed simultaneously while avoiding the tedious maintenance and analysis of animals. In another approach to use the most efficient shRNAs for target genes we used microRNA based shRNAs expressed from a RNA polymerase II promoter. After screening several promoters, we identified MND promoter is the most suitable promoter for RNAi experiments in ex-vivo erythropoiesis. The results of our experiments clearly identified the combination of the lentiviral promoter, design of shRNAs and the selection of the backbone to express the mir30 based shRNAs determine the efficiency of knockdown. Using a custom library that targets the genes identified by meta-analysis we identified novel pathways involved in human erythropoiesis. This is the first study to use lentiviral RNAi screen in ex-vivo erythropoiesis. Some of these pathways reported earlier using cell lines, and our study confirmed their roles in erythropoiesis. We also identified several genes that were not reported earlier in human erythropoiesis. It is important to perform further experiments to understand the role of the other genes with significant fold depletion of shRNAs.

This thesis has provided a platform for better understanding of the process of stem HSC self-renewal and differentiation. Identification of novel miRNAs (new in

erythropoiesis), signalling molecules as well as other transcription factors has helped in comprehending the complex process of erythropoiesis better. Use of high throughput shRNA screening libraries coupled with bioinformatics analysis has helped us in identifying several novel molecules which were not identified earlier in erythropoiesis. In addition to this, the long term erythroid culture system used has helped in carrying out the high throughput experiments successfully.

7.1 *HALLMARKS OF THE THESIS*

1. Identification of a suitable promoter which does not undergo silencing during erythroid differentiation.
2. Establishment of a seamless cloning procedure for cloning shRNAs and for creating a shRNA library.
3. Identification of pathways which were not reported earlier in human erythropoiesis.
4. Using CRISPR-Cas9 for the first time to disrupt miRNAs in HUDEP cells.

7.2 *FUTURE DIRECTIONS*

1. Ex-vivo erythropoiesis using the shRNAs against the genes identified by the RNAi screening followed by transcriptome analysis will help in understanding the mechanism through which these genes affect human erythropoiesis.
2. Further, mouse experiments can be performed for in-vivo validation of these data. Sca1+ mouse HSCs transduced with lentiviral shRNAs have been used for studying haematopoiesis and erythropoiesis earlier, and this model will

help us in our future studies to understand the role of individual genes in erythropoiesis.

3. In this thesis, the gene editing of miRNAs was performed in HUDEP cells. However, it is important to perform this in CD34+ HPSCs for other miRNAs. There are several other miRNAs for which the functional characterization could not be performed. A robust Cas9 vector that can help in efficient transduction in CD34+ cells needs to be developed for carrying out these experiments.

BIBLIOGRAPHY

Aerts S. (2012) Computational Strategies for the Genome-Wide Identification of cis-Regulatory Elements and Transcriptional Targets. In:Plaza S,Payre F (eds) Transcriptional switches during development, Elsevier Academic Press, San Diego,pp.121-145

Afshari P,Yao W. D. and Middleton F. A. (2017) Reduced Slc1a1 expression is associated with neuroinflammation and impaired sensorimotor gating and cognitive performance in mice: Implications for schizophrenia, *PLoS One* 12:1-41

Albers J, Danzer C, Rechsteiner M, Lehmann H, Brandt L P, Hejhal T, Catalano A, Busenhart P,Goncalves A F, Brandt S, Bode P K, Bode-Lesniewska B, Wild P J, Frew I J (2015) A versatile modular vector system for rapid combinatorial mammalian genetics, *J Clin Invest.*, 125:1603–1619.

Amendola M,Passerini L,Pucci F,Gentner B,Bacchetta R,Naldini L(2009) Regulated and multiple miRNA and siRNA delivery into primary cells by a lentiviral platform, *Mol Ther.*. 17:1039–1052.

An X,Schulz P. V.,Li J,Wu K,Liu J,Xue F,Hu J,Mohandas N,Gallagher P.G.(2014) Global transcriptome analyses of human and murine terminal erythroid differentiation, *Blood* 123:3466-3477.

Andolfo I,Falco L.D,Asci R,Russo R,Colucci S,Gorrese M,Zollo M,Iolascon A (2010) Regulation of divalent metal transporter 1 (DMT1) non-IRE isoform by the microRNA Let-7d in erythroid cells, *Haematologica* 95:1244–1252.

Andrews N. C. (1998) Molecules in focus the NF-E2 transcription factor, *Int J Biochem Cell Biol.* 30: 429–432.

Armisén R,Marcelain K,Simon F,Tapia J.C,Toro J,Quest A.F.G,Stutzin A (2011) TRPM4 enhances cell proliferation through up-regulation of the β-catenin signaling pathway, *J Cell Physiol.*, 226 :103–109.

Azzouzi I,Moest H,Winkler J,Fauchere J,Gerber A. P.,Wollscheid B,Stoffel M,Schmugge M,Speer O (2011) MicroRNA-96 directly inhibits γ-globin expression in human erythropoiesis., *PLoS One* 6: 228-238.

Becker P. B. and Workman J. L. (2013) Nucleosome remodeling and epigenetics, *Cold Spring Harb Perspect Biol.* 5 doi: 10.1101/cshperspect.a017905.

Bonifer C. and Bowen D. T. (2010) Epigenetic mechanisms regulating normal and malignant haematopoiesis: New therapeutic targets for clinical medicine, *Expert Rev Mol Med.* 12: 1–21.

Brendel C,Guda S,Renella R,Bauer D.E,Canver M.C,Kim Y,Heeney M.M,Klatt D,Fogel J,Milsom M.D,Orkin S.H,Gregory R.I,Williams D.A (2016) Lineage-specific BCL11A knockdown circumvents toxicities and reverses sickle phenotype, *J Clin Invest.*, 126:3868–3878.

Broxmeyer H. E. (2008) Chemokines in haematopoiesis, *Curr Opin Hematol.* 15: 49–58.

Byon J C H and Papayannopoulou T (2013) MicroRNAs : Allies or Foes in Erythropoiesis ?, *J Cell Physiol* 227: 7–13.

Cai Y,Yu X,Hu S and Yu J (2009) A Brief Review on the Mechanisms of miRNA Regulation. *Genomics Proteomics Bioinformatics* 7 : 147–154.

Chandra S, Vimal D, Sharma D, Rai V, Gupta S.C, Chowdhuri D.K (2017) Role of miRNAs in Development and Disease: Lessons Learnt from Small Organisms. *Life Sci.* 185:8-14

Chang H, Yi B, Ma R, Zhang X, Zhao H, Xi Y (2016) CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo, *Sci Rep.* . 6: 1–12.

Chang K, Marran K, Valentine A, Hannon G.J. (2013) Creating an miR30-based shRNA vector, *Cold Spring Harb Protoc.*, 8:631–635.

Chaudhuri A. A, So A.Y, Mehta A, Minisandram A, Sinha N, Jonsson V.D, Rao D S, Connell R.M, Baltimore D (2012) OncomiR miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proc Natl Acad Sci U S A,* 109: 4233–4238.

Chen C. Z,Li L,Lodish F.H,Bartel D.P. (2004) MicroRNAs Modulate Hematopoietic Lineage Differentiation *Science* (303):83–86.

Chen F. X, Smith E. R, and Shilatifard A. (2018) Born to run: Control of transcription elongation by RNA polymerase II, *Nat Rev Mol Cell Biol*. 19: 464–478.

Cheng X, Nostro M.C,Huber T.L,Gadue P,Keller M.G. (2008) Numb mediates the interaction between Wnt and Notch to modulate primitive erythropoietic specification from the hemangioblast, *Development 135: 3447-3458.*

Cho C. H, Lee Y S, Kim E, Hwang E.M, Park J Y,(2015) Physiological functions of the TRPM4 channels via protein interactions, *BMB Rep.* 48: 1–5.

Chotinantakul K and Leeanansaksiri W (2012) Hematopoietic Stem Cell Development, Niches, and Signalling Pathways. *Bone Marrow Res* 2012:270425

Congdon K. L, Voermans C, Ferguson E.C, DiMascio L.N, Uqoezwa M, Zhao C,Reya T (2008) Activation of Wnt Signaling in Hematopoietic Regeneration *Stem Cells*, 26(5): 1202–10.

Corrigan P. M,Dobbin E, Freeburn R.W, Wheadon H. (2008) Patterns of Wnt/Fzd/LRP Gene Expression During Embryonic Hematopoiesis. *Stem Cells and Dev.* 18: 759–772.

Cowell I. G. (1994) Repression versus activation in the control of gene transcription *Trends Biochem Sci* 19: 38–42.

Datta Nand Kontomichalou P (1965) Role of reduced glutathione in human erythrocytes *Nature* 206: Nature 239–241.

de Bruijn M. and Dzierzak E. (2017) Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* 129: 2061–2069.

Denny J B (2006) Molecular Mechanisms, Biological Actions, and Neuropharmacology of the Growth-Associated Protein GAP-43, *Current Neuropharmacol* 4:293–304.

DeVilbiss A.W,Sanalkumar R,Hall B.D,Katsumura K.R,Andrade I.F,Bresnick E.H (2015) Epigenetic Determinants of Erythropoiesis: Role of the Histone Methyltransferase SetD8 in Promoting Erythroid Cell Maturation and Survival *Mol Cell Biol* 35:2073–2087.

Ding M, Liu Y,Liao X,Zhan H,Liu Y,Huang W (2018) Enhancer RNAs (eRNAs): New insights into gene transcription and disease treatment *J Cancer* 9: 2334–2340.

Ding N,Xi J,Li Y,Xie X,Shi J,Zhang Z,Li Y,Fang F,Wang S,Yue W,Pei X,Fang X (2016) Global transcriptome analysis for identification of interactions between coding and non-coding RNAs during human erythroid differentiation, *Front Medicine*, 10:pp. 297–310.

Ding T,Zhou Y,Long R,Chen C,Zhao J,Cui P,Guo M,Liang G,Xu L(2019) DUSP8 phosphatase: Structure, functions, expression regulation and the role in human diseases, *Cell Biosci*. 9:1–13.

Dolores Delgado M. and León J. (2010) Myc roles in hematopoiesis and leukemia, *Genes and Cancer* 1: 605–616.

Dominguez J.R.A,Hu W,Yuan B,Shi J,Park S.S,Gromatzky A.A,Oudenaarden A.V,Lodish F.H (2014) Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation *Blood* 123: 570–581.

Dore L. C,Amigo J.D,Santos C,Zhang Z,Gai X,Tobias J.W,Yu D,Klein A.M,Dorman C,Wu W,Hardison R.C,Paw B.H,Weiss M.J (2008) A GATA-1-regulated microRNA locus essential for erythropoiesis., *Proc Natl Acad Sci U S A* 105: 3333–3338.

Dore L. C,Amigo J.D,Santos C,Zhang Z,Gai X,Tobias J.W,Yu D,Klein A.M,Dorman C,Wu W,Hardison R.C,Paw B.H,Weiss M.J (2008) A GATA-1-regulated microRNA locus essential for erythropoiesis., *Proc Natl Acad Sci U S A* 105: 3333–3338.

Doss J. F,Corcoran D.L,Jima D.D,Telen M.J,Dave S.S,Chi J (2015) A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes., *BMC Genomics*, 16: 952

Dvornyk V, Vinogradova O. and Nevo E. (2003) Origin and evolution of circadian clock genes in prokaryotes., *Proc Natl Acad Sci U S A*, 100: 2495–2500.

Dzierzak E and Philipsen S (2013) Erythropoiesis: Development and differentiation. *Cold Spring Harb Perspect Med* 1;3(4):a011601

Eblen S. T (2018) Extracellular Regulated Kinases: Signaling from Ras to ERK Substrates to Control Biological Outcomes Adv *Cancer Res*, 138:99-142

Fan A.X,Hossain A.M,Stees J,Gavrilova E,Bungert J (2015) Regulation of erythroid cell differentiation by transcription factors, chromatin structure alterations and non coding RNA.In:Fan A.X,Hossain A.M,Stees J,Gavrilova E,Bungert J (eds) Epigenetic Gene expression and regulation Elsevier Academic Press, San Diego,pp 237-264.

Fazi F,Rosa A,Fatica A,Gelmetti V,Marchis M L D,Nervi C,Bozzoni I (2005) A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis, *Cell* 123: 819–831.

Felli N,Fontana L,Pelosi E,Botta R,Bonci D,Fransesco F,Liuzzi F,Lulli V,Morsilli O,Santoro S,Valtieri M,Calin G A,Liu C,Sorrentino A ,Croce C.M,Peschle C (2005) MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation., *Proc Natl Acad Sci U S A* 102:18081–18086.

Felli N,Pedini F,Romania P,Biffoni M,Morsilli O,Castelli G,Santoro S,Chicarella S,Sorrentino A,Peschle C,Marziali G (2009) MicroRNA 223-dependent expression of LMO2 regulates nomal erythropoiesis., *Haematologica* 94: 479–86.

Fellmann C,Hoffmann T,Sridhar V,Hopfgartner B,Muhar M,Roth M,Lai D.Y,Barbosa I.A.M,Kwon J.S,Guan Y,Sinha N,Zuber J(2013) An optimized microRNA backbone for effective single-copy RNAi, *Cell Rep.* 5:1704–1713.

Filhol O,Ciais D,Lajaunie C,Charbonnier P,Foveau N,Vert J.P,Vandenbrouck Y(2012) DSIR: Assessing the Design of Highly Potent siRNA by Testing a Set of Cancer-Relevant Target Genes, *PLoS One* 7:1–11

Forsberg E.C,Prohaska S.S,Katzman S,Heffner G.C,Stuart J.M,Weismann I.L (2005) Differential expression of novel potential regulators in hematopoietic stem cells, *PLoS Genet.* 1:281–294.

Fowler D. K,Williams C,Gerritsen A.T,Washbourne P(2015) Improved knockdown from artificial microRNAs in an enhanced miR-155 backbone: A designers guide to potent multi-target RNAi, *Nucleic Acids Res*, 44:4-8

Fujiwara Y,Browne C.P,Cunniff K,Goff S C,Orkin S H (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1, *Proc Natl Acad Sci U S A* 93: 12355–12358.

Galeev R,Baudet A,Kumar P,Nilsson R A,Nilsson B,Soneji S,Torngren T,Borg A,Kvist A,Larsson J (2016) Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs, *Cell Rep.*14: 2988–3000.

Galeev R,Baudet A,Kumar P,Nilsson R A,Nilsson B,Soneji S,Torngren T,Borg A,Kvist A,Larsson J (2016) Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs, *Cell Rep.*14: 2988–3000.

GeestC. R and Coffer P. J. (2009) MAPK signaling pathways in the regulation of hematopoiesis, *J Leukoc Biol* 86: 237–250.

Gehring A. M, Walker J. E, and Santangelo T. J (2016) Transcription regulation in archaea, *J Bacteriol.* 198: 1906–1917.

Ginder G. D, Gnanapragasam M.N and Mian O. Y. (2008) The Role of the Epigenetic Signal , DNA Methylation , in Gene Regulation During Erythroid Development, *Current Topics in Dev Bio,* 82: 85–116.

Grass J. A,Boyer M E,Pal S,Wu J,Weiss M J,Bresnick E H (2003) GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling, , *Proc Natl Acad Sci U S A* 100: 8811–8816.

Grimm D,Streetz K.L,Jopling C.L,Storm T.A,Pandey K,Davis C.R,Marion P,Salazar F,Kay M.A(2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways, *Nature* 441:537–541

Guo S,Lu J,Schangler R,Zhang H,Wang J.Y,Fox M.C,Purton L.E,Fleming H.H,Cobb B,Merkenschlager M,Golub T.R,Scadden D.T (2010) MicroRNA miR-125a controls hematopoietic stem cell number, *Proc Natl Acad Sci U S A*, 107: 14229–14234.

Guschin, D. Y. et al. (2010) A rapid and general assay for monitoring endogenous gene modification, *Methods Mol Biol.* 649:247-56.

Ha M. and Kim V. N. (2014) Regulation of microRNA biogenesis., *Nat Rev Mol Cell Biol.*, 15: 509–524.

Hahn S. (2005) Structure and function of the RNA Polymerase II transcription machinery, *Nat Struct Mol Biol* 11:394–403.

Halene S, Wang L, Cooper RM, Bockstoce DC, Robbins PB, Kohn DB. Improved expression in hematopoietic and lymphoid cells in mice after transplantation of bone marrow transduced with a modified retroviral vector (1999). *Blood*. 15;94(10):3349-57.

Hattangadi S. M,Wong P,Zhang L,Flygare J,Lodish F.H. (2011) From stem cell to red cell: Regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications, *Blood* 118: 6258–6268.

Hawksworth J,Satchwell T J,Meinders M,Daniels D E,Regan F,Thornton N M,Wilson M C,Dobbe J G,Streekstra G J,Trakarnsanga K,Heesom K J,Anstee D J,Frayne J,Toye A M (2018) Enhancement of red blood cell transfusion compatibility using CRISPR-mediated erythroblast gene editing, *EMBO Mol Med* 10:e8454.

Heo H. S,Kim J H,Lee Y J,Kim S,Cho Y S,Kim C G (2005) Microarray profiling of genes differentially expressed during erythroid differentiation of murine erythroleukemia cells, *Mol Cells* 20: 57–68.

Herrera J. E,Sakaguchi K,Bergel K,Trieschmann L,Nakatani Y,Bustin M (1999) Specific Acetylation of Chromosomal Protein HMG-17 by PCAF Alters Its Interaction with Nucleosomes, *Mol Cell Biol.* 19:3466–3473.

Hino K,Tsuchiya K,Fukao T,Kiga K,Okamoto R,Kanai T,Watanabe M (2008) Inducible expression of microRNA-194 is regulated by HNF-1alpha during intestinal epithelial cell differentiation., *RNA* 14:1433–1442.

Hodges V. M,Rainey S,Lappin T.R,Maxwell A.P(2007) Pathophysiology of anemia and erythrocytosis, *Crit Rev Oncol Hematol 64(2):139-58.*

Hong S H,Kim K S,Oh I H (2015)Concise Review: Exploring miRNAs—Toward a Better Understanding of Hematopoiesis *Stem Cells* 33:1-7

Hope K. J,Cellot S,Ting S B,Macrae T,Mayotte N,Iscove N.N,Sauvageau (2010) An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity, *Cell Stem Cell* 7: 101–113.

Huang J,Liu X,Li D,Shao Z,Cao H,Zhang Y,Trompouki E,Bowman T V,Zon L I,Yuan G,Orkin S H,Xu J(2016) Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis, *Developmental Cell*,36 : 9–23.

Ingley E. (2012) Integrating novel signaling pathways involved in erythropoiesis., *IUBMB life*, 64: 402–410.

Ivanovs A,Rybtsov S,Ng E.S,Stanley E G,Elefanty A.G,Medvinsky A (2017) Human haematopoietic stem cell development: from the embryo to the dish, *Development*, 144: 2323–2337.

Jacobs D. K,Nakanishi N,Yuan D,Camara A,Nichols S.A,Hartenstein V(2007) Evolution of sensory structures in basal metazoa, *Integr Comp Biol*, 47:712–723.

Jafari M,Ghadami E,Dadkhah T,Niaki H A (2019) PI3k/AKT signaling pathway: Erythropoiesis and beyond, *J Cell Physiol*, 234: 2373–2385.

Jeanne Magram, Chada K, & Constantini (1985) Developmental regulation of a cloned adult β-globin gene in transgenic mice, *Nature*,315:338-340

Jeong HH, Kim SY, Rousseaux MWC, Zoghbi HY, Liu Z (2019). Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives. *Genome Res* 29(6):999-1008.

Jonkers I. and Lis J. T. (2015) Getting up to speed with transcription elongation by RNA polymerase II, *Nat Rev Mol Cell Biol* 16: 167–177.

Kaczynski J,Zhang J.S, Ellenrieder V,Conley A,Duenes T,Kester H,Vander Burg B,Urrutia R (2001) The Sp1-like Protein BTEB3 Inhibits Transcription via the Basic Transcription Element Box by Interacting with mSin3A and HDAC-1 Co-repressors and Competing with Sp1, *J Biol Chem*, 276: 36749–36756.

KadriZ,Lefevre C,Goupille O,Penglong T,Granger-Locatelli M,Fucharoen S,Maouche-Chretien L,Leboulch P,Chretien S(2015) Erythropoietin and IGF-1 signaling synchronize cell proliferation and maturation during erythropoiesis, *Genes Dev* 29:2603–2616.

KaikkonenM. U, Lam M. T. Y, and Glass C. K. (2011) Non-coding RNAs as regulators of gene expression and epigenetics, *Cardiovasc Res*., 90: 430–440.

Kang S. H,Perales O,Michaud M,Katz S.G (2017) Bok Promotes Erythropoiesis in a Mouse Model of Myelodysplastic Syndrome. *Ann Hematol*.,98:2089-2096.

Kato H. and Igarashi K. (2019) To be red or white: Lineage commitment and maintenance of the hematopoietic system by the "inner myeloid", *Haematologica* 104: 1919–1927.

Katsumura K. R,Mehta C,Hewitt K. J,Soukup A.A,Andrade I.F,Ranheim E.A,Johnson K.D,Bresnick E.H (2018) Human leukemia mutations corrupt but do not abrogate GATA-2 function, *Proc Natl Acad Sci U S A.* 2019, 115: 10109–10118.

Katsumura K. R. and Bresnick E. H. (2017) The GATA factor revolution in hematology, *Blood*, 129: 2092–2102.

Katzerke C,Madan V,Gerloff D,Hartmann D,Hartmann J,Wurm A A,Muller-Tidow C,Schnittger S,Tenen D G,Niederwieser D,Behre G (2013) Transcription factor C/EBPα-induced microRNA-30c inactivates Notch1 during granulopoiesis and is downregulated in acute myeloid leukemia, *Blood*, 122: 2433–2442.

Keller G,Kennedy M,Papayannopoulou T and Wiles M V (1993) Hematopoietic commitment during embryonic stem cell differentiation in culture., *Mol Cell Biol.*, 13:473–486

Kinkel S. A,Galeev R,Flensburg Christoffer,Keniry A,Breslin K,Gilan O,Lee S,Liu J,Chen K,Gearing L J,Moore D.L,Alexander W.S,Dawson M,Majewski I J,Oshlack A,Larsson J,Blewitt M.E(2015) Jarid2 regulates hematopoietic stem cell function by acting with polycomb repressive complex 2, *Blood* 125: 1890–1900.

Klein E,Vanky F,Ben-Bassat H,Neumann H,Ralph P,Zeuthen J.Polliack A (1976) Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia., *Int J Cancer.*, 18: 421–31.

Knott S. R. V,Maceli A,Erard N,Chang K,Marran K,Zhou X,Gordon A,Demerdash O.E,Wagenblast E,Kim S,Fellmann C,Hannon G J(2014) A Computational Algorithm to Predict shRNA Potency, *Mol Cell* 56:796–807.

Konopatskaya O,Matthews S.A,Harper M.T,Gilio K,Cosemans J M E M,Williams C.M,Navarro M.N,Carter D A,Heemskerk J W M,Leitges M,Cantrell D,Poole A W(2011) Protein kinase C mediates platelet secretion and thrombus formation through protein kinase D2, *Blood* 118:416–424.

Kotaki R,Koyama-Nasu R,Yamakawa N,Kotani A (2017) miRNAs in normal and malignant hematopoiesis, *Int J Mol Sci.*, 18:1-11.

Kowiański P, Lietzau G, Czuba E, Waśkow M, Steliga A, Moryś J (2018). BDNF: A Key Factor with Multipotent Impact on Brain Signaling and Synaptic Plasticity. Cell Mol Neurobiol.38(3):579-593.

Kulczyńska K. and Siatecka M. (2016) A regulatory function of long non-coding RNAs in red blood cell development, *Acta Biochim Pol*, 63:675–680.

KuritaR,Suda N,Sudo K,Miharada K,Hiroyama T,Miyoshi H,Tani K,Nakamura Y(2013) Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells, *PLoS One*, 8: doi: 10.1371/journal.pone.0059890.

Laurenti E and Gottgens B (2018)From haematopoietic stem cells to complex differentiation landscapes,*Nature*,553:418-426

Lechman E. R,Gentner B,Galen P.V,Giustacchini A,Saini M,Boccalatte F E,Hiramatsu H,Restuccia U,Bachi A,Voisin V,Bader G D,Dick J E,Naldini L (2012) Attenuation of miR-126 activity expands HSC in vivo without exhaustion, *Cell Stem Cell.*, 11: 799–811.

Lee Y,Kim M,Han J,Yeom K H,Lee S,Baek S H,Kim V N (2004) MicroRNA genes are transcribed by RNA polymerase II., *EMBO J,* 23:4051–4060.

LentoW,Congdon K,Voermans C,Kritzik M,Reya T (2013) Wnt Signaling in Normal and Malignant Haematopoiesis, *Cold Spring Harb Perspect Biol*, 5:1-10

Lessard S,Beaudoin M,Orkin S H,Bauer D E,Lettre G (2018) 14q32 and let-7 microRNAs regulate transcriptional networks in fetal and adult human erythroblasts, *Hum Mol Genet* 27:1411–1420.

Levesque J. Pand Winkler I. G. (2016) Cell Adhesion Molecules in Normal and Malignant Hematopoiesis: from Bench to Bedside, *Current Stem Cell Rep.*2:356–367.

LimL. P,Glasner M E,Yekta S,Burge C.B,Bartel D.P.(2003) Vertebrate microRNA genes., *Science* 299: 1540.

Ling K. W,Ottersbach K,Hamburg J P,Oziemlak A,Tsai F,Stuart O H,Ploemacher R,Hendricks R W,Dzierkak E (2004) GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells, *J Exp Med*, 200: 871–882

Liu J,Han X and An X. L (2015) Novel methods for studying normal and disordered erythropoiesis, *Sci China Life Sci.*, 58: 1270–1275.

Liu J,Li Y,Tong J,Gao J,Guo Q,Zhang L,Wang B,Zhao H,Wang H,Jiang E,Kurita R,Nakamura Y,Tanabe O,Engel J D,Bresnick E H,Zhou J,Shi L (2018) Long non-coding RNA-dependent mechanism to regulate heme biosynthesis and erythrocyte development, *Nat Commun*, 9:1–15.

LiY,Liu D,Zhang X,Li Z,Ye Y,Liu Q,Shen J,Chen Z,Huang H,Liang Y,Han X,Liu J,An X,Mohandas N,Xu X,(2018) miR-326 regulates HbF synthesis by targeting EKLF in human erythroid cells, *Exp Hematol,* 63:33-40.

Lu J,Guo S,Ebert B J,Zhang H,Peng X,Bosco J,Pretz J,Schlanger R,Wang J Y,Mak R.H,Dombkowski D M,Preffer F I,Scadden D T,Golub T R (2008) MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors., *Dev Cell.*, 14: 843–853.

Luis T. C,Ichii M,Brugman M H,Kincade P,Staal F.J.T(2012) Wnt signaling strength regulates normal hematopoiesis and its deregulation is involved in leukemia development, *Leukemia* 26: 414–421

Luis T. C,Naber B.A.E,Roozen P.P.C,Brugman M.H,Haas E.F.E,Ghazvini M,Fibbe W.E,Dongen J.J.M,Fodde R,Staal F.J.T (2011) Canonical wnt signaling regulates hematopoiesis in a dosage-dependent fashion, *Cell Stem Cell*,9: 345–356.

Lulli V,Romania P,Morsilli O,Cianciulli P,Gabbianelli M,Testa U,Giulani A,Marziali G (2013) MicroRNA-486-3p Regulates γ-Globin Expression in Human Erythroid Cells by Directly Modulating BCL11A, *PLoS One*, 8:1-11

Lulli V,Romania P,Morsilli O,Cianciulli P,Gabbianelli M,Testa U,Giulani A,Marziali G (2013) MicroRNA-486-3p Regulates γ-Globin Expression in Human Erythroid Cells by Directly Modulating BCL11A, *PLoS One*, 8:1-11

Lushchak V. I. (2012) Glutathione Homeostasis and Functions: Potential Targets for Medical Interventions, *J of Amino Acids* 2012:1–26.

Ma J. (2011) Transcriptional activators and activation mechanisms, Protein and Cell, 2: 879–888.

MacFarlane L.-A. and R. Murphy P. (2010) MicroRNA: Biogenesis, Function and Role in Cancer, *Curr Genomics*, 11: 537–561

Malik N, Sansom O. J, and Michie A. M. (2018) The role of mTOR-mediated signals during haemopoiesis and lineage commitment, Biochem Soc Trans, 46: 1313–1324.

Manley, G. et al. (2017) Developmental control of Polycomb subunit composition by GATA factors mediates a switch to non-canonical functions, *PLoS One*, 32 :736–740.

Maraia R. J. (2001) La protein and the trafficking of nascent RNA polymerase III transcripts, *J Cell Biol*, 153: 13–17.

Martin P. and Papayannopoulou T. (1982) HEL cells: A new human erythroleukemia cell line with spontaneous and induced globin expression, *Science*, 216 :1233–1235.

MaY,Wang B,Jiang F,Wang D,Liu H,Yan Y,Dong H,Wang F,Gong B,Zhu Y,Dong L,Yin H,Zhang Z,Zhao H,Wu Z,Zhang J,Zhou J,Yu J(2013) A Feedback Loop Consisting of MicroRNA 23a/27a and the β-Like Globin Suppressors KLF3 and SP1 Regulates Globin Gene Expression, *Mol Cell Biol* 33: 3994–4007.

McGinty R. K. and Tan S. (2015) Nucleosome structure and function, *Chem Rev.*, 115: 2255–2273.

Megraw M, Sethupathy P, Corda B, Hatzigeorgiou A.G (2007) miRGen: a database for the study of animal microRNA genomic organization and function., *Nucleic acids Res.*, 35:149-155.

MelloF. V, Land M.G.P, Costa E.S, Teodosio C, Sanchez M, Barcena P, Peres R.T,Pedreira C.E,Alves L.R,Orfao A(2019) Maturation-associated gene expression

profiles during normal human bone marrow erythropoiesis, *Cell Death Discov.* 28;5:69

Merryweather Clarke A.T, Atzberger A, Soneji S, Gray N, Clark K, Waugh C, Mcgowan S J, Taylor S,Nandi A.K,Wood W.G,Roberts D.J,Higgs D. R,Buckle V.J,Robson K.J.H (2011) Global gene expression analysis of human erythroid progenitors, *Blood*, 117: 96–108.

Michlewski G. and Cáceres J. F. (2019) Post-transcriptional control of miRNA biogenesis, *RNA*, 25:1–16

Mohammdai-Asl J,Ramezani A,Norozi F,Malehi A.S,Asnafi A.A,Far M.A.J,Mousavi S H,Saki N (2015) MicroRNAs in erythropoiesis and red blood cell disorders, *Front. Biol.*, 10: 321–332.

Molina C. A,Foulkes N.S,Lalli E,Sassone-Corsi P(1993) Inducibility and negative autoregulation of CREM: An alternative promoter directs the expression of ICER, an early response repressor, *Cell*, 75: 875–886.

Monkley S. J,Kostourou V,Spence L,Petrich B,Coleman B,Ginsberg M.H,Pritchard C.A,Critchley D.R (2011) Endothelial cell talin1 is essential for embryonic angiogenesis, *Dev Biol.* 349: 494–502.

Morozova N,Zinovyev A,Nonne N,Pritchard L,Gorban A.N,Harel-Bellan A (2012) Kinetic signatures of microRNA modes of action, *RNA*, 18:1635–1655.

Antoniou M. Induction of Erythroid-Specific Expression in Murine Erythroleukemia (MEL) Cell Lines. *Methods Mol Biol*. 1991;7:421-34..

Nimmo R. A, May G. E,and Enver T. (2015) Primed and ready: Understanding lineage commitment through single cell analysis, *Trends in Cell Biol.* 25:459–467.

Niu Y,Shen B,Cui Y,Chen Y,Wang J,Wang L,Kang Y,Zhao X,Si W,Li W,Xiang A P,Zhou J,Guo X,Bi Y,Si C,Hu B,Dong G,Wang H,Zhou Z,Li T,Tan T,Pu X,Wang F,Ji S,Zhou Q,Huang X,Ji W,Sha J (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos *Cell*, 156: 836–843.

OBrien J,Hayder H,Zayed Y,Peng C (2018) Overview of microRNA biogenesis, mechanisms of actions, and circulation, *Front Endocrinol (Lausanne)* , 9:1–12.

Oberlin E,Tavian M,Blazsek I,Peault B (2002) Blood-forming potential of vascular endothelium in the human embryo, *Development*, 129: pp. 4147–41

Ohishi K,Katayama N,Shiku H,Varnum-Finney B,Bernstein I.B (2003) Notch signalling in hematopoiesis, *Semin Cell Dev Biol*., 14: 143–150

Orkin S. H. and Zon L. I. (2008) Hematopoiesis: an evolving paradigm for stem cell biology., *Cell*, 132: 631–644.

Pangou E,Befani C,Mylonis I,Samiotaki M,Panayotou G,Simos G,Liakos P(2016) HIF-2α phosphorylation by CK1δ promotes erythropoietin secretion in liver cancer cells under hypoxia, *J Cell Sci* 129:4213–4226.

Papapetrou E. P, Korkola J. E, and Sadelain M. (2010) A genetic strategy for single and combinatorial analysis of miRNA function in mammalian hematopoietic stem cells, *Stem Cells*, 28 :287–296.

Patrick D. M,Zhang C.C,Tao Y,Yao H,Qi X,Schwartz R.J,Huang L.J,Olson E.N(2010) Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3zeta., *Genes Dev* 24:1614–1619.

Paulson RF, Ruan B, Hao S, Chen Y (2020). Stress Erythropoiesis is a Key Inflammatory Response. *Cells*. 6;9(3):634.

Pelossof R,Fairchild L,Huang C,Widmer C,Sreedharan V.T,Sinha N,Lai D, Guan Y, Premsrirut P.K,Tschaharganeh D.F, Hoffmann T, Thapar V,Xiang Q, Ralph J. Garippa R J,Rätsch G, Zuber J, Lowe S.W, Leslie C.S,Fellmann C (2017) Prediction of ultra-potent shRNAs with a sequential classification algorithm. *Nat Biotechnol* 35:350-353.

Petrillo C, Thorne LG, Unali G, Schiroli G, Giordano AMS, Piras F, Cuccovillo I, Petit SJ, Ahsan F, Noursadeghi M, Clare S, Genovese P, Gentner B, Naldini L, Towers GJ, Kajaste-Rudnitski A (2018). Cyclosporine H Overcomes Innate Immune Restrictions to Improve Lentiviral Transduction and Gene Editing In Human Hematopoietic Stem Cells. *Cell Stem Cell* 23:820-832.

Porcher C, Chagraoui H and Kristiansen M. S (2017) SCL/TAL1: A multifaceted regulator from blood development to disease, *Blood*, 129: 2051–2060.

Porrua O. and Libri D. (2015) Transcription termination and the control of the transcriptome: Why, where and how to stop, *Nat Rev Mol Cell Biol* 16: 190–202.

Pulido-Quetglas C,Aparicio-Prat E,Arnan C,Polidori T,Hermoso T,Palumbo E,Ponomarenko J,Guigo R,Johnson R (2017) Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion, *PLoS Comput Biol*, 13:1–18.

Rak J,Foster K,Potrzebowska K,Talkhoncheh M S,Miharada N,Komorowska K,Torngren T,Kvist A,Borg A,Svensson L,Bonnet D,Larsson J (2017) Cytohesin 1 regulates homing and engraftment of human hematopoietic stem and progenitor cells, *Blood*, 129: 950–958.

Rasmussen K. D,Simmini S,Abreu-Goodger C,Bartonicek N,Giacomo M D,Bilbao-Cortes D,Horos R,Lindern M.V,Enright A.OCarroll D J. (2010) The miR-144/451 locus is required for erythroid homeostasis., *J Exp Med*, 207: 1351–1358.

Ren L, Xiao RD, Zhang Q, Lou XM, Zhang ZJ, Fang XD (2018). [Synergistic regulation of the erythroid differentiation of K562 cells by KLF1 and KLF9]. *Yi Chuan*. 40(11):998-1006..

Rivkin N,Chapnik E,Mildner A,Barshtein G,Porat Z,Kartvelishvily E Dadosh T,Birger Y,Amir G,Yedgar S,Izraeli S,Jung S,Hornstein Eet al. (2017) Erythrocyte survival is controlled by microRNA-142, *Haematologica* 102:676–685.

Robb L and Begley C. G. (1997) The SCL/Tal1 gene:roles in normal and malignant haematopoiesis, *Bioessays*,19:607-613.

Robin C,Ottersbach K,Durand C,Peeters M,Vanes L,Tybulewicz V,Dzierzak E (2006) An Unexpected Role for IL-3 in the Embryonic Development of Hematopoietic Stem Cells, *Dev Cell*, 11:171–180.

Rodrigues N. P,Tipping A.J,Wang Z,Enver T (2012) GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia, *Int J Biochem Cell Biol*, 44: 457–460.

Romanello KS, Teixeira KKL, Silva JPMO, Nagamatsu ST, Bezerra MAC, DomingosIF, Martins DAP, Araujo AS, Lanaro C, Breyer CA, Ferreira RA, Franco-Penteado C,Costa FF, Malavazi I, Netto LES, de Oliveira MA, Cunha AF (2018). Global analysis of erythroid cells redox status reveals the involvement of Prdx1 and Prdx2 in the severity of beta thalassemia. *PLoS One* 13:0208316.

Rouzbeh S,Kobari L,Cambot M,Mazurier C,Hebert N,Faussat A,Durand C,Douay C,Lapillonne H(2015) Molecular signature of erythroblast enucleation in human embryonic stem cells., *Stem cells* 33 : 2431–2441.

Rupaimoole R and Slack F. J. (2017) MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases, *Nat Rev Drug Discov* 16: 203–221.

Sainsbury S, Bernecky C and Cramer P. (2015) Structural basis of transcription initiation by RNA polymerase II, *Nat Rev Mol Cell Biol.* 16:129–143.

Saito R,Tabata Y,Muto A,Arai K.I,Watanabe S (2005) Melk-Like Kinase Plays a Role in Hematopoiesis in the Zebra Fish, *Mol Cell Biol* 25:6682–6693.

Salviano-Silva A,Lobo-Alves S.C,Almeida R.C,Malheiros D,Petzl-Erler M.L (2018) Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis, *Non-Coding RNA*, 4: 1-30

Sangokoya C,Telen M. J,Chi J.-T. (2010) microRNA miR-144 modulates oxidative stress tolerance and associates with anemia severity in sickle cell disease., *Blood* 116:4338–4348.

Sankaran V. G,Menne T.F,Scepanovic D,Vergilio J,Ji P,Kim J,Thiru P,Orkin S.H,Lander E.S,Lodish H.F(2011) MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13 *Proc Natl Acad Sci U S A*; 108: 1519–1524.

Schramm L and Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters, *Genes Dev*, 16: 2593–2620.

Sertorio M,Du W,Amarachintha S,Wilson A,Pang Q (2017) In Vivo RNAi Screen Unveils PPARγ as a Regulator of Hematopoietic Stem Cell Homeostasis, *Stem Cell Rep.*8:1242–1255.

Shandilya J and Roberts S. G. E. (2012) The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*,1819: 391–400.

Shi G, Ozog S, Torbett BE, Compton AA. mTOR inhibitors lower an intrinsic barrier to virus infection mediated by IFITM3 (2018). *Proc Natl Acad Sci U S A*. 115(43):10069-10078.

Shi L,Lin Y,Sierant M.C,Zhu F,Cui S,Guan Y,Sartor M.A,Tanabe O,Lim K,Engel J D (2014) Developmental transcriptome analysis of human erythropoiesis., *Human molecular genetics*, 23: 4528–4542.

Siatecka M and Bieker J. J. (2011) The multifunctional role of EKLF / KLF1 during erythropoiesis, *Blood*, 118: 2044–2055.

Singh N, Sahu T.K, Rao A.R, Mohapatra T(2012) shRNAPred (version 1.0): An open source and standalone software for short hairpin RNA (shRNA) prediction, *Bioinformation* 8:629-633.

Singh V.K, Saini A, Kalsan M, Kumar N, Chandra R (2016) Stage-Specific Regulation of Erythropoiesis and Its Implications in Ex-Vivo RBCs Generation *J Stem Cells* 11:149-169.

Song B, Wang Y, Kudo K, Gavin E J, Xi Y, Ju J(2008) miR-192 Regulates dihydrofolate reductase and cellular proliferation through the p53-microRNA circuit, *Clinical Cancer Research* 14:8080–8086.

Spence JS, He R, Hoffmann HH, Das T, Thinon E, Rice CM, Peng T, Chandran K, Hang HC (2019). IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nat Chem Biol. 15(3):259-268.*

Stenvang J, Petri A, Lindow M, Obad S, Kauppinen S(2012) Inhibition of microRNA function by antimiR oligonucleotides, *Silence* 3:1

Tallack M. R, Magor G.W, Dartigues B, Sun L, Huang S, Fittock J.M, Fry S.V, Glazov E.A, Bailey T.L,Perkins A.C (2012) Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq, *Genome Research*, 22: 2385–2398.

Tavian M, Hallais M. F and Péault B. (1999) Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo, *Development*, 126: 793–803.

Thiel G, Lietz M and Hohl M. (2004) How mammalian transcriptional repressors work, *European Journal of Biochemistry*, 271: 2855–2862.

Thomas D, Vadas M and Lopez A. (2004) Regulation of haematopoiesis by growth factors – emerging insights and therapies, *Expert Opinion on Biological Therapy*, 4 :869–879.

Tondeleir D, Drogat B, Slowicka K, Bakkali K, Bartunkova S, Goossens S, Haigh JJ, Ampe C. (2013) Beta-Actin Is Involved in Modulating Erythropoiesis during Development by Fine-Tuning Gata2 Expression Levels. *PLoS One*. 26;8(6):e67855.

Treiber T, Treiber N and Meister G. (2019) Regulation of microRNA biogenesis and its crosstalk with other cellular pathways, *Nat Rev Mol Cell Biol*. 20: 5–20.

Tsai F. Y,Keller G,Kuo F.C,Weiss M,Chen J,Rosenblatt M,Alt F.W,Orkin S H (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2, *Nature*,371: 221–226.

Tsiftsoglou A. S, Vizirianakis I. S and Strouboulis J. (2009) Erythropoiesis: Model systems, molecular regulators, and developmental programs, *IUBMB Life*, 61: 800–830.

Utsugisawa T,Uchiyama T,Toki T,Ogura H,Aoki T,Hamaguchi I,Ishiguro A,Ohara A,Kojima S,Ohga S,Ito E,Kanno H(2016) Erythrocyte glutathione is a novel biomarker of Diamond-Blackfan anemia, *Blood Cells Mol Dis.*. 59:31–36.

Valinezhad Orang A, Safaralizadeh R and Kazemzadeh-Bavili M. (2014) Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation *Int J Genomics.*, 2014: 970607

Van Den Berg D. J.,Sharma A K,Bruno E,Hoffman R (1998) Role of members of the Wnt gene family in human hematopoiesis. *Blood* 92 :3189–3202.

Van Vuren A. J,Gaillard C.A.J.M,Eisenga M.F,Wijk R,Beers E.J(2019) The EPO-FGF23 signaling pathway in erythroid progenitor cells: Opening a new area of research, *Front Physiol.* 10:1–16.

Varma N,Janic B,Ali M,Iskander A,Arbab A (2011) Lentiviral based gene transduction and promoter studies in human hematopoietic stem cells (hHSCs), *J Stem Cells Regen Med*., 7: 41–53.

Vicente C,Conchillo A,Garcia-Sanchez M.A,Odero M.D (2012) The role of the GATA2 transcription factor in normal and malignant hematopoiesis, *Crit Rev Oncol Hematol.* 82: 1–17.

Volker K Haase (2014) Regulation of erythropoiesis by hypoxia inducible factors, *Blood Rev* 27: 41–53.

Wahid F,Shehzad A,Khan T,Kim Y.Y (2010) MicroRNAs: Synthesis, mechanism, function, and recent clinical trials, *Biochim Biophys Acta Mol Cell Res*,1803:1231–1243.

Wang Q,Huang Z,Xue H,Jin C,Ju X,Han J.J,Chen Y (2008) MicroRNA miR-24 inhibits erythropoiesis by targeting activin type I receptor ALK4. *Blood*, 111: 588–595.

Warren A. J,Colledge W H,Carlton M.B,Evans M.J,Smith A.J,Rabbitts T H (1994) The Oncogenic Cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development, *Cell* 78:45–57

Wery M,Kwapisz M and Morillon A. (2011) Noncoding RNAs in gene regulation, *Wiley Interdiscip Rev Syst Biol Med.* 3: 728–738.

Westholm JO, Lai EC. (2013) Mirtrons: microRNA biogenesis via splicing. *Biochimie*. 93(11):1897-904.

Wodnar-Filipowicz A. (2003) Flt3 ligand: role in control of hematopoietic and immune functions of the bone marrow., *News Physiol Sci* 18: 247–51.

Xu J. and Wong C. (2008) A computational screen for mouse signaling pathways targeted by microRNA clusters, *RNA* 14:1276–1283.

Yan B. X. and Ma J. X. (2012) Promoter-associated RNAs and promoter-targeted RNAs, *Cell Mol Life Sci.*, 69: 2833–2842.

Yang J,Ma J,Xiong Y,Wang Y,Jin K,Xia W,Chen Q,Huang J,Zhang J,Jiang N,Jiang S,Ma D(2018) Epigenetic regulation of megakaryocytic and erythroid differentiation by PHF2 histone demethylase, *J Cell Physiol*, 233: 6841–6852.

Yokomizo T,Hasegawa K,Ishitobi H,Osato M,Ema M,Ito Y,Yamamoto M,Takahashi S (2008) Runx1 is involved in primitive erythropoiesis in the mouse, *Blood* 111:4075–4080.

Yu D,Santos C,Zhao G,Jiang J,Amigo J.D,Khandros E,Dore L.C,Yao Y,DSouza J,Zhang Z,Ghaffari S,Choi J,Friend S,Tong W,Orange J.S,Paw B.H,Weiss M.J(2010) miR-451 protects against erythroid oxidant stress by repressing 14-3-3 z, *Genes Dev* 24: 1620–1633.

Yu J,Wang F,Yang G,Wang F,Ma Y,Du Z,Zhang J (2006) Human microRNA clusters: genomic organization and expression profile in leukemia cell lines., *Biochem Biophys Res Commun* 349: 59–68.

Yu X,Zhan X,DCosta J,Tanavde V.M,Ye Z,Peng T,Malehorn M.T,Yang X,Civin C I,Cheng L(2003) Lentiviral vectors with two independent internal promoters transfer high-level expression of multiple transgenes to human hematopoietic stem-progenitor cells, *Mol Ther*. 7:827–838.

Zermati Y,Fichelson S,Valensi F,Freyssinier J M,Rouyer Fessard P,Cramer E,Guichard J,Varet B,Hermine O(2000) Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors, *Exp Haematology* 28: 885–894.

Zhang L,Flygare J,Wong P,Lim B,Lodish H.F(2011) miR-191 regulates mouse erythroblast enucleation by down-regulating Riok3 and Mxi1., *Genes Dev.*, 25:119–24.

Zhang Y,Li L,Yu C,Senyuk V,Li F,Quigley J.G,Zhu T,Qian j(2018) MiR-9 upregulation leads to inhibition of erythropoiesis by repressing FoxO3, *Sci Rep.* 8:1–8.

Zhao H,Kalota A,Jin S,Gewirtz A.M (2009) The c-myb proto-oncogene and microRNA-15a comprise an active autoregulatory feedback loop in human hematopoietic cells, *Blood*, 113:505–516.

Zhu H,Foretz M,Xie Z,Zhang M,Zhu Z,Xing J,Leclerc J,Gaudry M,Viollet B,Zou M H(2014) PRKAA1/AMPKα1 is required for autophagydependent mitochondrial clearance during erythrocyte maturation, *Autophagy* 10:1522–1534.

Zivot A,Lipton J.M,Narla A,Blanc L(2018) Erythropoiesis : insights into pathophysiology and treatments, *Molecular Med.*, 24: 1–15.

Untitled

by Shaji

General metrics

88,552	13,123	563	52 min 29 sec	1 hr 40 min
characters	words	sentences	reading time	speaking time

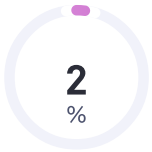
Score

This text scores better than 97% of all texts checked by Grammarly

Writing Issues

266	32	234
Issues left	Critical	Advanced

Plagiarism



24 sources

2% of your text matches 24 sources on the web or in archives of academic publications

Untitled

by Shaji

General metrics

87,193	13,369	648	53 min 28 sec	1 hr 42 min
characters	words	sentences	reading time	speaking time

Score

This text scores better than 97% of all texts checked by Grammarly

Writing Issues

272	34	238
Issues left	Critical	Advanced

Plagiarism

53 sources

4% of your text matches 53 sources on the web or in archives of academic publications

Untitled

by Shaji

General metrics

32,762	4,940	223	19 min 45 sec	38 min 0 sec
characters	words	sentences	reading time	speaking time

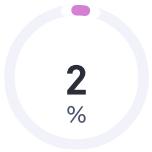
Score

This text scores better than 96% of all texts checked by Grammarly

Writing Issues

109	7	102
Issues left	Critical	Advanced

Plagiarism



10 sources

2% of your text matches 10 sources on the web or in archives of academic publications