



# **The Role of S100A4 in Acute Myeloid Leukaemia**

by

**Bader Alanazi**

November 2019

A thesis presented to Cardiff University in partial  
fulfilment of the requirement for the degree of  
Doctor of Philosophy

Department of Haematology, Division of Cancer & Genetics  
School of Medicine, Cardiff University,  
CF14 4XN Cardiff, United Kingdom

Supervisors:

Dr. Alex Tonks, Dr. Steven Knapper, and Professor Richard Darley

---

*This work is dedicated to;*  
*My Parents; Shayah and Fatimah Alanazi who always encouraged me to*  
*follow my dreams*  
*and*  
*My wife; Norah Alohlie who supported me and stood by myside throughout*  
*the course of this journey*

---

## **Acknowledgements**

I would like to express my deepest thanks and gratitude to my primary supervisor, Dr. Alex Tonks and secondary supervisor, Professor Richard Darley, for their continuous support, patience, and priceless advice. Thanks for being there in the most stretching time for discussion, great suggestions and guidance throughout my project. Indeed, I am fortunate to work with such a considerate and encouraging individuals. I would also like to extend my sincere thanks to my secondary clinical supervisor Dr. Steven Knapper for his support and guidance in some clinical aspects of this project despite his busy schedule with his patients.

I would also like to extend my sincere thanks to Dr. Paul Hole and Dr Namrata Rastogi, for their scientific advice and technical assistance around particular aspects of this project.

Great thanks to Sara Davies, Sandra Kreuser, and Sarah Baker for their training and all the technical assistance they have given me throughout the past four years.

Many thanks for Michelle Doyle, Sarah Baker, Dr. Amanda Gilkes, and Dr. Marie Hodge for their help in getting AML patients' samples from the AML Clinical Trials Research Tissue Bank.

I also would like to thank Dr. Charlotte Wilhelm-Benartzi and Dr. Robert Hills from the Clinical trial centre (Cardiff) for their help in the statistical data analysis.

I also would like to acknowledge Dr. Andy Tee, Dr. Paul Brennan, Dr. Ian Brewis (Cardiff University), Dr. Kate Heesom (University of Bristol Proteomic Facility) for their valuable advice on proteomic study design and data analysis.

I am particularly grateful to Dr. Nahar Alazemi, former CEO of Prince Mohammed Medical City (PMMC) for nominating me for this scholarship and his words of encouragements.

At the end, I would to thank Prince Mohammed Medical City (PMMC), Al-Jouf, Saudi Arabia for granting me this scholarship and funding this project.

---

## Forward

A Google Drive is provided below, which contains the following:

- LCMS raw proteomic data (Chapter 3, section 3.3.2.2);
- All AML patients' information who their samples used in this study (Chapter 3);
- LCMS raw and analysed proteomic data (cytoplasmic and nuclear) (Chapter 5, section 5.3.3.1);
- An electronic version of published article of this study

**Google Drive Link:**

<https://drive.google.com/open?id=1fGmrG1zVJq7ITW4s2lr3x1glFOFrScXr>

---

## **Publications and presentations**

### **Publications**

- Alanazi, B., Munje, C. R., Rastogi, N., *et al.* Integrated nuclear proteomics and transcriptomics identifies S100A4 as a therapeutic target in acute myeloid leukemia. *Leukemia*

### **Conferences proceedings:**

- Bader Alanazi<sup>1\*</sup>, Steven Knapper<sup>1</sup>, Richard Darley<sup>1</sup>, Alex Tonks<sup>1</sup>, “Acute myeloid leukemia cells are dependent on S100A4 for growth and survival”, EHA 2019; the 24<sup>th</sup> congress of European Hematology Association, Amsterdam, the Netherlands. 13-16 June 2019.

**(Poster presentation with oral presentation pitch in the main AML Session)**

- Bader Alanazi<sup>1\*</sup>, Andrew Pierce<sup>2</sup>, Anthony D. Whetton<sup>2</sup>, Steven Knapper<sup>1</sup>, Chinmay R. Munje<sup>3</sup>, Richard Darley<sup>1</sup>, Alex Tonks<sup>1</sup>, “The role of S100A4 in acute myeloid leukemia”, EHA 2018; the 23<sup>rd</sup> congress of European Hematology Association, Amsterdam, the Netherlands. 14-17 June 2018.

**(Accepted for publication only)**

### **Oral presentations**

- “The role of S100A4 in acute myeloid leukemia”, Division of Cancer and Genetics seminar, School of Medicine, April 2018.
- “The role of S100A4 in acute myeloid leukemia” *30<sup>th</sup> Annual School of Medicine Postgraduate Research Day*. January 2018.
- “The role of S100A4 in acute myeloid leukemia”, *Annual Division of Cancer and Genetics Postgraduate Research Day*, School of Medicine, October 2017.
- “The role of S100A4 in acute myeloid leukemia” *29<sup>th</sup> Annual School of Medicine Postgraduate Research Day*. January 2017.
- “The role of S100A4 in acute myeloid leukemia” *Haematology Dept. Seminar*. (March 2016, October 2016, May 2017, November 2017, February 2018)

### **Awards**

- Travel grant winner for the 24<sup>th</sup> congress of European Hematology Association (EHA 2019) in Amsterdam, the Netherlands.
- 2<sup>nd</sup> Place winner of 3 minutes PhD thesis presentation at *29<sup>th</sup> Annual School of Medicine Postgraduate Research Day*. January 2017.

---

## Abstract

Nuclear mislocalization of proteins can interfere with normal cellular function and cooperatively drive tumour development. To understand how this process mediates AML development, the nuclear proteome of AML blasts was analyzed in comparison with normal human CD34<sup>+</sup> cells to identify misregulated nuclear proteins. This study identified that S100A4 as the most significant and fold changing protein in AML blasts which has not been previously associated with AML. *S100A4* belongs to the S100 multi-gene family of calcium-binding proteins of the EF-hand type and has been implicated in tumour progression and metastasis in many solid tumours but little is known of its role in haematological malignancy. Using western blotting, S100A4 protein expression was observed in the nucleus of AML blasts FAB M1 (83%; 24/33) and 44% FAB M4 (4/9) whilst normal CD34<sup>+</sup> or CD14<sup>+</sup> differentiated monocytic controls have shown only cytosolic expression of S100A4. An independent dataset (TCGA) supports the overexpression of *S100A4* mRNA in AML and suggests that overexpression may confer a poor prognosis (p=0.0118). To determine whether ectopic expression of nuclear S100A4 can affect the growth and survival of normal hematopoietic cells, CD34<sup>+</sup> cells were infected with lentiviral vectors expressing nuclear-targeted S100A4. Overexpression of nuclear S100A4 could not be demonstrated in transduced CD34<sup>+</sup> cells or in normal differentiated cells (probably due to rapid degradation of ectopically expressed S100A4 in these cells). To examine functional significance of S100A4 expression on normal and leukaemic cells, S100A4 expression was knocked down. In CD34<sup>+</sup> cells, no significant effects were observed on the growth or lineage development of these cells suggesting S100A4 is not required for normal hematopoiesis. Conversely, knocking down S100A4 expression in AML lines (NOMO-1, TF-1, THP-1, and OCI-AML2) showed significant reduction in growth and induced cell death through apoptosis suggesting that AML cells are dependent on S100A4 for their growth and survival. Further, to identify the binding partners of S100A4 through which mediates its functions, a co-immunoprecipitation coupled with LC/MS was performed on cytoplasmic and nuclear extract of AML cell line (ME-1) under Ca<sup>2+</sup> enriched conditions. Heterogeneous nuclear Ribonucleoprotein M (hnRNPM) was identified as novel binding partner of S100A4 AML. These findings suggest that therapeutically targeting S100A4 would be an effective strategy in AML while sparing normal hematopoietic cells.

---

## List of Abbreviations

Ab	Antibody
ABL	Abelson murine leukemia viral oncogene homolog 1
ADH	Alcohol dehydrogenase
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
AML1	Acute myeloid leukemia 1 protein
ATP	Adenosine triphosphate
BCL	B-cell lymphoma
BSA	Bovine serum albumin
cDNA	Complementary DNA
CEB	Cell extraction buffer
CH	Clonal Hematopoiesis
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitor
C/N	Cytoplasmic/nuclear
DDX21	DEAD-box helicase 21
dH <sub>2</sub> O	Distilled H <sub>2</sub> O
DMEM	Dulbecco's Modified Eagles Medium
DMP	Dimethyl phthalate
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribose nucleoside triphosphate
DTT	Dithiothreitol
EGFP	Enhanced green fluorescent protein
FAB	French-American-British
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
FLT3 -ITD	FMS-like tyrosine kinase 3- Internal Tandem
FSC	Forward scatter
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
G-CSF	Granulocyte colony stimulating factor
GFP	Green fluorescent protein
GM-CSF	Granulocyte-macrophage colony-stimulating factor
HEPES	4-(2hydroxyethyl)-1-piperazine-ethanesulfonic acid
HIST1H2AA	Histone cluster 1 H2A family member A
hnRNPM	heterogeneous nuclear ribonucleoprotein M
HPLC	High-performance liquid chromatography
HRP	Horseshoe peroxidase
HSC	Haematopoietic stem cell
Ig	Immunoglobulin
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
IP	Immunoprecipitation
ITD	Internal tandem duplication

---

ITRAQ	Isobaric tags for relative and absolute quantitation
KD	Knockdown
KI	Knock-in
LDS	Lithium-dodecyl sulphate
LRP	Low density lipoprotein receptor-related protein
LSC	Leukemic stem cell
LT-HSC	Long term haematopoietic stem cell
LTR	Long terminal repeat
MACS	Magnetic-antibody cell sorting
MAK16	MAK16 homolog
MCM7	MiniChromosome Maintenance complex component 7
miRNA	microRNA
MRC	Medical Research Council
MS	Mass spectrometry/mass spectrometric
NLS	Nuclear localization signal
NPM1	Nucleophosmin
NUP	Nucleoporin
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	PBS Tween
PIC	Protease inhibitor cocktail
PML	Promyelocytic leukemia protein
PP	Protein phosphatase
PPI	Protein-Protein Interaction
RAG1	Recombination activating gene 1
RNA	Ribonucleic acid
RRP1B	Ribosomal RNA Processing 1B
SCF	Stem cell factor
SDF	Stromal cell-derived factor
SDS	Sodium dodecyl sulphate
shRNA	Short hairpin RNA
SF3B1	Splicing Factor 3B subunit 1
SUPT16H	SPT16 homolog, facilitates chromatin remodeling subunit
SSC	Side scatter
ST-HSC	Short term haematopoietic stem cell
SUMO	Small ubiquitin-like modifier
TBS	Tris-buffered saline
TBST	TBS Tween
TEAB	Triethylammonium bicarbonate
TGF	Transforming growth factor
TMPO	Thymopoietin Lamina-associated polypeptide 2, isoforms beta/gamma
UC	Universal container
UV	Ultra violet
WHO	World Health Organization

---

## Table of contents

1	Chapter 1 - Overview of haematopoiesis.....	2
1.1	Normal haematopoiesis .....	2
1.1.1	Human haematopoietic stem cells.....	3
1.1.2	Regulation of normal haematopoiesis.....	8
1.2	Leukaemia .....	9
1.2.1	Overview of leukaemia .....	9
1.2.2	Acute myeloid leukaemia (AML) .....	10
1.3	S100A family of proteins .....	22
1.3.1	Overview of S100A protein family.....	22
1.3.2	S100 proteins' structure .....	22
1.4	S100A4 protein.....	23
1.4.1	S100A4 function and structure .....	25
1.4.2	Regulation of expression of S100A4 .....	25
1.4.3	Role of S100A4 in cancer pathophysiology.....	26
1.5	Aims and objectives .....	28
2	Chapter 2 - Materials and Methods.....	30
2.1	Reagents .....	30
2.2	Cell Culture and cryopreservation.....	35
2.2.1	Cell culture of suspension cell lines.....	35
2.2.2	Cell culture of adherent cell lines.....	35
2.2.3	Cell culture of haematopoietic progenitor cells .....	37
2.2.4	Cell counting .....	38
2.2.5	Cryopreservation and resuscitation of cells. ....	38
2.3	Isolation of CD34 <sup>+</sup> haematopoietic progenitor cells (HSPCs) from human cord blood 38	
2.3.1	Cord blood collection.....	38

---

2.3.2	Isolation of Mononuclear cells (MNCs) from cord blood .....	38
2.3.3	Immunophenotyping MNCs .....	39
2.3.4	Isolation of CD34 <sup>+</sup> cells from MNCs.....	41
2.3.5	Cell surface phenotyping and differentiation analysis.....	42
2.4	AML patients' samples, FAB subtype determination and viability analysis ...	42
2.5	Preparation of recombinant plasmid DNA .....	42
2.5.1	Restriction enzyme conditions for creation of PINCO constructs .....	42
2.5.2	Agarose gel purification of insert DNA .....	43
2.5.3	Ligation of S100A4 DNA into retroviral or lentiviral vectors.....	44
2.5.4	Transformation of competent cells with plasmid DNA .....	44
2.5.5	Purification of plasmid DNA .....	44
2.5.6	Preparation of retro- and lentivirus .....	45
2.6	Determining the effect of nuclear mislocalised S100A4 on CD34 <sup>+</sup> cells and AML cell lines .....	46
2.6.1	Retro- and lentiviral infection of CD34 <sup>+</sup> cells .....	46
2.6.2	Retro- and lentiviral infection of AML cells lines .....	47
2.6.3	Effect of serum on proliferation of infected AML cell lines .....	47
2.7	Determining the effect of S100A4 knockdown on CD34 <sup>+</sup> cells and AML cell lines	48
2.7.1	Generating shRNA lentiviral vectors .....	48
2.7.2	Growth and apoptosis assays .....	48
2.8	Protein extraction and quantification.....	51
2.8.1	Total Protein extraction.....	51
2.8.2	Nuclear/Cytosol Extraction and Fractionation.....	51
2.9	Antibody optimisation and western blotting .....	52
2.9.1	S100A4 antibodies optimisation .....	52
2.9.2	His-Tag S100A4 recombinant protein .....	52

---

2.9.3	SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and electroblotting .....	52
2.9.4	Immunoblotting.....	55
2.10	Co-Immunoprecipitation (co-IP).....	56
2.10.1	Direct co-IP .....	57
2.10.2	Indirect co-IP.....	58
2.10.3	Florescent staining of SDS-PAGE gel.....	58
2.11	Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis.....	61
2.12	Flow cytometry and data analysis .....	62
2.13	Bioinformatics analysis and online databases .....	64
2.14	Statistical Analysis .....	64
3	Chapter 3 - Introduction.....	67
3.1	Overview .....	67
3.2	Aims and objectives .....	68
3.3	Results .....	69
3.3.1	S100A4 is detected in AML cell lines by three commercial antibodies ...	69
3.3.2	Confirmation of S100A4 mAb specificity .....	74
3.3.3	S100A4 expression levels and subcellular localisation in AML cell lines. 75	
3.3.4	S100A4 expression levels and subcellular localization in normal haematopoietic cells.....	79
3.3.5	S100A4 expression levels and subcellular localization in AML patients (FAB M1) and (FAB M4).....	84
3.3.6	Correlation of upregulated S100A4 to patients' survival, prognosis, and outcomes .....	89
3.4	Discussion .....	93
3.4.1	Optimisation of S100A4 protein detection .....	93

---

3.4.2	S100A4 expression levels and subcellular localisation in normal haematopoiesis and AML .....	95
3.4.3	Expression of <i>S100A4</i> does not correlate with patients' clinical outcome	98
4	Chapter 4 - Introduction.....	102
4.1	Overview .....	102
4.2	Aims .....	103
4.3	Results .....	103
4.3.1	Generating retro- and lentiviral vectors overexpressing S100A4.....	103
4.3.2	Validation of S100A4 ligation into PINCO and pHIV vectors and transduction efficiency in Phoenix cells .....	104
4.3.3	Phenotypic effects of overexpression of nuclear S100A4 on normal human CD34 <sup>+</sup> HSPCs.....	108
4.3.4	Phenotypic effects of over-expression of nuclear mislocalised S100A4 on AML cell lines. ....	111
4.3.5	S100A4 knock down in normal human CD34 <sup>+</sup> HSPC slows proliferation	114
4.3.6	S100A4 KD in normal CD34 <sup>+</sup> HSPCs does not affect cell proliferation, differentiation, and survival .....	117
4.3.7	Leukaemic cells are dependent on S100A4 for their growth and survival	120
4.4	Discussion .....	127
4.4.1	Generating retro- and lentiviral vectors that over-express S100A4 in the nucleus	127
4.4.2	CD34 <sup>+</sup> HSPCs could not be shown to overexpress S100A4.....	128
4.4.3	CD34 <sup>+</sup> cells do not depend on S100A4 for their growth, differentiation, and survival.....	129
4.4.4	Nuclear S100A4 over-expression does not affect AML cell lines proliferation .....	130
4.4.5	AML cell lines are dependent on S100A4 for their growth and survival	131

---

5	Chapter 5 - Introduction.....	134
5.1	Overview .....	134
5.2	Aims .....	135
5.3	Results .....	135
5.3.1	Generating cytosolic and nuclear protein lysates from AML cell lines..	135
5.3.2	Optimisation of co-IP of S100A4 and its binding partners complex in calcium enriched conditions for LC-MS analysis.....	137
5.3.3	LC-MS Proteomic analysis and data filtration of S100A4 binding partners 140	
5.3.4	Validation of S100A4 nuclear binding proteins using two-step co-IP and western blot.....	151
5.3.5	Functional analysis of S100A4 final binding partners candidates .....	158
5.3.6	Protein-protein interaction analysis between S100A4 and its binding partners	161
5.4	Discussion .....	164
5.4.1	Optimal protein extraction and co-IP conditions were achieved in ME-1 cell line	164
5.4.2	Mass Spectrometry proteomic analysis and data filtration strategy.....	166
5.4.3	S100A4 is bound to proteins that are implicated in AML .....	167
5.4.4	Validation of S100A4 candidate binding partners .....	169
5.4.5	Validated S100A4 binding partners (hnRNPM and HIST1H2AA) are identified in overrepresented MF and BP .....	170
5.4.6	Validated S100A4 binding partners (hnRNPM and HIST1H2AA) are highly interactive with AML-related proteins .....	171
6	Chapter 6 - General discussion .....	174
6.1	Summary and conclusion .....	174
6.2	Future directions.....	180
7	Appendices.....	199



---

## List of Figures

Figure 1-1: Diagram of normal haematopoietic hierarchy. ....	5
Figure 1-2: Schematic diagram of the S100 proteins' structure .....	24
Figure 2-1: Estimation of CD34 <sup>+</sup> cells in the MNCs population.....	40
Figure 2-2: Retroviral vector constructs based on PINCO.....	49
Figure 2-3: shRNA S100A4 pHIV vectors used in study.....	50
Figure 2-4: Diagram illustrates the distribution of His-Tag S100A4 protein serial molecular weights and serial dilutions of S100A4 mAb (CST, clone D9F9D) on dot blot. ....	54
Figure 2-5: Diagram of transblotting sandwich set up .....	55
Figure 2-6: Experimental design of co-IP enriched with Ca <sup>2+</sup> and depleted from Ca <sup>2+</sup> .....	56
Figure 2-7: Illustrative diagram of difference between direct and indirect co-IP. ....	60
Figure 2-8:LC/MS Proteomic data analysis strategy.....	65
Figure 3-1: Optimisation work-flow diagram for detection of S100A4 expression in AML. Illustrative diagram shows the optimisation work-flow of S100A4 detection conditions. WB; Western Blotting, IP; Immunoprecipitation, MS; Mass Spectrometry. ....	70
Figure 3-2:Optimisation of S100A4 protein detection in AML cell lines using western blot. ....	71
Figure 3-3:Expression of S100A4 in cytoplasmic and nuclear fractions of AML cell lines.....	72
Figure 3-4:Optimisation of His-Tag S100A4 protein detection for use as a positive control.....	73
Figure 3-5:S100A4 mAb specificity is validated by immunoprecipitation	76
Figure 3-6: Validation of S100A4 expression at 8kDa by mass spectrometry. ....	77

---

Figure 3-7: S100A4 mRNA expression levels and subcellular localisation in AML cell lines. ....	78
Figure 3-8: S100A4 mRNA Expression Patterns in normal Haematopoiesis. ....	80
Figure 3-9: Purification of individual lineages expressing S100A4 used in Affymetrix microarray analysis and western blot.....	81
Figure 3-10 : S100A4 protein is differentially expressed in normal haematopoietic lineage and restricted to the cytosol. ....	82
Figure 3-11: S100A4 Protein expression is restricted to the cytoplasm in bone marrow derived haematopoietic cells.....	83
Figure 3-12:S100A4 is overexpressed in AML sub-types. ....	86
Figure 3-13: Overexpression and mislocalization of S100A4 in the nucleus of AML patients. ....	87
Figure 3-14: Overexpression and mislocalization of S100A4 in the nucleus of AML patients. ....	88
Figure 3-15:Overexpressed <i>S100A4</i> is associated Poor Overall Survival Rate in AML Patients.....	90
Figure 3-16:Upregulated <i>S100A4</i> mRNA levels is associated poor overall survival rate in AML patients. ....	91
Figure 3-17:Overexpressed <i>S100A4</i> mRNA does not correlate with AML patient outcomes.....	92
Figure 4-1: Test digest of 3xNLS-S100A4 retro-and lentiviral vectors...	105
Figure 4-2: Confirmation of GFP and S100A4 protein expression in Phoenix cells transfected with PINCO retroviral vectors. ....	106
Figure 4-3: Confirmation of GFP and S100A4 protein expression in HEK293T cells transfected with pHIV lenti-viral vectors. ....	107
Figure 4-4: Human CD34 <sup>+</sup> HSPC did not stably over-express S100A4. .	109
Figure 4-5: S100A4 expression could not be achieved in the nucleus of CD34 <sup>+</sup> HSPCs using lentiviral vector.....	110
Figure 4-6: S100A4 is overexpressed in the nucleus of leukemic cells K562 and TF-1. ....	112

---

Figure 4-7: S100A4 nuclear overexpression does not affect proliferation of AML cell lines.....	113
Figure 4-8: NOMO-1 requires S100A4 expression for cell growth. ....	115
Figure 4-9: Re-designed shRNA S100A4 vector is validated in NOMO-1 cells.....	116
Figure 4-10: S100A4 was successfully knocked down in CD34 <sup>+</sup> cells....	118
Figure 4-11:S100A4 knockdown in CD34 <sup>+</sup> cells does not affect the development or survival of human myeloid haemopoietic cells. ....	119
Figure 4-12: OCI AML-2 cells are dependent on S00A4 for growth and survival.....	121
Figure 4-13: Gating strategy for apoptosis assay in OCI AML-2 cell line. ....	122
Figure 4-14 : S100A4 is required for cell survival in OCI AML-2 cell lines. ....	123
Figure 4-15: NOMO-1 cells are dependent on S00A4 for growth and survival.....	124
Figure 4-16: TF-1 cells are dependent on S00A4 for growth and survival ....	125
Figure 4-17: THP-1 cells are dependent on S00A4 for growth and survival ....	126
Figure 5-1: S100A4 is expressed in cytosolic and nuclear sub-cellular compartments of NOMO-1 and ME-1 cells.....	136
Figure 5-2:Immunoblots of direct and indirect S100A4 antibody binding co-IP of S100A4 binding partners. ....	138
Figure 5-3:Quality Control check of co-IP on Tricine SDS-PAGE gel stained with SYPRO Ruby Fluorescent Stain.....	139
Figure 5-4:LC-MS Chromatograms Showing the Relative Abundance of Peptides Detected in Cytoplasmic Co-IP.....	142
Figure 5-5: LC-MS Preliminary Peptide filtration of cytoplasmic co-IP.	144
Figure 5-6: LC-MS preliminary peptide filtration of nuclear co-IP. ....	147

---

Figure 5-7: S100A4 candidate binding partners are validated using co-IP. ....	154
Figure 5-8:hnRNPM is validated as a candidate binding partner of S100A4 using co-IP. ....	155
Figure 5-9: Validation of putative S100A4 candidate binding partners using reciprocal co-IP. ....	156
Figure 5-10: HIST1H2AA and hnRNPM are validated as S100A4 binding partners using reciprocal co-IP. ....	157
Figure 5-11: Functional analysis using molecular function enrichment for the final list of A100A4 binding partners’ candidates. ....	159
Figure 5-12: Functional analysis using biological process enrichment for the final list of S100A4 binding partners candidates. ....	160
Figure 5-13: String interaction network analysis of S100A4 and binding partners. ....	162
Figure 5-14: S100A4 nuclear candidate binding partners and AML-related proteins are highly interactive.....	163
Figure 6-1: Integrated working model summarising the investigation plan of role of S100A4 in AML carried out in this study.....	176
Figure 6-2: Graphical abstract summarising the findings of this study about the role of S100A4 in Acute Myeloid Leukaemia.....	179

## **List of Tables**

Table 1-1: Common functional genetic mutations and frequencies in AML. ....	13
Table 1-2: The French–American–British (FAB) classification system of acute myeloid. ....	15
Table 1-3: The 2016 World Health Organisation (WHO) classification of acute myeloid leukaemia and myeloid neoplasms.....	16
Table 1-4: The 2017 European LeukemiaNet (ELN) AML risk stratification by genetics.....	18

---

Table 1-5: A summary of targeted therapeutic agents in clinical trials for AML. ....	21
Table 2-1: The list of all antibodies used in this study. ....	30
Table 2-2: The list of all materials used in this study. ....	32
Table 2-3: The list of Recto- and lentivirus used in this study ....	34
Table 2-4: Cell lines used in this study and their growth conditions requirements. ....	36
Table 2-5: Concentrations of human cytokines required for CD34+ cells growth media mix ....	37
Table 2-6: co-IP's buffers composition. ....	59
Table 5-1: Number of Peptides Identified by LC-MS in Cytoplasmic and Nuclear Co-IP Samples ....	141
Table 5-2: Identification of reviewed proteins on UniProt KB database in the S100A4 co-IP sample. ....	147
Table 5-3: S100A4 nuclear candidate binding partners identified by LC-MS ....	150
Table 5-4: AML proteins have been detected in S100A4 in Ca <sup>2+</sup> enriched co-IP . ....	152

# Chapter 1

## General Introduction

## 1 Chapter 1 - Overview of haematopoiesis

### 1.1 Normal haematopoiesis

Mature blood cells carry out a variety of crucial functions such as immune response, gases exchange, wound healing and haemostasis. Blood cell production is continuously maintained by a process called “Haematopoiesis”. Haematopoiesis describes the commitment and differentiation processes by which all blood cells are produced. Terminally differentiated haematopoietic cells are derived from the haematopoietic stem cell (HSC). HSCs are defined by their dual capability of self-renewal and multi-lineage differentiation (Matsuzaki and Sato 2015). Adult haematopoiesis primarily occurs in the bone marrow (medullary) or alternatively, in other tissues such as the spleen, liver, and thymus (extramedullary).

Normal haematopoiesis is organized as a cellular hierarchy in which mature haematopoietic cells originates from a primitive HSC. These cells may differentiate into two main multi-potential progenitors called haematopoietic progenitor cells (HPCs); classified into myeloid or lymphoid progenitor cells (Figure 1-1). HPCs then differentiate into lineage-committed progenitor cells that can give rise to terminally differentiated haematopoietic cells in which each cell type has a distinctive biological function. Generally, lymphoid progenitors are responsible for producing B- and T-cells that mediate immune response. On the other hand, myeloid progenitors produce white blood cells, platelets and thrombocytes (Pouzolles *et al.* 2016). During haematopoiesis, a complex mixture of cytokines, growth factors and expression of transcription factors determines the differentiation of primitive HSC to certain haematopoietic cells type.

HSCs are relatively quiescent in the bone marrow (BM) during steady state haematopoiesis (Szade *et al.* 2018). However; upon entering cell cycle, HSCs generates a hierarchy of differentiating progenitor and lineage-committed precursors that undergo a proliferative expansion required to replace the predominantly short-lived and mature haematopoietic cells. HSCs can be divided into two categories; long-term reconstituting HSCs (LT-HSCs) and short-term reconstituting HSCs (ST-HSCs). LT-HSCs are responsible for maintaining self-renewal and multi-lineage differentiation potential throughout life. Whereas, ST-HSCs are derived from LT-HSCs and have limited self-renewal capacity ST-HSCs preserves the multipotency property by differentiating into multipotent progenitors (MPPs) (Pouzolles *et al.* 2016) (Figure 1-1).

These MPP populations have the ability to differentiate into restricted self-renewal progenitors known as “oligo-lineage” which ultimately give rise to functionally mature myeloid or lymphoid cells (Bonnet 2002). The common myeloid progenitors (CMPs) give rise to granulocyte-macrophage progenitors (GMPs), which differentiate into further restricted progenitors as monocyte progenitors (MPs) giving rise to precursors (macrophages and granulocytes) and granulocyte progenitors (GPs) (basophil, neutrophil, eosinophil). Moreover, CMPs produce megakaryocyte-erythroid progenitors (MEPs), which differentiate into megakaryocytes/erythrocytes. On the other hand, common lymphoid progenitors (CLPs) differentiate into B-lymphocytes, T-lymphocytes, and natural killer (NK) cells (Mendelson and Frenette 2014). Both myeloid and lymphoid differentiation produce dendritic cells (Figure 1-1). However, the classical haematopoietic differentiation model has been debated in recent years as several studies have shown that the differentiation of haematopoietic lineage committed progenitors are highly heterogeneous than previously suggested. Using single cell RNA-Seq (ScRNA-Seq), Yamamoto *et al.*, shown that unipotent myeloid lineage-committed progenitors such as megakaryocyte-erythroid progenitors (MEP) can directly arise from HSCs (Yamamoto *et al.* 2013). Similar findings were reported by Drissen, R. *et al.* supported the fact that mature blood cells may not follow downstream differentiation pathway as previously suggested. Results suggested that in early lineage bifurcation of the multi-potent HSCs, the separation of lineages is based on the expression of Gata1 in lineages (erythrocytes, megakaryocytes, mast cells, and eosinophils) and lineages that did not express Gata1 (monocytes, lymphocytes, and neutrophils) rather than differentiation downstream from GMP (Drissen *et al.* 2016). Another study suggested that distinct subsets of HSCs exist that are stably biased towards the generation of platelets (Sanjuan-Pla *et al.* 2013). These findings substantially change the view towards the haematopoietic hierarchical step wise model to a much more dynamic model that allow potential lineages to differentiate to desired mature blood cells. The subsequent work in this study is focussed on the myeloid lineage abnormality in the form Acute Myeloid Leukemia (AML).

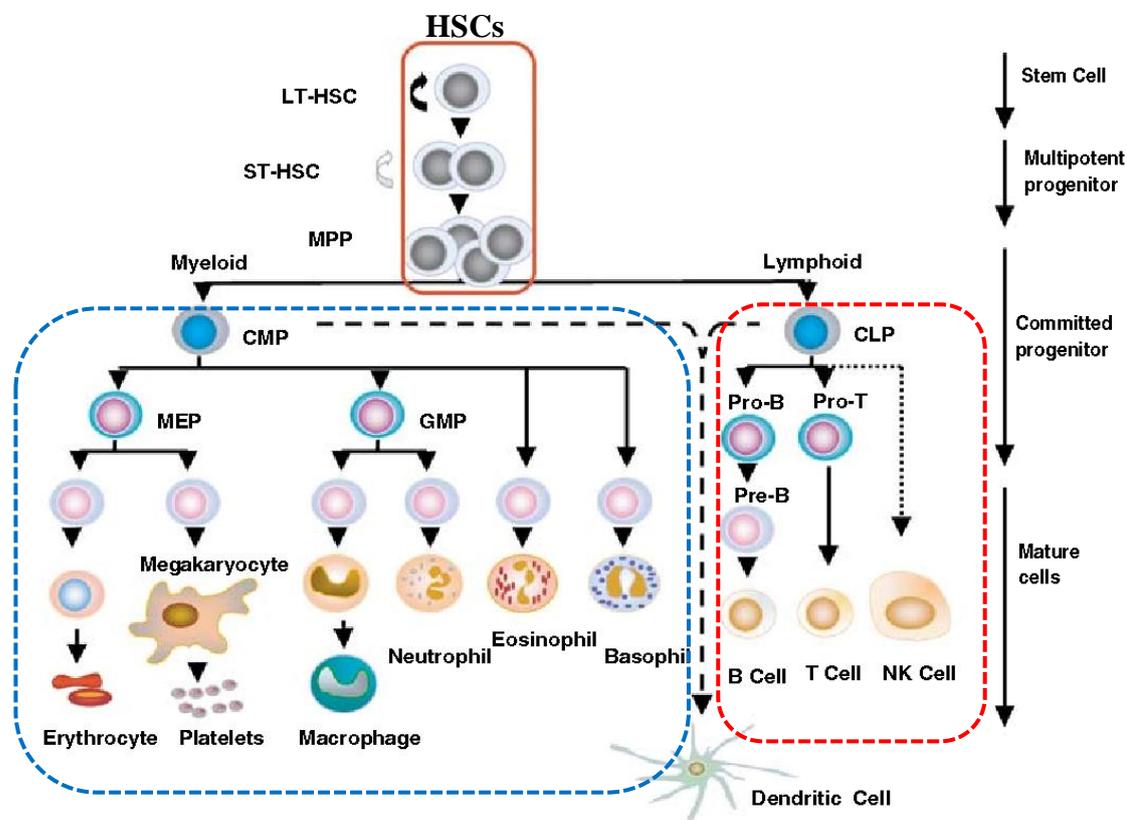
### **1.1.1 Human haematopoietic stem cells**

#### *1.1.1.1 Characteristics of HSCs*

HSCs are a rare population of cells which constitute an estimated 0.05% to up to 0.5% of cells in the adult bone marrow (reviewed in Bonnet 2002). HSCs can be found

in BM as immobilised cells (1 in every 100,000 cells), mobilised cells in peripheral blood (PB) forming 5-20% of cells and abundantly present in umbilical cord blood (CB) as mobilised cells 1 in 100 cells are HSCs (Mosaad 2014).

As described above, HSCs can be classified according to their self-renewal capacity and specific cell surface markers into; long-term, short-term and MPPs (Figure 1-1). LT- HSCs are known for their ability of both self-renewal and differentiation into lineage committed progenitors that reconstitute the entire blood and immune system for the lifespan of the organism. Conversely, ST-HSCs have limited self-renewal capability which then differentiate into multipotent MPPs that lack self-renewal capability (Ema and Nakauchi 2003). It has been proposed that there are two models in which LT-HSCs maintain their self-renewal capacity for the lifespan of the human life which can account for the recovery of the stem cell pool following haematosis or damage (Ema and Nakauchi 2003).



**Figure 1-1: Diagram of normal haematopoietic hierarchy.**

Classification of HSC into long-term, short-term, and MPP, based on surface expression markers. Shown are common myeloid progenitor (CMP) (left), common lymphocyte progenitor (CLP) (right), granulocyte-monocyte progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP), natural killer (NK) cells.

The first model proposes that there is a fixed number of LT-HSCs that maintain haematopoiesis for life and as HSCs proliferate they undergo a progressive loss in their self-renewal capacity. The second model proposes that there is a small number of LT-HSCs sustaining lifelong haematopoiesis due to their ability of self-renewal during cell division. Leaving a proportion of these cells to differentiate into lineage committed progenitors and mature blood cells and maturation.

#### 1.1.1.2 HSCs Microenvironment

The BM niche is the anatomical location where HSCs reside and the physiological microenvironment in which they self-renew and differentiate. The BM niche tightly regulates HSCs dynamic balance between self-renewal and differentiation in a non-autonomous manner (Sanchez-Aguilera and Mendez-Ferrer 2017). The microenvironment maintains HSCs at quiescence to sustain self-renewal and activates HSC for differentiation and/or injury repair (Crane *et al.* 2017). The BM niche consists of adherent cells called stromal cells and provides a shelter for HSCs, protecting them from physiological stress and restrains them from constant differentiation that could deplete the stem cell pool, and overproduction which could result in tumour formation (Kanji *et al.* 2011).

HSCs are mostly located in two distinctive niches in the bone marrow, the endosteal and perivascular niches. Both niches are complex microenvironments containing a wide range of BM cells that play different roles in HSC regulation (Tamma and Ribatti 2017). The endosteal niche is located in the inner surface of the bone marrow cavity lined by bone cells such as osteoclasts and osteoblasts. The osteoblast cells are bone-forming progenitor cells that produce a chemokine C-X-C motif ligand 12 (CXCL12) (also known as stromal cell-derived factor-1 [SDF-1]) (Lucas 2017). Thus, osteoblast cells are named as CXCL12-abundant reticular (CAR) cells. CXCL12 is essential for HPCs and quiescent HSCs pool retention in the BM. Further, CXCL12 has been reported to stimulate HSCs motility by stromal migration via activation of adhesion molecules (Greenbaum *et al.* 2013). Moreover, CXCL12 maintains the viability of HSCs via reacting to corresponding ligands on cell surface. On the other hand, the perivascular niche includes thin-walled vessels lined by a single layer of endothelium. The perivascular niche serves as a communication channel between the blood circulation and BM cavity and supply the BM with oxygen and nutrients (Tamma and Ribatti 2017).

The perivascular contain arteries that penetrate the BM compact bone via bone canal and branching into small arterioles which transit to venous sinusoids close to endosteum (reviewed in Wei and Frenette 2018). Analysis of unprocessed bones sections using improved immunofluorescence methods showed that quiescent HSCs are located in a perivascular niche near the BM sinusoids (Asada *et al.* 2017). Thus, significant numbers of quiescent HSCs are located close to small arterioles near the endosteal surface (Nombela-Arrieta *et al.* 2013). Moreover, quiescent HSCs can be located in the megakaryocyte niche which is associated with vWF<sup>+</sup> myeloid-biased HSCs. However, the arteriolar niche is associated with vWF<sup>-</sup> and regulates lymphoid-biased HSCs which suggests that lineage-biased HSCs location in the BM is not randomly distributed (Pinho *et al.*, 2018 ).

It has been reported that HSCs that reside near osteoblasts (endosteal niche) are LT-HSCs quiescent in nature. Whereas ST-HSCs are present at vascular niches and are actively differentiating into haematopoietic cell types to replenish circulating blood cells (Yu and Scadden 2016). The endosteal niche is hypoxic by nature with the lowest oxygen (O<sub>2</sub>) levels, whilst highest levels of O<sub>2</sub> are thought to be observed at the vascular niche. Hypoxia is an essential condition in which HSCs are maintained at quiescent state (Go stage) and thereby preventing exhaustion of their self-renewal capacity (Hoggatt and Pelus 2011). In contrast, soluble extracellular calcium is present at significantly high levels at endosteal niche. Adams, G. *et al* demonstrated that HSCs expressing calcium-sensing receptor (CaR) are attracted to soluble extracellular calcium (Ca<sup>2+</sup>). When CaR was knocked out in HSCs, mice had reduced numbers of HSCs within the BM niche whereas increased HSCs were observed in the peripheral blood. Interestingly, CaR-knockout HSCs poorly engrafted in haematopoietic transplantation experiments (Adams *et al.* 2006). Thus, these findings may suggest that high Ca<sup>2+</sup> concentration is essential signal for homing and retention of HSC in endosteal niche.

### 1.1.2 Regulation of normal haematopoiesis

The haematopoiesis is a tightly regulated process during steady-state, differentiation, and upon injury or inflammation response (Pouzolles *et al.* 2016). During steady-state, HSCs self-renewal capacity is maintained to insure lifelong supply of HSCs. Conversely, during differentiation and/or inflammation response HSCs undergoes proliferative expansion required to replace the predominantly short-lived mature blood cells (Kronenwett and Haas 2006). Thus, HSCs are constantly balancing maintenance between the undifferentiated HSCs and differentiating into cells of multiple lineages. This process is largely regulated by extrinsic and intrinsic factors. The extrinsic factors include cytokines, growth factors and other environmental factors which can activate certain cell signalling pathways that decide HSCs lineage fate (Endele *et al.* 2014).

#### 1.1.2.1 Regulation of haematopoiesis by cytokines and transcription factors

Cytokines are small proteins that can be either membrane-bound or secreted and have a specific effect on cell signalling (Oppenheim 2001). In normal haematopoiesis, cytokines are produced by various cell types and can exert their functions either within the BM microenvironment, within the bloodstream, or lymphatic vessels. These cytokines include, colony-stimulating factors, erythropoietin, Thrombopoietin (TPO), interferons, and interleukins (Endele *et al.* 2014). Haematopoietic cytokines are sub-classified into two classes of receptors including, tyrosine kinase receptors; such as FLT3 and c-KIT, or cytokine receptor superfamily; such as type I and type II cytokine receptors (Endele *et al.* 2014, Robb 2007). In steady-state, differentiation, and/or inflammation response, haematopoiesis is mediated mainly by cytokines. These cytokines can define HSC's fate via initiating lineage specific gene expression programme that lead to either self-renewal HSC or lineage committed progenitors (Pouzolles *et al.* 2016). Haematopoietic cytokines can influence lineage commitment either directly by cytokines instructed lineage-specific transcriptional programmes or through activation of transcription factor networks followed by a selective cytokine signalling (Song *et al.* 2016). On the other hand, transcription factors such as GATA-1, GATA-3, Notch-1, FOG-1, PU.1, C/EBP $\alpha$  and Pax5 are involved in defining the HSPCs cellular fate. For example, GATA-1 and FOG-1 are involved in the development of megakaryocytes and erythrocytes, and PU.1 and C/EBP $\alpha$  are involved in the myeloid progenitor cells commitment (Iwasaki 2013).

## 1.2 Leukaemia

### 1.2.1 Overview of leukaemia

Leukaemogenesis is characterised by recurring chromosomal aberrations and gene mutations that could lead to either over-activation of the growth program or interfere with haematopoietic terminal differentiation (Chen *et al.* 2010). These chromosomal and molecular abnormalities alter the normal primitive cells differentiation programme or cause selective expression of certain genes that limits the self-renewal capability of normal haematopoietic cells (Chen *et al.* 2010). Consequently, undifferentiated leukaemic blasts overcrowd normal haematopoietic cells in the BM and blood stream. However, gradual acquisition of genetic mutations in HSCs may lead to haematological malignancies such as acute myeloid leukemia (AML) in adults (Jan *et al.* 2012). Recently, genomic analysis of PB DNA of healthy individuals identified leukemia-associated genetic mutations that are strongly associated with age (Jaiswal *et al.* 2014). Further, large chromosomal deletions and mutations not associated with leukemia are observed in other clones (Shlush *et al.* 2014). Thus, this phenomenon represents a risk factor for leukaemia particularly in AML and known as clonal hematopoiesis (CH). Although HSCs are quiescent and protected by cell intrinsic properties, these cells can have an increased mutational burden with age (Beerman 2017). It has been estimated that human HSPCs can develop  $1.3 \pm 0.2$  exonic mutations per decade of individual's life. Thus, by age of 50, it can be estimated that an individual could accumulate an average of five coding mutations in each HSPC (Welch *et al.* 2012). Most common mutations are reported in epigenetic regulators such as DNMT3A, TET2, and ASXL1. However, Other mutations are also observed in genes regulating DNA damage (TP53, PPM1D), RNA splicing (SRSF2, SF3B1), and signaling (JAK2) (Moran-Crusio *et al.* 2011, Challen *et al.* 2011). Although healthy individuals can have CH, it is not considered as a disease. However, having CH can be a significant risk factor for future haematological malignancies such as AML ~10% of CH cases will progress to AML (Jaiswal *et al.* 2014, Abelson *et al.* 2018). The accumulation of genetic mutations in HSCs overtime lead to development of stem cell-like cells known as leukaemic stem-like cell (LSC). Functionally, LSCs can initiate leukemia when transplanted into NOD/SCID mice models using in xenotransplantation assays (Feuring-Buske *et al.* 2003). Further, LSCs have an unlimited self-renewal capacity similar to normal HSCs and higher proliferative potential than normal haematopoietic cells (Reinisch *et al.* 2015). Recent study shows that LSCs

can selectively survive chemotherapy and evade immune surveillance by preventing the activation of cytotoxic lymphocytes such as natural killer (NK) cells through lacking the expression of stress-induced ligands for NKG2D (Paczulla *et al.* 2019). Transplantation studies shows that human NKG2DL<sup>-</sup> AML cells infiltrate more efficiently the mouse bone marrow than NKG2DL<sup>+</sup> AML cells and were associated with poor survival. Further, NKG2DL<sup>-</sup> engrafted mice showed chemo resistance when treated with a chemotherapy drug compared to NKG2DL<sup>+</sup> cells which showed reduction of bone marrow infiltration (Hilpert *et al.* 2012).

Leukaemia maybe classified based on the haematopoietic cell type affected and on how rapidly the disease progresses. Thus, myeloid leukaemia (also known as myelocytic or myelogenous) develop in early myeloid progenitor cells. Whilst, lymphocytic leukaemia (also known as lymphoid or lymphoblastic) are derived from lymphocytes in the BM. The acute form of leukaemia is defined by a rapid increase in immature blasts differentiation, whilst the chronic form is characterised by slow accumulation of immature blasts. Thus, there are four main types of leukaemia, chronic lymphocytic leukaemia (CLL), chronic myeloid leukaemia (CML), acute lymphocytic leukaemia (ALL) and acute myeloid leukaemia (AML).

### **1.2.2 Acute myeloid leukaemia (AML)**

AML is a malignant disorder of HSC characterized by a differentiation block in the myeloid lineage coupled with an increased proliferation of leukaemic blast cells. As a result, immature leukaemic blasts accumulate in the BM and often infiltrate to the peripheral blood and soft tissues (Reviewed in Short *et al.* 2018).

#### *1.2.2.1 Diagnosis and incidence of AML*

In the United Kingdom, leukaemia was the 15<sup>th</sup> most common cancer (amongst the top 20 most common cancers) and accounts for 2% of all cancers in 2017 (CancerResearchUK 2014). AML is the most common acute type of leukaemia in adults, accounting for less than 1% of all cancers. It has been estimated that around 3,000 new cases are diagnosed with AML annually and AML is responsible for around 2,600 deaths in the UK every year with the highest mortality rates in people aged 85 to 89 (2014-2016) (CancerResearchUK 2014). Although the disease can occur in all age groups, acute myeloid leukaemia predominantly occurs in older adults, with a median age of diagnosis (>68 years of age) and is more common in males compared to females.

AML is often presented with a rapid onset of symptoms that can be fatal within weeks or months if left untreated. Abnormal accumulation of immature leukaemic blasts leads to infiltration of these blasts to tissues and replaces normal blood cells. Patients diagnosed with AML often present with nonspecific clinical signs and symptoms such as fatigue, haemorrhage, shortness of breath, frequent infections and fever due to severe decrease in red blood cells (anaemia), platelets (thrombocytopenia) or neutrophils (neutropenia). Moreover, leukaemic blasts infiltration to other tissues such as spleen could cause splenomegaly, liver (hepatomegaly), lymph nodes (lymphadenopathy), skin (leukaemia cutis), bone (bone pain) and central nervous system (CNS) (Khwaja *et al.* 2016, Dohner *et al.* 2015). Diagnosis requires at least the presence of  $\geq 20\%$  blasts count (which includes Monoblasts, Myeloblasts, and megakaryo blasts) in the BM aspirate or circulating blood (Dohner *et al.* 2017). However, cytogenetic and mutation analysis is important, particularly as next generation technologies (NGS) become cheaper and quicker. Detection of known chromosomal abnormalities (including translocations and inversions) is critically important for diagnosis, disease classification, and treatment decision-making. For example, detection of t(16;16)(p13.1;q22), t(8;21)(q22;q22.1), inv(16)(p13.1q22), or PML-RARA fusion transcripts allow the diagnosis of AML to be made, (Taylor *et al.* 2017). However, in AML patients whom have normal cytogenetics, often have CEBPA, Nucleophosmin 1 (NPM1), and fms-like tyrosine kinase 3- internal tandem duplication (FLT3-ITD) mutations that predict response to induction and consolidation chemotherapy.

#### 1.2.2.2 Pathophysiology and molecular abnormalities in AML

Most individuals diagnosed with AML presented with no predisposing risk factors. Nevertheless, the risk of developing AML increases with exposure to DNA-damaging agents, such as ionizing radiation (due to therapeutic radiotherapy), cigarette smoke, benzene, and cytotoxic chemotherapy agents (reviewed in Khwaja *et al.* 2016). Clinically, AML secondary to cytotoxic chemotherapy and/ or ionizing radiation is called therapy-related AML. In many types of leukaemia, first-degree relatives of patients have an increased risk of developing similar disease. It has been reported that 5-7 fold increased risk of myeloproliferative neoplasms (MPN). However, no strong evidence has been reported that genetic predisposition is a risk factor in myeloid malignancies (AML and MDS). In addition, certain inherited disorders can be a pre-disposing risk

factors for developing AML such as Down syndrome, Bloom syndrome, ataxia-telangiectasia, Schwachman–Diamond syndrome, severe congenital neutropaenia, Fanconi anaemia, and Diamond–Blackfan anaemia (Seif 2011). Frequently mutated genes are classified into 9 functional categories include: transcription factor fusions, tumour suppressor genes, DNA methylation-related genes, the NPM1 gene, signalling genes, chromatin-modifying genes, myeloid transcription factor genes, spliceosome complex genes, and cohesin complex genes, a summary is provided in Table 1-1 (Naoe and Kiyoi 2013, Khwaja *et al.* 2016).

**Table 1-1: Common functional genetic mutations and frequencies in AML.**

Adopted from Khwaja et al. 2016.

Functional groups	Examples of mutated genes in AML	Frequency in Patients with AML (%)
Signalling pathways	FLT3, KIT, KRAS, NRAS and serine/threonine kinases	59
DNA methylation	DNMT3A, TET2, IDH1 and IDH2	44
Chromatin modifiers	MLL fusions, ASXL1 and EZH2	30
Nucleophosmin	NPM1	27
Myeloid transcription factors	RUNX1 and CEBPA	22
Transcription factors	PML–RARA, MYH11–CBFB and RUNX1–RUNX1T1	18
Tumour suppressors	TP53, WT1 and PHF6	16
Spliceosome complex	SRSF2 and U2AF1	14
Cohesin complex	STAG2, RAD21, SMC1 and SMC3	13

### 1.2.2.3 Classification of AML

Classifying AML often presented challenges due to the disease heterogeneity in terms of blast morphology, immunophenotypic features, cytogenetic and molecular abnormalities. In 1976, a cooperative group of scientists devised a new classification system in attempt to subdivide AML based on leukaemic blast cellular morphology and cytochemistry, called the French–American–British (FAB) classification as shown in Table 1-2 (Bennett *et al.* 1976). The FAB classification system subdivides AML into eight subtypes, M0 - M7 based on the degree of maturation of leukaemic blasts and direction of differentiation from which the leukaemia developed (Bennett *et al.* 1976). However, in 2002, the FAB AML classification system was superseded by a new classification system revised by a group of pathologists and clinicians under the umbrella of the World Health Organization (WHO) (Vardiman *et al.* 2002) and subsequently updated in 2008 and lastly in 2016 (Arber *et al.* 2016). Although the WHO classification system retained FAB main classification criteria (i.e. morphology, immunophenotypic features, and clinical features), it incorporates recurring cytogenetic and molecular genetic abnormalities, summarised in Table 1-3. Thus, incorporating the molecular abnormalities of AML in the WHO classification system enabled precise diagnosis and prognosis of the disease compared to FAB classification.

### 1.2.2.4 Prognostic factors and treatment of AML

#### **Prognosis**

Prognosis is the prediction of the course of a disease and patient's recovery chances based on pre-treatment signs and symptoms. In AML, prognostic factors such as cytogenetics and molecular genetics abnormalities can be independent predictors of AML outcomes which allow assigning AML patients to appropriate treatment plans tailored to their risk groups. The European Leukemia Net (ELN) guidelines is a widely used AML risk stratification guideline (Dohner *et al.* 2017). The ELN guidelines categorise AML patients into four groups based on cytogenetic and molecular abnormalities; favourable, intermediate, and adverse (Table 1-4). This categorisation of AML risk factors have useful prognostic and therapeutic implications. For example, AML patients who carry certain chromosomal translocations such as; t(8;21), t(15;17), and inv(16)/t(16;16) are considered among the favourable group.

**Table 1-2: The French–American–British (FAB) classification system of acute myeloid.**

Type	Name
M0	acute myeloblastic leukemia, minimally differentiated
M1	acute myeloblastic leukemia, without maturation
M2	acute myeloblastic leukemia, with granulocytic maturation
M3	promyelocytic, or acute promyelocytic leukemia (APL)
M4	acute myelomonocytic leukemia
M4eo	myelomonocytic together with bone marrow eosinophilia
M5	acute monoblastic leukemia (M5a) or acute monocytic leukemia (M5b)
M6	acute erythroid leukaemia's, including erythroleukemia (M6a) and very rare pure erythroid leukemia (M6b)
M7	acute megakaryoblastic leukemia

**Table 1-3: The 2016 World Health Organisation (WHO) classification of acute myeloid leukaemia and myeloid neoplasms.**

Adopted from Arber et al. 2016.

<b>Acute myeloid leukemia (AML) and related neoplasms</b>
<b>AML with recurrent genetic abnormalities</b>
<p>AML with t(8;21)(q22;q22.1);RUNX1-RUNX1T1</p> <p>AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);CBFB-MYH11</p> <p>APL with PML-RARA</p> <p>AML with t(9;11)(p21.3;q23.3);MLLT3-KMT2A</p> <p>AML with t(6;9)(p23;q34.1);DEK-NUP214</p> <p>AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM</p> <p>AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);RBM15-MKL1</p> <p>Provisional entity: AML with BCR-ABL1</p> <p>AML with mutated NPM1</p> <p>AML with biallelic mutations of CEBPA</p> <p>Provisional entity: AML with mutated RUNX1</p>
<b>AML with myelodysplasia-related changes</b>
<b>Therapy-related myeloid neoplasms</b>
<b>AML, Not Otherwise Specified (NOS)</b>
<p>AML with minimal differentiation (FAB M0)</p> <p>AML without maturation (FAB M1)</p> <p>AML with maturation (FAB M2)</p> <p>Acute myelomonocytic leukemia (FAB M4)</p> <p>Acute monoblastic/monocytic leukemia (FAB M5)</p> <p>Pure erythroid leukemia (FAB M6)</p> <p>Acute megakaryoblastic leukemia (FAB M7)</p> <p>Acute basophilic leukemia</p> <p>Acute panmyelosis with myelofibrosis</p>
<b>Myeloid sarcoma</b>
<b>Myeloid proliferations related to Down syndrome</b>
<p>Transient abnormal myelopoiesis (TAM)</p> <p>Myeloid leukemia associated with Down syndrome</p>
<b>Blastic plasmacytoid dendritic cell neoplasm</b>
<b>Acute leukaemia's of ambiguous lineage</b>
<p>Acute undifferentiated leukemia</p> <p>Mixed phenotype acute leukemia (MPAL) with t(9;22)(q34.1;q11.2); BCR-ABL1</p> <p>MPAL with t(v;11q23.3); KMT2A rearranged</p> <p>MPAL, B/myeloid, NOS</p> <p>MPAL, T/myeloid, NOS</p>

Recently, several studies have shown that the coexistence of Nucleophosmin1 (NPM1) or biallelic CEBPA mutations with chromosomal abnormalities does not modify the prognostic effect of these mutations (Haferlach *et al.* 2009, Schlenk *et al.* 2013). Further, several studies suggested that AML patients with NPM1 mutations and FLT3-ITD (with low allelic ratio  $<0.5$ ) have a similar favourable prognosis to patients with NPM1 mutations with no FLT3-ITD which makes both groups considered under favourable outcomes. Conversely, patients who have FLT3-ITD (with high allelic ratio  $\geq 0.5$ ) and carry wild-type NPM1 have a poor prognosis and categorised under the adverse-risk group (Gale *et al.* 2008, Linch *et al.* 2014). Therefore, the relapse rate and overall survival (OS) associated with FLT3-ITD depends largely on the allelic ratio of ITD (Dohner *et al.* 2017). On the other hand, the intermediate risk group considered as a very heterogeneous due to inclusion of patients who have normal karyotype (NK) with rare chromosomal aberrations. The final group of adverse outcomes carry a poor prognosis for patients with TP53 mutations, complex karyotype, and monosomal karyotype (Estey 2018).

### **Conventional treatment of AML**

Treatment of AML is traditionally divided into three phases: remission-induction chemotherapy usually followed by post-remission consolidation chemotherapy with or without HSC transplantation (HSCT) and maintenance therapy (Dombret and Gardin 2016). Initially, AML patients are evaluated whether they can be fit as candidate for intensive induction chemotherapy. This step is crucially important to assess the risk of treatment-related mortality post intensive therapy especially in older patients (above the age of 65 years). However, treatment-related mortality rates has declined in recent years due to improvement made in supportive care in older patients (Krug *et al.* 2010, Giles *et al.* 2007). Further, few randomized trials suggested that medically fit older patients may benefit from intensive induction therapy (Burnett 2018). Thus, age should not be the only determinant to decide patient's fitness for therapy but should include other factors such as performance status and pre-treatment co-morbidities (Dohner *et al.* 2017).

**Table 1-4: The 2017 European LeukemiaNet (ELN) AML risk stratification by genetics.**

Adopted from Dohner et al. 2017.

Risk category*	Genetic abnormality
Favourable	t(8;21)(q22;q22.1); RUNX1-RUNX1T1 inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 Mutated NPM1 without FLT3-ITD or with FLT3-ITD <sup>low</sup> Biallelic mutated CEBPA
Intermediate	NK and NPM1 <sup>-</sup> / FLT3 <sup>-</sup> NK and NPM1 <sup>+</sup> / FLT3 <sup>+</sup> NK and NPM1 <sup>-</sup> / FLT3 <sup>+</sup> t(9;11)(p21.3;q23.3); MLLT3-KMT2A (Cytogenetic abnormalities not classified as favourable or adverse)
Adverse	t(6;9)(p23;q34.1); DEK-NUP214 t(v;11q23.3); KMT2A rearranged t(9;22)(q34.1;q11.2); BCR-ABL1 inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2,MECOM(EVI1) 25 or del(5q); 27; 217/abn(17p) Complex karyotype, monosomal karyotype NK and NPM1 <sup>+</sup> / FLT3 <sup>+</sup> Mutated RUNX1 Mutated ASXL1 Mutated TP53

NK, Normal Karyotype

The main goal of remission-induction chemotherapy in AML treatment is to reduce leukaemic blast numbers to undetectable levels in the blood, BM, and extramedullary sites and to recover normal haematopoiesis (Dombret and Gardin 2016). Achievement of a complete remission (CR) following the induction phase is an essential step in AML treatment. However, during remission phase, few remaining leukaemic blasts are likely to persist and may lead to relapse (Bose *et al.* 2017). Further, several trials suggested that patients who did not receive post-remission chemotherapy may relapse within six to nine months. Thus, consolidation chemotherapy is required with high-dose chemotherapy and/or allogeneic HSCT to achieve durable disease control (Cornelissen *et al.* 2012). For younger patients (< 60 years), the traditional remission-induction treatment is “7+3” chemotherapy with cytarabine and an anthracycline. The complete remission (CR) rates achieved in this group are between 60% to 70% in large trials using the combination of cytarabine and anthracycline (Luskin *et al.* 2016). However, the Eastern Cooperative Oncology Group (ECOG) study has reported that using daunorubicin 90 mg/m<sup>2</sup> compared to 45 mg/m<sup>2</sup> daily for three days in younger patients with previously untreated AML has significantly increase in CR rate (71% vs. 57%) and median OS (24 vs. 16 months). Further, in older patients (50-60 years of age) with FLT3-ITD or NPM1 have also benefitted from high-dose daunorubicin (Luskin *et al.* 2016). In older AML patients, conventional therapy achieved lower CR rates and very few long-term OS compared to younger patients which led to rapid advancement in the field of anti-leukaemic drugs. Thus, older patients who are unfit for intensive chemotherapy induction, receive low intensity chemotherapy with hypomethylating agents (HMA) such as Vidaza<sup>®</sup> and Dacogen<sup>®</sup>, or low dose cytarabine as induction therapy (Dombret and Gardin 2016). However, in order to achieve an optimal treatment for AML patients, precise diagnosis of the disease and identifying prognostic factors are needed to determine patients’ response to therapy and survival.

### **Novel treatment of AML**

Given that CR rates achieved in older and unfit AML patients is still low, as well as no standard therapy for relapsed AML, the need for novel and targeted therapy in AML is higher than ever before. Recently, the frontline treatment of AML is rapidly advancing towards targeted molecular inhibitors such as kinase inhibitors (FLT3-ITD inhibitors). FLT3-ITD is one of the most common mutations in AML account for approximately 20–

30% of patients with AML and often associated with poor clinical outcomes (Yokota *et al.* 1997). Thus, in November 2018, a new drug called gilteritinib (Xospata<sup>®</sup>) has been approved by the FDA which can be used alone in adult AML patients with FLT3 mutation who don't respond to initial treatment or have relapsed (Kayser and Levis 2018, FDA 2018). Moreover, AML novel therapies has been advancing toward other molecular targets such as pro-apoptotic agents, cell cycle regulators, epigenetic modifiers and metabolic therapies are currently in clinical trials; a summary is provided in Table 1-5 (Reviewed in Kayser and Levis 2018, Khwaja *et al.* 2016). Further Immunotherapies, cell-based therapies, and immune checkpoint inhibitors are also being evaluated in current clinical trials. For instance, Chimeric antigen receptor (CAR)-T cells are engineered T-cells containing an extracellular single-chain variable fragment (scFv) from antigen-specific binding domain and an intracellular T-cell signalling domain (Miliotou and Papadopoulou 2018). Moreover, CAR-T cells are capable of targeting surface antigens of leukaemic cells in their native conformation independent from MHC cells (Holzinger *et al.* 2016). Recent studies show that CAR-T cell therapy targeting CD19 antigen has been successfully achieved in acute lymphoblastic leukemia (ALL) with low off-tumour toxicity (Badar and Shah 2020) . In AML, expression of specific antigens are not exclusive to AML blasts and LSCs. These antigens overlap with normal haematopoiesis, which can lead to severe haemato-toxicity of antigen-targeting therapies (Abreu *et al.* 2020). However, few antigens are being assessed in pre-clinical studies in AML such as CD33, CD123, FLT3 (CD135), and Lewis Y (LeY) as CAR-T cells antigen targets in AML (Reviewed in Hofmann *et al.* 2019). Although CAR-T cell therapy in AML still need to find specific leukaemic antigens and improve off-tumour toxicity, in vivo studies showed promising results which can enhance AML patients' therapeutic options of targeted therapy. On the other hand, a variety of therapeutic antibodies in the form of naked, conjugated and bispecific monoclonal antibodies directed against AML antigenic targets such as; CD33 and CD123. For example, using anti-CD33 monoclonal antibody -drug conjugate in combination with chemotherapy improved relapse-free and OS in older patients group with previously untreated AML (Kung Sutherland *et al.* 2013). In addition, targeting CD123 with CSL362 antibody was found to be well tolerated and safe as maintenance therapy for patients in CR and have high risk of relapse (reviewed in Beyar-Katz and Gill 2018).

**Table 1-5: A summary of targeted therapeutic agents in clinical trials for AML.**  
Summary table adopted from Khwaja *et al.* 2016.

Target category	Specific target	Examples of agent (or agents) in trials
<b>Signalling pathways</b>		
Tyrosine kinases	FLT3	Midostaurin, AC220 (also
	JAKs	known as quizartinib),
	KIT	pacritinib and crenolanib
	SYK	Ruxolitinib
	AXL	Dasatinib
Serine/threonine kinases	MEK	Selumetinib, binimetinib and
	MAPK	E6201
	PIM	BVD-523
	AKT	LGH447
	mTOR	GSK2141795
Lipid kinases	PI3K	BYL719, BKM120 and TGR-1202
Cell cycle regulators	PLK1	Rigosertib
	Aurora kinases	Alisertib and AZD1152
	WEE1	MK-1775
<b>Epigenetic regulators</b>		
DNA methylation	DNMTs	SGI-110 and CC-486
	IDH1 and IDH2	AG-120 and AG-221
Histone methylation	DOT1L	EPZ-5676
	LSD1	GSK2879552 and ORY-1001
Histone acetylation	HDACs	Vorinostat and panobinostat
Epigenetic 'readers'	BETs	GSK525762, OTX015 and
<b>Monoclonal antibodies</b>		
Leukaemia targets	CD33	SGN-CD33A, actimab-A and
	CD123	lintuzumab
	CD47	CSL362 and KHK2823
	CD56	CC-90002 and Hu5F9-G4
<b>Others</b>		
Immune checkpoints	PD1	Nivolumab
	CTLA4	Ipilimumab
Apoptosis regulators	BCL-2	ABT-199
	IAPs	Birinapant
	MDM2	RO5503781
Developmental pathways	Hedgehog	LDE225 (also known as
	WNT	sonidegib) and PF-04449913

## 1.3 S100A family of proteins

### 1.3.1 Overview of S100A protein family

The S100 proteins' family is a group of acidic and relatively small proteins that constitutes the largest sub-group of EF-hand-containing  $\text{Ca}^{2+}$  binding proteins (Donato *et al.* 2013). This family is comprised of 21 proteins that have a low molecular weight of around 8-12 kDa (Bresnick *et al.* 2015). Of the 21 human S100 proteins, 19 (S100 proteins, group A) are clustered on the chromosome 1q21 (Marenholz *et al.* 2004). Moreover, S100 proteins are exclusively expressed in vertebrates and often expressed in a tissue and cell-specific manner (Donato *et al.* 2013). Interestingly, the S100 proteins have different subcellular distributions in the cell. S100 proteins are given this name as they are soluble in 100% ammonium sulphate solution at neutral pH (Donato 1986). The first member of the S100 family was purified from bovine brain by (Moore 1965).

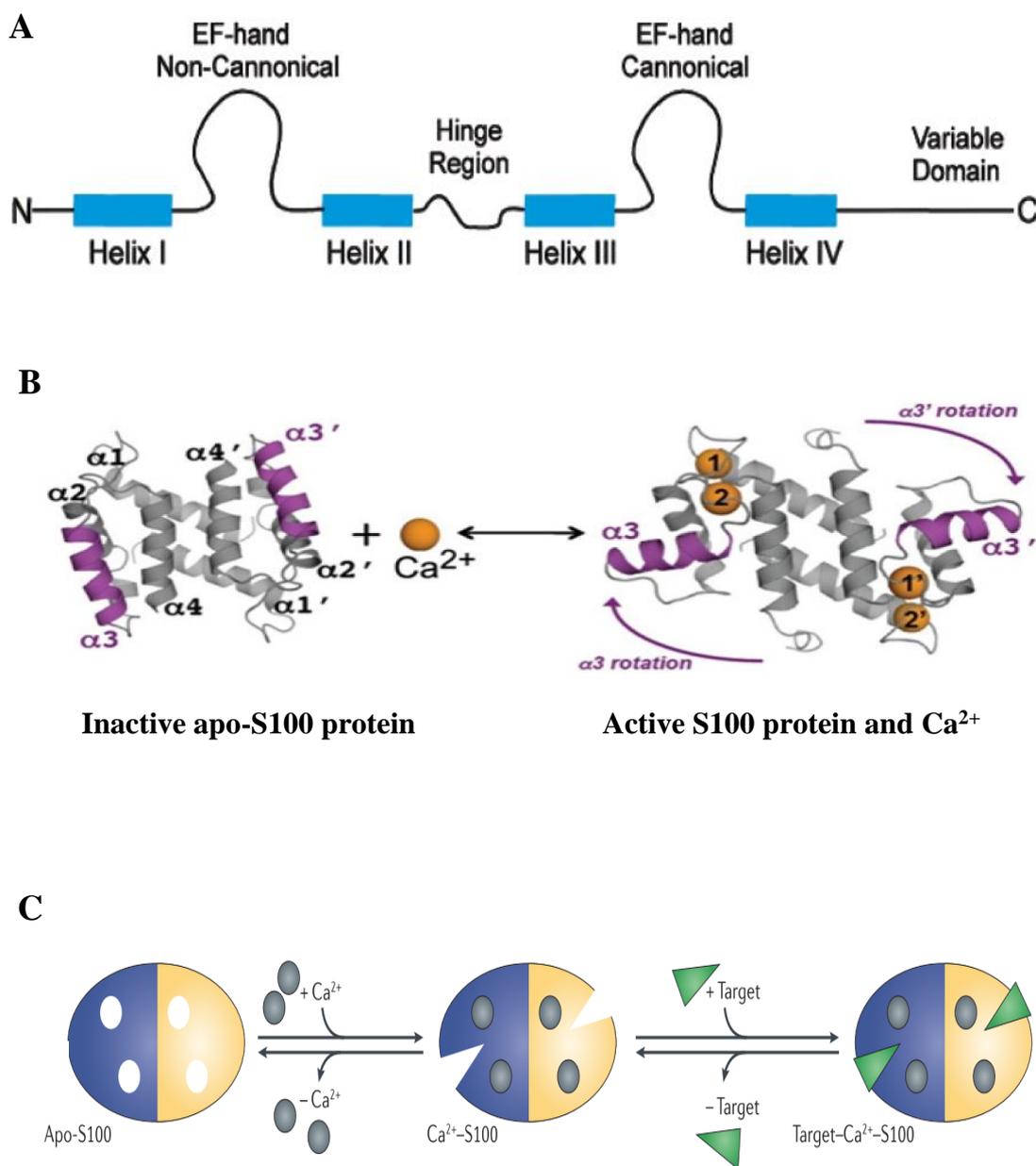
### 1.3.2 S100 proteins' structure

The structure of S100 proteins are symmetrically dimers in nature containing four  $\alpha$ -helices (Bresnick *et al.* 2015). EF-hand  $\text{Ca}^{2+}$  binding motifs consist of E- and F-helices in the form of helix-loop-helix (Heizmann *et al.* 2002). Typically, these EF-hands exist in pairs and are held together by short anti-parallel  $\beta$ -strands as well as hydrophobic interactions (Donato *et al.* 2013). The EF-hand  $\text{Ca}^{2+}$  binding motifs have two forms; canonical and pseudo EF-hands (Figure 1-2A). The canonical EF-hand, which is referred to as a "cEF hand", consists of 12-residues which are highly conserved in many  $\text{Ca}^{2+}$ -binding proteins such as troponin C (TnC) and calmodulin (CaM). The canonical EF-hand has six or seven oxygen ligands in the backbone which utilize five residues as  $\text{Ca}^{2+}$ -binding domains at positions 1, 3, 5, 7, and 12 of the helix-loop-helix (Strynadka and James 1989). On the other hand, the pseudo EF-hand, which is referred to as a "pEF hand", is unique to the S100 family and consists of 14 residues. The pEF hand has carbonyl oxygen atoms instead of oxygen atoms in the backbone and binds to  $\text{Ca}^{2+}$  at several of the ligands, which include Asn, Asp, Gln, or Glu residues (Strynadka and James 1989). The difference between the S100 proteins is in the position of H3 in the inactive form (apo-state) and that is due to amino acid sequence variations in the C-terminus of H4 and in L2 (the hinge region) (Heizmann *et al.* 2002). The C-terminal loop of S100 proteins is crucial for target protein binding. Thus, the lack of sequence homology in C-terminal loops suggests that S100 proteins

interact with binding partners in specific manner (Zhang *et al.* 2005). Studies show that the cEF has a higher binding affinity to  $\text{Ca}^{2+}$  with a dissociation constant of 27–250 mM compared to pEF 250–16,700 mM (Gross *et al.* 2014). S100 proteins are dimeric in structure and unlike other EF-hand containing proteins such as TnC and Cam, the dimeric structure prevents the movement of the exiting helix (helices 4, 4') upon  $\text{Ca}^{2+}$ -binding. This make only the entering helix (helices 3, 3') rotates at 90 degrees upon binding to  $\text{Ca}^{2+}$  (Donato *et al.* 2013), as shown in Figure 1-2B. Upon binding to  $\text{Ca}^{2+}$ , S100 proteins undergo conformational change to expose a hydrophobic residues in the first helix of the C-terminal loop which is necessary for interacting with target proteins Figure 1-2C (Moravkova *et al.* 2016). During this process, the affinity towards  $\text{Ca}^{2+}$  increases up to 300 times (Malashkevich *et al.* 2008).

## 1.4 S100A4 protein

S100A4 belongs to the *S100* multigene family of  $\text{Ca}^{2+}$  -binding proteins of the EF-hand type (reviewed in Boye and Maelandsmo 2010). S100A4 is also known as metastasin (Mts1), fibroblast-specific protein (FSP1), 18A2, pEL98, p9Ka, CAPL, and calvasculin. However, the new nomenclature of S100 proteins has renamed *S100* genes based on their genomic loci (i.e. S100A designated for genes clustered on 1q21 and the rest of S100 proteins located on the other chromosomes carry S100 followed by a letter) (Marenholz *et al.* 2004). S100A4 is a small protein consisting of 101-amino acids (Boye and Maelandsmo 2010). S100A4 is reported to be involved in a variety of intra- and extra-cellular processes including cell cycle, proliferation, and differentiation (Koshelev Iu *et al.* 2008). Moreover, overexpression of S100A4 is widely associated with poor prognosis, metastasis and progression in various types of solid tumours such as pancreatic and breast tumours (Helfman *et al.* 2005, Ismail *et al.* 2008). Further, S100A4 has been shown to negatively regulate key proteins in known pathological signalling pathways in cancer such as p53 and positively regulates of other protein that promote metastases and tumour progression such as myosin and actin (Binz *et al.* 2004, van Dieck *et al.* 2009).



### Figure 1-2: Schematic diagram of the S100 proteins' structure

**A)** Schematic diagram shows the overall structure of S100 protein family member which consists of four  $\alpha$ -helix domains, a central hinge region, two calcium-binding EF hands (one non-canonical EF hand, and one canonical), and N- and C-terminal domains. Figure is adopted from (Eckert *et al.* 2004). **B)** S100 undergoes conformational change upon binding to  $\text{Ca}^{2+}$  to expose hydrophobic residues in on the C-terminal loop to interact with its target proteins. This figure is adopted from (Donato *et al.* 2013). **C)** Illustrative cartoon shows Apo-S100 protein in blue and yellow subunits which upon  $\text{Ca}^{2+}$  (grey circles) binding induces conformation structural rearrangement to expose hydrophobic sites to allow target binding (green triangles). Figures is adopted from (Bresnick *et al.* 2015).

### 1.4.1 S100A4 function and structure

S100A4 has no known enzymatic activity and it exerts its functions via protein-protein interactions (Donato *et al.* 2013). S100A4 is involved in many intra- and extracellular functions. Intracellularly, S100A4 interacts with the heavy chain of non-muscle myosin II, F-actin, and tropomyosin and regulates cell motility, migration, and adhesion (Donato *et al.* 2013). Further, S100A4 promotes cell proliferation and survival via binding to p53 and induces MDM2-dependent degradation of p53 protein (Orre *et al.* 2013). Extracellular functions of S100a4 includes extracellular matrix remodelling, angiogenesis, migration, and invasion (Koshelev Iu *et al.* 2008).

S100A4 has a symmetric homodimer structure which has an overall topology similar to other S100 proteins in the family (Pathuri *et al.* 2008). Each monomer contains four  $\alpha$ -helices (H1, H2, H3, and H4) connected by three short loops called “hinge” [L1, L2, and L3] (Gingras *et al.* 2008). Moreover, S100A4 shares 50% similarity of its amino acid residuals with other member of S100 family and has a molecular mass of 10 kDa. However, S100A4 is different to other S100 proteins as the C-terminal loop (Phe89-Lys101) is long and very basic (Gingras *et al.* 2008). Further, the hinge (L2) that links helices H2 and H3, shares a less similar sequence homology with other S100 proteins. The C-terminal loop in S100A4 structure is important for metastasis-inducing properties. In a study by Zhang *et al.*, mutant S100A4 which has deletion of 15 amino acids residues in the C-terminal loop, has reduced  $\text{Ca}^{2+}$  binding by 26%, reduced motility/invasion, and impaired the interaction with its molecular target (NMMHC IIA) *in vitro* (Zhang *et al.* 2005).

### 1.4.2 Regulation of expression of S100A4

Several studies have reported that S100A4 is expressed in a variety of normal human cell types. These cells include fibroblasts (Schmidt-Hansen *et al.* 2004), monocytes, neutrophilic granulocytes, T lymphocytes (Cabezón *et al.* 2007), macrophages (Takenaga *et al.* 1994, and endothelial cells (Semov *et al.* 2005). Initial cloning studies found that S100A4 is highly expressed in growth-stimulated cultured cells (Goto *et al.* 1988) and metastatic tumour cell lines (Ebralidze *et al.* 1989). High S100A4 expression was also observed in adult mouse and rat tissues such as; bone marrow, spleen, thymus, neutrophils, T-lymphocytes, and macrophages (Grigorian *et al.* 1993).

Similar studies observed higher expression of S100A4 in embryonic macrophages and in differentiating mesenchymal tissues during mouse development (Garrett *et al.* 2006). In addition, the expression levels of S100A4 are up-regulated during oncogenic transformation in many solid tumours (Zhou *et al.* 2018, Li *et al.* 2013). S100A4 is predominantly expressed in the cytoplasmic compartment of the cell, however; it can be translocated into the nucleus by post-translational modifications (PTM) such as sumoylation process. In a study by Miranda *et al.*, they have shown that intracellular S100A4 translocated into the nucleus by SUMO-1 protein upon interleukin-1 (IL-1) stimulation (Miranda *et al.* 2010). Two sumoylation sites at Lys22 and Lys96 have been identified on the S100A4 molecule (Miranda *et al.* 2010). This study suggests that translocating S100A4 into the nucleus is crucial for the S100A4-driven metastasis phenotype by binding to matrix metalloproteinase-13 (MMP-13) to promote matrix remodelling. Moreover, not only does the overexpression of S100A4 linked poor prognosis and metastasis phenotype, but also nuclear overexpression of S100A4 is linked to poor prognosis in colorectal cancer (Boye *et al.* 2010) and aggressiveness in epithelial ovarian carcinoma (Kikuchi *et al.* 2006).

### 1.4.3 Role of S100A4 in cancer pathophysiology

#### 1.4.3.1 Role in metastasis

The overexpression of S100A4 has been linked to poor prognosis and metastatic potential in multiple solid tumours such as breast (Rudland *et al.* 2000), liver (Taylor *et al.* 2002), brain (Taylor *et al.* 2002) and prostate (Saleem *et al.* 2006). It has been reported that S100A4 interacts with cytoskeletal proteins including non-muscle myosin-heavy chain IIA (NMMHC IIA) and directly binds to it and as a result induces metastasis-associated cellular motility (Li and Bresnick 2006). Further, in colon cancer, upregulated  $\beta$ -catenin/T-cell factor complex upregulated the expression of S100A4 which induces migration and invasion *in vitro* and metastasis *in vivo* (Stein *et al.* 2006). Subsequently, knocking down  $\beta$ -catenin downregulates S100A4 expression and reduces cell migration and invasion (Stein *et al.* 2011). Alternatively, S100A4 has been shown to be involved in matrix remodelling by up-regulating the expression of MMP-13 translocating sumoylated intracellular S100A4 by SUMO-1 protein into the nucleus to bind to MMP-13 (Miranda *et al.* 2010).

### 1.4.3.2 Role in haematopoiesis and leukaemogenesis

The pathophysiological role of S100A4 in solid tumour is well studied and defined. However, little is known about the role of S100A4 in normal haematopoiesis and leukaemia. Early studies in paediatric AML patients show that *S100A4* mRNA expression was found to be higher in healthy controls by 3-fold (Steinbach *et al.* 2007). A study by Xu *et al.*, showed that overexpression of repressor of retinoic acid signalling (PRAME) downregulates the expression of S100A4 and increase the activity of p53 which subsequently increased apoptosis in the AML cell line KG-1 (Xu *et al.* 2016). Interestingly, when PRAME is knocked down in K562 (CML cell line), S100A4 expression is restored and p53 is decreased (Xu *et al.* 2016). Moreover, S100A4 has reported to promote chemoresistance in CML cell line. K562 cells with low expression levels of S100A4 showed chemosensitivity towards treatment with 4-hydroperoxy-cyclophosphamide compared to KU812 which has higher S100A4 expression levels (He *et al.* 2017). Further, it has been reported that S100A4 is expressed by mesenchymal stem cells (MSCs) in the BM. Thus, S100A4 is likely to be involved in the MSC proliferation and differentiation (Bresnick *et al.* 2015, Atlasi *et al.* 2016). Moreover, other member of S100 family are implicated in AML such as S100A8 and S100A9. S100A8 and S100A9 have previously been shown to be abundant in myeloid cells and associated with poor prognosis in AML; these studies focused on total expression and not subcellular expression (Nicolas *et al.* 2011, Laouedj *et al.* 2017).

## 1.5 Aims and objectives

In attempt to identify mislocalised and dysregulated proteins networks in AML blasts compared to normal human CD34<sup>+</sup> haematopoietic cells, our group analysed the AML nuclear proteome. The analysis of the proteome identified S100A4 as the highest differentially expressed protein in AML nuclei compared with normal CD34<sup>+</sup> nuclei (not previously implicated in AML) (Alanazi *et al.* 2019). Whilst S100A4 is associated with proliferation and metastasis of various types of solid tumours, the pathophysiological role of S100A4 as well as its subcellular distribution in haematological malignancies (as well as normal haematopoiesis) is unknown. The functional implication of altered S100A4 expression, subcellular localization and mechanisms of action in leukaemia remain unidentified. Further, the significance of expression and localization in normal haematopoietic development and to the pathogenesis of AML has not been studied. Thus, the main objective of this study is to gain an understanding of the role of S100A4 in haematopoiesis and AML. This project will achieve this objective through the following aims:

1. Determine the expression level and subcellular localisation of S100A4 during normal haematopoiesis, in differentiated myeloid subpopulations, and in normal bone marrow controls;
2. Determine the expression level and subcellular localisation of S100A4 in AML blasts;
3. Determine the effect of nuclear overexpression of S100A4 or S100A4 knock down on normal human haematopoietic cell growth, differentiation and development;
4. Determine the effect of nuclear overexpression of S100A4 or S100A4 knock down on AML cell growth, proliferation and apoptosis.
5. Identify potential S100A4 binding proteins in the cytoplasm and nucleus of ME-1 using the optimised protocol established above coupled with LC/MS.

# Chapter 2

## Materials and Methods

## 2 Chapter 2 - Materials and Methods

### 2.1 Reagents

#### Antibodies

The following antibodies were used in western blotting (2.9.3), flow cytometry (2.12), and immunoprecipitation (2.10) experiments.

**Table 2-1: The list of all antibodies used in this study.**

FC, flow cytometry; WB, western blotting; IP, immunoprecipitation; NA, not applicable; CST, Cell Signaling Technology; NEB, New England Biolab.

Antibody	Clone	Technique	Dilution/Conc.	Company	Lot No.
Anti CD36 biotin (1mg/mL)	NA	FC	NA	Ancell, Bayport, MN, USA	185104
Anti-GAPDH (Mouse) mAb	6C5	WB	1:20,000	Santa Cruz Biotechnology, Heidelberg, Germany	H2144
Anti-Histone 1 (H1) mAb (Mouse)	AE-4	WB	1:40,000	BioRad, Watford, UK	SAB3701132
Anti-Histone 3 (H3) mAb (Rabbit)	NA	WB	1:1,000	CST, London, UK	9715
Anti-Mouse IgG Horseradish Peroxidase linked whole Antibody	NA931	WB	1:1,000	GE Healthcare, Little Chalfont, UK	9729340
Anti-Rabbit IgG, Horseradish Peroxidase linked whole Antibody	NA934	WB	1:5,000	GE Healthcare, Little Chalfont, UK	9653374
Anti-RRP1B pPAb (Rabbit)	ab123397	WB/IP	1:500/250 µg/mL	Abcam, Cambridge, UK	ab123397
Anti-SUPT16H pPAb (Rabbit)	ab117081	WB/IP	1:500/1 mg/mL	Abcam, Cambridge, UK	ab117081
CD11b-PE	2LPM19C	FC	NA	DAKO, Ely, UK	000333469
CD13-APC	WM15	FC	NA	Biolegend, London, UK	B135747
CD14-PE	HCD14	FC	NA	Biolegend, London, UK	B174251

CD15-PE	W6D3	FC	NA	Biolegend, London, UK	B139265
CD34-APC	581	FC	NA	Biolegend, London, UK	B167218
CD34-PE	B13574 7	FC	NA	BD Biosciences, Oxford, UK	5119628
CD45-APC	H130	FC	NA	Biolegend, London, UK	B167510
CD45-FITC	H130	FC	NA	Biolegend, London, UK	B146601
DDX21 pAb (Rabbit)	NB100- 1716	WB/IP	1:2,000/0. 2 mg/mL	Novus Biological, USA	NB100- 1716
hnRNPM mAb (Mouse)	2A6	WB/IP	1:1,000/	Abnova, Taipei City, Taiwan	MAB233 9
IgG XP <sup>®</sup> Isotype control mAb (Rabbit)	DA1E	IP	2.5 mg/mL	CST, London, UK	3900
IgG1-APC	NA	FC	NA	DAKO, Ely, UK	38468
IgG1-FITC	NA	FC	NA	DAKO, Ely, UK	72359
IgG1-PE	MOPS- 21	FC	NA	Biolegend, London, UK	B202580
IgG2B-PE	MPC- 11	FC	NA	Biolegend, London, UK	B169497
MAK16 pAb (Rabbit)	NB100- 60425	WB/IP	1:2,000/0. 2 mg/mL	Novus Biological, USA	NB100- 60425
MCM7 mAb (Rabbit)	D10A1 1	WB/IP	1:1000/64 µg/mL	CST, London, UK	3735
PerCP Cy 5.5 Streptavidin	NA	FC	NA	BD Biosciences, Oxford, UK	03159
S100A4 mAb (Rabbit)	D9F9D	WB/IP	1:1,000/7 6 µg/mL	CST, London, UK	13018
S100A4 mAb (Rabbit)	1F12- 1G7	WB	1:1,000	Novus Biological, USA	NA
S100A4 mAb (Rabbit)	ab1830 92	WB	1:1,000	Abcam, Cambridge, UK	Ab18309 2
Ubiquityl-Histone H2A (Lys119) mAb (Rabbit)	D27C4	WB/IP	1:2000/61 6 µg/mL	CST, London, UK	8240

**General reagents****Table 2-2: The list of all materials used in this study.**

<b>Material</b>	<b>Company</b>	<b>Lot No.</b>
0.05% v/v Trysin EDTA	Gibco, Paisley, UK	1814197
10x Buffer for T4 DNA Ligase with 10mM ATP	New England Bio Labs (NEB), Ipswich, UK	0021506
1kb DNA Ladder	NEB, Ipswich, UK	1171310
50 µL vial of One Shot Stbl3 cells	Invitrogen, Bisley, UK	1807059
7AAD (1mg/mL)	Invitrogen, Bisley, UK	1795453
Agarose	Invitrogen, Bisley, UK	0000238148
Ampicillin (100mg/mL)	Sigma-Aldrich, Poole, UK	023M4080V
Annexin V Apoptosis Detection Kit APC	ThermoFisher Scientific, UK	88-8007
<i>Bam</i> HI HF (20000 U/mL)	NEB, Ipswich, UK	0941110
BD FACS Lysing Solution	BD Biosciences, Oxford, UK	79786
<i>Bgl</i> II (10000 U/mL)	NEB, Ipswich, UK	0421203
Bradford Reagent	Sigma-Aldrich, Poole, UK	SLBS2204V
Buffer TE	Qiagen, Manchester, UK	154021562
Dead Cell Removal kit	Miltenyi Biotec, Bisley, UK	130-090-101
ECL Prime Western blotting Detection Reagent	GE Healthcare, Little Chalfont, UK	9766187
<i>Eco</i> RI (20000 U/mL)	NEB, Ipswich, UK	0321201
<i>Eco</i> RI Buffer	NEB, Ipswich, UK	0011106
<i>Eco</i> RI HF (20000 U/mL)	NEB, Ipswich, UK	0321201
Ficoll-Paque Premium	GE Healthcare, Little Chalfont, UK	10249168
Gel Loading Dye Blue (6X)	NEB, Ipswich, UK	0021302
Gentamycin (50 mg/mL)	Gibco, Paisley, UK	1811725
Hanks' Balanced Salt Solution (1X)	Gibco, Paisley, UK	1413908
Heparin Sodium 1000 U/mL	Wockhardt, Wrexham, UK	NA
Hepes 1M	Invitrogen, Bisley, UK	804834
HiSpeed Plasmid Maxi Kit	Qiagen, Manchester, UK	154028153
His-Tag S100A4	R&D Systems	4137-S4
IMDM	Sigma	RNBF6090
Indirect CD34 Micro Bead Kit Human	MACS Miltenyi Biotec, Bisley, UK	5131122004
LB Broth	Sigma-Aldrich, Poole, UK	BCBQ1867V
LB Broth with Agar	Sigma-Aldrich, Poole, UK	SLBR5882V
Lipofectamine 3000 Transfection Kit	Invitrogen, Bisley, UK	1811507

Methanol	ThermoFisher Scientific, UK	1706141
Ethanol	ThermoFisher Scientific, UK	1706122
Isopropanol	ThermoFisher Scientific, UK	1705151
MS columns	Miltenyi Biotec, Bisley, UK	5161216101
NE Buffer 3.1 10x	NEB, Ipswich, UK	0161302
<i>Not1</i> 10,000U/ml	NEB, Ipswich, UK	0551602
Novex 10-20% Tricine gel	Invitrogen, Bisley, UK	16972540
Nuclear/Cytosol Fractionation Kit	Bio Vision, Cambridge, UK	1J140266
NuPage Antioxidant	Invitrogen, Bisley, UK	1771573
Opti-MEM	Gibco, Paisley, UK	1750005
PeqGreen DNA/RNA dye (20000X in water)	Peglab, Shanghai, China	1021004
Poly-L-lysine solution	Sigma-Aldrich, Poole, UK	RNBF6040
PVDF membrane (0.45µm)	Invitrogen, Bisley, UK	1802799
QIA prep Spin Miniprep Kit	Qiagen, Manchester, UK	154015847
QIAquick Gel Extraction Kit	Qiagen, Manchester, UK	142349618
QIAquick PCR purification kit	Qiagen, Manchester, UK	148010062
RetroNectin (1mg/mL)	Takara, Paris, France	AFY9002
RPMI 1640	Sigma-Aldrich, Poole, UK	RNBF6996
Sample reducing agent	Invitrogen, Bisley, UK	1762058
SeaKem GTG Agarose	Lonza, Nottingham, UK	0000511489
Soc medium	Invitrogen, Bisley, UK	1723796
T4 DNA Ligase 2,000,000U/mL	NEB, Ipswich, UK	1151506
Tricine SDS Running Buffer	Invitrogen, Bisley, UK	1772855
Tricine SDS Sample Buffer	Invitrogen, Bisley, UK	1692686
Tris Glycerine Transfer Buffer	NEB, Ipswich, UK	1772799
Tris-Acetate-EDTA Buffer 10X	Sigma-Aldrich, Poole, UK	SLBM3638V
Tris-Borate-EDTA Buffer 5X	Sigma-Aldrich, Poole, UK	SLBM8210V
ZAP-Oglobin II Lytic Reagent	Beckman Coult Inc., Galway, Ireland	129016

**Retro- and lentiviral vectors**

The following vectors were used in packaging cells transfection (2.5.6), CD34+ cells (1.2.5.1), and in AML transduction (2.6.2) experiments.

**Table 2-3: The list of Recto- and lentivirus used in this study**

Vector	Target Sequence information	Source
PINCO	PINCO GFP control	Eurofins MWG (Ebersberg, Germany)
PINCO	S100A4 PINCO GFP	Eurofins MWG (Ebersberg, Germany)
PINCO	1x NLS-S100A4 PINCO GFP	Eurofins MWG (Ebersberg, Germany)
PINCO	3x NLS-S100A4 PINCO GFP	Eurofins MWG (Ebersberg, Germany).
pHIV	pHIV-EGFP control	Addgene (Watertown, USA)
pHIV	S100A4 pHIV-EGFP	Eurofins MWG (Ebersberg, Germany)
pHIV	3x NLS-S100A4 pHIV-EGFP	Eurofins MWG (Ebersberg, Germany)
shRNA Control	Scramble shRNA	Mission® (Sigma, UK) #
shRNA S100A4 clone 1	S100A4 shRNA-pLKO.5 ( <a href="https://trcn.org/clone/S100A4/shRNA-pLKO.5/437516">TRCN0000437516</a> )	Mission® (Sigma, UK) #
shRNA S100A4 clone 2	S100A4 shRNA-pLKO.5 ( <a href="https://trcn.org/clone/S100A4/shRNA-pLKO.5/438093">TRCN0000438093</a> )	Mission® (Sigma, UK) #
shRNA S100A4 clone 3	S100A4 shRNA-pLKO.5 ( <a href="https://trcn.org/clone/S100A4/shRNA-pLKO.5/416498">TRCN0000416498</a> )	Mission® (Sigma, UK) #
shRNA S100A4 clone 4	S100A4 shRNA-pLKO.5 ( <a href="https://trcn.org/clone/S100A4/shRNA-pLKO.5/446826">TRCN0000446826</a> )	Mission® (Sigma, UK) #
shRNA S100A4 clone 5	S100A4 shRNA-pLKO.5 ( <a href="https://trcn.org/clone/S100A4/shRNA-pLKO.5/053608">TRCN0000053608</a> )	Mission® (Sigma, UK) #
pLV[shRNA S100A4]	pLV-EGFP:T2A:Puro-U6 ( <a href="https://trcn.org/clone/pLV-EGFP-T2A-Puro-U6/416498">TRCN0000416498</a> )	VectorBuilder, CA, USA
pLV[shRNA Control]	EGFP:T2A:Puro Control Scramble shRNA	VectorBuilder, CA, USA

#<http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna/library-information/vector-map.html#pLKO>

## **2.2 Cell Culture and cryopreservation**

All tissue culture work was performed in a Class II biosafety cabinet and liquid waste was disinfected with bleach or autoclaved. Cells were cultured under aseptic conditions at 37°C in 5% CO<sub>2</sub> in air unless otherwise specified.

### **2.2.1 Cell culture of suspension cell lines**

Each cell line was cultured according to supplier standard cell culturing guidelines and biohazard risk. Cells were cultured at a cell density of between 1 x 10<sup>5</sup> and 5x10<sup>5</sup> cells/mL and passage number did not exceed 20. Cell density was determined prior to subculturing by haemocytometer or flow cytometry. Cell culture growth medium and solutions were pre-warmed to room temperature (RT) prior to sub-culturing unless otherwise stated. The media used to culture each cell line are listed in Table 2-4.

### **2.2.2 Cell culture of adherent cell lines**

The viral packaging cell lines Phoenix and HEK293T were examined microscopically for cell confluence prior to subculturing. These adherent cells were detached from the tissue culture plastic ware using 1 mL / 25 cm<sup>2</sup> trypsin-EDTA for 3 minutes at room temperature (RT). Subsequently, an equal volume of appropriate growth medium was added to neutralise the trypsin activity and the cells were transferred into a fresh tube. If required, cells were counted before centrifugation for 10 minutes at 200 x g. Supernatant was discarded and the cell pellet was resuspended with fresh growth medium and seeded according to the culture discussions in Table 2-4.

**Table 2-4: Cell lines used in this study and their growth conditions requirements.**

Cell Line	Seeding density (cells/mL)	Medium Conditions	Source
HL60	$1 \times 10^5$	RPMI-1640 (Sigma-Aldrich, Dorset, UK) + 10 % FBS (Biosera, UK) + L-Glutamine + 1:2 k Gentamycin (50 mg/mL)	ATCC (USA)
K-562	$1 \times 10^5$		ECACC (UK)
U937	$1 \times 10^5$		ATCC (USA)
HEL	$1 \times 10^5$		ECACC (UK)
THP-1	$1 \times 10^5$		ECACC (UK)
NOMO-1	$1 \times 10^5$		DSMZ (Germany)
TF-1	$1 \times 10^5$		ATCC (USA)
ME-1	$1 \times 10^5$		RPMI-1640 (Sigma-Aldrich, Dorset, UK) + 20 % FBS (Biosera, UK) + L-Glutamine + 1:2 k Gentamycin (50 mg/mL)
Kasumi-1	$1 \times 10^5$	DSMZ (Germany)	
OCI-AML 2	$1 \times 10^5$	$\alpha$ -MEME (Sigma-Aldrich, Dorset, UK) + 20 % FBS (Biosera, UK) + L-Glutamine + 1:2 k Gentamycin (Sanofi, UK) + 10 ng/mL GM-CSF [For OCI-AML-5]	DSMZ (Germany)
OCI-AML 5	$1 \times 10^5$		DSMZ (Germany)
KG-1	$1 \times 10^5$	IMDM (Sigma-Aldrich, Dorset, UK) + 20 % FBS (Biosera, UK) + L-Glutamine + 1:2 k Gentamycin (50 mg/mL)	DSMZ (Germany)
MV4;11	$1 \times 10^5$	IMDM (Sigma-Aldrich, Dorset, UK) + 10 % FBS (BioWest, UK) + L-Glutamine + 1:2 k Gentamycin (50 mg/mL)	ATCC (USA)
Phoenix & HEK293T	$2.5 \times 10^6$	DMEM (Sigma-Aldrich, Dorset, UK) + 10 % FBS (BioWest, UK) + L-Glutamine + 1:2 k Gentamycin (50 mg/mL)	ATCC (USA)

### 2.2.3 Cell culture of haematopoietic progenitor cells

Freshly isolated CD34<sup>+</sup> cells (2.3) were cultured at a density of  $1 \times 10^5$  /mL in “36S<sup>hi</sup>” media mix Iscove modified Dulbecco medium (IMDM) (Sigma), containing (1% v/v BSA fraction V (Sigma), 20% v/v FCS, 45 pM beta-mercaptoethanol (Sigma), 360 pg/mL 30% iron-saturated human transferrin (Roche), 100 U/mL penicillin, 100 pg/mL streptomycin (Invitrogen)) supplemented with appropriate growth factors for each culture stage, as described below in Table 2-5. Following isolation of CD34<sup>+</sup> cells (defined as day 0), cells were cultured (days 0-3) in “36S<sup>hi</sup>” supplemented with the addition of human growth as shown below (Table 2-5A). Following 3 days in culture, cells were subsequently cultured in media mix supplemented with the addition of lower concentrations of cytokines: 5ng/mL of IL-3, SCF, GM-CSF and C-CSF, subsequently defined as “3S<sup>lo</sup>G/GM” (Table 2-5B).

**Table 2-5: Concentrations of human cytokines required for CD34<sup>+</sup> cells growth media mix**

Growth Factors	Stock (µg/mL)	Target conc. (ng/mL)	Vol/ 5 mL (µL)
<b>A. 36S<sup>hi</sup>G/GM (0-3 days)</b>			
IL-3	5	50	50
IL-6	10	25	12.5
SCF	20	50	12.5
G-CSF	5	25	25
GM-CSF	5	25	25
Flt3	5	50	50
1x PBS/BSA	-	-	25
TOTAL	-	-	200 µL
<b>B. 3S<sup>lo</sup>G/GM media mix (3- onwards)</b>			
Cytokines	Stock (µg/mL)	Target conc. (ng/mL)	Vol/ 1 mL (µL)
IL-3	5	50	1
SCF	5	50	1
G-CSF	5	25	1
GM-CSF	5	25	1
TOTAL	-	-	4 µL

### 2.2.4 Cell counting

Estimation of cell number was determined using haemocytometer counting chambers with improved Neubauer ruling (Hawksley, Brighton, UK). Briefly, 10  $\mu\text{L}$  of cell suspension was aliquoted from culture and directly pipetted under a counting chamber cover slip. The cellularity of the sample in cells/mL was given by the average number of cells per cuboid multiplied by  $1 \times 10^4$  cell/mL, as counted using Eclipse TS100 light transmission microscope (Nikon, Surrey, UK).

### 2.2.5 Cryopreservation and resuscitation of cells.

To freeze cells for continued culture later, between  $1-10 \times 10^6$  cells were collected by centrifugation at  $200 \times g$  for 5 minutes and resuspended in the appropriate growth medium into 1.8 mL cryopreservation vials (Nunc). An equal volume of freezing mix (50% v/v growth medium (RPMI or DMEM), 30% (v/v) FCS (Labtech), 20% (v/v) DMSO) was added dropwise to the cell suspension and immediately placed inside Nalgene Mr. Frosty™ freezing container (Thermoscientific, UK) that have been filled with 100% (v/v) isopropanol (Sigma-Aldrich, UK). The freezing container then placed at  $-80^\circ\text{C}$  overnight and the tubes were transferred to liquid nitrogen ( $\text{LN}_2$ ) for long-term storage.

## 2.3 Isolation of $\text{CD34}^+$ haematopoietic progenitor cells (HSPCs) from human cord blood

### 2.3.1 Cord blood collection

Neonatal cord blood (CB) samples were collected upon informed consent during elective caesarean sections at the Women units, University Hospital Wales, Cardiff. Cord blood was collected in 50 mL Falcon tubes containing 200  $\mu\text{L}$  heparin.

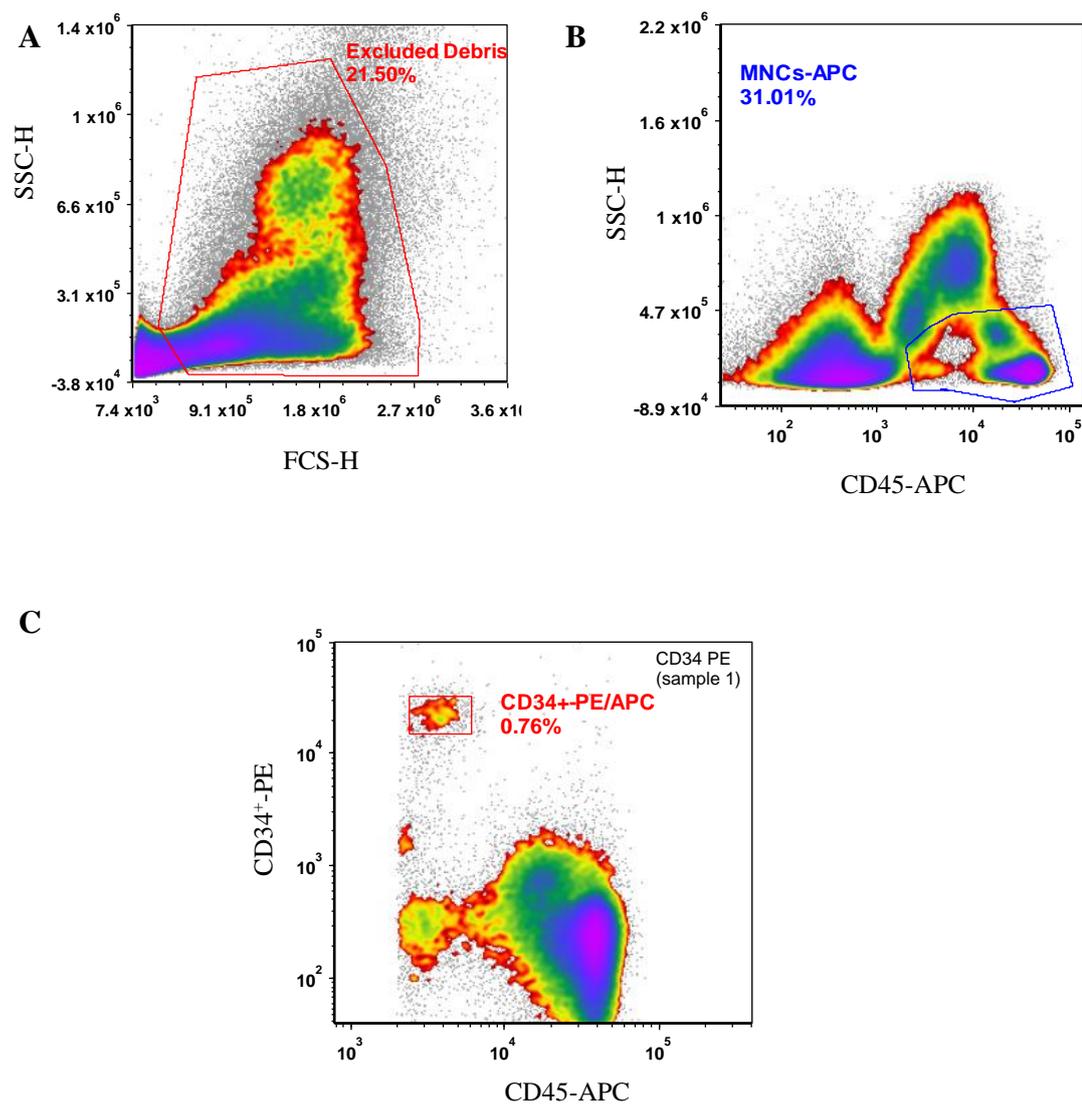
### 2.3.2 Isolation of Mononuclear cells (MNCs) from cord blood

Mononuclear cells (MNCs) were separated on Ficoll-Hypaque (Sigma-Aldrich, Poole Dorset, UK). Initially, 100  $\mu\text{L}$  aliquots of CB were analysed to estimate the number of isolated  $\text{CD34}^+$  cells (2.12). Subsequently, CB was diluted with Hanks solution 1:1 (v/v). Eight mLs were carefully layered on top of 5 mL of Ficoll in universal containers (UC). Samples were subsequently centrifuged at RT at  $400 \times g$  for 40 minutes. During

the centrifugation, 15 mL of RPMI medium (RPMI 1640 serum free medium) were pipetted in UCs. Following centrifugation, the plasma layer and the layer of mononuclear cells were carefully collected with a syringe and quill and ejected into UCs containing RPMI wash. Cells were centrifuged at 200 x g for 10 minutes. The supernatant was discarded without disrupting the MNCs pellet leaving approximately 1mL of RPMI at the bottom of the UCs as the pellet became loose tube. Subsequently, all cellular pellets were combined into one UC and filled up to 20 mL with RPMI wash. For counting the isolated MNCs, 10  $\mu$ L of cells pellet were removed and resuspended into 190  $\mu$ L of RPMI and 1.5  $\mu$ L Zaponin (to lyse red cells). The remaining sample was centrifuged at 200 x g for 10 minutes. The pellet was washed 3 times or till the supernatant was clear of platelets. Finally, the cell pellet was resuspended at a density of approximately  $5 \times 10^7$  cell/mL per vial and the MNCs were stored in liquid nitrogen as described above (2.2.5).

### 2.3.3 Immunophenotyping MNCs

To estimate the number of isolated CD34<sup>+</sup> cells in each sample processed, MNCs were immunophenotyped of CD34 positivity by flow cytometry. To each tube of 100  $\mu$ L of CB, 20  $\mu$ L of IgG1-PE (or anti-CD34<sup>+</sup>-PE) was added with 5  $\mu$ L CD45 APC and incubated for 30 minutes at 4°C. Meanwhile, 1x lysis buffer was prepared by diluting 1mL of BD FACS Lysis buffer with 9 mL dH<sub>2</sub>O. Upon the end of the incubation time, 5mL diluted FACS Lysis buffer were added. Samples were incubated for 10 minutes at RT. Following incubation, isotype control and CD34<sup>+</sup> stained samples were washed with 5mL 1x phosphate buffered saline (PBS) and centrifuged at 200 x g for 10 minutes. Subsequently, the supernatant was discarded leaving a residual volume of buffer of approximately 200  $\mu$ L to resuspend the pellet. The samples were then transferred to appropriate tubes and analysed with the flow cytometer (2.12).



**Figure 2-1: Estimation of CD34<sup>+</sup> cells in the MNCs population.**

A) Example of non-debris gating strategy to remove debris from the analysis. B) Using the non-debris gate applied, CD45 expression was used to identify different population of white cells including the MNCs. C) Upon gating on the “MNC cell gate”, CD34 positivity was determined using IgG-PE to set the background autofluorescence in the PE channel as described in (2.12).

### 2.3.4 Isolation of CD34<sup>+</sup> cells from MNCs

Human CD34<sup>+</sup> cells (> 95% pure) were isolated using MiniMACS<sup>®</sup> CD34<sup>+</sup> Progenitor Cell Isolation Kit (Miltenyi Biotec, Camberley, Surrey, UK). MNCs were removed from LN<sub>2</sub> and rapidly thawed in 37°C then transferred to a UC (2ml max per UC). A volume of 1 mL of FCS and 20 µL of DNase (100µg/mL) were added. An equal volume of RT 1x PBS with magnesium chloride (MgCl)(MACS buffer) was added dropwise so that volume doubles over 3 minutes (mixed gently after each addition). This step was repeated twice, subsequently, UCs were centrifuged at 200 x g for 10 minutes. The supernatant was carefully aspirated and 10 µL of sterile DNase was added to each UC. Cells were resuspend by tapping pellet in tube in 400 µL cold MACS buffer per 10<sup>8</sup> cells. 100 µL of solution A1 and 100 µL of solution A2 were added per 10<sup>8</sup> cells. UCs were mixed gently and incubated for 15 minutes at 4°C. Next, 5 mL of MACS buffer per 10<sup>8</sup> cells was added and centrifuged as above. Buffer was subsequently aspirated, and cells resuspended in 400 µL of MACS buffer per 10<sup>8</sup> cells. A volume of 100 µL of reagent B per 10<sup>8</sup> cells was added and cells were incubated for 15 minutes at 4°C. Next, 5 mL of column buffer were added per 10<sup>8</sup> cells. Cells suspension were passed through cell strainer and rinsed strainer with 1ml buffer to recover residual cells. The flow through cells were centrifuged at 200 x g for 5 minutes. Subsequently, MACS buffer was removed without disturbing pellet and washed with 500 µL of column buffer per 10<sup>8</sup> cells and applied to the first column with a flow restrictor in place. Once column flow had stopped, flow restrictor was removed and a further 500 µL of MACS buffer were added. This step was repeated 3 times. When last wash was complete, column was removed from the magnet and 1 mL of MACS buffer was added to top of the column; push plunger firmly into column top to release bound CD34<sup>+</sup> cells into clean tubes. Cells from final elution were counted on haemocytometer. Purity of CD34<sup>+</sup> cells was determined by flow cytometry (2.12). Then, CD34<sup>+</sup> cells either frozen in LN<sub>2</sub> for future analysis as described (2.2.5) grown to day six in appropriated supplemented primary cell medium and used in transduction experiments (2.6.1). Differentiated myeloid sub-lineages (monocytes, erythrocytes, and granulocytes) were isolated by positive selection process using MiniMACS<sup>®</sup> (Miltenyi Biotec, Camberley, Surrey, United Kingdom) as previously described by (Tonks *et al.* 2007).

### **2.3.5 Cell surface phenotyping and differentiation analysis**

To evaluate CD34<sup>+</sup> cells purity and differentiation status, cell surface phenotyping using flow cytometry was performed. Cells were stained with CD13-APC (Biolegend, London, UK) combined with CD36-biotin (Ancell, Bayport, MN, USA) and one of the following PE-labelled antibodies: CD34<sup>+</sup>-PE (BD Biosciences, Oxford, UK), CD11b-PE, CD14-PE, CD15-PE (Biolegend, London, UK). CD36 was subsequently labelled with SA-PerCP (Pharmingen BD Biosciences, Oxford, UK). All incubations were carried out at 4°C for 30 minutes. All reactions were controlled with the appropriate isotype-matched antibody, as described in 2.12.

## **2.4 AML patients' samples, FAB subtype determination and viability analysis**

Peripheral blood (PB) or bone marrow (BM) AML patients' samples (n = 33) were obtained with ethical approval from the AML Clinical Trials Research Tissue Bank at the Haematology Department, Cardiff University. AML blast viability and cell surface phenotype were analyzed by flow cytometry to support FAB classification. AML patient blast FAB-M1 subtype was confirmed using early surface marker CD34<sup>+</sup>-PE and differentiation markers CD14-PE and CD15-PE (Biolegend, London, UK) for FAB-M4 subtype. In this study the experimental design was restricted to minimally differentiated FAB-M1 since this subtype has little developmental heterogeneity and was developmentally matched to normal human HSPCs derived from neonatal CB. All AML samples analysed in this study were  $\geq 80\%$  viable cells. Cells viability was analysed manually using Trypan Blue solution 4% (1:1) (Sigma-Aldrich, Poole, UK) and confirmed by flow cytometric viability stain propidium iodide (PI) (2.12). Viable cells were recovered from samples with  $\leq 80\%$  viable cells using Dead Cell Removal kit (Miltenyi Biotec, Baisley, UK).

## **2.5 Preparation of recombinant plasmid DNA**

### **2.5.1 Restriction enzyme conditions for creation of PINCO constructs**

The PINCO expression vector co-expressing S100A4 and GFP from an internal CMV promoter was created by directionally sub cloning S100A4 (NM\_002961.2) into

the *Bam*H1/*Eco*R1 restriction sites as previously described by (Grignani *et al.* 1998, Tonks *et al.* 2005). To target expression of S100A4 to the nucleus, an additional vector was prepared where the N-terminus encoded a 1x and 3x nuclear localisation sequence (NLS) “GATCCAAAAAAGAAGAGAAAGGTA” (Fischer-Fantuzzi and Vesco 1988) as shown in Figure 2-2. In this way four cultures were generated; control vector expressing GFP alone, 1xNLS-S100A4, 3x NLS-S100A4 overexpressing S100A4 in the nucleus, and the forth construct expresses S100A4 without a nuclear targeting sequence. Restriction enzyme (RE) digest for the creation of PINCO expression plasmids was performed in a total reaction volume of 100  $\mu$ L using 5–10  $\mu$ g of plasmid DNA containing appropriate NEB (New England Biolabs Ltd., Hitchin, UK) buffer, RE enzymes, BSA and water. Thus, pEX-A2 were digested with *Bam*H1 and *Not*I RE. For both enzymes, a 100 U enzyme was used with a compatible NEB4 buffer. All reactions were mixed gently by pipetting and incubated at 37°C in a water bath for 1 hour. Aliquot of digested DNA was electrophoresed using 0.8% (*w/v*) agarose gel and 1xTAE buffer at 80 V for 45 minutes and visualised using PeqGreen (20000X) loading dye.

### 2.5.2 Agarose gel purification of insert DNA

To purify DNA inserts generated by RE, preparative agarose gel was prepared by creating a 0.8% (*w/v*) Seakem agarose solution in 1x Tris Acetic acid and EDTA (TAE)buffer (or 3% (*w/v*) agarose). DNA samples for loading contained 5-10  $\mu$ g digested DNA, 200  $\mu$ M bromophenol blue and 120  $\mu$ M xylene cyanol per lane. Samples were electrophoresed in 1X TAE buffer at 40 V for 2 hours. The size of DNA fragments or plasmids was estimated using a 1 kb DNA Ladder. Upon completion of electrophoresis, agarose gel was stained in 10  $\mu$ L of PeqGreen (20000X) for 20 minutes. Gel was de-stained using water for 20 minutes and DNA was visualised by long wavelength UV transillumination using a LAS-3000 digital imaging device (Fujifilm UK Ltd., Bedford, UK) before excising the desired DNA band by using a clean scalpel. All apparatus used was initially wiped with 70% (*v/v*) ethanol. The excised gel was then placed inside a sterile 15 mL tube and subjected to further purification using the QIAquick Gel Extraction kit protocol (Qiagen®).

### 2.5.3 Ligation of S100A4 DNA into retroviral or lentiviral vectors

For directional subcloning, PINCO GFP vector (Grignani *et al.* 1998) and insert DNA fragments of pEX-A2 ‘carrying’ 1x and 3xNLS S100A4 were purified according to 2.5.2) prior to ligation. A final volume of 10  $\mu$ L ligation reaction was used, containing insert DNA combined with vector DNA using a 5:1 ratio of insert:vector ligation, 200 cohesive end units/ $\mu$ L T4 DNA ligase (1  $\mu$ L) and 1X T4 DNA ligase reaction buffer (1  $\mu$ L) (New England Biolabs Ltd., Hitchin, UK). In order to determine the level of background ligation due to self-ligation of the linearised vector plasmid, a background reaction tube was also set up as above, but lacking insert DNA. All reaction tubes were incubated for between 4 - 6 hours at 16°C then stored at -20°C before performing the transformation into competent *E.coli*.

### 2.5.4 Transformation of competent cells with plasmid DNA

Each plasmid DNA to be propagated by transformation was thawed on ice alongside one 50  $\mu$ L vial of OneShot® Stbl3 or OneShot® Top10 chemically competent *E. coli* (Invitrogen™). For each reaction, 6  $\mu$ L of DNA sample was aseptically transferred directly into the respective vial of competent cells and tapped gently to mix, followed by 30 minute incubation on ice. Following this, the cells were further incubated at 42°C for 30 seconds, without agitation, before replacing on ice and adding 250  $\mu$ L of pre-warmed SOC medium (Invitrogen™) to each vial to maximise transformation efficiency. All vials were secured in a shaking incubator at 37°C for 1 hour at 225 revolutions per minute (rpm). Once the incubation had elapsed, between 50  $\mu$ L to 100  $\mu$ L of transformation mixture was spread onto pre-warmed LB-agar plates containing 100  $\mu$ g/mL ampicillin (or kanamycin, using sterile plastic spreaders to generate a film of bacterial growth, and incubated at 37°C overnight. Transformed colonies were selected on the following day for large scale purification of plasmid DNA as described below.

### 2.5.5 Purification of plasmid DNA

Plasmid DNA was prepared from an *E.coli* strain using HiSpeed™ Maxi prep kit (Qiagen®) based on manufacturer’s recommendations. Firstly, colonies of interest were picked from LB-agar plates and a single colony was inoculated into 5 mL of LB-broth containing 5  $\mu$ L of ampicillin or kanamycin (stock 100 mg/mL) by using a sterile inoculation loop and incubated in 37°C incubator with 225 rpm shaker for 8 hours.

Following that, the starter culture was diluted 1 in 500 into LB-broth with appropriate antibiotic and further incubated in 37°C with 225 rpm shaker for 16 hours. After overnight incubation, the bacterial pellet was harvested from 150 mL culture by centrifugation at 6000 x g for 15 minutes at 4°C. Briefly, pelleted bacteria were resuspended in 10 mL Buffer P1 before mixing with 10 mL of Buffer P2, and tube was allowed to stand for 5 minutes at RT. Next, 10 mL of Buffer P3 was added followed by incubation on ice for 15 minutes, and centrifugation at 20,000 x g at 4°C for 30 minutes. Supernatant containing plasmid DNA was removed and centrifuged again, as above, to remove insoluble material. During this process, QIAGENtip-500 was equilibrated by adding 10 ml Buffer QBT onto the tip. The bacterial supernatant was transferred to the equilibrated QIAGEN-tip where plasmid DNA was bound to the column and washed twice with 30 ml Buffer QC. Subsequently, DNA was eluted from the column using 15 ml Buffer QF and precipitated by the addition of 10.5 ml isopropanol and centrifugation at 15,000 x g for 30 minutes at 4°C. The DNA pellet was finally washed using 5 mL of 70% (v/v) ethanol before final centrifugation at 15,000 x g for 10 minutes. The supernatant was decanted with care, and the purified DNA pellet was air-dried for 10 minutes before resuspended in 1 mL of water and concentration of DNA was quantified using the NanoDrop<sup>®</sup> and stored at -20°C.

### **2.5.6 Preparation of retro- and lentivirus**

Phoenix cells were used as the viral packaging cell line for retroviruses and HEK293T for lentiviruses and cultured as described above (2.2.2). These cells were subcultured at a density of  $1 \times 10^6$  cells/mL in 15 mL media. The following day, for preparation of retro- or lentiviruses, cDNA (45 µg) cell transfection was achieved using Lipofectamine<sup>®</sup> 3000 (Life Technologies, UK). The Lipofectamine<sup>®</sup> 3000 lipid-DNA complexes were prepared according the manufacturer's instructions. Opti-MEM 1 reduced Serum Medium was brought to room temperature. The Opti-MEM 1 medium was reduced in each flask to 6 mL. Lipid DNA complexes were carefully added, gently mixed, and flasks were incubated for 6 hours at 37°C, 5% CO<sub>2</sub>. At 6 hours post-transfection, flasks mediums were replaced with (9 mL for [F75] and 3 mL for [F25] flasks) of pre-warmed retro- or lentivirus packaging mediums. Then, flasks were returned to incubator boxed in a containers. On day 2 post-transfection, viral infection was evaluated under a fluorescent microscope. To harvest viruses, supernatants from flasks

were removed and transferred to labelled UCs. Next, UCs containing viral supernatant were centrifuged at 200 x g for 10 minutes. After centrifugation, 0.5 mL of virus were aliquoted in 1 mL cryotubes and snapped frozen in LN<sub>2</sub> and stored at -80 °C. The medium in flasks were replaced with fresh medium (7.5 mL for [F75] and 2.5 mL for [F25] flasks). Flasks were then incubated at 37°C, 5% CO<sub>2</sub>. Retroviruses were harvested and snap frozen on two consecutive days.

## **2.6 Determining the effect of nuclear mislocalised S100A4 on CD34<sup>+</sup> cells and AML cell lines**

### **2.6.1 Retro- and lentiviral infection of CD34<sup>+</sup> cells**

In order to infect CD34<sup>+</sup> cells, cells were thawed and revived in 36S<sup>hi</sup>-G/GM (Table 2-5A) and incubated at 37°C in 5% CO<sub>2</sub> overnight. To infect cells with 1x and 3xNLS S100A4 PINCO or pHIV vectors, a 24-well plate was coated with 500 µL retronectin (30 µg/mL) (Takara Bio) and incubated for 2 hours at RT. The following day, retronectin was aspirated and replaced with 250 µL of 1% v/v BSA (Sigma) in 1x PBS (Invitrogen) and left for 30 min at RT. The BSA was then aspirated and replaced with 1 mL retroviral or lentiviral supernatant. The culture plate was centrifuged at 2200 x g for 2 hours at 12°C. The supernatant was carefully removed from the wells and replaced with CD34<sup>+</sup> cells (2 x 10<sup>5</sup> cells/mL). In case of retroviral infection the process was repeated on the second day, cells were removed to a UC and placed into an incubator (37°C, 5% CO<sub>2</sub>), the wells were treated with retrovirus as above and the cells returned to the wells and incubated at 37°C, 5% CO<sub>2</sub> overnight. On day 3 post infect, CD34<sup>+</sup> cells were harvested from the plate and resuspended with 3S<sup>lo</sup>G/GM (Table 2-5B). Cells expressing GFP were assessed alongside with negative control treated similarly but without viral infection (mock cultures) by flow cytometry (2.12) to measure infection rates. Cells were centrifuged at 200 x g for 5 minutes, washed in 1x PBS (Invitrogen), centrifuged at 200 x g for 5 minutes resuspended in 1x PBS and analysed by flow cytometry (2.12). Following infection, CD34<sup>+</sup> cells were analysed for lineage commitment and phenotypic analysis of early cell surface differentiation markers at three-time points; day 6, 8, and 13.

### 2.6.2 Retro- and lentiviral infection of AML cells lines

A 24-well was treated with Retronectin (30 µg/mL) for 2 hours at RT. The BSA was aspirated from all wells and 1 mL of either retro- or lentivirus was added. The plate was centrifuged at 2200 x g for 2 hours. During centrifugation, AML cells were diluted in fresh growth medium to  $4 \times 10^5$  cells/mL. Following centrifugation, the virus was pipetted off and discarded. Subsequently, 1 mL of cells were immediately aliquoted to their designated wells and incubated at 37°C in 5% CO<sub>2</sub> at day of infection (defined as Day 0). The cells were left to rest on day1 post infection. On day 2 of culture the supernatant was discarded. The cells were resuspended in 1mL fresh medium and transferred into a labelled UC containing 5 mL fresh medium. Cells were pelleted by washing in fresh complete medium and centrifugation at 200 x g for 5 minutes and resuspended to a density of  $4 \times 10^5$  cells/mL in fresh medium. Cell viability was assessed using 7AAD (1 µg/mL) coupled with flow cytometric analysis (2.12) and analysed in (2.6.3)

### 2.6.3 Effect of serum on proliferation of infected AML cell lines

To determine the effect of nuclear overexpression of S100A4 on the proliferation of AML cells, a proliferation/viability assay was conducted. Cells in log phase growth ( $4-8 \times 10^5$  cells/mL) were pelleted in a UC and washed in 10 mL of serum free medium. Following centrifugation (as above), the medium was discarded and the wash with serum free medium was repeated. Subsequently, 90 µL of cells were aliquoted into designated well of a 96 U-well plate and each plate contained three replicates. Three plates were set up for analysis over the next three days. The serum dilutions were added of 10 µL to each well containing cells (0%, 1%, 3% and 10% v/v serum), then the plate was mixed by vortexing. Plates were incubated at 37°C in 5% CO<sub>2</sub>. Cells were analysed for viability and proliferation change after 24 hours, 48 hours and 72 hours using PI (1 µg/mL). To analyse cells, 10 µL of PI were added to 90 µL of the cells were transferred into FACS tubes. Then, cells were analysed on the flow cytometer (2.12).

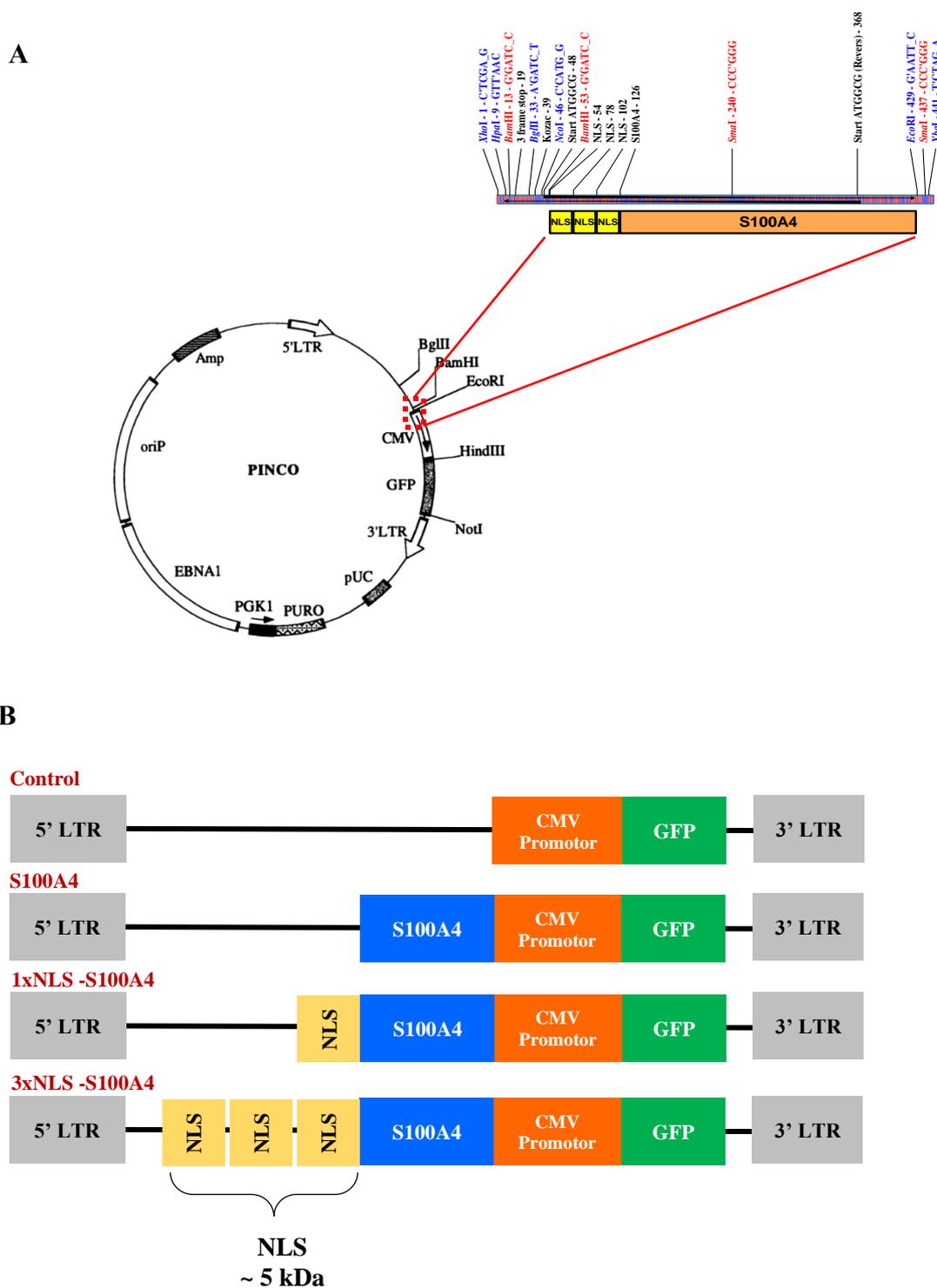
## 2.7 Determining the effect of S100A4 knockdown on CD34<sup>+</sup> cells and AML cell lines

### 2.7.1 Generating shRNA lentiviral vectors

For knock down studies, Mission<sup>®</sup> sRNA vectors based on TRC(1)2-pLKO.5-puro(S100A4 shRNA and non- mammalian shRNA control) were purchased om Sigma-Aldrich, Poole, UK, as shown in Table 2-3. Five independent shRNA vectors were used to knockdown S100A4 in NOMO-1 cells. Subsequently, the TCR(1)2-pLKO.5-puro (s100A4 shRNA [[TRCN0000416498](#)] and non-mammalian shRNA control) were redesigned into a new vector containing an EGFP marker, as shown in Figure 2-3B. The S100A4 with EGFP (Table 2-3) was purchased from VectorBuilder (California, USA). This vector shRNA ([TRCN0000416498](#)) was used in all S100A4 KD in CD34<sup>+</sup> cells and AML cell lines.

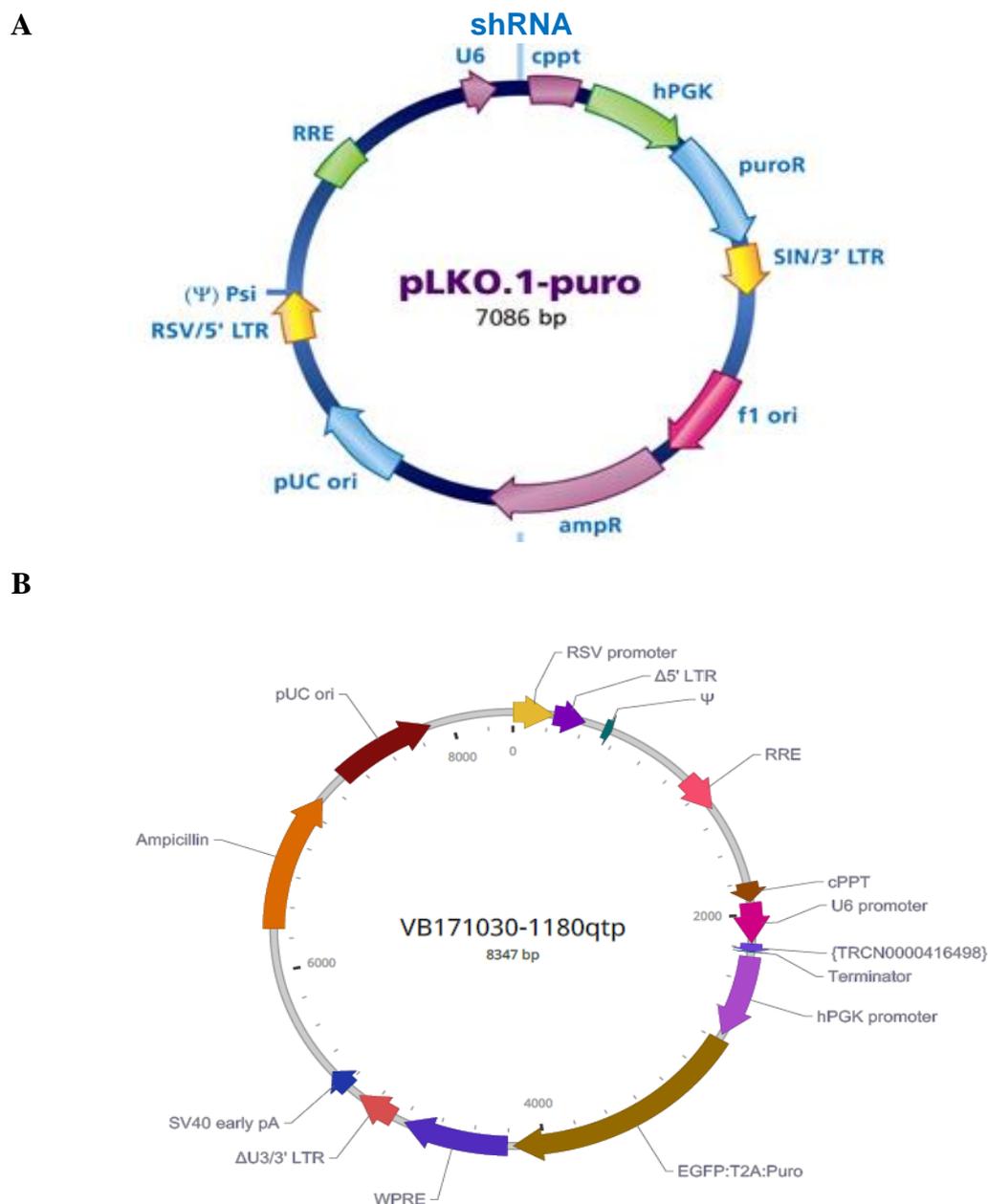
### 2.7.2 Growth and apoptosis assays

Following infection of CD34<sup>+</sup> cells and AML cell lines (2.6.1 and 2.6.2), cells were seeded at a density of  $1 \times 10^5$  cells/mL on day 3 post infection in 12-well plate (mock, shRNA control, and shRNA S100A4) lines. Cells were grown in culture over 5 days post infection. On each day, cells were resuspended in a plate and 100  $\mu$ L of cells were removed from culture of each infected cells into FACS tubes. Cells were washed with 1x PBS then resuspended with 90  $\mu$ L of staining buffer (1xPBS+ 1% v/v BSA, 1% v/v Na-azide) and 10  $\mu$ L of viability stain PI was added. Cells were counted using flow cytometry and viability status was daily (2.12). Cells were resuspend in 1x binding buffer at  $1-5 \times 10^5$  cells/mL and 5  $\mu$ L of Annexin V-APC fluorochrome-conjugated were added to 100  $\mu$ L of the cell suspension and incubated 10-15 minutes at RT in the dark. After incubation, cells were washed once in 1X binding buffer and resuspend in 200  $\mu$ L of 1x binding buffer and 5  $\mu$ L of PI were added and analyzed by flow cytometry.



**Figure 2-2: Retroviral vector constructs based on PINCO.**

**A)** Limited restriction enzyme map of the PINCO retroviral expression vector. Generation of S100A4 and NLS-S100A4 by directionally cloning into the BamHI and EcoRI restriction enzyme sites. Long terminal repeat, LTR; cytomegalovirus, CMV; green fluorescent protein, GFP; 3x nuclear localisation sequence (NLS; GATCAAAAAAAGAAGAGAAAGGTA) (figure modified from Grignani et al. 1998). **B)** Graphical representation of the retroviral vectors created for this study.



**Figure 2-3: shRNA S100A4 pHIV vectors used in study.**

shRNA vector maps showing sequences of interest **A**) p-LKO.2-puro (Mission®, Sigma-Aldrich, Poole, UK). **B**) pHIV-EGFP:T2A:Puro-U6 (VectorBuilder, California, USA). U6 (RNA polymerase III U6 promoter), cppt (central polypurine tract), hPGK (human phosphoglycerate kinase eukaryotic promoter), puroR (puromycin resistance), SIN3'LTR (3' self inactivating long terminal repeat), f1 ori (f1 origin of replication), ampR (ampicillin resistance), pUC ori (pUC origin of replication), RSV 5'LTR (Rous Sarcoma virus 5' long terminal repeat), ΨPsi (RNA packaging signal), RRE (Rev response element), RSV promoter (Rous sarcoma virus enhancer/promoter), Δ5' LTR (Truncated HIV-1 5' long terminal repeat), Ψ (HIV-1 packaging signal), U6 promoter Terminator (Pol III transcription terminator), hPGK promoter (Human Phosphoglycerate kinase 1 promoter), EGFP:T2A:Puro (EGFP and Puro linked by T2A), WPRE (Woodchuck hepatitis virus posttranscriptional regulatory element), ΔU3/3' LTR (Truncated HIV-1 3' long terminal repeat), SV40 early pA (Simian virus 40 Early polyadenylation signal), Ampicillin (Ampicillin resistance gene), pUC ori (pUC origin of Replication)

## 2.8 Protein extraction and quantification

### 2.8.1 Total Protein extraction

Frozen cell pellets were thawed on ice in the presence of 1  $\mu$ L of DNase (1  $\mu$ g/mL) (Sigma) for 5 minutes with regular tapping the tube to ensure cell pellet thoroughly mixed with DNase. Cells ( $1 \times 10^6$  cells/mL) were resuspend in 50  $\mu$ L of homogenisation buffer (composed of 10 mM HEPES-KOH (Invitrogen), 10 mM 2-mercaptoethanol (Sigma), 1 mM magnesium acetate (Fisons Scientific Equipment), 0.5 mM EDTA (Sigma), 0.5 mM EGTA (Sigma), 0.25 M sucrose (Sigma), 1 mM  $\text{Na}_3\text{VO}_4$  (Sigma), 1 complete mini EDTA, protease inhibitor cocktail [PIC] tablet per 10 mL (Roche), 1% v/v x100-Triton (Sigma)), incubated on ice for 30 minutes with occasional vortexing. The cell lysate was centrifuged at 13,000 g for 5 min at 4°C. The supernatant (cell lysate) was aspirated into a clean and pre-chilled 1.5 mL Eppendorf tube. Later, protein concentration of the supernatant determined by Bradford assay (2.8).

### 2.8.2 Nuclear/Cytosol Extraction and Fractionation

Nuclear and cytosol protein extraction was achieved using the Nuclear/Cytosol Fractionation kit (Biovision, California, USA). Cells ( $2 \times 10^6$  cells/mL) were pelleted and washed twice with 20 mL Tris-buffered saline (TBS), followed by centrifugation for 10 min at 200 x g. Cells were subsequently resuspended in 200  $\mu$ L cytosol extraction buffer A (CEB-A) containing 1x protease inhibitor cocktail (PIC) (Sigma) and 1mM dithiothreitol (DTT) and vortexed for 15 seconds and incubated on ice for 10 min. This buffer contains PIC and DTT to reduce the samples and inhibit protease activity during protein extraction. The supernatant was carefully aspirated and discarded. 11  $\mu$ L of ice-cold CEB-B was added and the sample was vortexed for 5 seconds and incubated on ice for 1 minute. Following the incubation, cells were centrifuged for 8 minutes at 10,000 x g at 4°C.

## 2.9 Antibody optimisation and western blotting

### 2.9.1 S100A4 antibodies optimisation

Three commercial primary S100A4 monoclonal antibodies (mAb) were chosen and tested for optimal binding and specificity for S100A4: clone D9F9D (Cell Signaling Technology [CST], Danvers, USA); clone 1F12-1G7 (Novus Biological, USA); clone Ab183092 (Abcam, UK) (Table 2-1). These antibodies were tested on western blot (Figure 3-2) using the cytoplasmic fraction of a small cohort of AML cell lines (NOMO-1, ME-1, and THP-1). A dilution factor of 1:1000 was used as per manufacturer instructions. Subsequent experiments used S100A4 mAb from CST (clone D9F9D ; optimised between 1:1000 and 1:10,000 on cytoplasmic/nuclear fractions of ME-1 cells).

### 2.9.2 His-Tag S100A4 recombinant protein

A recombinant human S100A4 protein derived from *E. coli*- (Ala2-Lys101), with a C-terminal 6-Histidine tag (R&D Systems, Abingdon, UK) was purchased and used as a positive control for S100A4 M.W. on western blot (Figure 3-4). A range of serial dilutions were analysed by dot blot to determine the maximum and minimum loading weights of His-Tag S100A4 for western blot analysis. Subsequently, His-tag S100A4 was probed with serial dilutions of mAb to S100A4 (clone D9F9D) between 1:1,000 and 1:50,000 as shown in Figure 2-4. Further, the loading concentration of His-Tag S100A4 protein was optimised for loading weights of 20 and 30 ng coupled with increasing S100A4 mAb dilution (1:1,000, 1:10,000, and 1:50,000).

### 2.9.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and electroblotting

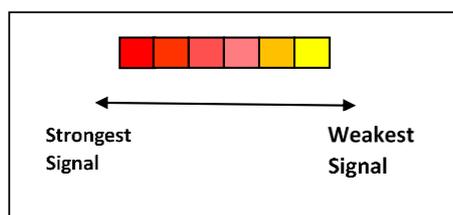
Protein electrophoresis for western blotting was performed using the NuPAGE<sup>®</sup> 12% Bis-Tris Tricine GEL electrophoresis and electroblotting systems (ThermoFisher Scientific). A maximum of 20 µg of protein extracts were denatured by incubation for 10 minutes at 90°C in the presence of 2x *sodium dodecyl sulphate* (SDS) sample buffer and 10x antioxidant. Following centrifugation at 10,000 xg at 4°C for 1 minute. In order to estimate the M.W. of proteins a 10-fold dilution of WesternSure<sup>®</sup> Pre-Stained Chemiluminescent Protein Ladder was used. The Tricine SDS-PAGE gel was left equilibrate to RT. Tricine running buffer (1x) was prepared from 10x stock. 500 µL

antioxidant reagent was added to 200 mL running buffer and was filled in the inner chamber of a NOVEX electrophoresis tank. The samples were centrifuged for 5 minutes at 4°C prior loading on gel. The Tricine gel was placed in the NOVE X-Cell II™ mini-cell containing 1x Tricine running buffer in the lower chamber. The electrophoresis was performed at 125 mA for 70 minutes.

During electrophoresis, 1x Tricine transfer buffer containing 20% (v/v) methanol, 1x NuPAGE® Tricine transfer Buffer and 1 mL antioxidant was prepared and used to soak blotting pads for 60 minutes. The Polyvinylidene fluoride or polyvinylidene difluoride (PVDF) membrane (0.45 µm pore size) was also pre-soaked for 5-10 minutes in 100% methanol. Following membrane incubation, the PVDF was washed with transfer buffer for 1 minute. The gel was subsequently removed from the tank and moistened with transfer buffer. During this process, a blotting pad (or sponge) was soaked with transfer buffer and was carefully layered onto the moistened gel avoiding any air bubbles. The other side of the SDS-PAGE gel was also moistened with transfer buffer, before layering a pre-soaked PVDF membrane ensuring all trapped air bubbles were removed. The sandwiched between pre-soaked blotting pads, the SDS-PAGE gel and PVDF membrane (illustrated in Figure 2-5) was then inserted into an XCell II™ Blot Module. The chamber was filled with 1x Tricine transfer buffer and the outside of the tank filled with dH<sub>2</sub>O and protein transfer was performed over 1 hour at 30 V. Following transblotting process, the membrane was washed with 20 mL of dH<sub>2</sub>O on a plate shaker for 5 min (x2) to remove residual gel piece and transfer buffer traces. The water was drained off and the blot and proteins visualised using Ponceau S solution for 30 seconds (Sigma), in order to facilitate cutting of membrane if required. The Ponceau S solution was drained off the membrane and the membrane was washed with dH<sub>2</sub>O as above.

Subsequently, PVDF membrane was incubated on a plate shaker for 60 minutes at RT for the blocking step in TBS-T (Tris-*buffered* saline [TBS], 0.1% v/v Tween-20) and 2.5% w/v milk powder (Marvel), then washed for 15 minutes in TBS-T and then a further three times for 5 minutes in TBS-T. Afterwards, membrane was subsequently used for immunoblotting (2.9.4).

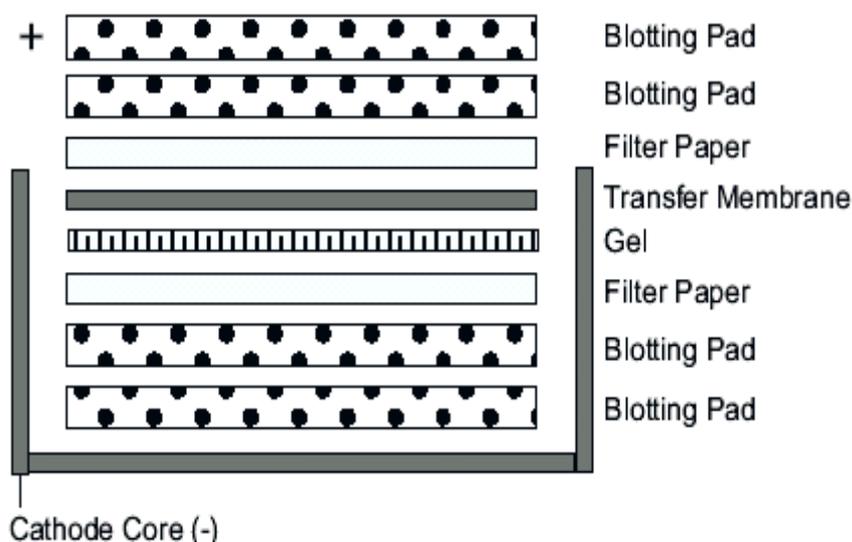
	His-Tag S100A4 Protein							
	150 ng	15 ng	1.5 ng	150 pg	15 pg	1.5 pg	0.15 pg	
S100A4 mAb (CST, clone: D9F9D)	1:1 k	A1	B1	C1	D1	E1	F1	G1
	1:5 k	A2	B2	C2	D2	E2	F2	G2
	1:15 k	A3	B3	C3	D3	E3	F3	G3
	1:30 k	A4	B4	C4	D4	E4	F4	G4
	1:50 k	A5	B5	C5	D5	E5	F5	G5



**Figure 2-4:** Diagram illustrates the distribution of His-Tag S100A4 protein serial molecular weights and serial dilutions of S100A4 mAb (CST, clone D9F9D) on dot blot.

### 2.9.4 Immunoblotting

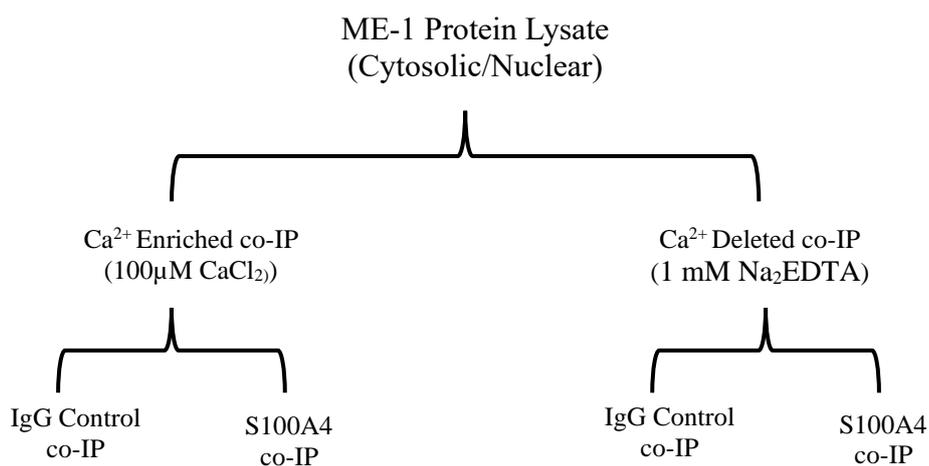
Membranes were incubated overnight at 4°C, in 10 mL TBS-T containing 2.5% *w/v* milk powder, containing primary antibodies at concentrations described in (Table 2-1). Membranes were washed with 10 mL of TBS-T buffer for 15 minutes on a plate shaker followed by three subsequent 10 mL washes for 5 minutes. Subsequent incubation of the membrane occurred at RT on a plate shaker in the presence of TBS-T containing 1% *w/v* milk powder and secondary anti-mouse or anti-rabbit horse radish peroxide conjugated antibody (GE Healthcare), diluted 1:5000, for 60 minutes. Following incubation, membranes were washed four times in TBS-T as above and target proteins were visualised using the ECL Advance Detection Kit (GE Healthcare), according to the manufacturer's instructions. Chemiluminescence was detected using a LAS-3000 image analyser (Fujifilm UK Ltd) and analysed using AIDA Image Analyser Version 4.19 (Fujifilm UK Ltd). To confirm equal loading of proteins the PVDF membrane was washed and reprobed as above using 1:20,000 dilution of anti-GAPDH (Clone 6C5, Santa Cruze, USA) for whole cell protein extraction, or anti-GAPDH for cytosol protein extraction and 1:40,000 dilution of anti-histone H1 (Clone AE-4, BioRad, UK) or 1:1,000 of H3 (Clone 9715, CST, UK) for nuclear protein extraction (Table 2-1).



**Figure 2-5: Diagram of transblotting sandwich set up**

## 2.10 Co-Immunoprecipitation (co-IP)

To identify S100A4 binding partners in leukaemia cells, ME-1 cells were fractionated into cytosolic/nuclear as described above (2.8.2). To achieve this, two co-IP approaches were used as detailed below. In order to retain Ca<sup>2+</sup>-based interactions between S100A4 and its protein binding partners, all co-IP buffers are supplemented with 100  $\mu$ M CaCl<sub>2</sub>. To provide negative control for Ca<sup>2+</sup>-based interactions, a similar set of co-IP experiments were conducted and all buffers used were supplemented with metal free chelator buffer (1 mM Na<sub>2</sub>EDTA) to strip Ca<sup>2+</sup> ion. As illustrated below in Figure 2-6.



**Figure 2-6: Experimental design of co-IP enriched with Ca<sup>2+</sup> and depleted from Ca<sup>2+</sup>**

### 2.10.1 Direct co-IP

The direct co-IP is based on pre-immobilization of the antibody to the magnetic beads followed by incubation with protein lysate mixture (as illustrated in Figure 2-7A). Firstly, S100A4 (CST, clone D9F9D) or IgG1 isotype control (CST, clone DA1E mAb) antibodies were crosslinked to the beads. To do this, the protein A Dynabeads (ThermoFisher Scientific, UK) were vortexed and 165  $\mu$ L of beads (enough for 5 immunoprecipitations) were placed into a 1.5 mL protein LoBind Eppendorf tubes, a magnet was applied and the supernatant (storage buffer) was removed. The beads were then re-suspended in 165 $\mu$ L of PBST (*v/v* PBS + 0.02% Tween-20), 25  $\mu$ g of S100A4 antibody was added and the mixture was made up to 1 mL in PBST. The tube was incubated at RT for 2 hours with rotation. Following incubation, a magnet was applied to the tube and the supernatant was discarded. The antibody-bound beads were washed three times in PBST with inverting. The beads were then washed three times in coupling buffer (PBST + 0.2 M triethanolamine pH 9) before incubation with 1 mL freshly prepared dimethyl phthalate (DMP) solution (coupling buffer + 2 mM DMP) for 30 minutes at RT with rotation. Then, a magnet was applied, and the supernatant was discarded before repeating the incubation with fresh DMP solution. The DMP solution was discarded and the beads were incubated for 30 min at RT with 1 mL of quenching buffer (PBST + 50mM ethanolamine). The beads were washed three times in elution buffer (0.2 M glycine pH. 2.5, 0.01% Tween-20 pH. 2.5 dH<sub>2</sub>O) and then re-suspended in 165  $\mu$ L of TBST and stored at 4° C. All steps unless otherwise specified were performed on ice.

Following crosslinking of antibodies to beads, immunoprecipitation of S100A4 and its binding partners was conducted. First a preclearing step was conducted. 33  $\mu$ L of IgG1 isotype control antibody-bound beads were placed in an Eppendorf tube and the supernatant was removed and replaced with 33 $\mu$ L of co-IP lysis buffer. The protein lysates (1000  $\mu$ g/mL) of were added and made up to 500 $\mu$ L in co-IP lysis buffer (1x PBS + 0.015% *v/v* IGEPAL-40). The tube was incubated for 6 hours in a cold room with rotation. A magnet was applied and the supernatant (precleared lysate) was transferred to a clean tube. Thirty-three  $\mu$ L of antibody-bound beads was added to the lysate and the tube was incubated overnight in the cold room with rotation. The next day, a magnet was applied to the tube and the supernatant was transferred to a clean tube and stored at -80° C for use in western blotting analysis of immunoprecipitation efficiency (2.9). The beads

were washed 6 times with 500 $\mu$ L of TBST wash buffer with inverting and incubation on ice for 5 minutes. To elute S100A4 and its binding partners from the beads, they were boiled in 2x SDS buffer (NuPAGE<sup>®</sup> 2xSDS, 50 nM NuPAGE<sup>®</sup> reducing agent and dH<sub>2</sub>O) at 95°C for 5 min. The eluted protein was stored at -20° C ready for LC/MS analysis (2.11).

### 2.10.2 Indirect co-IP

Alternatively, an indirect co-IP approach was undertaken where the antibody was incubated with a “protein mixture bait” to form an immune complex with the target protein antigen then retrieved by the magnetic beads (as illustrated in Figure 2-7B). Protein lysates (1000  $\mu$ g/mL) were prepared in co-IP lysis buffer (1x PBS/0.015% v/v IGEPAL-40) placed into a 1.5 mL protein LoBind Eppendorf tubes and incubated with 25  $\mu$ g of S100A4 mAb (CST, clone D9F9D) or IgG isotype matched control antibody (IgG mAb XP, clone: DA1E, CST) for 3 hours at 4 °C. The magnetic beads (Dynabeads<sup>®</sup> Protein A, ThermoFisher Scientific, UK) were washed with 300  $\mu$ L of wash buffer 3x times. Following the final wash, the magnetic beads were blocked with 50  $\mu$ L of co-IP lysis buffer + 1mg/mL BSA and incubated at RT for 15 minutes with rotation. Subsequently, both S100A4 and IgG control antibodies were cross-linked with magnetic beads and incubated for 2 hours at 4°C with rotation. Following incubation, unbound proteins in S100A4 and IgG samples were washed off the beads and analysed by western blot to assess IP efficiency. Bound S100A4 was eluted off the beads using 20 ml elution buffer (NuPAGE<sup>®</sup> 2xSDS, 50nM NuPAGE<sup>®</sup> reducing agent and dH<sub>2</sub>O) and samples were incubated at 95°C for 5 minutes. The eluted protein was stored at -20° C.

All steps unless otherwise specified were performed on ice. All samples were electrophoresed and analysed by western blot as described above (2.9) using clone D9F9D to detect S100A4 expression. To validate S100A4 co-IP candidate binding partners, reciprocal co-IP was conducted under same conditions. S100A4 candidate binding partners were pulled down and confirmed whether S100A4 is pull down with them. Co-IP efficiency and reciprocal co-IP validation were confirmed by western blot.

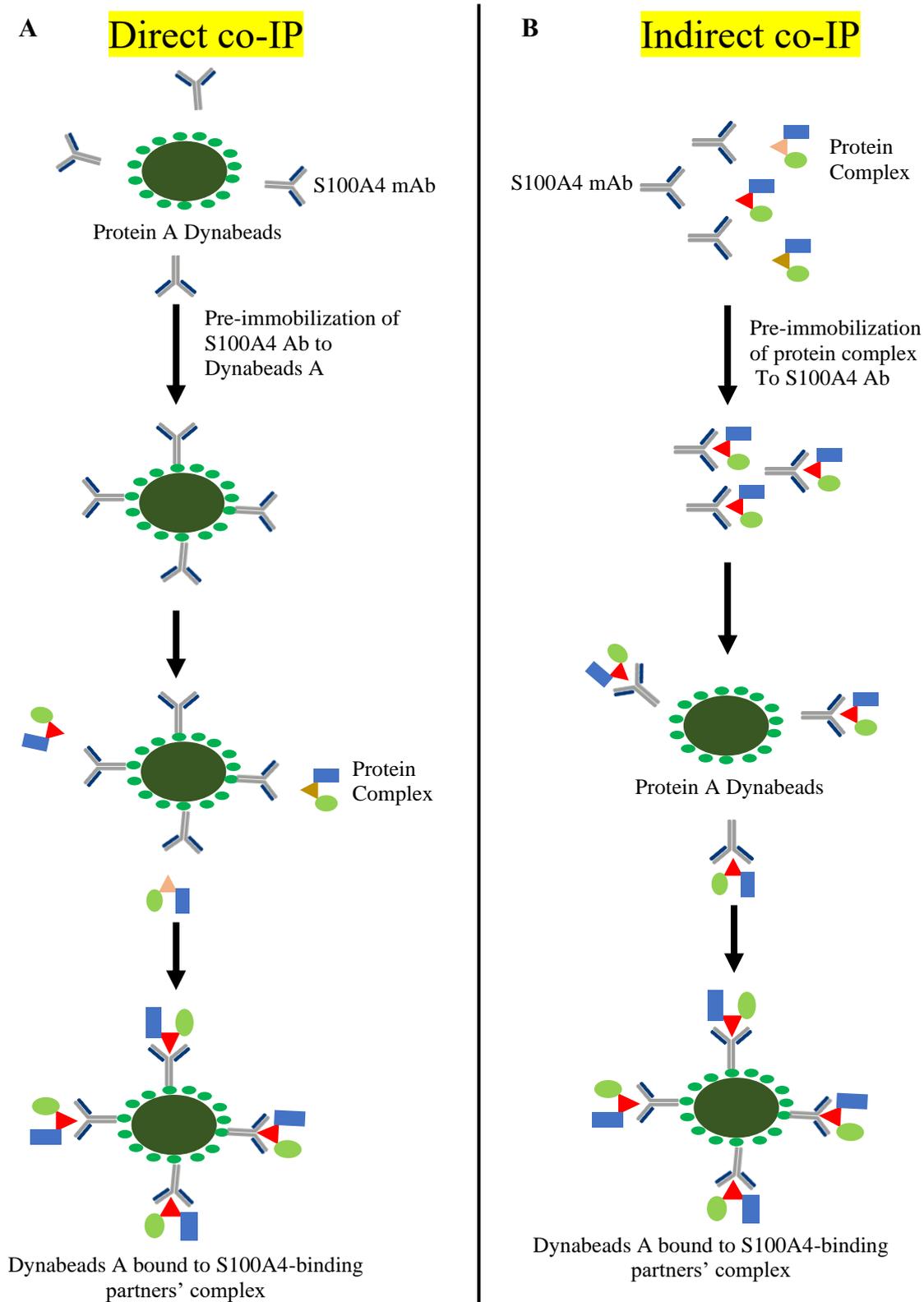
### 2.10.3 Florescent staining of SDS-PAGE gel

To evaluate the co-IP efficiency, protein complexes were subsequently separated using SDS-PAGE gel and visualised within gel using SYPRO<sup>®</sup> Ruby (Sigma-Aldrich,

UK). All washing and staining steps were performed by floating the gel face down on the solution with gentle agitating (on an orbital shaker at 50 rpm). After electroblotting proteins, the Tricine gel is placed face down in a container with 100 mL of fix solution (7% v/v acetic acid, 10% v/v methanol) and incubated for 15 minutes. Following fixing, gel was washed 3x in ultrapure water for 10 minutes each, before proceeding to the staining step. Then gel was transferred to a clean container and 60 mL of SYPRO<sup>®</sup> Ruby gel stain was added agitate gently on an orbital shaker overnight. Following staining, the Tricine gel was transferred to a clean container and washed in 100 mL of wash solution (1xPBS + 0.02% v/v Tween-20. PBS pH7.4) for 30 minutes. The transfer step helps minimize background staining irregularities and stain speckles on the gel. Before imaging, the gel was rinsed in ultrapure water, a minimum of 2x times for 5 minutes each to prevent possible corrosive damage to the imager. SYPRO<sup>®</sup> Ruby protein blot stain has two excitation maxima, one at ~280 nm and one at ~450 nm and has an emission maximum near 618 nm. The Tricine gel stained with the dye was visualized using a 300 nm UV transilluminator.

**Table 2-6: co-IP's buffers composition.**

Buffer	Compositions
<b>A. Direct co-IP solutions</b>	
Co-IP lysis buffer	1x PBS/0.015% v/v IGEPAL-40
Wash buffer (PBST)	1xPBS + 0.02% v/v Tween-20. PBS pH7.4.
Coupling buffer	PBST + 0.2 M triethanolamine. pH 9
Dimethyl phthalate (DMP) solution	coupling buffer + 2 mM DMP
Quenching buffer	PBST + 50mM ethanolamine
Elution buffer	0.2M glycine pH2.5, 0.01% Tween-20 pH.2.5 dH <sub>2</sub> O
<b>B. Indirect co-IP solutions</b>	
Co-IP lysis buffer	1x PBS + 0.015% v/v IGEPAL-40
Wash buffer (PBST)	1xPBS + 0.02% v/v Tween-20. PBS pH7.4.
Elution buffer	2x Tricine SDS loading buffer and sample reducing agent



**Figure 2-7: Illustrative diagram of difference between direct and indirect co-IP.**

**A)** Direct co-IP, antibody pre-immobilised with magnetic beads first then incubated with protein mixture. **B)** Indirect co-IP, protein mixture incubated with antibody first then bound Ab-protein mixture is pre-immobilised to magnetic beads.

## 2.11 Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

Following co-IP of S100A4 and binding partners, samples were sent for LC/MS analysis at the proteomic facility at Bristol University. The complete co-IP/LCMS pipeline analysis is detailed in Figure 2-8. Samples were separated using SDS-PAGE, with each sample being run approximately 1 cm into the separating gel. Each gel lane was then excised and subjected to in-gel tryptic digestion using a DigestPro automated digestion unit (Intavis Ltd.). The resulting peptides were fractionated using an Ultimate 3000 nano-LC system in line with an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). In brief, peptides in 1% (v/v) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (v/v) acetonitrile 0.1% (v/v) formic acid peptides were resolved on a 250 mm × 75 μm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1min., 6-15% B over 58 minutes, 15-32%B over 58 minutes, 32-40%B over 5 minutes 40-90%B over 1 minute, held at 90%B for 6min and then reduced to 1%B over 1 minute) with a flow rate of 300 nL/minute. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid.

Peptides were ionized by nano-electrospray ionization at 2.1 kV using a stainless-steel emitter with an internal diameter of 30 μm (Thermo Scientific) and a capillary temperature of 250°C. Tandem mass spectra were acquired using an LTQ-Orbitrap Velos mass spectrometer controlled by Xcalibur 2.1 software (Thermo Scientific) and operated in data-dependent acquisition mode. The Orbitrap was set to analyse the survey scans at 60,000 resolution (at m/z 400) in the mass range m/z 300 to 2000 and the top twenty multiply charged ions in each duty cycle selected for MS/MS in the LTQ linear ion trap. Charge state filtering, where unassigned precursor ions were not selected for fragmentation, and dynamic exclusion (repeat count, 1; repeat duration, 30s; exclusion list size, 500) were used. Fragmentation conditions in the LTQ were as follows: normalized collision energy, 40%; activation q, 0.25; activation time 10ms; and minimum ion selection intensity, 500 counts. The raw data files were processed and quantified using Proteome Discoverer software v1.4 (Thermo Scientific) and searched against the UniProt Human database (140000 sequences) using the SEQUEST algorithm. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.8Da. Search criteria included carbamidomethylation of cysteine (+57.0214) as a fixed

modification and oxidation of methionine (+15.9949) as a variable modification. Searches were performed with full tryptic digestion and a maximum of 1 missed cleavage was allowed. The reverse database search option was enabled, and all peptide data was filtered to satisfy false discovery rate (FDR) of 5% or 1%.

## **2.12 Flow cytometry and data analysis**

For flow cytometric analysis, a minimum of 10,000 - 50,000 events were collected using Accuri C6 instrument (BD Biosciences, Oxford, UK) and data analysed using FCS express v6 (De Novo, California, USA). Cells to be analysed by flow cytometry were obtained from culture and centrifugated at 5000 rpm for 5 minutes. Subsequently, cells washed in 1 ml staining buffer (1xPBS+ 1% v/v BSA, 1% v/v Na-azide) per  $1 \times 10^6$  cells, then centrifugated as above. Next, cells were resuspended in 25  $\mu$ L of staining buffer and incubated with primary stage antibodies as per experimental protocol. For secondary stage staining, cells were washed as above and resuspended in 25  $\mu$ L of secondary antibody and incubated. All reactions were controlled with the appropriate isotype-matched antibody. Generally, forward scattered light (FSC) was used to determine size and side scattered light (SSC) the granularity. Bi-variate gating strategy applied to allow the exclusion of cell debris and other sub-cellular size detected events. The threshold for GFP positivity was determined from the autofluorescence of identically treated mock-transduced cultures. Gating strategy was indicated in figure's legends in appropriate results sections. Immunophenotypic analysis of infected CD34<sup>+</sup> cells were performed at the time points indicated above 2.3.5).

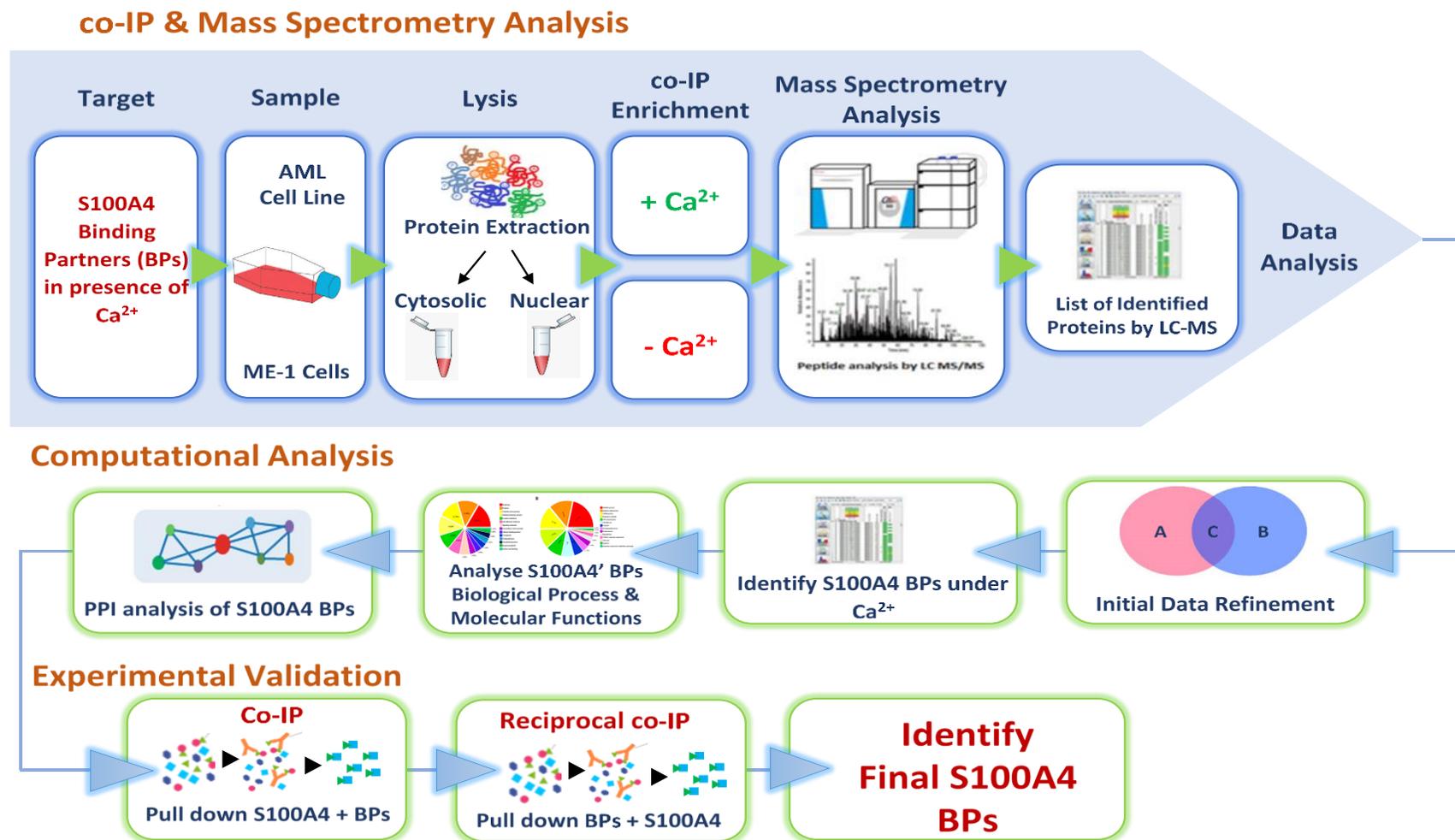


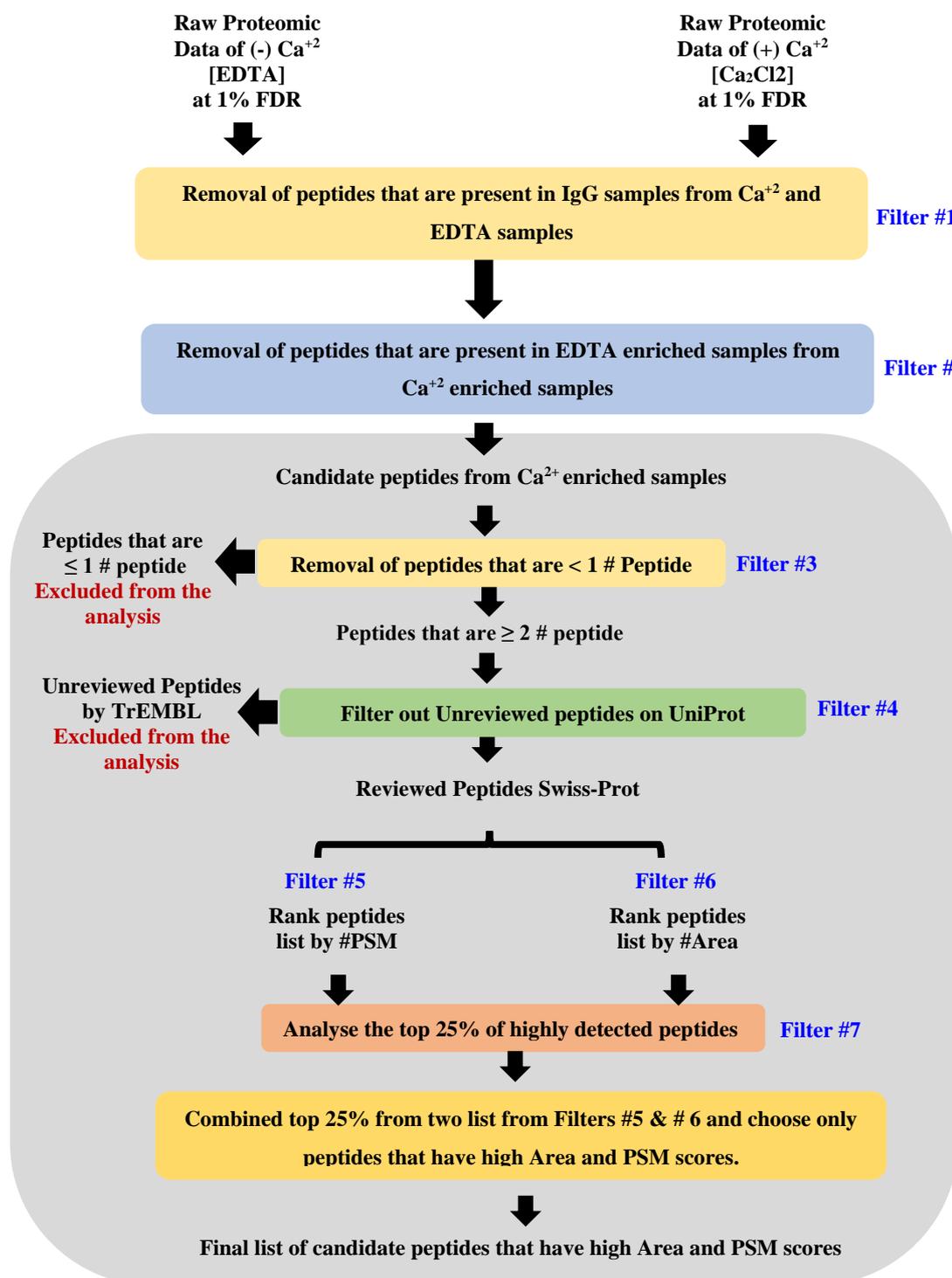
Figure 2-8: Graphical pipeline of co-IP/LCMS analysis of S100A4 binding partners in AML cell line (ME-1)

### 2.13 Bioinformatics analysis and online databases

The *S100A4* mRNA expression data were analysed in normal haematopoiesis and AML abnormalities using the RNAseq data by The Cancer Genome Atlas (TCGA) data set utilised through BloodSpot database (Tomczak *et al.* 2015, Bagger *et al.* 2016) (URL: <http://servers.binf.ku.dk/bloodspot/>). The proteomic data generated by LC/MS were analysed as per the data filtration strategy detailed below in Figure 2-9. The true peptide sequences identified by LC/MS were validated using “UniProt Knowledge” database (URL: <https://www.uniprot.org/>). The Gene Ontology (GO) molecular functions (MF) and biological processes (BP) of S100A4 candidate binding partners were analysed using PANTHER14.1, the 2019\_04 release of (based on **UniProt Release 2019\_04, Ensembl release 95 and Ensembl Genome release 42**). (URL: <http://www.pantherdb.org/>). Protein-protein interaction (PPI) network analysis performed using STRING (version 11.0) (URL: [https://string-db.org/cgi/input.pl?sessionId=T6th4zhhWCHE&input\\_page\\_show\\_search=on](https://string-db.org/cgi/input.pl?sessionId=T6th4zhhWCHE&input_page_show_search=on)).

### 2.14 Statistical Analysis

Significance of difference was determined using the Student’s paired t-test. Values of  $p < 0.01$  and  $< 0.05$  were considered significant differences. All statistical analyses were conducted using GraphPad Prism (ver. 8) (GraphPad Software, California, USA). Appropriate statistical tests used are labelled in figure legends.



**Figure 2-9: LC/MS Proteomic data analysis strategy.**

FDR; False Discovery Rate, #Peptide; displays the number of distinct peptide sequences in the protein group., #PSM; displays the protein abundance in the sample, #Area; displays the highest detected peptide signal, UniProt; The Universal Protein Resource (is a comprehensive online resource for protein sequence and annotation data), TrEMBL; Translated European Molecular Biology Laboratory Nucleotide Sequence Database; (is a computer-annotated unreviewed peptide sequence), Swiss-Prot; Translated Nucleotide Sequence Database; (is a manually reviewed peptide sequence on UniProt database).

# Chapter 3

**The expression level and subcellular localisation of S100A4 protein in AML blasts and normal human CD34<sup>+</sup> haematopoietic cells**

### 3 Chapter 3 - Introduction

#### 3.1 Overview

Recent studies have reported that dysregulation of key nuclear Transcription Factors (TF) are critical to AML development and progression (Dohner *et al.* 2015). Using iTRAQ coupled with mass spectrometry, our group previously analysed the nuclear proteome of minimally differentiated AML (subtype FAB M1) which has less developmental heterogeneity and can be developmentally matched to normal haematopoietic progenitor cells (CD34<sup>+</sup> cells) (Kumar and Agrawal 2014). Data from this study showed that several proteins (including TF such as CEBPA and WT1) were significantly dysregulated in the nuclear proteome.

Our group identified S100A4 as one of the most statistically significantly increased and highest fold changing proteins in AML blasts that are mislocalized compared to normal CD34<sup>+</sup> nuclei. S100A4 was significantly upregulated in the nuclei of 11/15 AML patients with an average fold increase of 4.4 (across all samples) when compared to controls (Alanazi *et al.* 2019). Given that this protein was also found to be a downstream target of CEBPA it was thus identified as a top candidate for further study (Liu *et al.* 2014).

S100A4 is a member of the calcium binding super-family of EF-hand type which comprises 20 proteins clustered on chromosome 1q21 (Bresnick *et al.* 2015). S100A4 is reported to be involved in a variety of intra- and extra-cellular processes including transcription and differentiation (Donato *et al.* 2013). Moreover, overexpression of S100A4 is associated with tumour metastasis and tumour progression in pancreatic and breast cancers (Helfman *et al.* 2005, Ismail *et al.* 2008). To exert its function, S100A4 undergoes a conformational change upon Ca<sup>2+</sup> binding leading to a change in structure which exposes protein binding sites that influence various intra- and extra- cellular signalling pathways (Lewit-Bentley and Rety 2000, Santamaria-Kisiel *et al.* 2006). Moreover, S100A4 has been shown to interact with key proteins in known pathological signalling pathways in cancer such as p53 (Orre *et al.* 2013, van Dieck *et al.* 2009). Aberrant expression of S100A4 proteins in solid tumours has been associated with either poor clinical outcome or metastasis (Helfman *et al.* 2005). Further, several studies have linked mislocalised nuclear expression of S100A4 in solid tumours such as breast,

ovarian, and colorectal cancers to the aggressiveness, progression, and poor outcomes of these types of cancer (Kikuchi *et al.* 2006, Boye *et al.* 2010, Egeland *et al.* 2017).

Whilst S100A4 is widely accepted in the literature to have a key role in mediating proliferation and metastasis of various types of solid tumours, fewer studies have shown that overexpression of S100A4 plays a role in haematological malignancies. In AML, Preferentially Expressed Antigen in Melanoma (PRAME) is a tumour associated antigen found to induce apoptosis by negatively regulating S100A4 expression and positively regulating p53 in AML cell line [KG-1] (Xu *et al.* 2016). Moreover, overexpression of S100A4 has been shown to reduce chemosensitivity in CML (He *et al.* 2017). However, the pathophysiological role of S100A4 as well as its subcellular distribution in haematological malignancies (as well as normal haematopoiesis) remains largely unknown.

## 3.2 Aims and objectives

The main objective of experiments described in this chapter was to determine the expression level and subcellular localisation of S100A4 during normal haematopoiesis and in AML blasts, principally:

- Optimisation and validation of the detection of S100A4 by western blot.
- Determination of the expression level and subcellular localisation of S100A4 during normal haematopoiesis, in differentiated myeloid subpopulations, and in normal bone marrow controls.
- Determination of the expression level and subcellular localisation of S100A4 in AML blasts.
- Investigation of the correlation between S100A4 expression and clinical patient outcomes

### 3.3 Results

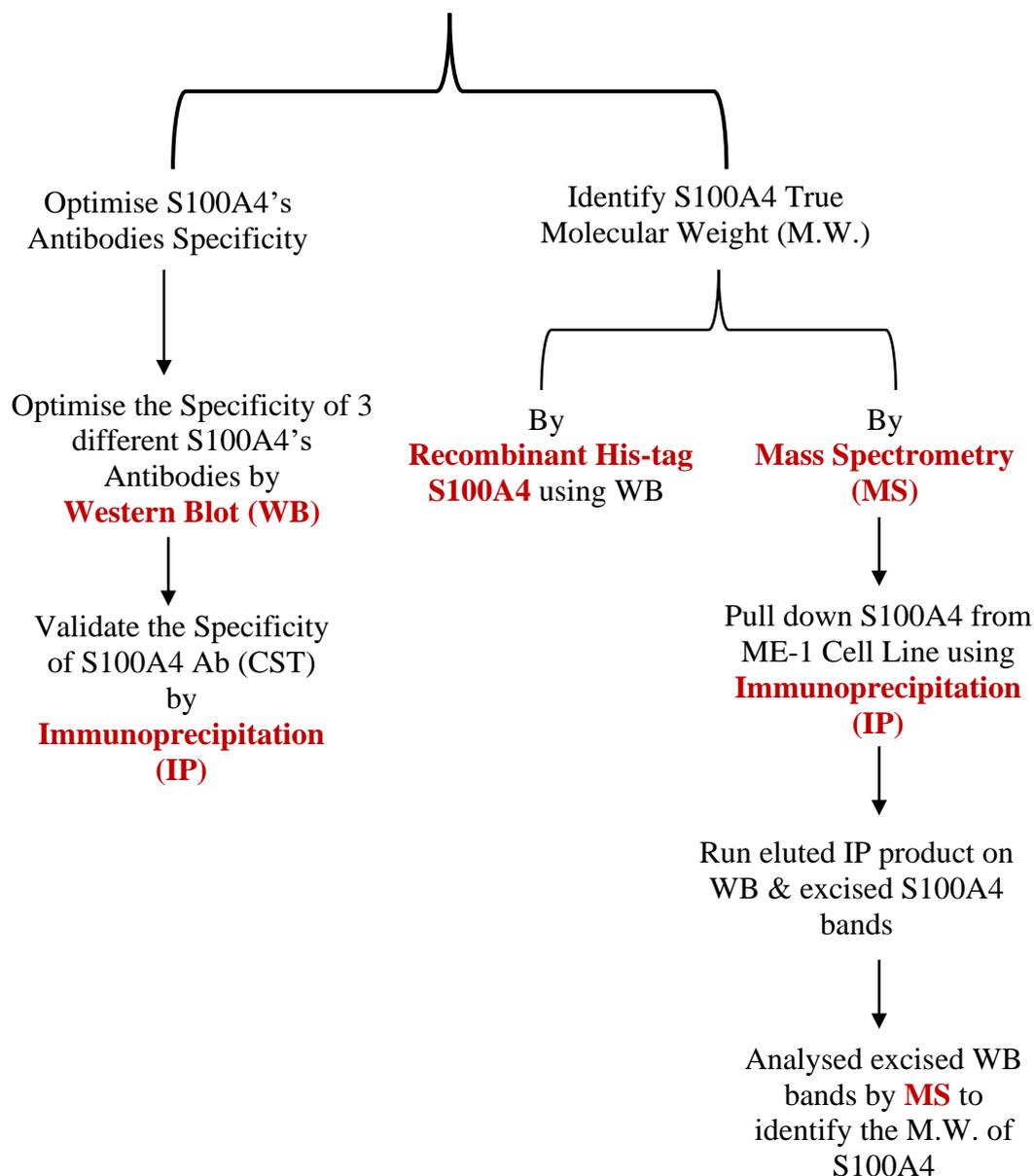
#### 3.3.1 S100A4 is detected in AML cell lines by three commercial antibodies

In order to detect S100A4 protein expression in AML blasts, all the detection conditions were initially optimised as illustrated in Figure 3-1. Firstly, three commercial monoclonal antibodies were assessed and compared using the cytoplasmic fraction of a small cohort of AML cell lines (Figure 3-2A). Clone: D9F9D from Cell Signalling Technology (CST) showed the greatest signal to noise ratio when compared with two alternative clones. A dilution of 1:1,000 of D9F9D as per manufacturer recommendation was selected to detect S100A4 protein expression (Figure 3-2B).

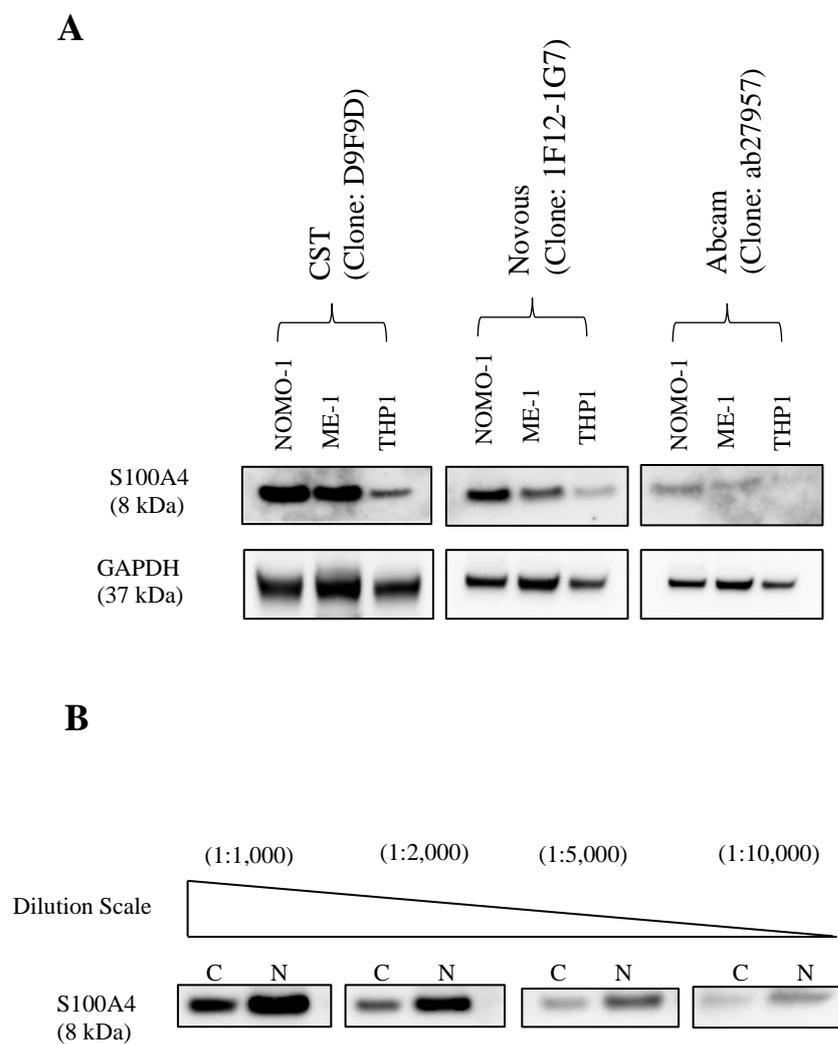
A preliminary analysis of S100A4 protein expression in cytoplasmic and nuclear fractions of AML cell lines was performed (Figure 3-3). Interestingly, a heavier 15 kDa band was visible in the nuclear compartment of ME-1, HEL, and K562 raising doubts as to the specificity of the mAb. In order to provide a positive peptide control and adequately control for non-specific binding, a recombinant human His-Tag S100A4 protein fragment was also used. Initially His-tag S100A4 loading protein weight on the PAGE was optimised using dot blotting coupled with electrophoresis. A concentration of 30 ng of HisTag-S100A4 protein was found to be the optimal amount of protein to load on the gel when detecting with 1:1,000 D9F9D mAb (Figure 3-4A & B). His-tag S100A4 migrated at an apparent weight of 8 kDa.

These results suggest that S100A4 antibody (clone D9F9D, CST) has optimal binding capacity when used at a 1:1,000 dilution coupled with 30 ng of His-Tag S100A4 as positive control.

## Optimisation of S100A4 detection methods

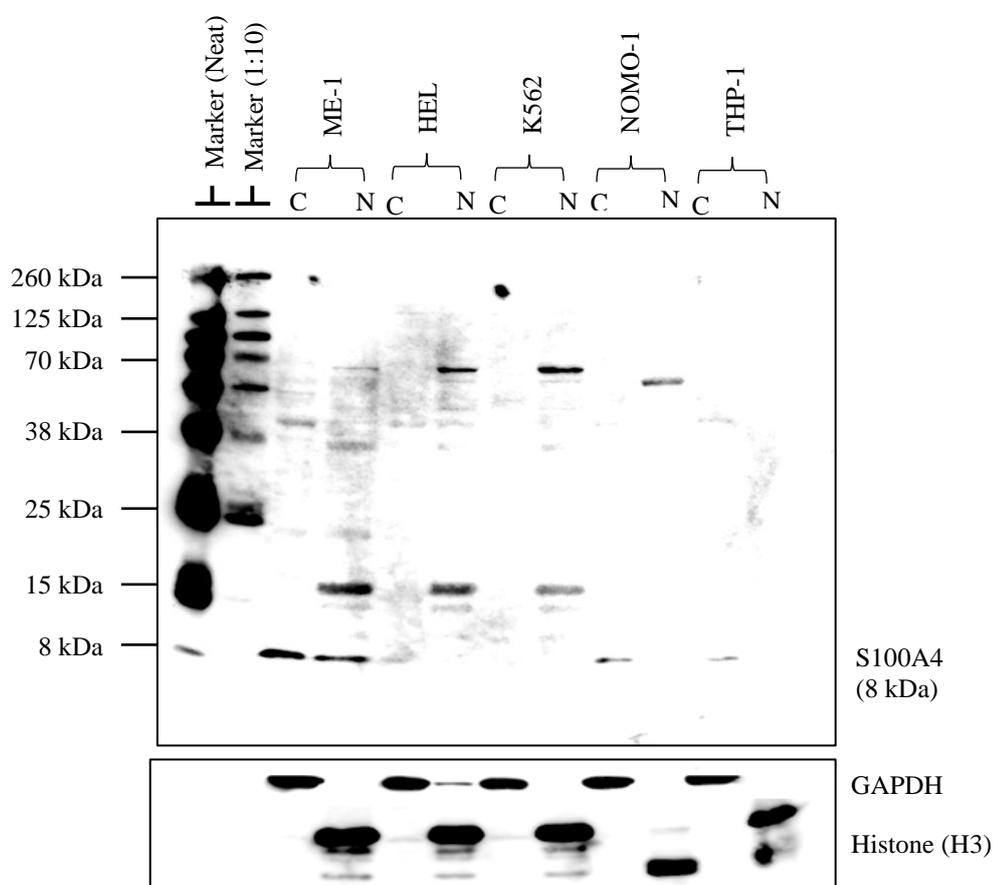


**Figure 3-1: Optimisation work-flow diagram for detection of S100A4 expression in AML.** Illustrative diagram shows the optimisation work-flow of S100A4 detection conditions. WB; Western Blotting, IP; Immunoprecipitation, MS; Mass Spectrometry.



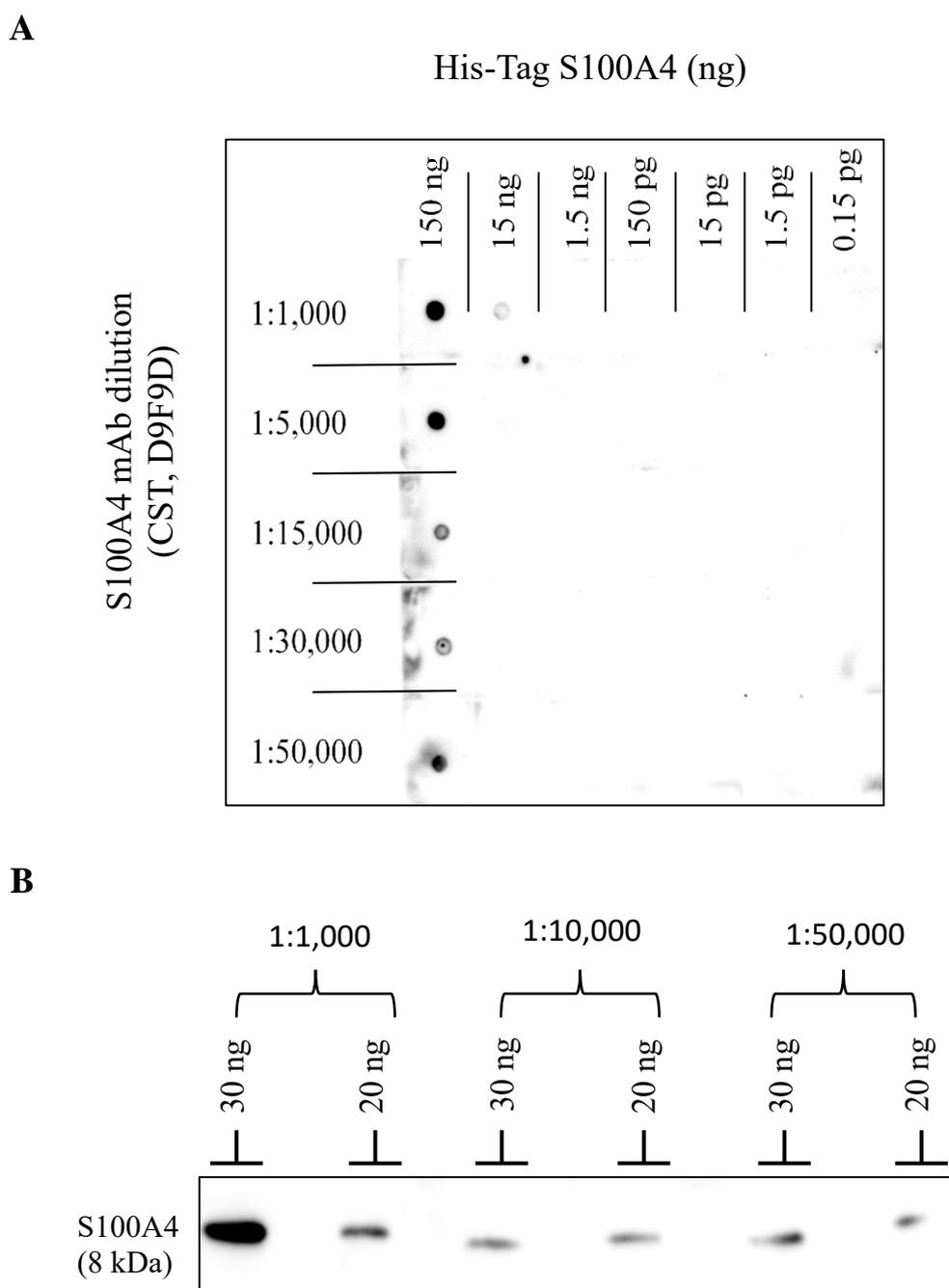
**Figure 3-2: Optimisation of S100A4 protein detection in AML cell lines using western blot.**

**A)** Expression of cytoplasmic S100A4 protein in three AML cell lines using commercial antibodies. All antibodies in this comparison were used as per the supplier recommended concentration. **B)** Example immunoblots illustrating ME-1 cytosolic (C) and nuclear (N) S100A4 expression using increasing dilution of anti-S100A4 mAb (clone: D9F9D, CST). The housekeeping protein GAPDH was used as a loading control.



**Figure 3-3: Expression of S100A4 in cytoplasmic and nuclear fractions of AML cell lines.**

Example immunoblot of S100A4 expression in AML cell lines fractionated into cytoplasmic (C) and nuclear (N) subcellular compartments. S100A4 was immunoblotted with 1:1000 CST, clone: D9F9D. Membranes were reprobed with GAPDH and histone (H3) as loading control and to assess fractionation efficiency.



**Figure 3-4: Optimisation of His-Tag S100A4 protein detection for use as a positive control.**

**A)** Representative dot blot showing serial loading concentrations of His-Tag S100A4 using increasing dilution factors of anti-S100A4 mAb. **B)** Optimization of His-Tag S100A4 protein loading weight coupled with increasing S100A4 mAb concentration (CST, clone: D9F9D).

### 3.3.2 Confirmation of S100A4 mAb specificity

#### 3.3.2.1 Validation of immunoprecipitation of S100A4

To provide further evidence that mAb CST clone D9F9D specifically detects S100A4 at 8 kDa in our electrophoresis system, IP coupled with mass spectrometry was performed. Using ME-1, an AML cell line with high levels of cytoplasmic and nuclear S100A4 protein, IP of the nuclear and cytoplasmic protein fraction was performed, and validation of this protocol was confirmed by western blot. As shown in Figure 3-5, the level of S100A4 protein expression at 8 kDa is reduced in the nuclear fraction in the IP supernatant and is associated with an increase in S100A4 expression in the IP eluted sample. This suggests that the IP worked as most of the nuclear S100A4 was eluted off the beads. Thus, the upper ambiguous band was not pulled down with S100A4 Ab and it can still be seen at 15 kDa in the IP supernatant lane (Figure 3-5). However, in the eluted lane, S100A4 protein was detected by S100A4 Ab (CST, clone; D9F9D) as a single band runs at 8 kDa (Figure 3-5).

In contrast, IP was less efficient in the cytoplasmic fraction. As expected, in the IgG control reaction, no detectable S100A4 protein was eluted from the IgG supernatant sample (Figure 3-5). Furthermore, light and heavy chains (25 and 50 kDa respectively) provide evidence that the protein elution has occurred and S100A4 and IgG antibodies had dissociated from the magnetic beads. Therefore, the binding specificity of S100A4 antibody from CST (clone D9F9D) was confirmed by immunoprecipitation as the antibody bound specifically to S100A4 which was detected on western blot as a single band to apparent weight of 8 kDa.

#### 3.3.2.2 S100A4 expression determined by mass spectrometry

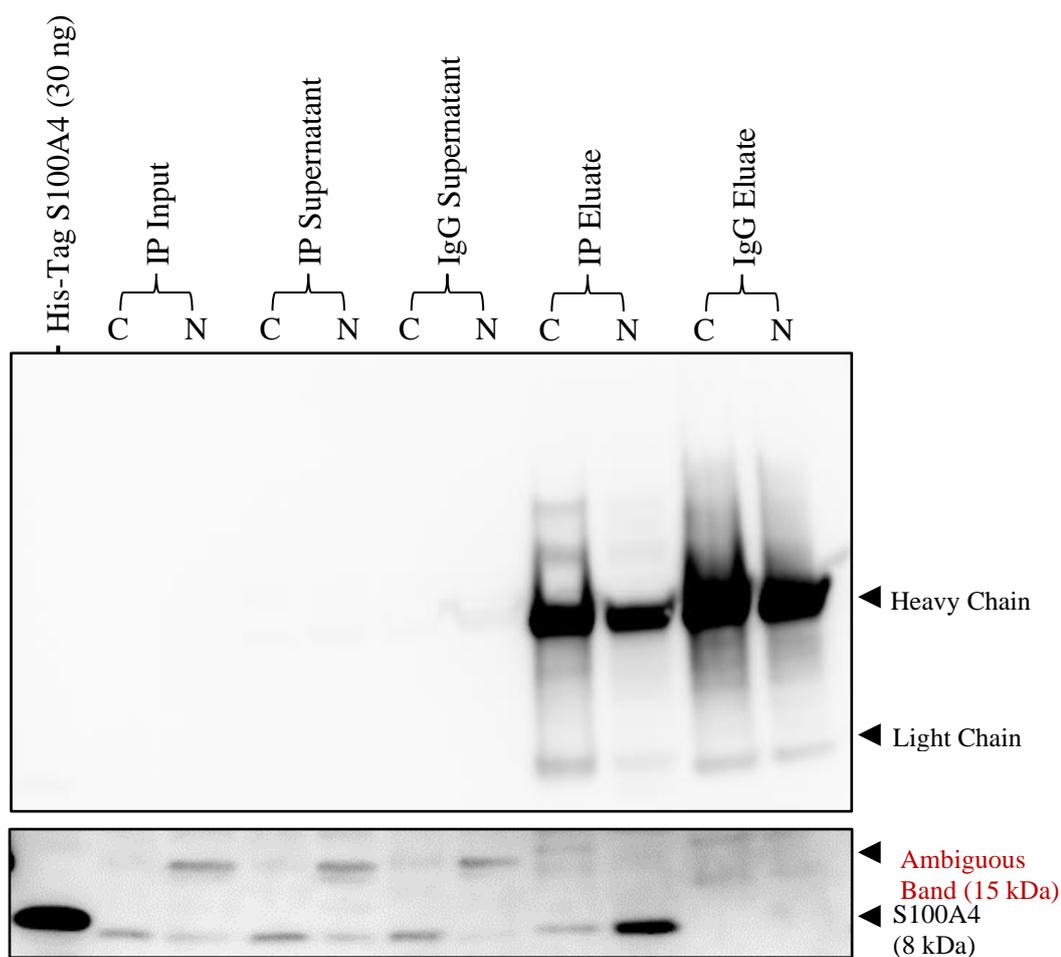
Proteins within the IP fractions above were separated using Tricine SDS PAGE and visualised with Coomassie Blue (Figure 3-6A). Three protein bands in the region 8-15kDa were detected in the nuclear IP Eluate fraction (Figure 3-6A). These three bands were excised from the gel and analysed by LC-MS. Using a 1% FDR, band #1 had the highest abundance of S100A4 (PSM Score = 36) as compared to bands #2 and #3 (PSM Scores = 11 and 8 respectively) (Figure 3-6B) (full MS data are presented in attached CD).

Therefore, mass spectrometry analysis provided further confidence of the true molecular weight of S100A4 as shown by the abundance of the 8 kDa band. Further, LC-MS analysis identified “SUMO-conjugating enzyme” (UBC9) in the band #3 dataset (UniProt accession # P63279). UBC9 enzyme functions by catalysing the covalent attachment of ubiquitin-like proteins such as SUMO1, SUMO2, SUMO4, and SUMO5 to the target protein in coordination with E3 Ligase (Wilson 2017). The identification of UBC9 in band #3 may explain the higher weight shift detected in previous western blots (Figure 3-3).

### **3.3.3 S100A4 expression levels and subcellular localisation in AML cell lines.**

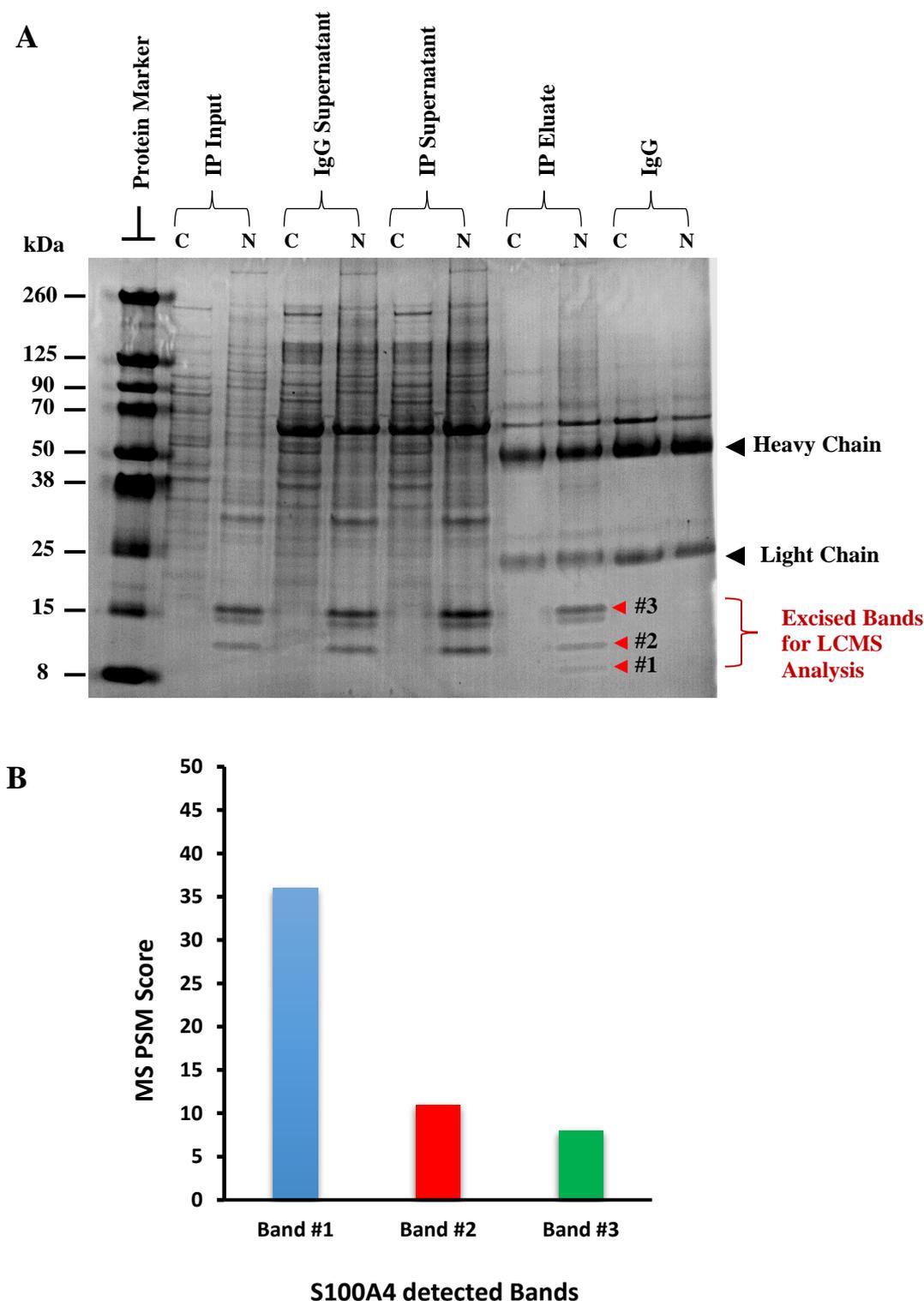
Using the Cancer Cell Lines Encyclopaedia (CCLE), the overall levels of *S100A4* mRNA in AML cell lines were examined. *S100A4* is highly expressed in most of the AML cell lines used in this study as shown in Figure 3-7A. However, mRNA levels do not always correspond to protein levels nor to the subcellular localisation as discussed in 3.4.2.2.

Having optimised and validated the expression of S100A4 in a small cohort of AML cell lines, this study next analysed the expression levels and subcellular localisation of S100A4 in a larger cohort of 10 AML cell lines by western blot (Figure 3-7B). Nearly all of the cell lines analysed expressed S100A4 (9/10) with nuclear S100A4 protein expression being demonstrated on 6/10 (60%). ME-1, HEL, NOMO-1, THP-1, OCI AML-2, and U937 cell lines showed expression of S100A4 in both nuclear and cytoplasmic compartments. However, Mv4;11 and HL-60 cell lines showed expression of S100A4 only in the cytoplasmic compartment, whereas the Kasumi-1 cell line showed less overall expression in both compartments. S100A4 expression was undetectable in KG-1 cells (Figure 3-7B). These data show that these cell lines can provide potential models for subsequent functional studies to investigate the effects of overexpressed/knocked down S100A4 on proliferation, progression, and apoptosis (4.3.4 and 4.3.5).



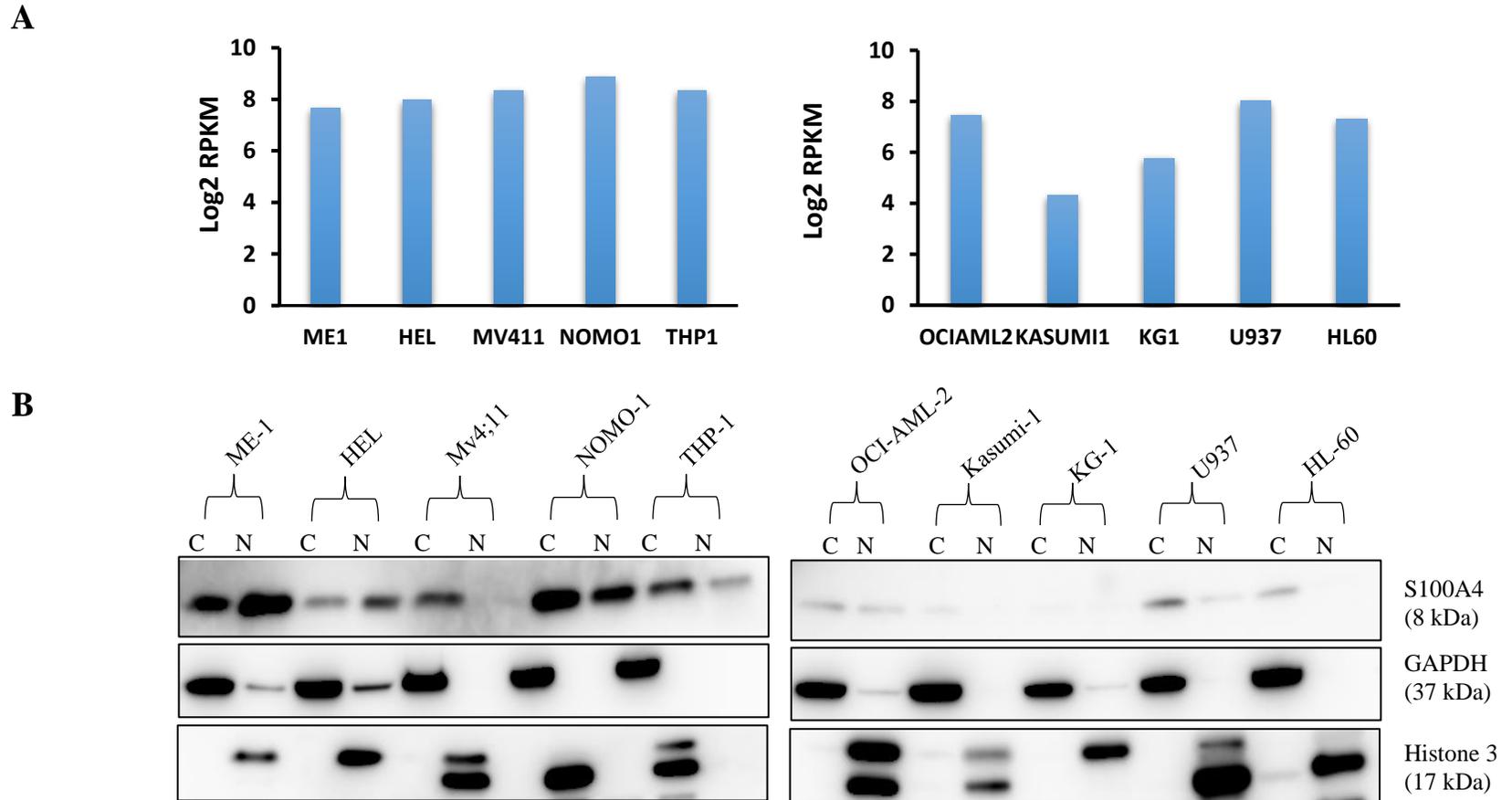
### Figure 3-5: S100A4 mAb specificity is validated by immunoprecipitation

Immunoblot demonstrates the immunoprecipitation of S100A4 protein with S100A4 antibody (CST, clone: D9F9D) from ME-1 Cytosolic (C) and Nuclear (N) compartments. Upper blot image shows the breakdown of IgG control and S100A4 antibodies (Heavy and Light Chains) upon IP elution. Lower blot image shows S100A4 Ab (CST, clone D9F9D) specificity as it was bound to S100A4 in the “IP Eluate” lane detected as a single band migrating at an apparent molecular weight of 8 kDa. Upper blot image exposure (30 seconds) due to image oversaturation, lower image exposure (20 minutes).



**Figure 3-6: Validation of S100A4 expression at 8kDa by mass spectrometry.**

**A)** Immunoprecipitated proteins from (Figure 3-5) using S100A4 antibody were electrophoresed using 20% Tricine-SDS PAGE and visualised with Coomassie Blue. Three distinct bands (see (A)) in the IP Eluate lane were extracted from the gel and analysed by LC-MS. **B)** LC-MS identified that the highest abundance of S100A4 in band #1 ran at 8 kDa. S100A4 abundance is represented in MS by PSM score (This displays the total number of identified peptide sequences) of S100A4 detected in each band.



**Figure 3-7: S100A4 mRNA expression levels and subcellular localisation in AML cell lines.**

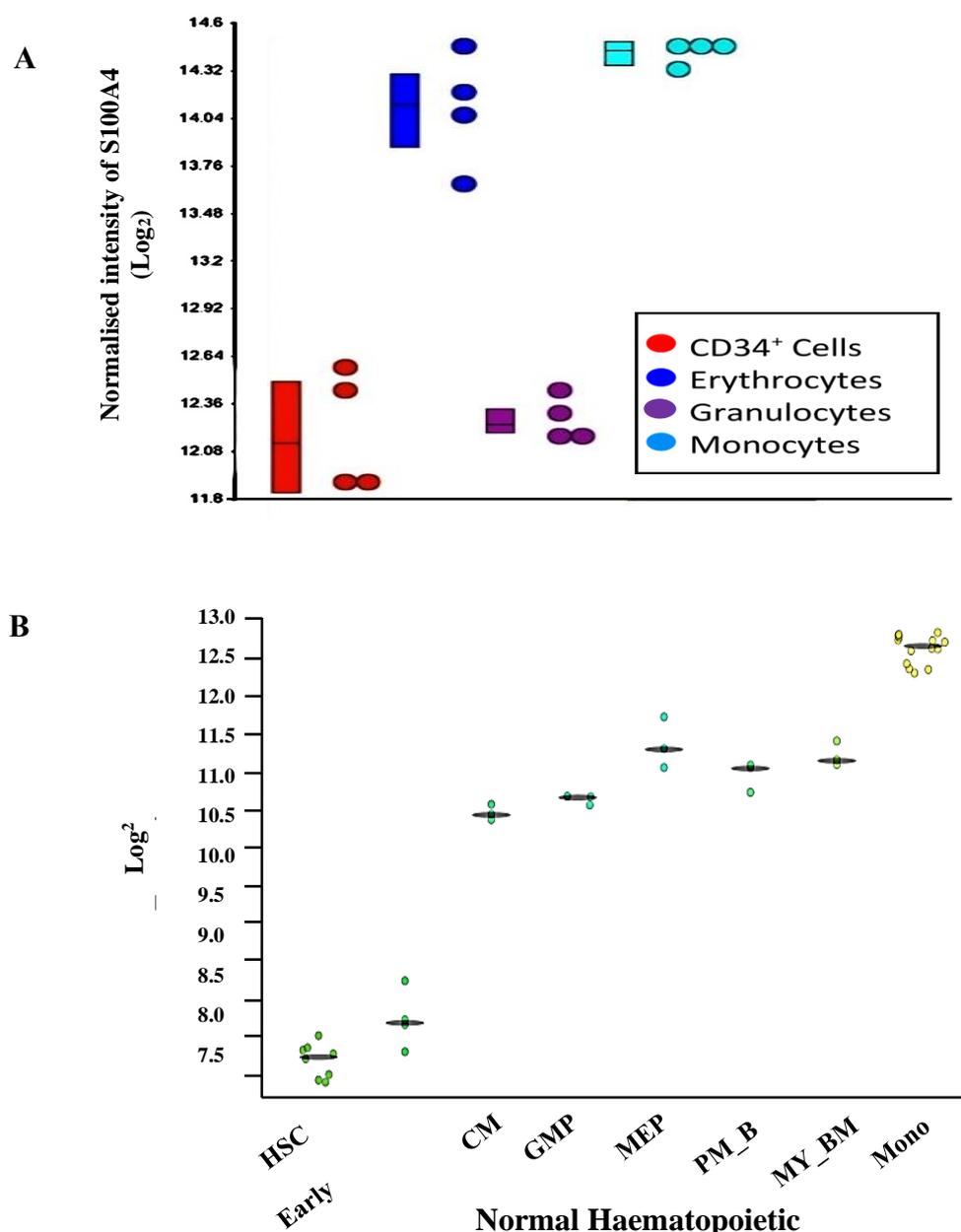
**A)** RNA-seq data showing that S100A4 mRNA levels are upregulated in most AML cell lines. Y-axis RPKM (Reads Per Kilobase Million). **B)** Example western blot showing subcellular expression of S100A4 from AML cell lysates; Cytosolic (C) Nuclear (N). Membranes were probed with S100A4 mAb (clone D9F9D, CST) and reprobbed with GAPDH and histone (H3) to ensure equal loading and assess fractionation efficiency.

### 3.3.4 S100A4 expression levels and subcellular localization in normal haematopoietic cells.

To establish the endogenous expression levels of *S100A4* in normal haematopoiesis, mRNA expression data from haematopoietic cells was analysed. Analysis of transcriptomic data that has been generated in a previous study (Tonks *et al.* 2007) confirmed that *S100A4* is differentially expressed in normal human CD34<sup>+</sup> cells compared to unipotent progenitors derived from cord blood CD34<sup>+</sup> cells (Figure 3-8A). Erythrocytes and monocytes upregulate *S100A4* expression; conversely, CD34<sup>+</sup> cells and granulocytes have less expression (Figure 3-8A). The publicly available database “BloodSpot” (Bagger *et al.* 2016) in combination with The Cancer Genome Atlas (TCGA) data set (Tomczak *et al.* 2015) also showed that expression of S100A4 appears to increase with HSC differentiation (Figure 3-8B). Although transcriptomic data provide insights into *S100A4* transcriptional patterns in normal haematopoiesis, they do not delineate the subcellular distribution of *S100A4* expression.

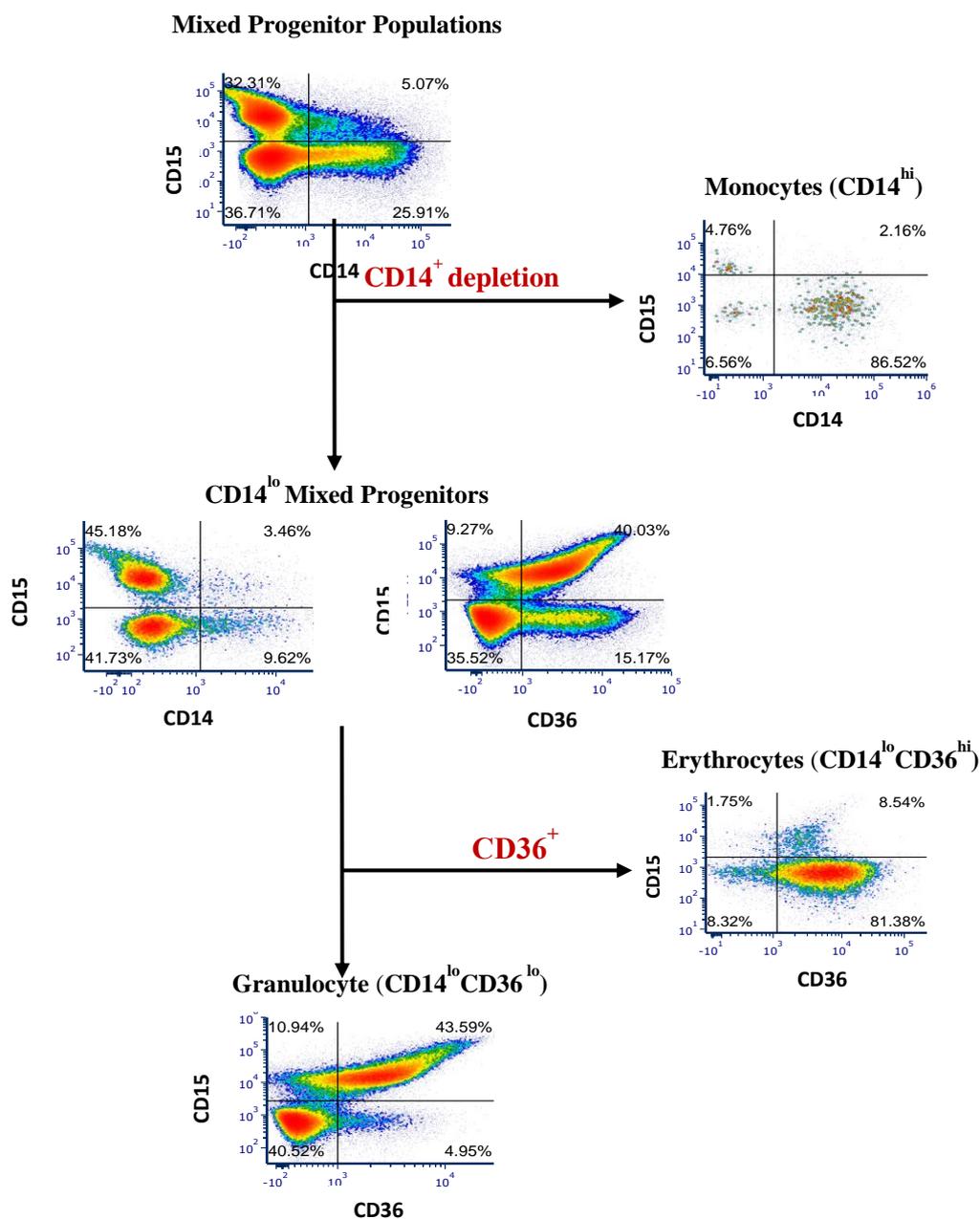
To further investigate S100A4 subcellular localisation, normal CD34<sup>+</sup> cells were isolated from neonatal cord blood and fractionated into cytoplasmic/nuclear lysates at day zero. Further, CD34<sup>+</sup> cells were grown to day six in appropriately supplemented primary cell medium. Unilineage progenitors (monocytes, erythrocytes, and granulocytes) were isolated by a positive selection process (Figure 3-9). Western blot analysis of S100A4 expression and subcellular localization in CD34<sup>+</sup> cells shows significant cytosolic expression levels with undetectable nuclear expression (Figure 3-10A). Similarly, Monocytes and erythrocytes showed cytosolic expression of S100A4 with no detectable S100A4 nuclear expression (Figure 3-10B). In support of the transcriptomic data, granulocyte progenitors appear to have less *S100A4* expression which is restricted to the cytosol. Analysis of normal human bone marrow also showed that S100A4 was expressed only in the cytosol (Figure 3-11).

Taken together, these data suggest that S100A4 is transcriptionally regulated during differentiation, but that protein expression remains restricted to the cytoplasm in normal human haematopoietic cells.



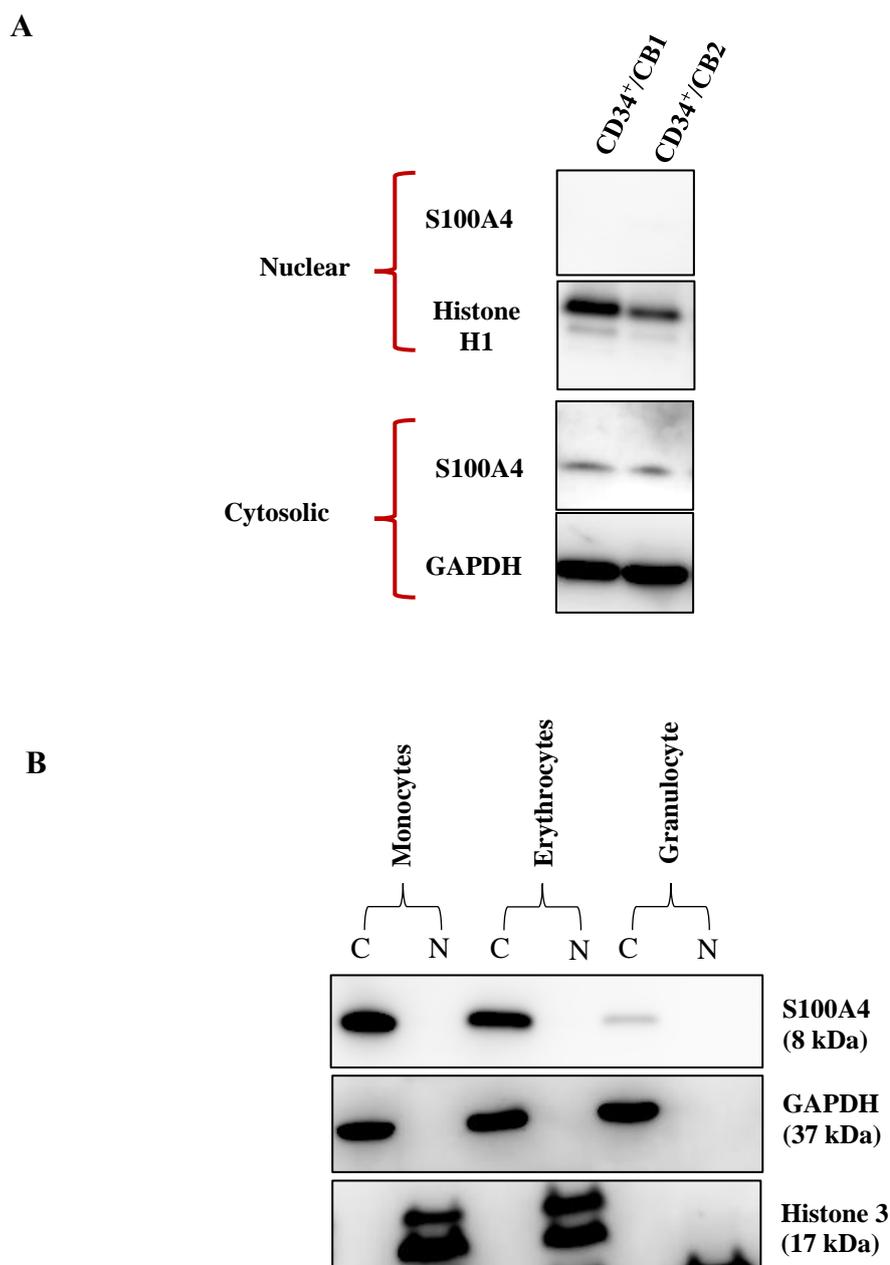
**Figure 3-8: *S100A4* mRNA Expression Patterns in normal Haematopoiesis.**

Normalised microarray data showing log<sub>2</sub> mRNA *S100A4* expression in human haematopoietic cells. **A**) *S100A4* expression in normal human CD34<sup>+</sup> cells and MACS isolated haematopoietic progenitor subsets (day 6 of culture from CD34<sup>+</sup> cells); Monocytes (CD14<sup>hi</sup>, CD36<sup>lo</sup>), Erythrocytes (CD36<sup>hi</sup>, CD14<sup>lo</sup>), and Granulocytes (CD36<sup>lo</sup>, CD14<sup>lo</sup>) (represented by 22 283 probe sets by Tonks *et al.* 2007) (n=4). **B**) *S100A4* mRNA expression in normal haematopoiesis using the TCGA dataset probe set: 203186\_s\_at, (Bagger *et al.* 2016). *S100A4* mRNA levels are lower in bone marrow derived HSCs compared to committed myeloid progenitors HSC\_BM; Hematopoietic stem cell from Bone Marrow, early HPC\_BM; Hematopoietic progenitor cell from Bone Marrow, CMP; Common myeloid progenitor cell, GMP; Granulocyte monocyte progenitors, MEP; Megakaryocyte-erythroid progenitor cell, PM\_BM; Promyelocyte from bone marrow, MY\_BM; Myelocyte from bone marrow, and Mono; Monocytes.



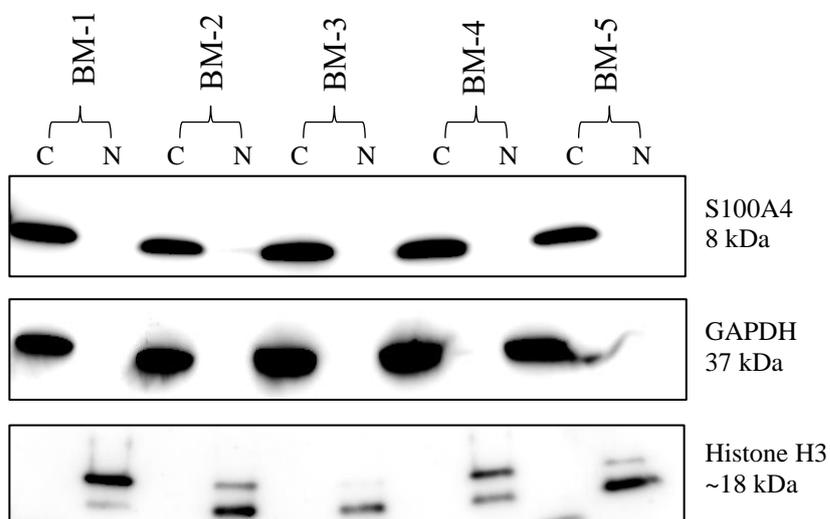
**Figure 3-9: Purification of individual lineages expressing S100A4 used in Affymetrix microarray analysis and western blot.**

Representative density plots of human progenitor cells enriched for specific lineages on day 6 of culture using a MiniMACS depletion strategy. Quadrants delimit background isotype staining. Monocytes\_(CD14<sup>hi</sup>, CD36<sup>lo</sup>), Erythrocytes (CD36<sup>hi</sup>, CD14<sup>lo</sup>), and Granulocytes (CD36<sup>lo</sup>, CD14<sup>lo</sup>).



**Figure 3-10 : S100A4 protein is differentially expressed in normal haematopoietic lineage and restricted to the cytosol.**

Example western blots showing S100A4 protein expression in human haematopoietic cells. **A)** S100A4 expression in cytosol and nuclear fractions of normal human CD34<sup>+</sup> cells (n=2). **B)** Expression patterns of S100A4 in differentiated sub-lineages of HSPCs derived from cord blood grown to day 6; Monocytes (CD14<sup>hi</sup>, CD36<sup>lo</sup>), Erythrocytes (CD36<sup>hi</sup>, CD14<sup>lo</sup>), and Granulocytes (CD36<sup>lo</sup>, CD14<sup>lo</sup>). Immunoblots were reprobbed with GAPDH and histone (H3) to ensure equal loading and assess fractionation efficiency. Cytosolic (C) Nuclear (N). Cord Blood (CB)



**Figure 3-11: S100A4 Protein expression is restricted to the cytoplasm in bone marrow derived haematopoietic cells.**

Western blot showing S100A4 expression and subcellular localisation in normal human bone marrow samples (n=5), Cytosolic (C), Nuclear (N). Immunoblot was reprobbed with GAPDH and histone (H3) to ensure equal loading and assess fractionation efficiency.

### 3.3.5 S100A4 expression levels and subcellular localization in AML patients (FAB M1) and (FAB M4)

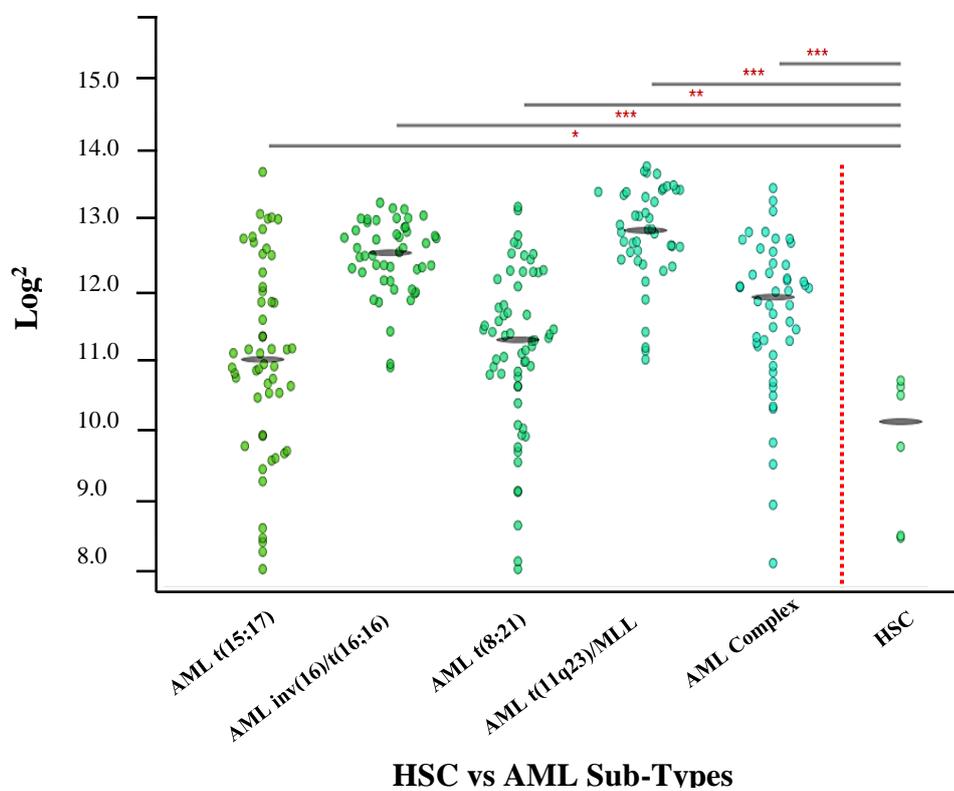
In the previous section, I showed that S100A4 expression is only detected in the cytoplasmic compartments of normal CD34<sup>+</sup> cells and differentiated myeloid progenitors. Thus, the overexpression and mislocalisation of S100A4 to the nucleus of tAML patient samples may suggest that S100A4 plays a role in AML development or progression. Initially, to determine whether S100A4 is overexpressed in AML, I analysed S100A4 mRNA expression levels across different AML subtypes in comparison to normal haematopoietic stem cells (HSC). To do this analysis, the BloodSpot<sup>®</sup> database was utilised using mRNA data derived from a high quality AML data set derived from The Microarray Innovations in LEukemia (MILE) study (Kohlmann et al. 2008, Haferlach et al. 2010). As shown in Figure 3-12, S100A4 is significantly overexpressed across multiple AML molecular subtypes (2-8 fold) compared to normal HSCs. For example, S100A4 is significantly overexpressed in these AML molecular abnormalities such as t(15;17) p<0.05, inv(16)/t(16;16) p<0.001, t(8,21) p<0.01, t(11q23)/MLL p<0.001, and AML with complex cytogenetics p<0.001 (Figure 3-12).

However, examining the expression levels of S100A4 on the transcriptomic level is not informative enough in terms of subcellular distribution which in this case is very important. Thus, S100A4 expression and subcellular localisation is confirmed at the protein level. AML blasts from two differentiation stages: minimally differentiated (FAB M1) and myelomonocytic (FAB M4) were analysed by western blot with their matching controls; CD34<sup>+</sup> cells and CD14<sup>+</sup> cells respectively. Given the clear potential for FAB subtypes to be misapplied, FAB designations were confirmed by flow cytometry using double stains of lineage-specific markers (CD14<sup>+</sup>/CD15<sup>+</sup>) and CD34<sup>+</sup> (Figure 3-13A and Figure 3-14A respectively).

Minimally differentiated AML blasts (FAB M1) which have less developmental heterogeneity and can be an appropriate match to normal CD34<sup>+</sup> cells were analysed by western blot for S100A4 expression and subcellular localisation. As mentioned in section (1.5), I was able to confirm the preliminary findings that S100A4 is not only overexpressed but also mislocalised into the nucleus in AML patients' samples. A total of 24 AML patients samples (FAB M1) were fractionated into cytoplasmic and nuclear lysates and analysed by western blot. The results show that S100A4 is overexpressed in 75 % of AML patient samples (18/24) (Figure 3-13B). However, one of main

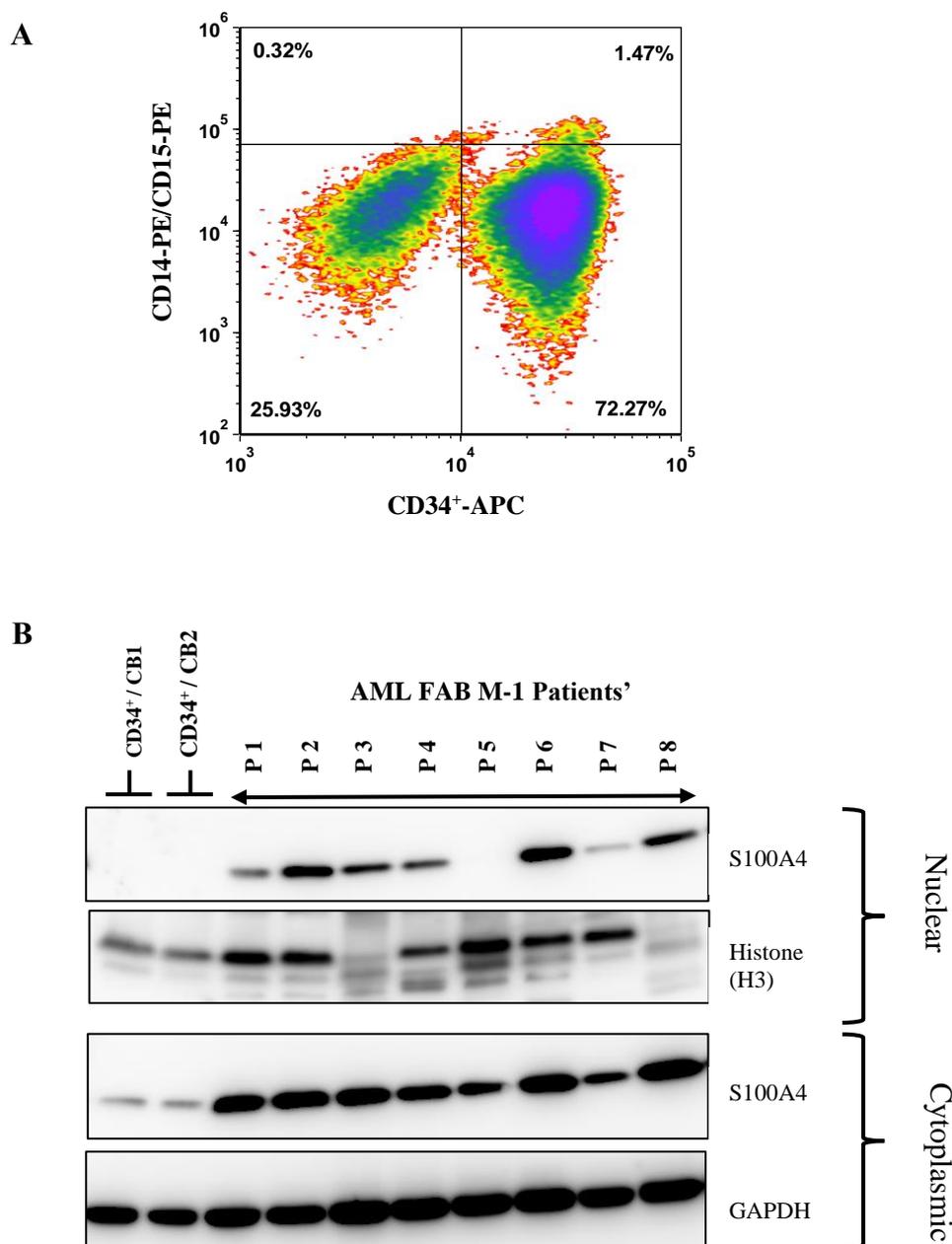
observations of this study is that S100A4 is mislocalised to the nucleus as compared to normal CD34<sup>+</sup> cells which show only cellular expression of S100A4. In the AML samples analysed in this study, 83% of AML patients (20/24) showed mislocalised expression of S100A4 to the nucleus (Figure 3-13B). This study further analysed AML samples from myelomonocytic AML blasts which were characterised as FAB M4. Similarly, nuclear S100A4 protein was mislocalised in 44% of AML patient samples (4/9) when compared to normally-differentiated monocytes (CD14<sup>+</sup>) (Figure 3-14B).

Taken together these results validated the preliminary nuclear proteome study outcomes (Tonks et al. 2007) which suggested that S100A4 is upregulated in the nuclear compartments. However, western blot analysis shed some light on the importance of mislocalisation of S100A4 in the nucleus to AML development.



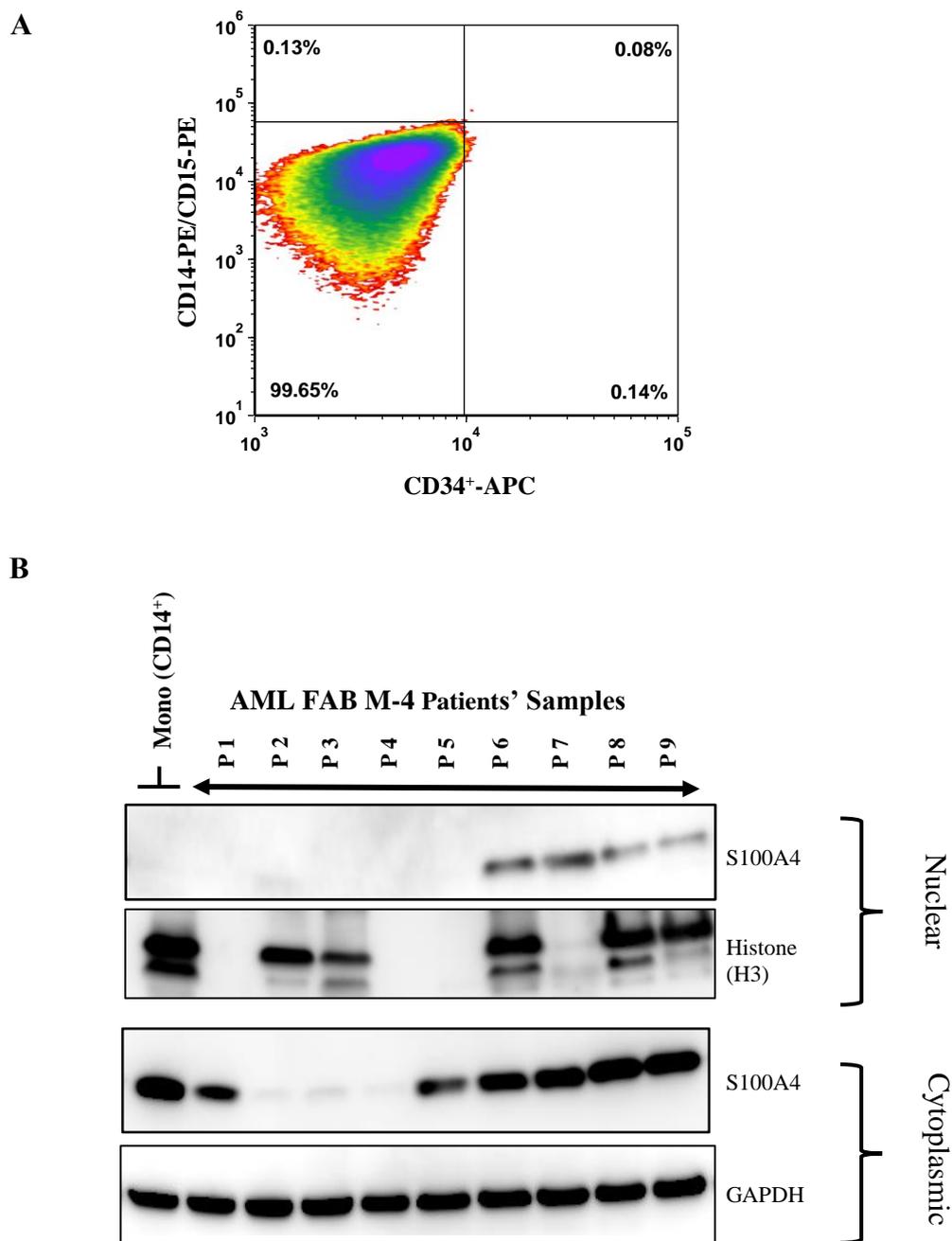
**Figure 3-12: S100A4 is overexpressed in AML sub-types.**

S100A4 is significantly overexpressed across multiple AML molecular abnormalities (n=296) as compared to normal HSC (student t-test \* denotes  $p < 0.05$ ; \*\* denotes  $p < 0.01$ ; \*\*\* denotes  $p < 0.001$ ). (TCGA dataset; Normal hematopoiesis with AMLs probe set 203186\_s\_at). (Bagger et al. 2016) Figure adopted from <http://www.bloodspot.com.eu/>; Data Accessed on 20/01/2019



**Figure 3-13: Overexpression and mislocalization of S100A4 in the nucleus of AML patients.**

**A)** Example density plot of FAB M-1 subtypes classification by flow cytometry. FAB subtype determined for minimal differentiation by examination of cell surface markers (CD34<sup>+</sup>) and (CD14<sup>+</sup>/CD15<sup>+</sup>) normally associated with myeloid differentiation. Quadrants delimit background isotype staining. **B)** Example of western blot for S100A4 protein expression and mislocalisation in AML blasts FAB M-1 (n= 24). AML blasts FAB M-1 blotted with differentially matched control CD34<sup>+</sup> cells. CD34<sup>+</sup>; Haematopoietic Progenitor Cells, CB; Cord Blood. GAPDH and Histone demonstrate the purity/relative loading of cytosolic and nuclear fractions respectively



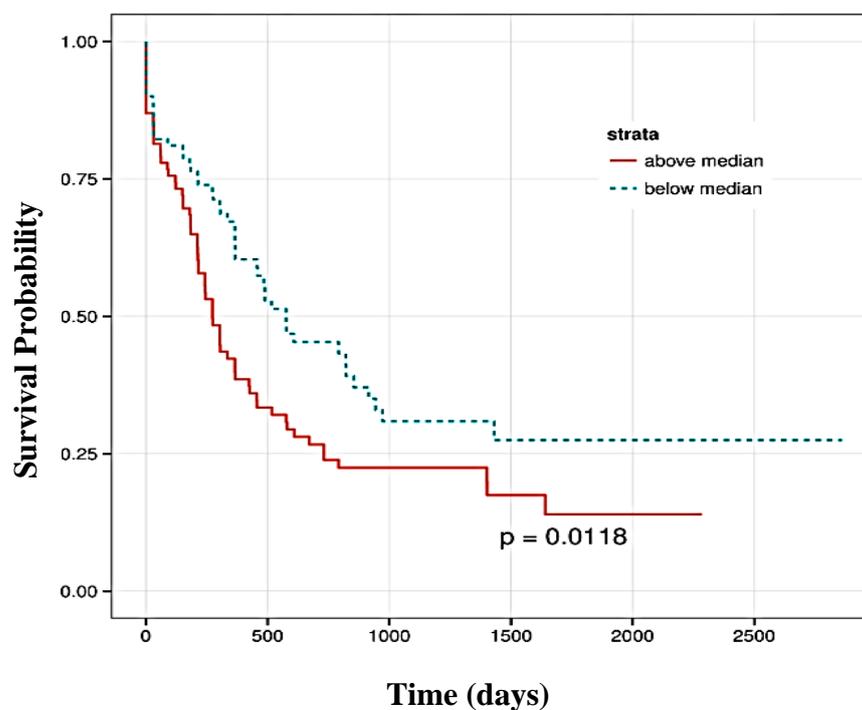
**Figure 3-14: Overexpression and mislocalization of S100A4 in the nucleus of AML patients.**

**A)** Example density plot of FAB M-4 subtypes classification by flow cytometry. FAB subtype determined for minimal differentiation by examination of cell surface markers (CD34<sup>+</sup>) and (CD14<sup>+</sup>/CD15<sup>+</sup>) normally associated with myeloid differentiation. Quadrants delimit background isotype staining. **B)** Example of western blot S100A4 protein expression and mislocalisation in AML blasts FAB M4 (n= 9). AML blasts FAB M4 blotted with differentially matched control CD14<sup>+</sup> cells isolated from cord blood. CD34<sup>+</sup>; Haematopoietic Progenitor Cells, Mono; Monocytes (CD14<sup>+</sup>). GAPDH and Histone demonstrate the purity/relative loading of cytosolic and nuclear fractions respectively.

### 3.3.6 Correlation of upregulated *S100A4* to patients' survival, prognosis, and outcomes

Aberrant *S100A4* expression has been associated with poor patient survival and outcomes in many solid tumour types (Zhao *et al.* 2013, Kho *et al.* 2012). *S100A4* mRNA data from the MILE study, BloodSpot database was utilised to examine whether *S100A4* mRNA expression can be an independent predictor of patients' overall survival (OS). Results show that upregulated *S100A4* is associated with poor survival rates in AML patients ( $p= 0.0118$ ) (Figure 3-15). However, Bloodspot survival analysis does not take into account those patients with favourable risk AML such as t(15;17) and t(8;21) and inv(16) AML nor the type of treatment they received to stratify them into intensively and non-intensively treated patients. Here, we utilised the TCGA data (Ley *et al.* 2013) to analyse the effect of upregulated *S100A4* mRNA on overall survival of AML patients with exclusion of AML patients who have t(8;21) and t(15;17) abnormalities from the dataset. As shown in Figure 3-16, Kaplan-Meier plot shows that higher *S100A4* expression is associated with poor overall survival in AML patients ( $p= 0.01$ ).

Further, the overall survival analysis does not take into account several confounding variables that are known to be associated with clinical outcome including cytogenetics, WBC, and Age (Dohner *et al.* 2015). Therefore, we used a transcriptomic data set derived from the MILE study group, of which Cardiff University's sample collection forms part, (Haferlach *et al.* 2010) to determine the association of *S100A4* expression with clinical outcomes in combination with these factors. The effects of *S100A4* mRNA levels on AML patient outcomes was assessed by conducting a correlation analysis test. As shown in Figure 3-17, our analysis shows that there is no correlation between overexpressed *S100A4* mRNA and patient outcomes ( $p\text{-value}= 0.7$ ). However, the sample size of our mRNA data set is small ( $n=128$ ); to see such an effect on patients' outcomes the sample size has to be increased. In Summary, *S100A4* mRNA expression can be used as an independent prediction tool to assess the overall survival rates in AML patients. To gain insight on whether *S100A4* expression correlates to patients' outcomes, the sample size may need to be increased to detect such correlation.

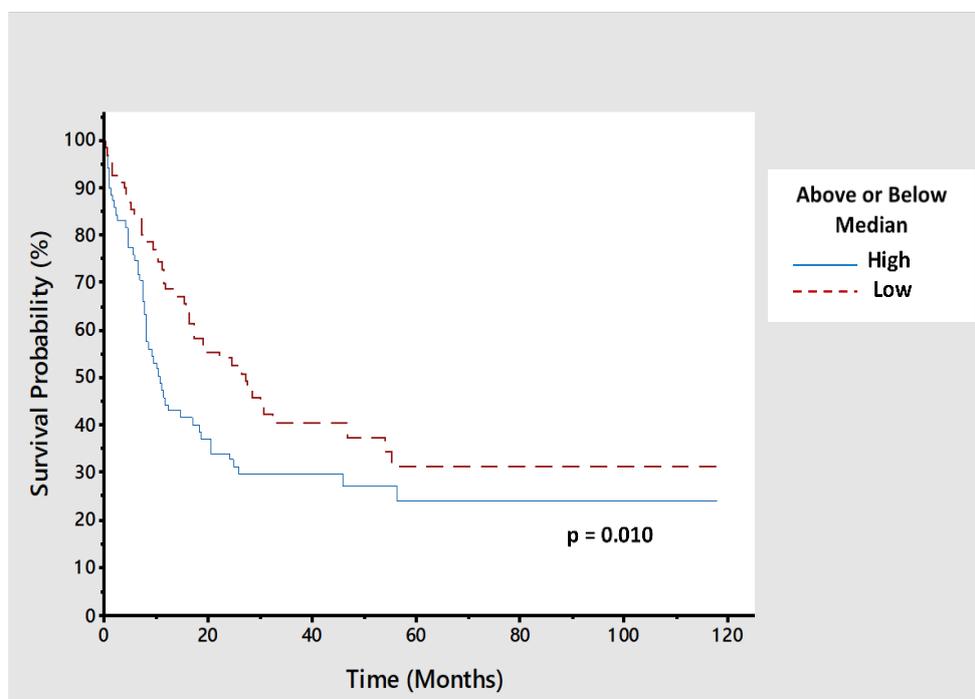


#### Numbers at risk

	0	500	1000	1500	2000	2500
<b>Above Median</b>	88	25	13	6	2	0
<b>Below Median</b>	86	35	15	8	2	1

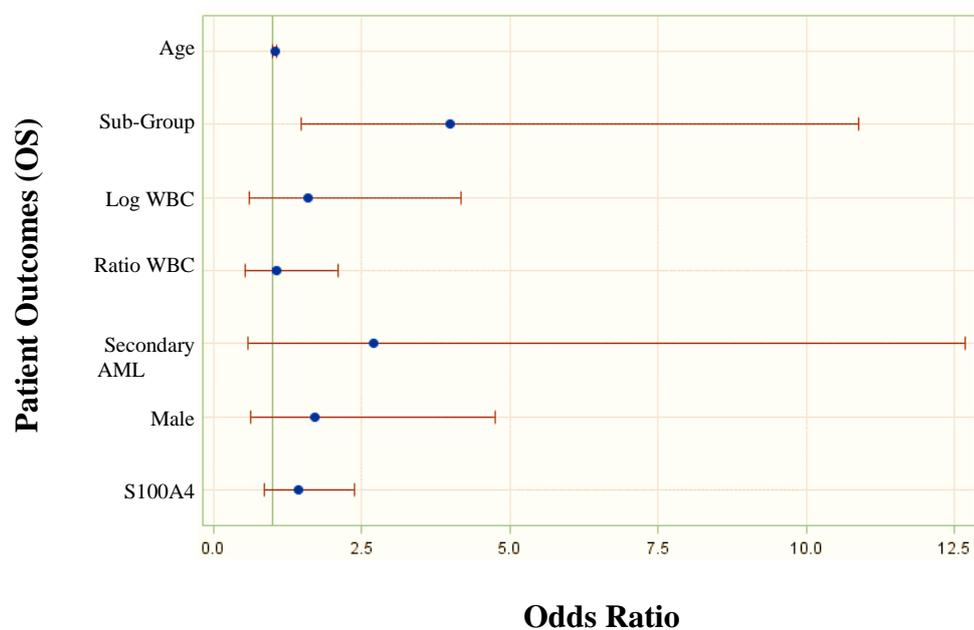
**Figure 3-15: Overexpressed *S100A4* is associated Poor Overall Survival Rate in AML Patients.**

Kaplan–Meier plot of the overall survival in leukaemia patients who have high *S100A4* gene expression vs low *S100A4* gene expression. (TCGA dataset; Normal haematopoiesis with AMLs probe set 203186\_s\_at) from the MILE study. (Haferlach *et al.* 2010). Figure adopted from [http://www.bloodspot.com.eu](http://www.bloodspot.com.eu;);



**Figure 3-16: Upregulated *S100A4* mRNA levels is associated poor overall survival rate in AML patients.**

Kaplan–Meier plot of the overall survival in AML patients who have high *S100A4* gene expression vs low expression of *S100A4* gene expression. (p-value = 0.01). *S100A4* mRNA is derived from TCGA data (Ley et al. 2013) from cbioportal.org database



**Figure 3-17: Overexpressed *S100A4* mRNA does not correlate with AML patient outcomes.**

Forest plot of correlation of *S100A4* mRNA of upregulated *S100A4* with patients' outcomes (n=128), odd ratio with CI 95%, (p-value= 0.7). Patients' outcomes; *S100A4* mRNA, Gender: male, Secondary AML, Ratio of WBC count, Log WBC, Sub-Group, and Age. Figure kindly provided by Prof. Robert Hills (Head of Clinical Trial Centre)

## 3.4 Discussion

The main objective of this chapter was to determine the expression level and subcellular localisation of S100A4 during normal haematopoiesis and in AML blasts. To achieve this aim it was necessary to optimise our western blotting protocols to enable the analysis of such a small molecular weight protein as S100A4. This section discusses the approach taken to optimise the conditions for western blot as well as some of the issues that were encountered throughout the process. The subsequent discussion in the chapter focuses on the main abnormalities observed in AML with regard to S100A4 expression, S100A4 mislocalisation to the nucleus of AML blasts and discussion of whether S100A4 expression affects patient outcome.

### 3.4.1 Optimisation of S100A4 protein detection

S100A4 is a small protein of 8-11 kDa and shares key structural features with other family members such as EF hand. More specifically, the structure of Ca<sup>2+</sup>-bound S100A4 (active form) displays a similar protein topology to the other members of the S100 family (Pathuri et al. 2008). Therefore, it is important to validate an antibody that has high sensitivity/specificity capacity to detect S100A4 with outcross reacting with structurally closely related members of the S100 family. Hence, S100A4 clone: D9F9D (CST) emerged as the most sensitive antibody amongst the other antibodies tested. D9F9D is a monoclonal antibody that binds to a specific S100A4 epitope. Monoclonal antibodies are known for less batch-to-batch variability compared with polyclonal antibodies which are also more likely to bind non-specifically as they bind to many epitopes (Hanack et al. 2016), although polyclonal antibodies do tend to have a higher overall binding affinity for their targets in IP (Marcon et al. 2015). Ideally, the specificity of the antibody can be determined by the presence of single band in a complex biological sample at the expected molecular weight in a western blotting system (Signore et al. 2017).

Therefore, we applied our optimised conditions (3.3.1) to the subsequent analysis of S100A4 expression in a small random cohort of AML cell lines fractionated into nuclear and cytoplasmic lysates. Whilst the cytosolic S100A4 protein appeared as a single band, nuclear lysates demonstrated two bands that migrated to approximate molecular weights of 8 kDa and 15 kDa (Figure 3-3). The inclusion of a recombinant human S100A4 peptide provided further evidence of the specificity of S100A4 mAb as

the resulting band migrated to an apparent weight of 8 kDa. To resolve the ambiguity around the expression of S100A4 appearing at either 8 or 15kDa, immunoprecipitation followed by mass spectrometric (IP-MS) validation was performed on extracts of AML cell line (ME-1). The ME-1 cell line was chosen due to its prominent expression of S100A4 in the nucleus (Figure 3-7). IP-MS analysis provided confirmation that S100A4 mAb (CST, Clone: D9F9D) bound specifically by pulling down S100A4 efficiently (Figure 3-6A). Analysis of these “pull downs” using PAGE coupled with Coomassie Brilliant Blue to visualise the proteins showed 3 bands (Band# 1= at ~ 8 kDa, Band# 2= at ~ 12 kDa, and Band# 3= at ~ 15 kDa) (Figure 3-6A). The antibody specificity was determined by the relative abundance of S100A4 represented by MS PSM score (displays the total number of identified peptide sequences) in each of the three bands excised from the gel. Thus, MS results showed that S100A4 was predominantly detected at ~ 8 kDa (Band# 1, PSM score= 36) which confirms the specificity S100A4 mAb.

To provide further confidence of the true molecular weight of S100A4 amongst the three bands detected, mass spectrometry analysis of the abundance of S100A4 in each band represented by PSM score was performed. MS results showed that band# 1 (detected at ~ 8 kDa) has the highest abundance of S100A4 (PSM score= 36) (Figure 3-6B) which indicates the true molecular weight of S100A4. Further evidence of S100A4 mAb (CST, Clone: D9F9D) specificity was elucidated by functional studies. When S100A4 was ectopically overexpressed in AML cell lines, the S100A4 mAb bound specifically to overexpressed S100A4 as well as to endogenous S100A4 (Figure 4-6). Conversely, when S100A4 was knocked down, the S100A4 mAb did not bind to other members of the S100A family in the absence of S100A4 (Figure 4-15). Given that S100A4 has been detected at higher molecular weights (band #2 at ~12 kDa and band #3 at ~15 kDa) by S100A4 Ab (CST, clone: D9F9D) this raises the question of what causes the S100A4 molecular weight shift in electrophoretic mobility. There are several possible explanations in this case; one possible reason could be that S100A4 is post-translationally modified (PTM) and in turn the PTM group adds an extra molecular weight that can be detected at a different weight level. Using PTM prediction tools online such as PhosphoSitePlus.org identified several PTMs in S100A4 such as phosphorylation, sumoylation, ubiquitination, and acetylation sites.

Thus, it has been reported that intracellular S100A4 is sumoylated by SUMO1 at two sumoylation sites (Lys<sup>22</sup> and Lys<sup>96</sup>) in order to be translocated to the nucleus to

regulate the expression of MMP-13 upon IL1 $\beta$  stimulation (Miranda et al. 2010). Our MS analysis of the 3 bands excised from the gel identified UBC9 (also known as SUMO E2) a SUMO catalytic enzyme in the dataset of band #3 only suggesting that S100A4 may be sumoylated. However, MW of SUMO proteins can range between 10-15 kDa which may add a substantial weight to S100A4 MW (8-12 kDa) that may run to a much higher weight level than observed at 12 and 15 kDa. Thereby, S100A4 sumoylation is unlikely to be the reason for the higher bands detected by S100A4 Ab on western blot.

Another possible explanation could be that the secondary bands detected at 12 and 15 kDa are dimer forms of Ca<sup>2+</sup> bound S100A4 altering the electrophoretic mobility. Although S100A4 can be present in monomeric form, binding the EF-hand to Ca<sup>+</sup> promotes dimer formation in the S100A4 structure which is necessary to expose target binding domains in an antiparallel orientation on the face of the dimer (Donato et al. 2013). The dimer forms of S100A4 may thus appear on western blot with molecular weights approximately twice the expected molecular weight. A further technical explanation could be that S100A4 can still preserve multimeric forms even after SDS and heat treatment. It has been reported that proteins that have hydrophobic patches (such as S100A4) may exhibit abnormal migration on SDS-PAGE due to differential SDS-binding capacity to protein (Miranda et al. 2010). Taken together, our optimization concluded that S100A4 mAb (CST, Clone: D9F9D) has the highest binding specificity and optimal binding capacity when used at a 1:1,000 dilution. Mass spectrometry results provided further evidence that S100A4 migrates at an apparent molecular weight ~8 kDa in our electrophoretic system and AML cell lines.

### **3.4.2 S100A4 expression levels and subcellular localisation in normal haematopoiesis and AML**

The CCLE RNA-Seq data show that S100A4 is upregulated in AML cell lines. S100A4 mRNA levels in AML cell lines shown in Figure 3-7A are in relative agreement with the protein expression levels shown in Figure 3-7B. However, mRNA expression profiling is not a powerful prediction tool for protein expression (Vogel and Marcotte 2012). Further, transcriptomic analysis alone does not indicate the aberrant localisation of proteins to the nucleus which can be critical to cancer development (Wang and Li 2014). Thus, transcriptional profiling is often coupled with an alternative proteomic analysis such as western blot or mass spectrometry. Here, using the optimized conditions

for western blot analysis of S100A4 in a panel of 10 AML cell lines (nuclear and cytoplasmic) showed that five AML cell lines (ME-1, NOMO-1, THP-1, HEL, OCI-AML-2) expressed S100A4 in both compartments with 3 cell lines expressing S100A4 in the cytosolic compartment only (Figure 3-7B). The cell lines that expressed S100A4 in the nucleus can provide more information on the effect of silencing S100A4 expression by shRNA on survival and growth. Conversely, cells that did not express S100A4 in the nucleus could potentially be used as a model to study the effect of overexpressed S100A4 (4.3.4).

#### 3.4.2.1 *S100A4 expression is restricted to cytosolic compartments in normal haematopoietic cells*

In order to investigate the patho-physiological role of S100A4 in AML development or progression, the expression levels and sub-cellular distribution has to be established in the normal haematopoietic progenitor cells as well as the differentiated subpopulation. In this study, cord blood derived CD34<sup>+</sup> cells were used as a matched normal control to AML patient samples (FAB M1). Freshly isolated human CD34<sup>+</sup> cells from neonatal cord blood have similar transcriptional profiles to minimally differentiated AML blasts (FAB M-1) (Munje *et al.* 2015). First, to examine *S100A4* gene expression profile in normal haematopoiesis and AML, we utilised publicly available gene expression data on BloodSpot<sup>®</sup>. *S100A4* mRNA expression patterns suggest that *S100A4* is minimally expressed in undifferentiated HSCs and as these cells differentiate *S100A4* expression increases (Figure 3-8A).

Our microarray analysis of CD34<sup>+</sup> cells and myeloid progenitors (monocytes, erythrocytes and granulocytes) was consistent with the gene expression analysis from BloodSpot in which *S100A4* becomes upregulated with differentiation. Normal CD34<sup>+</sup> cells had low mRNA levels whilst upregulated gene expression of *S100A4* was observed in monocytes and erythrocytes. Much less increase was observed in granulocytes in our dataset and unfortunately there is no direct comparator in the Bloodspot data to verify this (Figure 3-7B). *S100A4* is reported to be expressed in early stages of development of normal tissues as it is involved in a plethora of cell differentiation-related functions (Koshelev Iu *et al.* 2008).

Taken together, these findings may suggest that *S100A4* is expressed in small quantities in early stages of development and as cells become actively involved in cell

division and differentiate towards committed sublineages of the blood, *S100A4* expression increases. However, mRNA transcriptional profiles alone are not a powerful tool to predict the effects of *S100A4* expression on AML development or progression. Often, protein mislocalisation between cytoplasm and nucleus may interfere with normal cellular functions by interacting with key oncoproteins or tumour suppressors which cooperatively may lead to tumour development and metastasis (Wang and Li 2014). Thus, western blot was performed to analyse the subcellular localisation S100A4 in normal CD34+ cells and myeloid progenitors. Blots demonstrated that S100A4 is normally expressed only in the cytosol and no expression was observed in the nucleus either in CD34+ cells (Figure 3-10A) or in uni-lineage progenitors with the exception of granulocytes (Figure 3-10B). The band present in the cytoplasmic compartment of granulocytes (Figure 3-10B) could be accounted for by contamination from unbound flow through of monocytes or erythrocyte fractions (< 10% and < 5% respectively) (Figure 3-9). Further, normal bone marrow samples have also shown that expression of S100A4 is cytosolic with no nuclear expression observed (Figure 3-11). In summary, in normal haematopoiesis, S100A4 expression is restricted to the cytoplasmic compartment only and it is developmentally regulated.

#### 3.4.2.2 *S100A4 is mislocalised to the nucleus in AML patients*

Nuclear mislocalisation of S100A4 has been previously linked to aggressiveness in epithelial ovarian carcinoma and thus poor survival rates ( $p = 0.0045$ ) (Kikuchi et al. 2006). Another study suggested that nuclear expression of S100A4 is correlated with advanced stages in primary colorectal adenocarcinoma ( $p = 0.002$ ) whilst cytoplasmic expression of S100A4 was not significant (Flatmark et al. 2003). Prior to this study, S100A4 nuclear mislocalisation was not reported in AML patient samples. Initially, interrogating curated gene expression data on the BloodSpot<sup>®</sup> database (Bagger et al. 2016) provided useful insights on *S100A4* gene expression patterns in normal haematopoietic stem cells versus AML subtypes. The analysis showed that *S100A4* mRNA is significantly overexpressed across all AML subtypes as compared to normal HSC (Figure 3-12). The transcriptomic data analysis findings derived from the Bloodspot database consolidated our previous analysis of the nuclear proteome of AML blasts (Tonks et al. 2007) which showed S100A4 upregulation in AML, however one of the

limitations of gene expression analysis is that it is not informative in terms of protein subcellular distribution.

S100A4 expression and subcellular localisation was analysed in AML patient samples to give us insights on abnormal subcellular location in AML. In this study, AML patient samples from minimally differentiated leukemic blasts (FAB M1) were chosen to minimise developmental variability compared to matching normal control CD34<sup>+</sup> cells. Similarly, AML FAB M4 leukemia blasts were chosen to analyse S100A4 in myelomonocytic leukemia blasts compared to matching control CD14<sup>+</sup> monocyte progenitors. In minimally differentiated AML (subtype FAB M1), S100A4 was not only mislocalised in the nucleus but also overexpressed as compared to CD34<sup>+</sup> cells (Figure 3-13). The expression patterns of S100A4 in AML blasts (FAB M-1 and M-4), where S100A4 was found to be mislocalised to the nucleus, suggests that nuclear mislocalisation of S100A4 could play a pathological role in AML development. Taken together, these findings were consistent with the results previously shown in section 3.4.2.1, that S100A4 is predominantly a cytosolic protein in normal CD34<sup>+</sup> cells and differentiated myeloid sublineages. However, the presence of S100A4 in the nucleus of AML patient samples may suggest that S100A4 could be translocated to the nucleus to interact with key regulatory proteins that are involved in AML and cooperatively drive AML development and/or progression.

In general, S100 family members have no known enzymatic activity and it is believed that they exert their functions by interacting with their binding partners (Donato et al. 2013). S100A4 nuclear function is still to be described, however; a recent study demonstrated an intra-nuclear association between S100A4 and suppressor protein p53 which led to its degradation and induced cell survival (Orre et al. 2013). In conclusion, understanding the nuclear function of S100A4 could be the key of the underlying mechanism in which S100A4 or its potential binding partner(s) play their role in AML development.

### **3.4.3 Expression of *S100A4* does not correlate with patients' clinical outcome**

Aberrant expression of *S100A4* has been associated with either poor clinical outcome or metastasis in several solid tumours (Helfman *et al.* 2005, Zhao *et al.* 2013). Using the TCGA microarray data set on the BloodSpot<sup>®</sup> database, Kaplan-Meier analysis of survival shows that the group of AML patients with above median *S100A4* mRNA

values have lower survival probability than the group of patients who are below median *S100A4* mRNA ( $p=0.0118$ ) (Figure 3-15). To further analyse the effect of *S100A4* mRNA on patients' survival, the TCGA data were analysed (Ley *et al.* 2013) excluding intensively treated patients and patients who have favourable risk abnormalities such as  $t(8;21)$  RUNX1/RUNX1T1 and  $t(15;17)$  PML-RAR $\alpha$  from the analysis. Kaplan-Meier analysis of survival shows that higher levels of *S100A4* mRNA are associated with poor OS of AML patients ( $p=0.01$ ) (Figure 3-16).

These findings suggest that overexpressed *S100A4* confers survival disadvantage for those patients who have high levels of S100A4. A recent studies suggested that overexpression of S100A4 mediated the favourable prognosis of PRAME in paediatric AML via deactivating the tumour suppressor p53 (Xu *et al.* 2016). Unfortunately, our correlation analysis of *S100A4* mRNA did not show any correlation between overexpressed S100A4 and patients' outcomes ( $p\text{-value}=0.7$ ) (Figure 3-17). The significant difference between the OS and correlation analysis could be explained by two reasons. First, to show the effect of upregulated *S100A4* on patients' outcomes the sample size may need to be increased to detect the effect of S100A4 on patients' outcomes. Secondly, the OS analysis is a multivariate analysis whereas the correlation analysis is bivariate. Further, as mentioned previously, mRNA profiling has poor correlation with protein expression in AML patients.

Moreover, mRNA profiles do not distinguish the subcellular localisation of S100A4 which is the main pathological observation in this project. Thus, a survival analysis may need to be performed on fractionated AML patients' samples to see whether or not the subcellular distribution of S100A4 correlates to patients' outcomes. However, this may not be feasible at this stage given the high number of samples required to show such correlation. Whilst these results provide information about *S100A4* gene expression and its correlation to patient outcomes, the pathophysiological role of S100A4 (on protein level) as well as its subcellular distribution in haematological malignancies (as well as normal haematopoiesis) could be elucidated further by functional studies in AML cell line models. Previous studies of knockdown of S100A4 in anaplastic thyroid cancer (ATC) showed significant reduction in proliferation and increased apoptosis in ATC cells *in vitro* and similar results were achieved in an *in vivo* model (Zhang *et al.* 2016). Therefore, in the next chapter, functional studies including nuclear targeted

overexpression of S100A4 and knocking down its expression will be used to elucidate the underlying mechanism of S100A4's role in AML development and/or progression.

# Chapter 4

**The effect of over-expressing and knockdown of S100A4 on normal human haematopoietic development, AML growth and survival**

## 4 Chapter 4 - Introduction

### 4.1 Overview

In the previous Chapter, I established that S100A4 protein expression in the nucleus was observed in approximately 86% of AML blasts. In contrast, nuclear expression of S100A4 was undetectable in normal human CD34<sup>+</sup> HSPC controls. Interestingly, S100A4 expression was also increased in the cytoplasm of AML blasts versus normal controls. Moreover, data derived from TCGA suggests that S100A4 overexpression may confer a poor prognosis. Indeed, overexpression of S100A4 is widely linked to the tumour aggressiveness and metastatic phenotype in many solid tumours (Li *et al.* 2013, Zhou *et al.* 2018) and can be utilised as a biomarker in early diagnosis of several solid tumours at treatable stages (Boye *et al.* 2016).

Nuclear expression or mislocalisation of S100A4 has also been associated with tumour stage in colorectal cancer, aggressiveness, and metastasis in ovarian cancers (Boye *et al.* 2010, Kikuchi *et al.* 2006). Studies knocking down S100A4 in an *in vitro* model in colorectal cancer blocks cells growth, migration and invasion activities and induces apoptosis (Huang *et al.* 2012). Further, knocking down S100A4 in human prostate cancer xenograft mice resulted in significant reduction in tumour vascularity and inhibited tumour growth (Ochiya *et al.* 2014). Therefore, dysregulation of S100A4 expression is common in many solid tumours which seem to rely on expression of this protein. However; little is known about the functional role of S100A4 in human haematopoiesis as well as in leukaemogenesis. This Chapter determines whether ectopic expression of nuclear S100A4 or S100A4 knock down (KD) can affect the growth and survival of human CD34<sup>+</sup> HSPCs. I also examined the consequences of knocking down S100A4 expression in leukaemia cell lines by assaying effects on cell proliferation, differentiation, and drug sensitivity analysis.

## 4.2 Aims

The main objective of this Chapter is to gain an understanding of the role of S100A4 in haematopoietic development and its possible contribution to the development of AML.

This chapter will achieve this objective through the following aims: -

1. Determine the effect of nuclear overexpression of S100A4 or S100A4 knock down on normal human haematopoietic cell growth, differentiation and development;
2. Determine the effect of nuclear overexpression of S100A4 or S100A4 knock down on AML cell growth, proliferation and apoptosis.

## 4.3 Results

### 4.3.1 Generating retro- and lentiviral vectors overexpressing S100A4

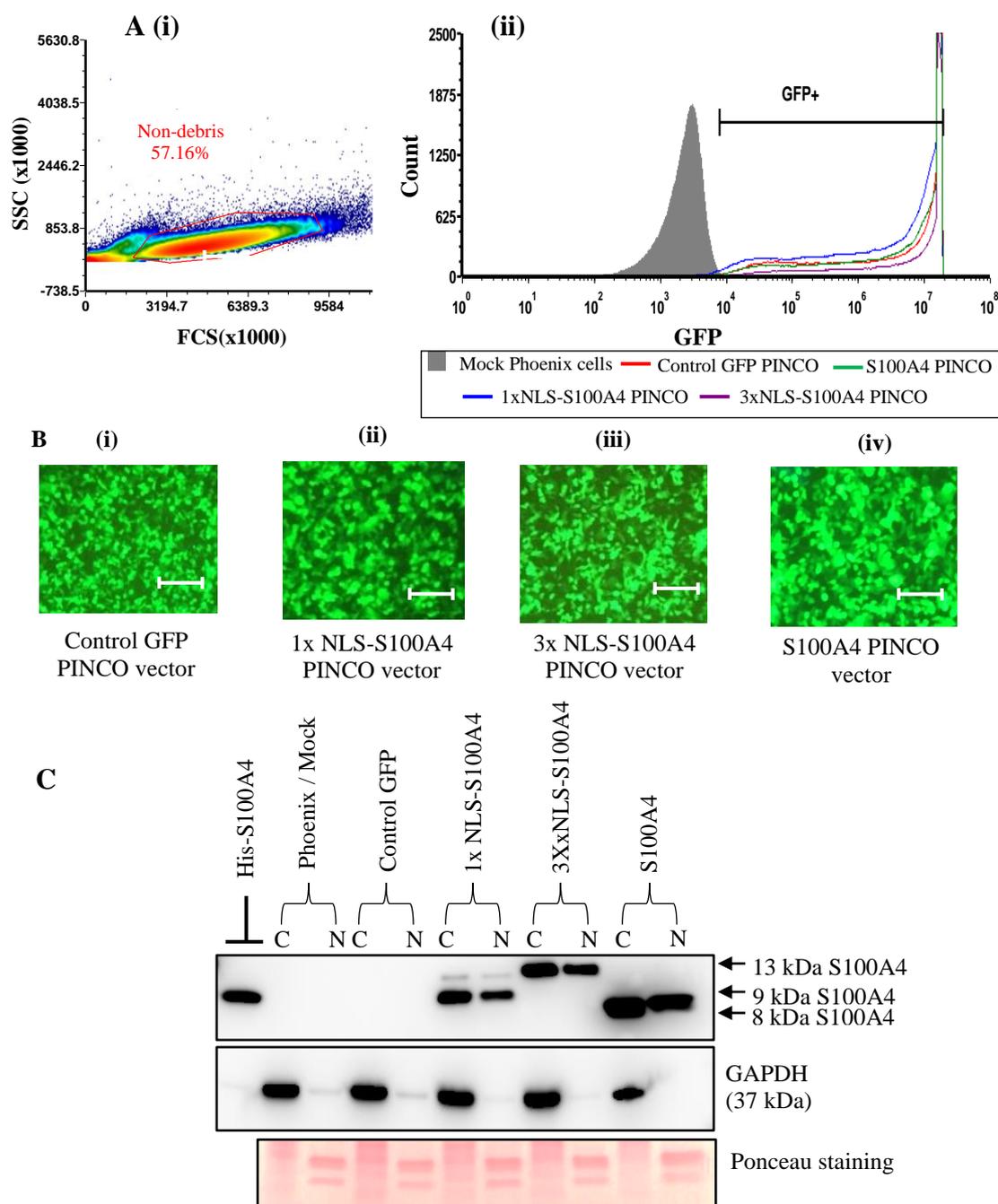
In order to determine the effect of nuclear mislocalised S100A4 on growth and differentiation in normal human HSPCs and AML cell lines, retro- and lentiviral vectors were generated. Retroviral (PINCO) and lentiviral (pHIV) vectors carrying a nuclear localisation signal (NLS) were constructed that would facilitate the expression of S100A4 in the nucleus (Grignani *et al.* 1998, Park *et al.* 2015). These vectors also co-expressed eGFP as a selectable marker. Four PINCO vectors were generated for this purpose; (i) control vector expressing GFP alone, (ii) 1xNLS-S100A4, (iii) 3xNLS-S100A4 to stably overexpress S100A4 in the nucleus, and (iv) S100A4 without a nuclear targeting sequence. To corroborate the retroviral nuclear transduction of S100A4, additional lentiviral vectors were also generated: (i) control vector expressing GFP alone and (ii) 3xNLS-S100A4 (Figure 2-2). Using a directional subcloning approach, 1xNLS S100A4 and 3xNLS S100A4 cDNA were ‘released’ from the expression plasmid (pEX-A2) using *Bam*HI and *Eco*RI restriction enzyme digestion and subsequently ligated into PINCO and pHIV vectors. To validate subcloning, initial test digests were performed using restriction enzymes. As per the predicted DNA migrating patterns following agarose gel electrophoresis (Figure 4-1B), the test digests validated successful subcloning of both PINCO and pHIV digested DNA fragments (Figure 4-1C). Subsequently, DNA sequencing across the subcloning site validated DNA sequence against published S100A4 NCBI sequence (NM\_002961.2) (data not shown).

### 4.3.2 Validation of S100A4 ligation into PINCO and pHIV vectors and transduction efficiency in Phoenix cells

To generate and functionally validate the retro and lenti-vectors above, DNA transfection of Phoenix and HEK 293T cells was performed respectively. Firstly, Phoenix cells were transfected with PINCO control vector DNA, 1xNLS S100A4, 3xNLS S100A4 or S100A4. Following harvest of the retrovirus, GFP expression was analysed by flow cytometry. Cells transfected with these vectors showed approximately 98% GFP positivity (Figure 4-2A). To confirm that transiently transfected Phoenix cells also expressed S100A4 in the nucleus, cells were fractionated into cytoplasmic and nuclear lysates and western blot was performed. Control cells showed little or no expression of S100A4 in both nuclear and cytosolic subcellular compartments (Figure 4-2C). In contrast 1xNLS-S100A4 and 3xNLS-S100A4 expressed S100A4 in both the cytosol and nucleus (Figure 4-2C). As expected, expression of 3xNLS-S100A4 migrated with an apparent protein molecular weight of ~13 kDa; consistent with the additional size of the 3xNLS signal (5 kDa). Further, S100A4 lacking NLS was overexpressed in the cytosolic and nuclear compartments and migrated to an approximate molecular weight of 8 kDa consistent with S100A4 M.W. (Figure 4-2C).

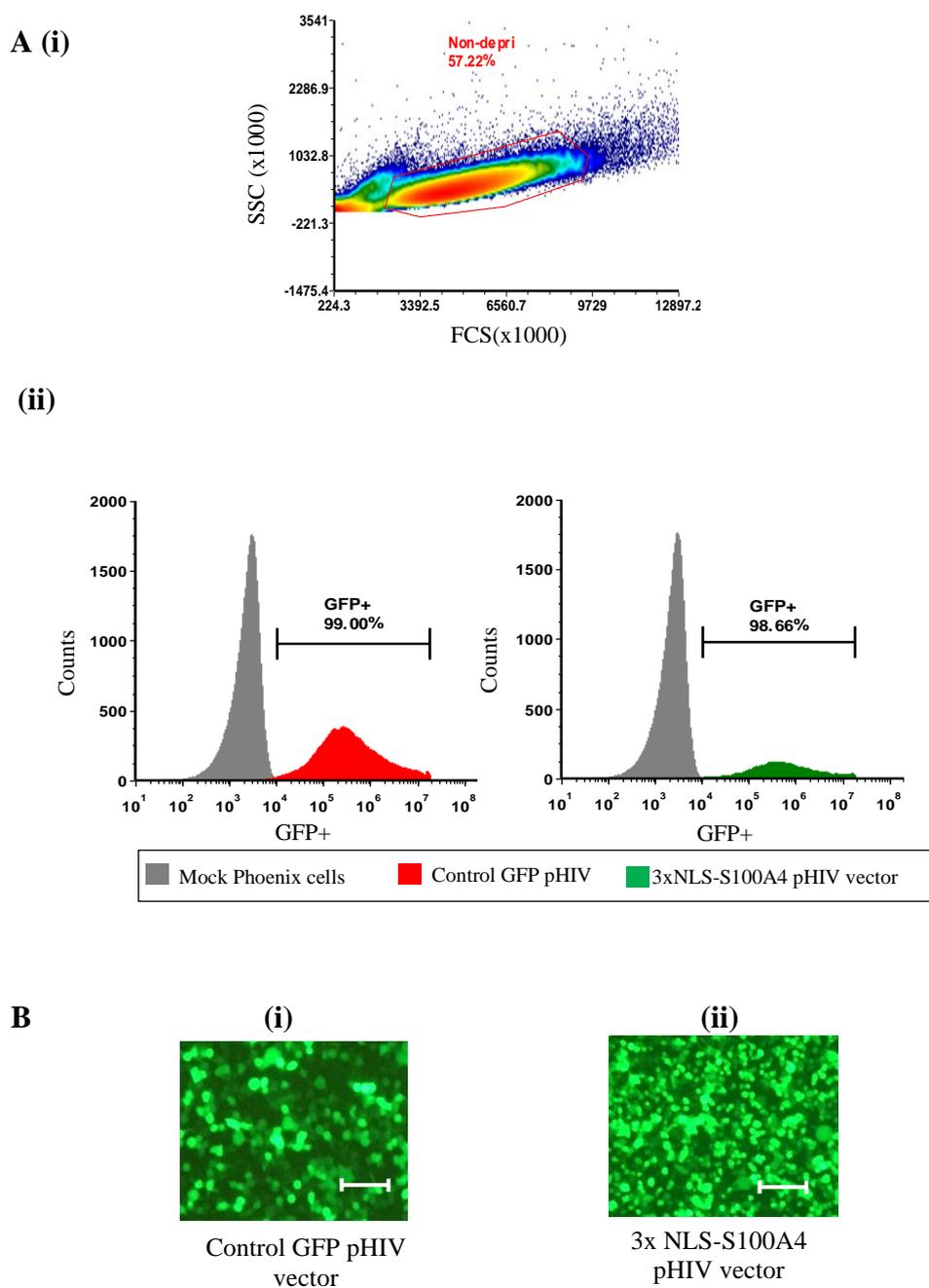
Alternatively, HEK 293T transfected with control, and 3xNLS-S100A4 pHIV vectors also showed more than 98% GFP positivity (Figure 4-3A). Cells transfected with control GFP pHIV vector showed no expression of S100A4. However, HEK 293T cells transfected with 3xNLS-S100A4 vector showed overexpression of S100A4 in cytoplasmic and nuclear compartments with a corresponding molecular weight shift to 13 kDa (data not shown). Taken together both retro- and lentiviral vectors are functional in terms of expressing GFP and S100A4 in the context of adherent packaging cells.





**Figure 4-2: Confirmation of GFP and S100A4 protein expression in Phoenix cells transfected with PINCO retroviral vectors.**

A) GFP expression of transfected packaging Phoenix cells was quantitated by flow cytometric analysis. (i) Example of non-debris gating strategy. (ii) The grey histogram shows the mock transfected control (cells 'transfected' without DNA). The coloured histogram overlays are Phoenix cells transfected with control expressing GFP only, vector expressing S100A4 only, 1xNLS-S100A4 or 3xNLS-S100A4. Note: the fluorescence of most cells is off-scale. B) GFP fluorescence microscopy images of Phoenix cells transfected with above PINCO vectors. Magnification is  $\times 20$  and scale bar indicates  $10\mu\text{m}$ . C) Western blot analysis of S100A4 in transfected Phoenix cells with control (PINCO expressing GFP alone), 1xNLS-S100A4 or 3xNLS-S100A4 or S100A4 without NLS sequence. Phoenix cells were fractionated into cytosol (C) and nucleus (N). S100A4 is visualised at 8kDa. 1xNLS-S100A4 is visualised at 9 kDa and 3xNLS-S100A4 at 13 kDa. GAPDH and Ponceau indicate the purity of each subcellular fraction and relative loading.



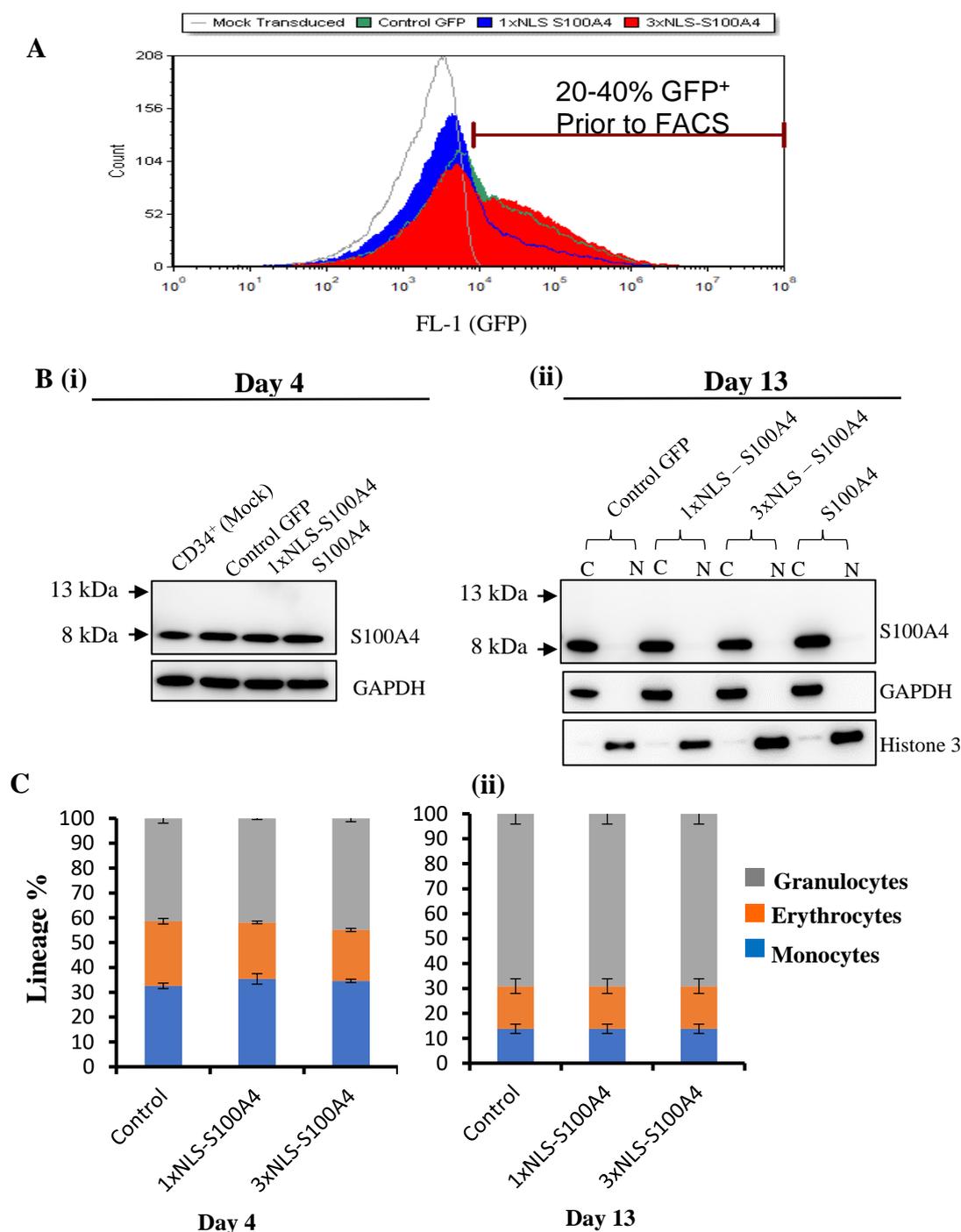
**Figure 4-3: Confirmation of GFP and S100A4 protein expression in HEK293T cells transfected with pHIV lenti-viral vectors.**

**A)** GFP expression of transfected packaging HEK293T cells was quantitated by flow cytometric analysis. **(i)** Example of non-debris gating strategy. **(ii)** The grey histogram shows the mock transfected control (cells ‘transfected’ without DNA). The coloured histogram overlays are HEK293T cells transfected with control expressing GFP only and 3xNLS-S100A4. **B)** GFP fluorescence microscopy images of HEK293T cells transfected with above PINCO vectors. Magnification is  $\times 20$  and scale bar indicates  $10\mu\text{m}$ .

### 4.3.3 Phenotypic effects of overexpression of nuclear S100A4 on normal human CD34<sup>+</sup> HSPCs.

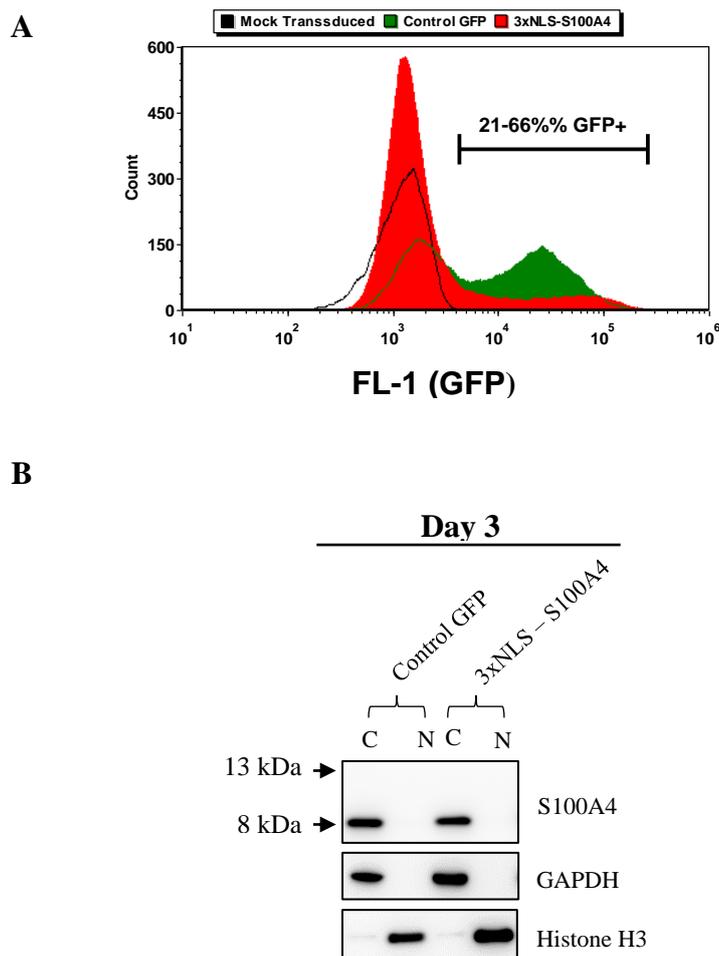
#### 4.3.3.1 Normal CD34<sup>+</sup> HSPCs cannot be shown to overexpress S100A4

Human CD34<sup>+</sup> HSPCs have a unique capacity to reproduce haematopoietic cells due to their self-renewal and pluripotency properties (Wang *et al.* 2009). Thus, CD34<sup>+</sup> cells represent an important model to study normal hematopoiesis. In order to determine the functional effects of over-expressed and nuclear mislocalised S100A4 on primary human CD34<sup>+</sup> HSPC differentiation and proliferation, cells were infected with retroviral (PINCO) vectors described above (4.3.1). Cells infected with control (expressing only GFP), 1xNLS-S100A4 and 3xNLS-S100A4 or S100A4 showed 20%, 33%, 40%, and 30% GFP positivity respectively (Figure 4-4A). However, whilst S100A4 was detectable in human CD34<sup>+</sup> HSPCs, there was no observable over-expression of S100A4 or NLS tagged constructs following infection (Figure 4-4B). Cells fractionated into cytosol and nuclear proteins on day 13 of culture, also failed to show cytosol or nuclear NLS tagged S100A4 protein expression. Unsurprisingly, S100A4 nuclear transduced CD34<sup>+</sup> HSPCs differentiated normally into myeloid committed progenitors compared to control counterparts (Figure 4-4C). I next used an alternative vector to corroborate the above findings. Lentiviral pHIV vectors were used to attempt to overexpress S100A4 in the nucleus of CD34<sup>+</sup> HSPC. As above, on day 3 post infection, cells expressed GFP (with infection frequencies of control pHIV vector; 66% and 3xNLS –S100A4 pHIV vector; 21%) (Figure 4-5A) but did not detectably overexpress S100A4 in the nucleus when analysed by western blot (Figure 4-5C). Given the successful validation of these constructs, these results suggest either that human CD34<sup>+</sup> HSPC may rapidly degrade overexpressed S100A4 protein or that the level of overexpression was too low to be detected in the fraction of the culture that was transduced.



**Figure 4-4: Human CD34<sup>+</sup> HSPC did not stably over-express S100A4.**

**A)** Example flow cytometric histogram showing CD34<sup>+</sup> cells expressing GFP. Histograms were gated on non-debris cell population based on FSC/SSC. Background auto fluorescence was established using mock transduced CD34<sup>+</sup> HSPCs treated in similar retroviral infection conditions except they were not subjected to retrovirus (grey line). The black marker line represents proportion of cells showing fluorescence greater than background auto fluorescence. **B)** Western blot analysis of transduced CD34<sup>+</sup> cells with 1x and 3xNLS-S100A4 on days 4 and 13 of culture. **C)** The bar chart showing the proportion of cells committed to myeloid sub-lineage (monocytes, erythrocytes, and granulocytes). GAPDH and Histone were used as loading controls and as verification of the subcellular fractionation.



**Figure 4-5: S100A4 expression could not be achieved in the nucleus of CD34<sup>+</sup> HSPCs using lentiviral vector.**

**A)** Example flow cytometric histogram showing CD34<sup>+</sup> HSPCs expressing GFP. Histograms were gated on non-debris cell population based on FSC/SSC. Background auto fluorescence was established using mock transduced CD34<sup>+</sup> HSPCs treated in similar lentiviral infection conditions except they were not subjected to lentivirus (black line). The black marker line represents proportion of cells showing fluorescence greater than background auto fluorescence. **B)** Western blot analysis of transduced CD34<sup>+</sup> HSPCs with control and 3xNLS-S100A4 on days 3 of culture. GAPDH and Histone were used as loading controls and as verification of the subcellular fractionation.

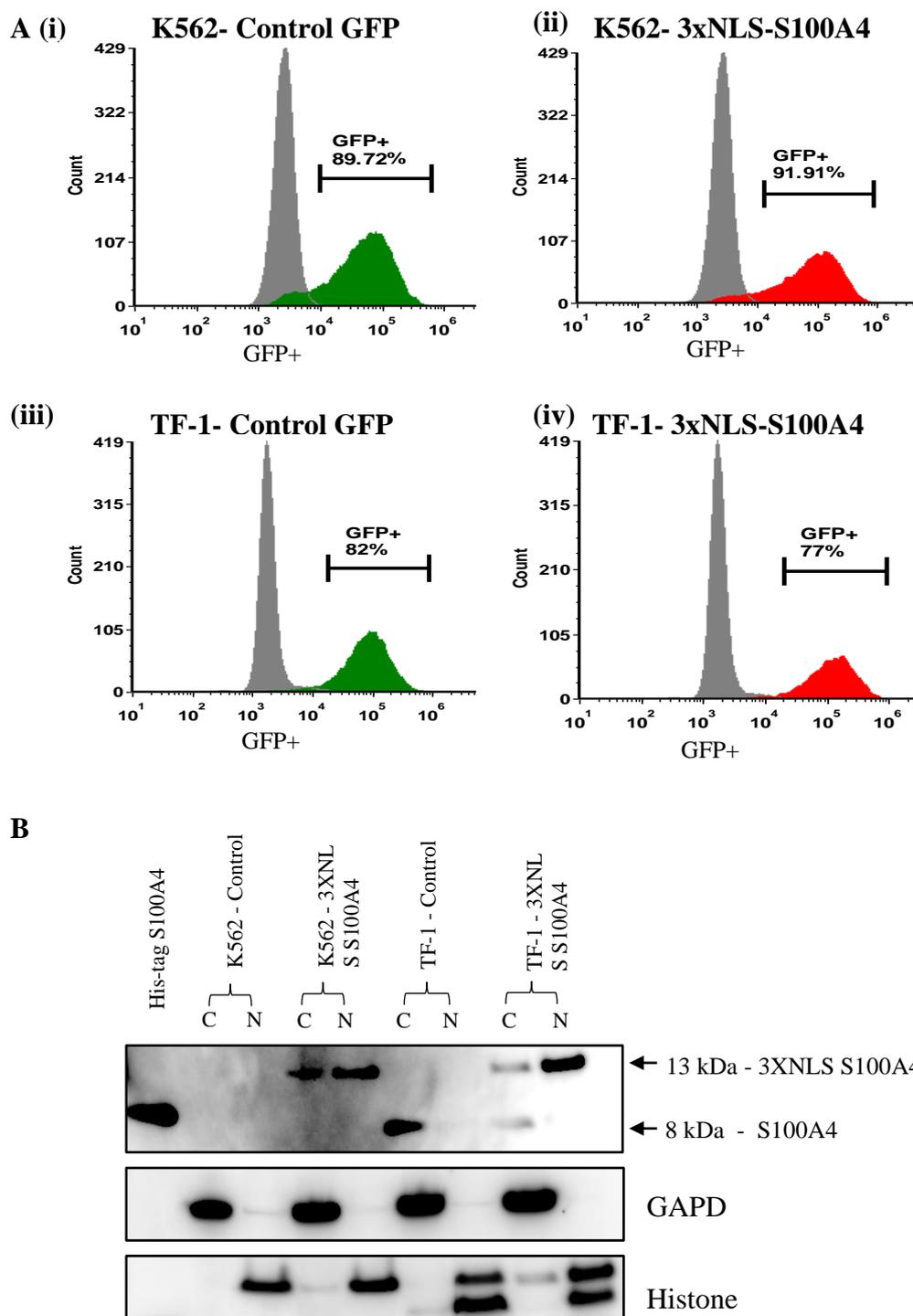
#### **4.3.4 Phenotypic effects of over-expression of nuclear mislocalised S100A4 on AML cell lines.**

##### *4.3.4.1 Generating AML cell lines that stably express S100A4 in the nucleus*

To study the effect of nuclear S100A4 expression in leukaemia cell lines, I used two lines that have minimal or no expression of S100A4 in the nucleus (K562 and TF-1). These cells were infected with control and 3xNLS S100A4 pHIV vectors and infection rates (GFP positivity) were determined flow cytometry. Flow cytometric analysis shows that K562 cells have high infection rates over 90% in both control and 3xNLS-S100A4 (Figure 4-6A). Likewise, TF-1 cells had 82% GFP<sup>+</sup> in control and 77% in 3xNLS-S100A4 (Figure 4-6A). Further, western blot analysis shows that forcing the expression of S100A4 in the nucleus was successfully achieved in both K562 and TF-1 cell lines (Figure 4-6B).

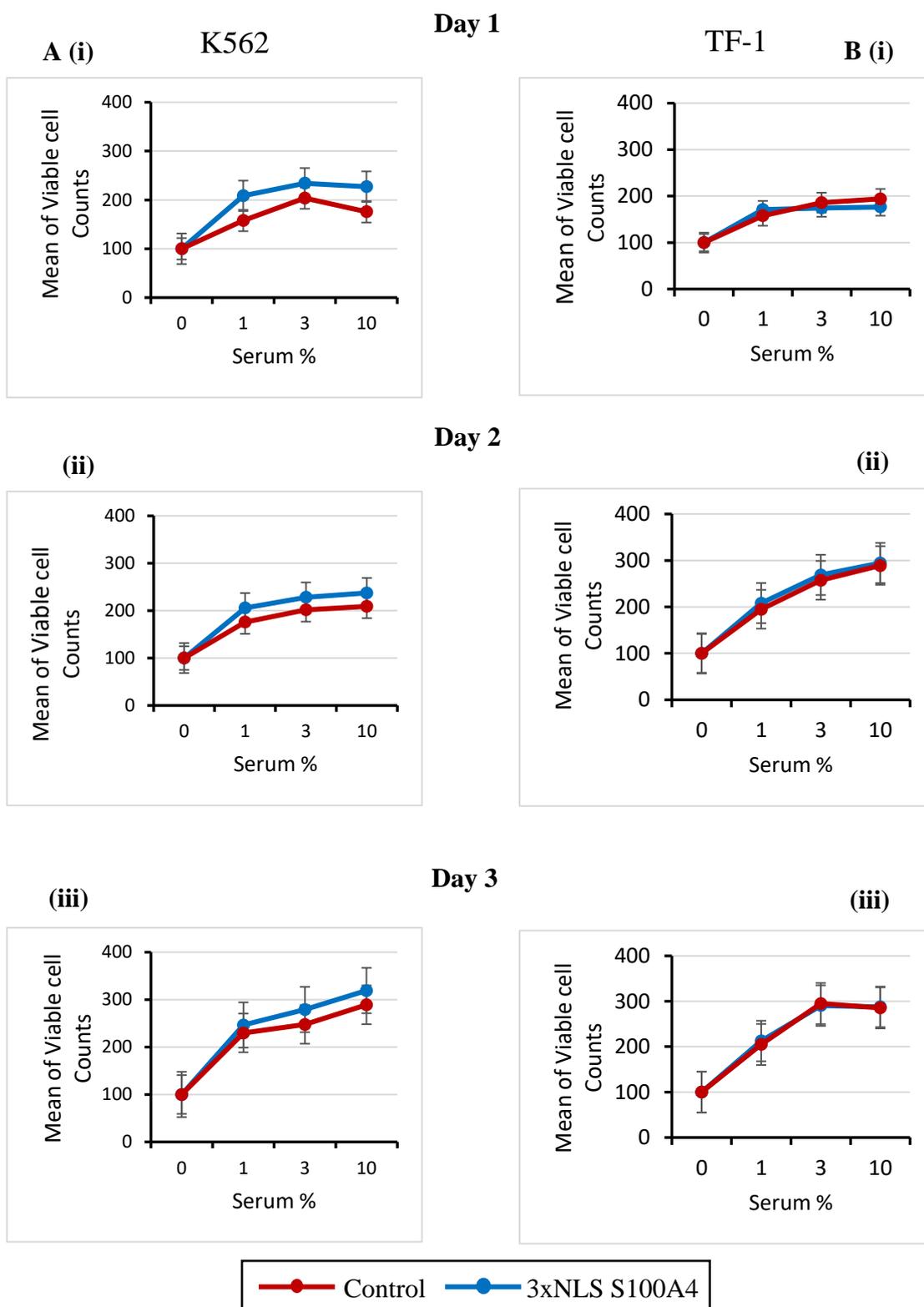
##### *4.3.4.2 S100A4 overexpression in the nucleus does not affect K562 and TF-1 proliferation under serum deprivation conditions*

To determine whether nuclear S100A4 over expression in leukemia cell lines promotes cell survival, K562 and TF-1 cell lines were grown under serum deprivation conditions. Cell viability were analysed on three consecutive days using flow cytometry. Proliferation was measured by viable cell count using viability stain Propidium Iodide (PI). There was no statistical difference in the growth and survival of S100A4 expressing cells compared to control over 3 days of culture in 0-10% v/v serum. (Figure 4-7A). Similar observations were observed in TF-1 cell expressing nuclear S100A4 (Figure 4-7B).



**Figure 4-6: S100A4 is overexpressed in the nucleus of leukemic cells K562 and TF-1.**

A) Example flow cytometric plots showing transduction efficiency in K562 and TF-1 cell lines infected with control or 3xNLS-S100A4. B) Example western blot showing nuclear overexpression of S100A4 to the nucleus of K562 and TF-1 cells. His-tag S100A4 peptide is used as a positive molecular weight control. GAPDH and Histone were used as loading controls and verification of the subcellular fractionation.



**Figure 4-7: S100A4 nuclear overexpression does not affect proliferation of AML cell lines.**

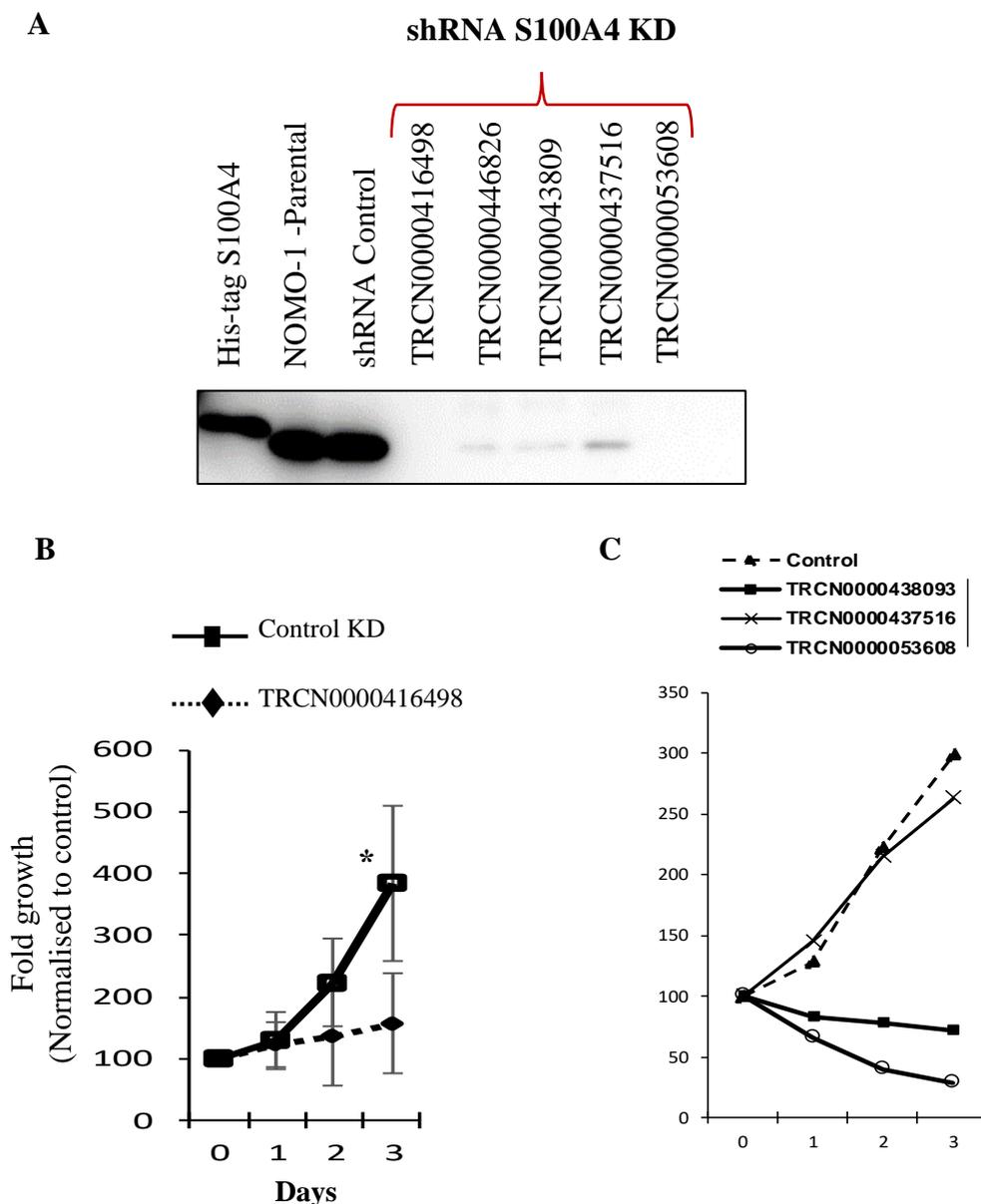
Flow cytometric analysis of cell proliferation in control and 3xNLS S100A4 K562 and TF-1 cells, on day 1 (top), day 2 (middle) and day 3 (bottom). Each data point corresponds to the average cell count from three replicates. The starting density of both K562 and TF-1 control were  $1 \times 10^5$ /mL, and  $1 \times 10^5$ /mL for K562 and TF-1 3xNLS S100A4. Viable cells are counted using flow cytometric viability stain Propidium Iodide (PI).

### 4.3.5 S100A4 knock down in normal human CD34<sup>+</sup> HSPC slows proliferation

#### 4.3.5.1 Generating shRNA control and S100A4 knockdown constructs

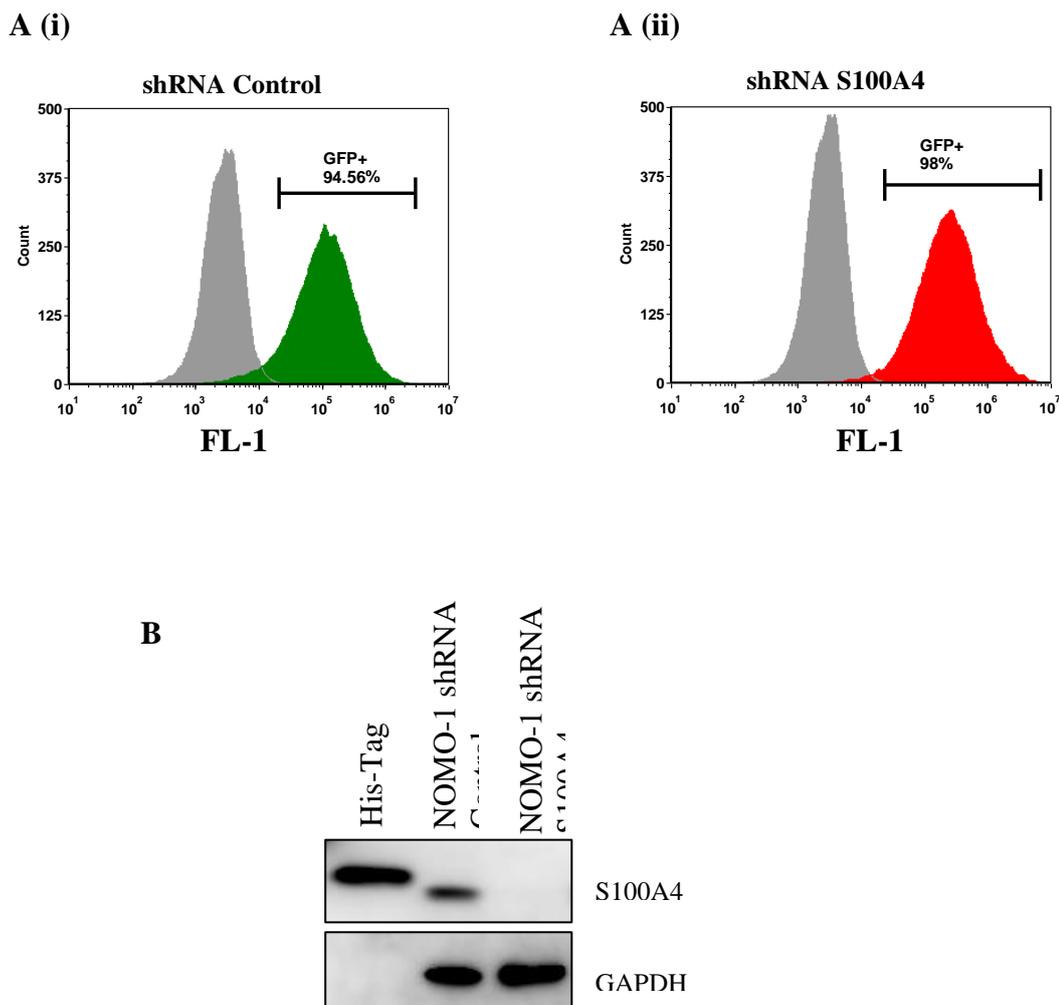
In order to study the effect of S100A4 KD on proliferation, differentiation and survival in normal CD34<sup>+</sup> HSPCs and AML cell lines, S100A4 was KD using shRNA. For this purpose, five short hairpin RNAs (shRNA) vectors were purchased from Mission<sup>®</sup> based on TRC(1)2-pLKO.5-puro (S100A4 shRNA and non-mammalian shRNA control), Table 2-3. Initially, S100A4 KD efficiency of these shRNA constructs were tested using NOMO-1; cell line expressing high S100A4 level. As shown in Figure 4-8A, the most efficient KD sequences of S100A4 shRNA were (TRCN0000416498 and TRCN000053608) with knocked down below detectable western blot sensitivity. The remaining three shRNA constructs partially knocked down S100A4. Mission<sup>®</sup> shRNA vectors do not contain GFP. Therefore, in order to follow growth of transduced cells, the most efficient S100A4 shRNA construct (TRCN0000416498) was re-designed with VectorBuilder<sup>®</sup> to include EGFP selectable marker. The analysis of transduced cells was based on the population of cells expressing GFP marker. As can be seen in Figure 4-8B, NOMO-1 cells with S100A4 KD (TRCN0000416498) showed significant reduction in growth compared to controls. Furthermore, the infection efficiency of re-designed shRNA S100A4 vector (TRCN0000416498) with EGFP selectable marker was validated in NOMO-1 cells. As shown in Figure 4-9A, the infection efficiency was over 95% in both lines infected with shRNA control KD and shRNA S100A4 KD (TRCN0000416498). Further, western blot confirmed S100A4 knockdown in NOMO-1 cells (Figure 4-9B). Three other shRNA S100A4 constructs were also used to knockdown S100A4 in NOMO-1 cells. As shown in Figure 4-8C, significant reduction in growth of NOMO-1 cells was also observed. However, no difference in growth was observed in NOMO-1 cells infected with shRNA S100A4 (TRCN0000437516) compared to control line.

In summary, Mission<sup>®</sup> shRNA S100A4 vectors were validated using AML cell line “NOMO-1”. Three shRNA S100A4 vectors consistently affected the growth and survival capacities of NOMO-1 cells. An improved shRNA S100A4 vector (TRCN0000416498) with EGFP selectable marker was chosen to carry out subsequent studies of the effect of S100A4 KD on normal HSPCs and AML cells lines.



**Figure 4-8: NOMO-1 requires S100A4 expression for cell growth.**

A) Western blot showing S100A4 KD in NOMO-1 cells using a panel of Mission<sup>®</sup> shRNA vectors (Sigma). Target sequence from TRCN0000416498 was used in subsequent studies due to its S100A4 KD efficiency. His-tagged S100A4 and mock infected NOMO-1 (parental) were used as positive controls. B) TRCN0000416498 were re-designed with VectorBuilder<sup>®</sup> to include EGFP marker. The summary data showing growth of NOMO-1 cells with S100A4 KD (TRCN0000416498) compared to control over 3 days of growth following infection. Data points indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \* P<0.05; analyzed by paired t-test. C) Data showing growth of NOMO-1 cells using a panel of Mission<sup>®</sup> shRNA vectors (Sigma). S100A4 KD was compared to control over 3 days of growth following infection. TCRN numbers identify target shRNA sequences used. Data points indicates mean (n=1). This experiment (part C) was conducted by Professor. Richard Darley (Cardiff University).



**Figure 4-9: Re-designed shRNA S100A4 vector is validated in NOMO-1 cells.**

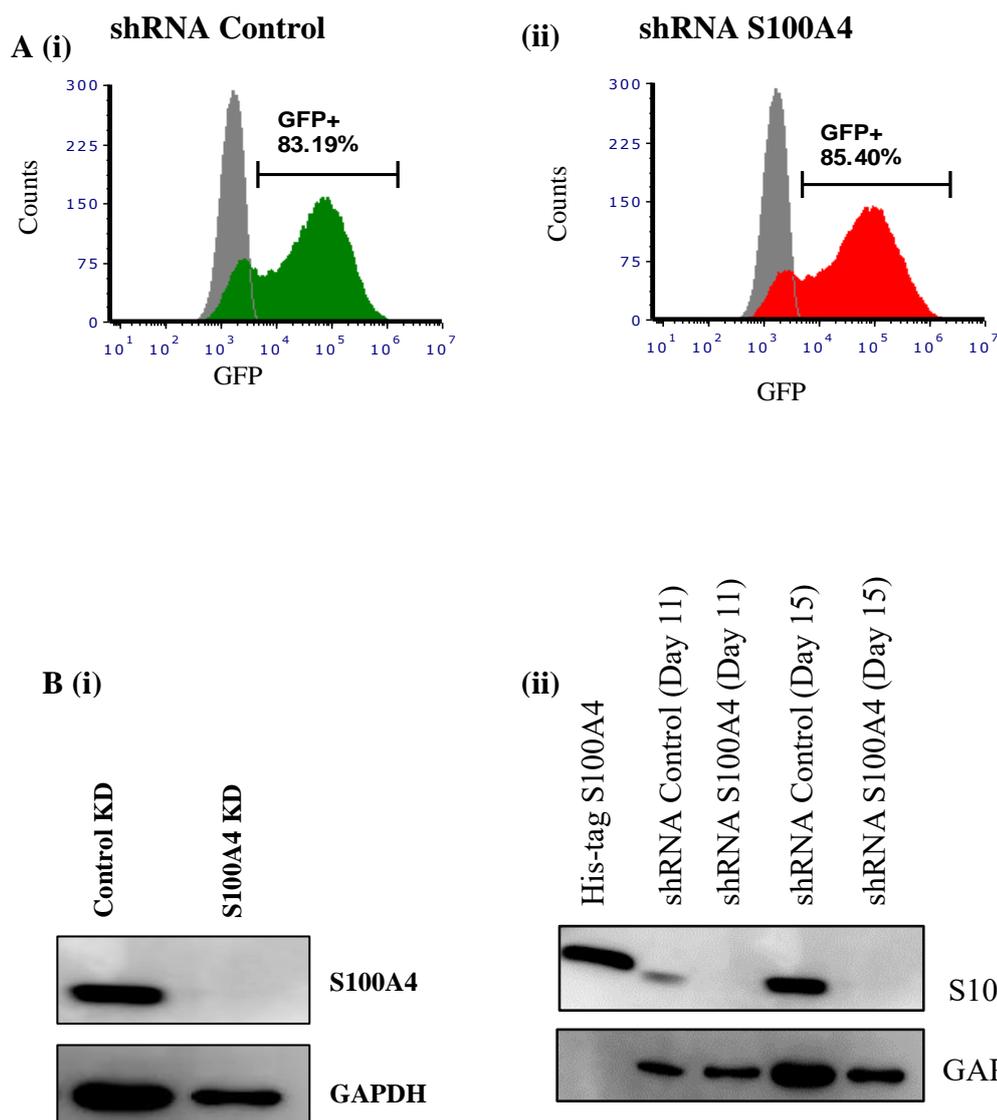
**A)** Example of flow cytometry analysis of infection efficiency rates of NOMO-1 cells infected with shRNA S100A4 (TRCN0000416498) with EGFP selectable marker (VectorBuilder®) **(i)** Control cells were infected with shRNA targeting a non-mammalian gene. Control shRNA infection rate in NOMO-1 cells 95% and **(ii)** S100A4 shRNA 98%. **B)** Western blot confirmation of S100A4 KD in total lysate of NOMO-1 cells. GAPDH was used as loading controls.

#### **4.3.6 S100A4 KD in normal CD34<sup>+</sup> HSPCs does not affect cell proliferation, differentiation, and survival**

To determine the effect of S100A4 KD on proliferation and differentiation in CD34<sup>+</sup> HSPCs, S100A4 was knocked down using shRNA (construct; TRCN0000416498). Infection rates of CD34<sup>+</sup> HSPCs with shRNA control and shRNA S100A4 KD constructs were determined using flow cytometry (GFP<sup>+</sup>). As shown in Figure 4-10A, over 80% of CD34<sup>+</sup> HSPCs are GFP<sup>+</sup> in both control and S100A4 KD lines on day 3 post infection. S100A4 KD was confirmed by western blot. As shown in Figure 4-10B, S100A4 was not detected in CD34<sup>+</sup> HSPCs infected with shRNA S100A4 compared to control. Further, CD34<sup>+</sup> HSPCs harbouring silenced S100A4 gene maintained no expression of S100A4 during differentiation.

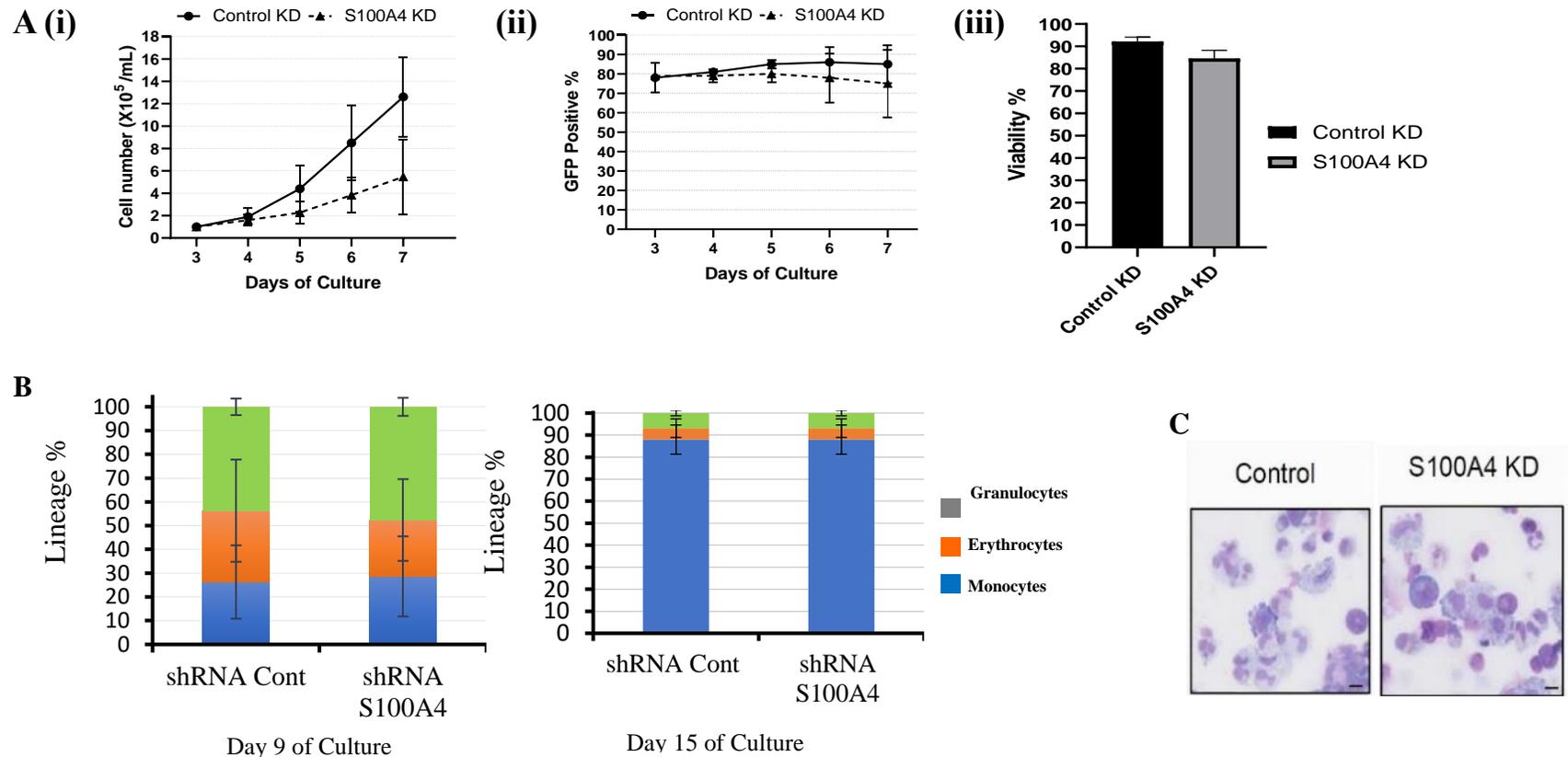
Subsequently, CD34<sup>+</sup> HSPCs control and S100A4 KD lines were analysed for the effect of S100A4 KD on growth, differentiation, and survival. As shown in Figure 4-11A, the growth rate of shRNA S100A4 transduced CD34<sup>+</sup> cells was reduced 2-fold compared to control. However, the effect of S100A4 KD on cell proliferation was not statistically significant ( $p$ -value= 0.104). Further, more than 80% of cells in control and S100A4 KD cultures were GFP<sup>+</sup> which indicates that the transduction is highly efficient and the majority of cells in the S100A4 KD culture are lacking S100A4 (Figure 4-11). Further, using PI, no significant change in cell viability was observed in S100A4 KD compared to control over 15 days of culture (Figure 4-11A). These data suggest that the loss of S100A4 in CD34<sup>+</sup> HSPCs marginally affects cell growth but does not affect cell viability.

To determine the effects on differentiation, cells were analysed for lineage markers (CD13 and CD36) as well as developmental cell surface markers (CD34, CD14, CD11b, CD15). S100A4 KD does not affect CD34<sup>+</sup> HSPC development (Figure 4-11C). Taken together, S100A4 was successfully knocked down in CD34<sup>+</sup> HSPCs but these cells do not require S100A4 for their growth and differentiation.



**Figure 4-10: S100A4 was successfully knocked down in CD34<sup>+</sup> cells.**

**A)** CD34<sup>+</sup> cells were transduced with shRNA non-mammalian gene (control) and shRNA S100A4 (TRCN0000416498) (Vector Builder<sup>®</sup>) encoding EGFP and grown in cytokine-driven culture (IL-3, SCF and G/GM-CSF). Representative flow cytometric histograms showing infection efficiency of CD34<sup>+</sup> cells analyzed on day 3 of culture. Mock transduced cultures (grey) delimits background auto-fluorescence. **B)** Western blot of S100A4 expression in CD34<sup>+</sup> cells grown under myeloid differentiating conditions. **(i)** Western blot example of Day 4 post-infection with shRNA control and shRNA S100A4 **(ii)** Western blot example of Day 11 and 15 post-infection with shRNA control and shRNA S100A4. Western blot was conducted on total protein lysates. GAPDH was used as loading controls.



**Figure 4-11: S100A4 knockdown in CD34<sup>+</sup> cells does not affect the development or survival of human myeloid haemopoietic cells.**

**A)** Summary data showing growth of CD34<sup>+</sup> HSPCs with S100A4 KD compared to control over 4 days of growth following infection **(i)** Expansion of shRNA transduced CD34<sup>+</sup> cells; fold change ( $p=0.104$ ). **(ii)** Proportion of GFP expression over 5 days post infection. Data indicated mean  $\pm$  1SD ( $n=3$ ). **(iii)** shRNA transduced CD34<sup>+</sup> cells viability. **B)** Analysis of the differentiation of transduced CD34<sup>+</sup> by four-colour cytometric analysis. GFP<sup>+</sup> myeloid cells were gated on CD13<sup>hi</sup>, CD36<sup>lo</sup> (granulocytes), CD13<sup>hi</sup>, CD36<sup>hi</sup> (monocytes) and CD13<sup>lo</sup>, CD36<sup>lo</sup> (erythrocytes). Bar chart showing the proportion of cells committed to sub-lineage development during culture. Data from 1 experiment; mean  $\pm$  1SD from five intra experimental repeats. **C)** Cytochemistry preparations, stained with Wright-Giemsa following 15 days in culture, showing myeloid morphology in the control (Left) and S100A4 KD (Right)-transduced cultures. Magnification is  $\times 20$  and scale bar represents 10 $\mu$ m.

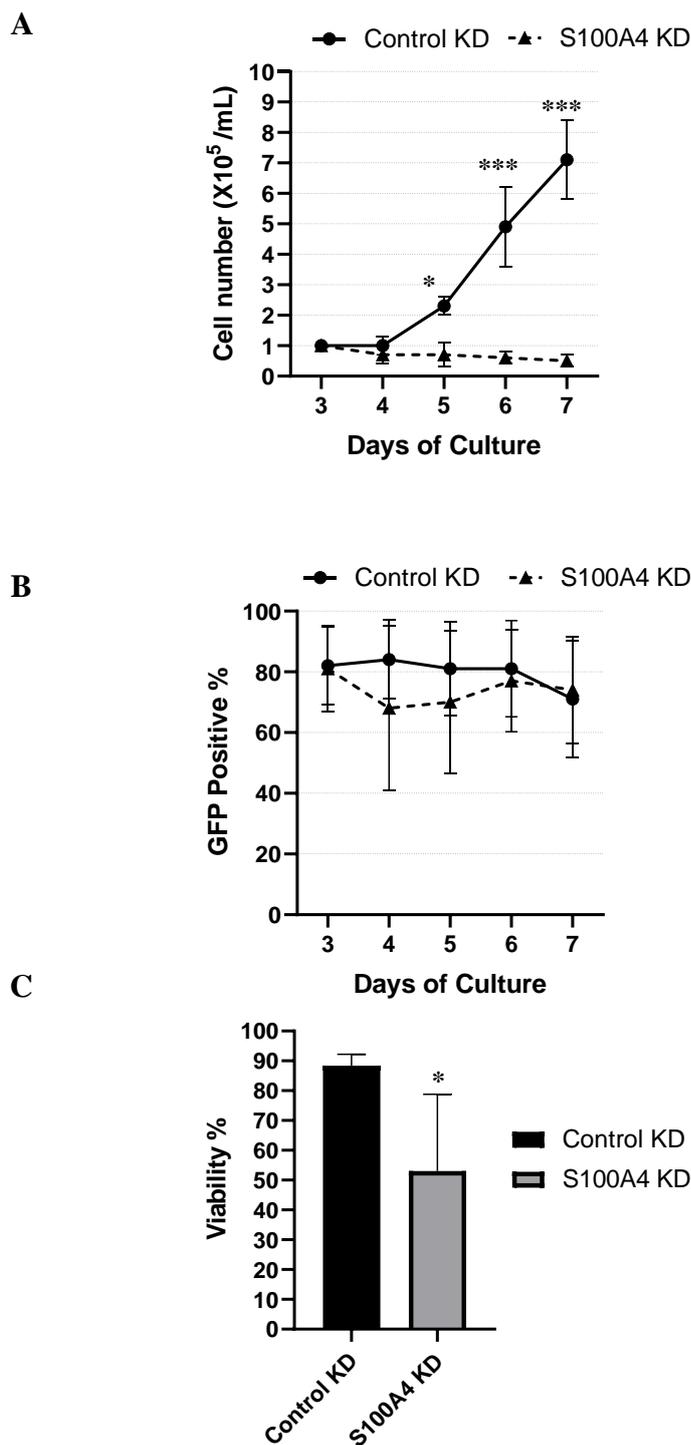
### 4.3.7 Leukaemic cells are dependent on S100A4 for their growth and survival

#### 4.3.7.1 *S100A4 knockdown in AML cell lines reduces cell proliferation and promotes apoptosis*

Following S100A4 KD in normal human CD34<sup>+</sup> HSPCs, the effect of S100A4 KD on proliferation and survival was determined in leukaemia cell lines. Using the same shRNA constructs, S100A4 was knocked down in leukaemia cell lines (NOMO-1, TF-1, THP-1, and OCI AML-2). S100A4 KD in these lines was validated by western blot (Figure 4-15, Figure 4-16, and Figure 4-17 respectively). Cell proliferation was significantly reduced in OCI AML-2 on 2<sup>nd</sup> and 3<sup>rd</sup> days of culture ( $p=0.0004$  and  $p=0.0001$  respectively), (Figure 4-12A). Similarly, cell proliferation was significantly reduced in NOMO-1 on 3<sup>rd</sup> day of culture (Figure 4-15B), TF-1 on the 1<sup>st</sup> and 3<sup>rd</sup> days of culture (Figure 4-16B), and THP-1 on 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> days of the culture (Figure 4-17B).

Next, these lines were assayed for apoptosis using Annexin V binding in combination with PI exclusion (Figure 4-14A). The most markedly affected cell line was OCI AML 2 among the other AML cell lines analysed ( $p < 0.001$ ). OCI-AML 2 cells exhibited apoptosis on the second day of culture post infection (Figure 4-14B). A similar trend was observed in TF-1 and THP-1 cell lines where apoptotic cells started forming following 48 hours post S100A4 KD. Conversely, NOMO-1 cells had a delayed apoptotic effect during same culture duration post S100A4 KD (Figure 4-14A). In all cell lines tested, loss of S100A4 expression induced Annexin V positivity suggesting that the lack of cell growth observed above was a result of programmed cell death.

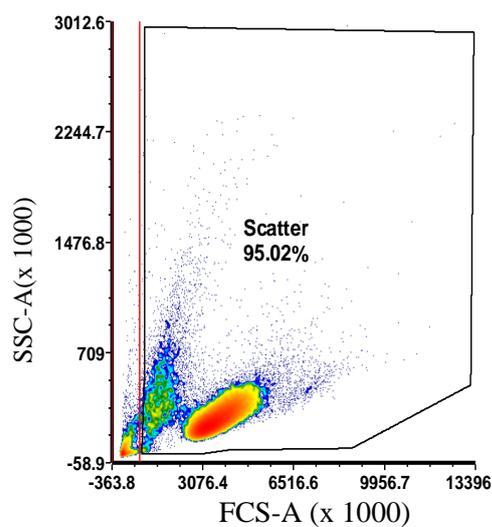
Taken together, these data infer that S100A4 is required for AML cells' survival but not for normal cells suggesting that targeting S100A4 would be an effective strategy in this disease.



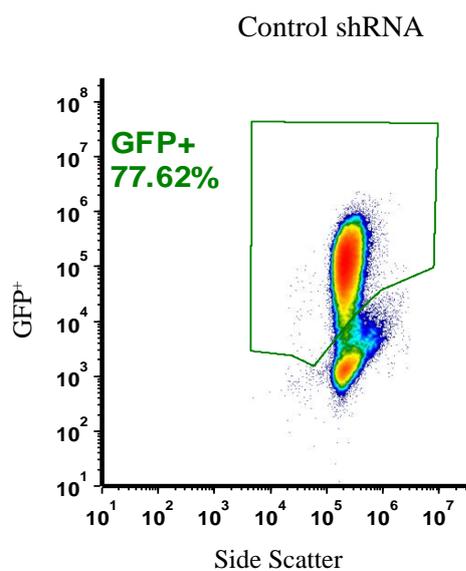
**Figure 4-12: OCI AML-2 cells are dependent on S00A4 for growth and survival.**

A) Summary data showing growth of OCI AML-2 cells with S100A4 KD compared to control over 3 days of growth following infection. B) Summary data of OCI AML-2 cells population expressing GFP over the course of growth assay. C) Summary data of OCI AML-2 cells viable counts of shRNA S100A4 compared to control shRNA. Propidium Iodide (PI) was used for viability staining. Data indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \*  $P < 0.05$ ; \*\*  $< 0.001$ , and \*\*\*  $< 0.0001$  analyzed by paired t-test.

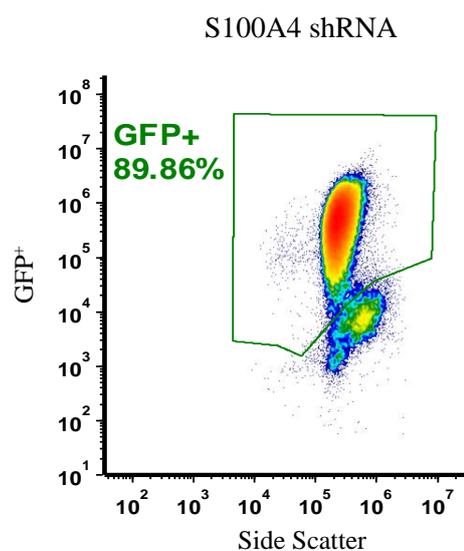
A



B (i)

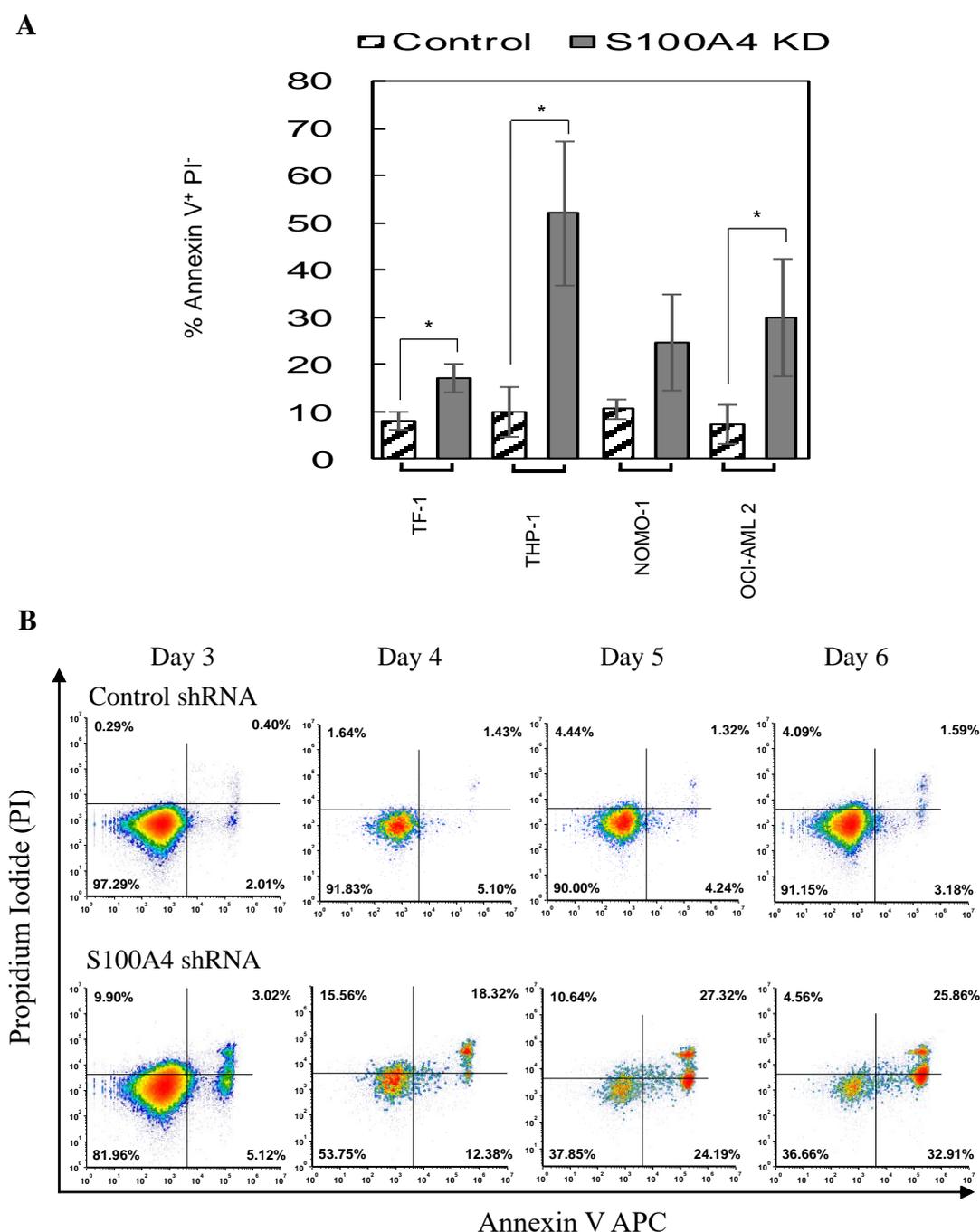


(ii)



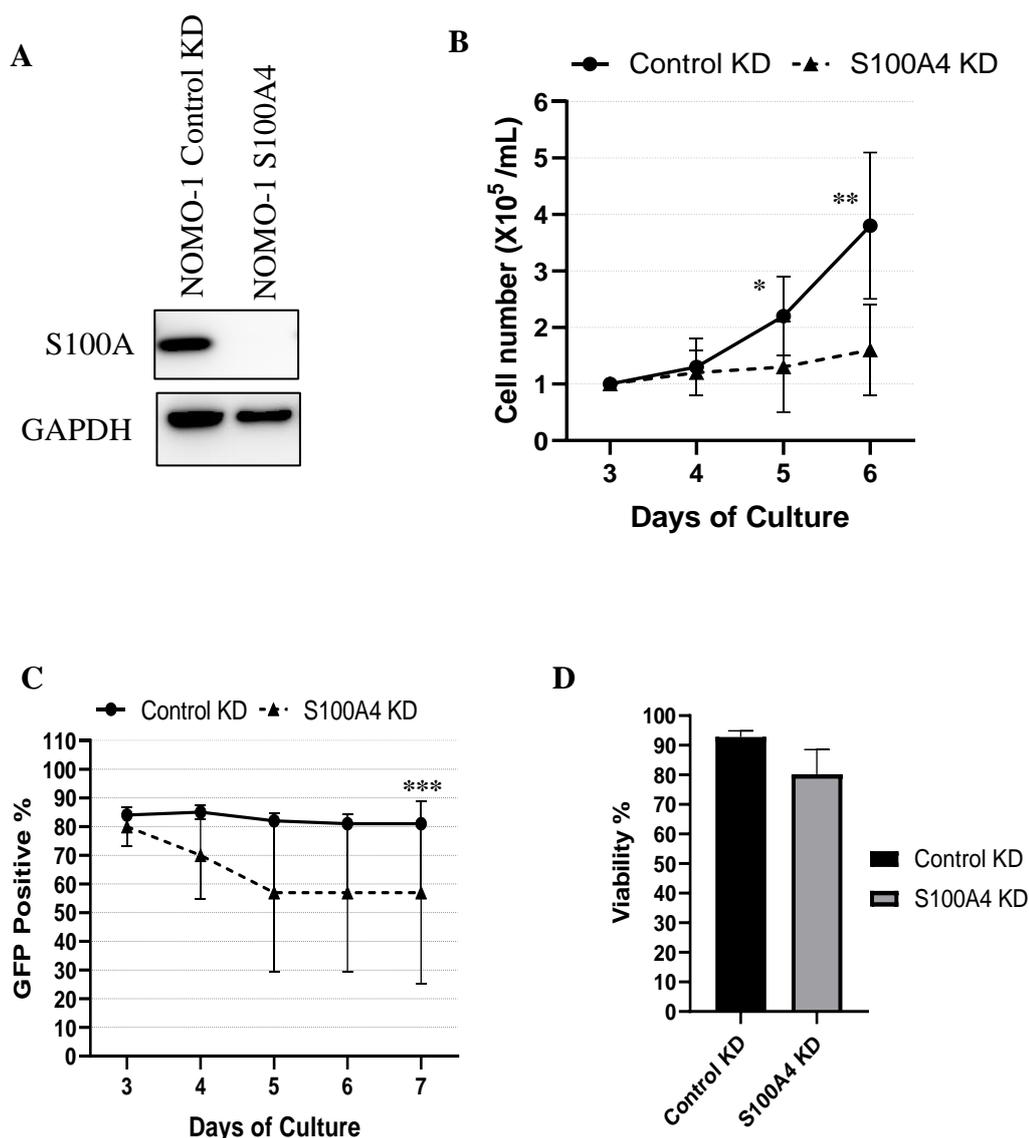
**Figure 4-13: Gating strategy for apoptosis assay in OCI AML-2 cell line.**

A) GFP expression of infected cells with control and S100A4 shRNA were confirmed by flow cytometric analysis. A) Example of gating strategy in which apoptotic bodies are included. B) Example of GFP positive cell population were gated on SSC & FL1. (i) OCI AML-2 control shRNA expressing GFP only (78%) and (ii) S100A4 shRNA (90%).



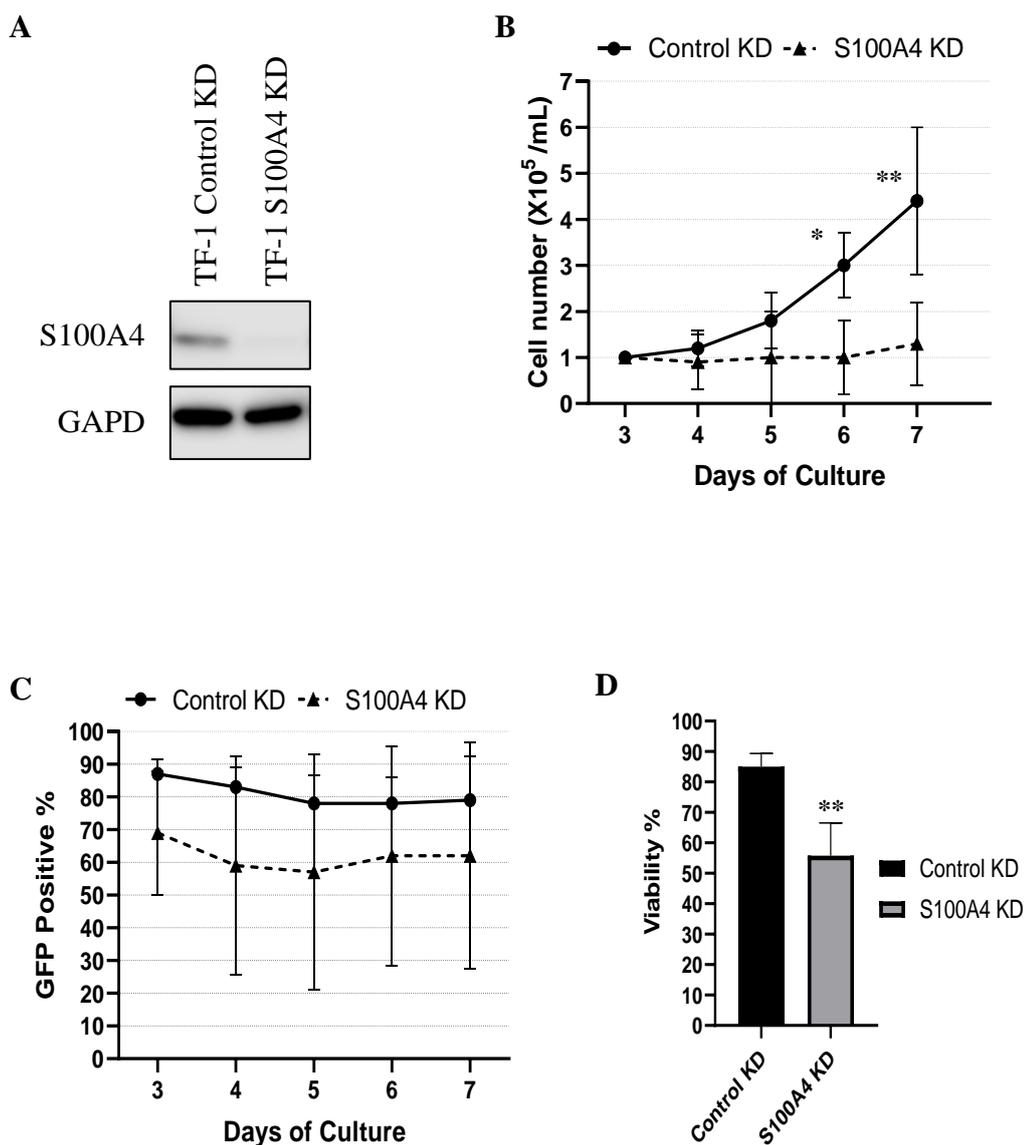
**Figure 4-14 : S100A4 is required for cell survival in OCI AML-2 cell lines.**

**A)** Summary data showing the effect of S100A4 KD on Annexin V staining in leukemia cell lines following 48 hours post infection. **B)** Example plots of S100A4 KD compared to control using OCI-AML2 are shown. Annexin V<sup>neg</sup> and PI<sup>neg</sup> (lower - left quadrant), annexin V<sup>pos</sup> and PI<sup>neg</sup> (lower - right quadrant) and both annexin V<sup>pos</sup> and PI<sup>pos</sup> (upper - right quadrant) cells were considered as the viable, early-phase apoptotic, late-phase apoptotic/necrotic cells, respectively. Data indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \* P<0.05; \*\*<0.001 analyzed by paired t-test.



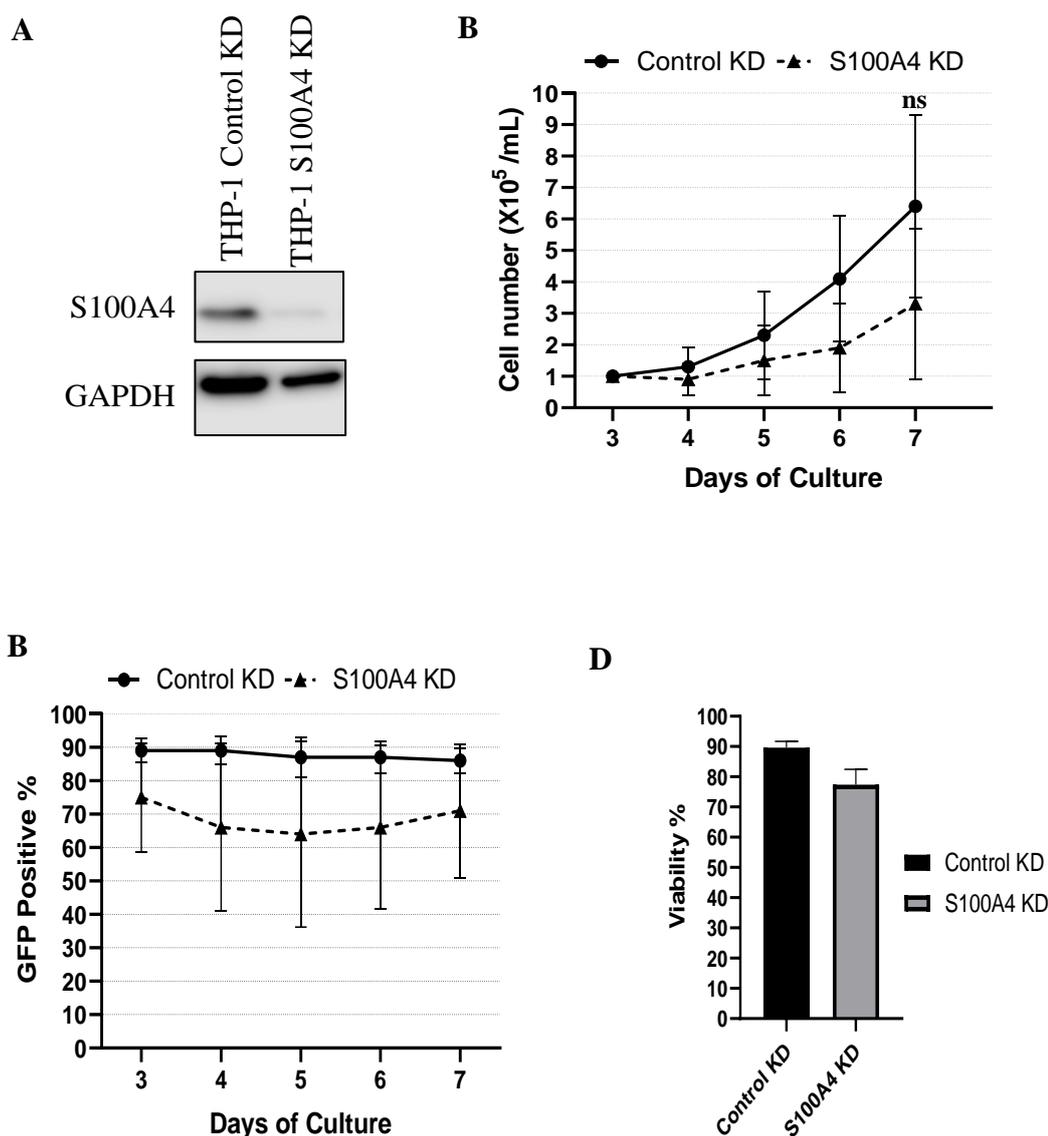
**Figure 4-15: NOMO-1 cells are dependent on S100A4 for growth and survival.**

**A)** Example western blot showing S100A4 expression in NOMO-1 cells with S100A4 KD compared to control (targeting non-mammalian gene) using shRNA. **B)** Summary data showing growth of NOMO-1 cells with S100A4 KD compared to control over 6 days of growth following infection. **C)** Summary data of OCI AML-2 cells population expressing GFP over the course of growth assay. **D)** Summary data of OCI AML-2 cells viable counts of shRNA S100A4 compared to control shRNA. Propidium Iodide (PI) was used for viability staining. Data indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \*  $P < 0.05$ ; \*\*  $< 0.001$  analyzed by paired t-test.



**Figure 4-16: TF-1 cells are dependent on S00A4 for growth and survival**

**A)** Example western blot showing S100A4 expression in TF-1 cells with S100A4 KD) compared to control (targeting non-mammalian gene) using shRNA. **B)** Summary data showing growth of TF-1 cells with S100A4 KD compared to control over 3 days of growth following infection. **C)** Summary data of TF-1 cells population expressing GFP over the course of growth assay. **D)** Summary data of TF-1 cells viable counts of shRNA S100A4 compared to control shRNA. Propidium Iodide (PI) was used for viability staining. Data indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \*  $P < 0.05$ ; \*\*  $< 0.001$  analyzed by paired t-test.



**Figure 4-17: THP-1 cells are dependent on S100A4 for growth and survival**

**A)** Example western blot showing S100A4 expression in THP-1 cells with S100A4 KD) compared to control (targeting non-mammalian gene) using shRNA. **B)** Summary data showing growth of THP-1 cells with S100A4 KD compared to control over 7 days of growth following infection. **C)** Summary data of THP-1 cells population expressing GFP over the course of growth assay. **D)** Summary data of THP-1 cells viable counts of shRNA S100A4 compared to control shRNA. Propidium Iodide (PI) was used for viability staining. Data indicates mean  $\pm$  1SD (n=3). Statistical significance is denoted by \*  $P < 0.05$ ; \*\*  $< 0.001$  analyzed by paired t-test.

## 4.4 Discussion

The main objective of this Chapter was to determine the functional significance of nuclear over-expressed and S100A4 KD on proliferation, differentiation, and survival of primary CD34<sup>+</sup> HSPCs as well as in leukaemia cell lines. To achieve this, S100A4 was attempted to be ectopically overexpressed in the nucleus of CD34<sup>+</sup> HSPCs and leukaemia cell lines (K562 and TF-1). Primary CD34<sup>+</sup> HSPCs did not overexpress of S100A4 and possibly degraded it. Further, overexpressing S100A4 in the nucleus of leukaemia cell lines had little or no effect on proliferation under serum deprivation conditions. Conversely, S100A4 KD in CD34<sup>+</sup> HSPCs possibly slowed their growth compared to control but does not affect their survival. Finally, this chapter showed that AML cell line (OCI AML-2, NOMO-1, TF-1, and THP-1) are dependent on S100A4 for their growth and survival. Taken together, this data suggests that S100A4 could be a potential targeted therapy that targets only leukemic cell while sparing normal CD34<sup>+</sup> HSPCs.

### 4.4.1 Generating retro- and lentiviral vectors that over-express S100A4 in the nucleus

Retro- and lentiviruses have different preferences regarding their cellular tropism and genomic integrations sites. Many subtypes of retroviruses are dependent on the degeneration of the nuclear membrane during cell division to enter the nucleus and genomically integrate their DNA (Miller 2014). Conversely, lentiviruses have the ability to transduce dividing and non-dividing cells efficiently via nuclear pores (Sakuma *et al.* 2012). Accordingly, lentiviral vectors such as pHIV have the advantages over the retroviral vectors in terms of the ability to infect both dividing and non-dividing cells (Cai and Mikkelsen 2016) and have the ability to stably and irreversibly integrate their DNA into the host cell genome (Liu and Berkhout 2014). Therefore, 3xNLS S100A4 pHIV vector was used should high transduction efficiency could not be achieved by PINCO retroviral vectors in primary or cell lines models. Given that the abnormal localisation of S100A4 expression to the nucleus is a common feature of leukaemia as shown in Figure 3-13, nuclear expression of S100A4 is forced by introducing up to 3x nuclear localisation signals sequence (NLS; e.g. GATCCAAAAAAGAAGAGAAAGGTA, SV40 large T-antigen NLS). The NLS is a short sequence of amino acid residues which are recognised by a nuclear protein transport called importins  $\alpha$  and/or  $\beta$  that mediate trafficking proteins

into the nucleus via nuclear envelope (Christophe *et al.* 2000). The NLS sequence can be attached artificially either at the N- or C-terminus of target protein. However, the C-terminus of S100A4 is crucial for target proteins binding such as p53 (Ismail *et al.* 2008) and important for metastasis-inducing properties through interaction with NMMHC IIA (Zhang *et al.* 2005). Therefore, to ensure the NLS did not interfere with the binding capacity of S100A4, the NLS sequence was attached at the N-terminus of S100A4 CDS cDNA.

In summary, retro- and lentiviral vectors containing the nuclear localisation signal and S100A4 have been successfully confirmed in the packaging cells to overexpress S100A4 in the nucleus at the expected M.W. of 13 kDa.

#### **4.4.2 CD34<sup>+</sup> HSPCs could not be shown to overexpress S100A4**

This study used a retroviral vector to target dividing CD34<sup>+</sup> HSPCs to induce overexpression of S100A4. DNA transfection efficiency was high in Phoenix cells (measured by GFP expression) and use of 1x or 3xNLS-S100A4 PINCO vectors resulted in protein expression of S100A4 in the nucleus of the packaging cells. However, protein expression could not be replicated in CD34<sup>+</sup> HSPC model. Most importantly, 1x and 3xNLS S100A4 transduction efficiency was only 20% and 33% respectively. Although an improved retroviral gene transduction protocol has been used to transduce CD34<sup>+</sup> HSPCs in this study (Tonks *et al.* 2005), these low transduction efficiencies cannot be explained but is likely due to low viral titres. Thus, CD34<sup>+</sup> HSPCs were transduced with lentiviruses; control pHIV expressing GFP only and 3xNLS-S100A4 pHIV vectors; however, this did not give higher transduction frequencies and again overexpression of S100A4 could not be demonstrated in transduced CD34<sup>+</sup> HSPCs with 3xNLS-S100A4 pHIV vector in the cytosol or nucleus.

These findings are consistent with the previous results in Section 3.3.4 which shows that normal CD34<sup>+</sup> cells and differentiated progenitors express S100A4 in the cytoplasmic compartment only and no nuclear expression of S100A4 was observed. It has been reported that S100A4 expressed predominantly in the cytoplasm. For example, the Mandinova *et al.* have shown that S100A4 is predominantly expressed in the cytosol of smooth muscle cell lines derived from human aorta and intestines. Whereas, other members of the family such as S100A2 and S100A6 were located primarily in the nucleus suggesting distinctive localisation pattern of S100 protein family (Mandinova *et al.*

1998). Further, it has been reported that S100A4 is located in the plasma membrane and involved in  $\text{Ca}^{2+}$  trafficking and regulation (Boye and Maelandsmo 2010). Other studies suggested that changes in the extracellular concentration of  $\text{Ca}^{2+}$  triggers the release of S100A4 into extracellular space of artery smooth muscle cells (Donato 2007). However, there is no published evidence that S100A4 is expressed in the nucleus of normal cells. In contrast, nuclear expression of S100A4 is often linked to metastasis and poor outcomes in colorectal (Boye *et al.* 2010), liver (Fabris *et al.* 2011), ovarian (Kikuchi *et al.* 2006), and breast cancer (Egeland *et al.* 2017). This suggests that  $\text{CD34}^+$  cells may recognise the nuclear expression of S100A4 as abnormal and therefore rapidly degrade it (hence the failure to detect its nuclear expression). Since nuclear localisation of S100A4 could not be demonstrated in  $\text{CD34}^+$  HSPCs, the functional consequences of mislocalised S100A4 to nucleus could not definitively be determined.

#### **4.4.3 $\text{CD34}^+$ cells do not depend on S100A4 for their growth, differentiation, and survival**

Survival rates of AML patients (under 60's years) has improved greatly over the last thirty years 20-75%. However, in patients over 60 years 3-5 year survival rate is still around 10% (Seval and Ozcan 2015). Consequently, the need for alternative targeted treatments or adjuncts to current treatment programmes is more needed than ever. Thus, to determine whether  $\text{CD34}^+$  cells are dependent on S100A4 for their growth, differentiation, and survival, S100A4 was knocked down using shRNA S100A4 lentiviral vector. S100A4 KD in  $\text{CD34}^+$  cells shows that loss of S100A4 expression in  $\text{CD34}^+$  cells slowed the growth rate by 2-fold, however; this reduction was not statistically significant ( $p= 0.103$ ). Moreover, S100A4 KD did not affect the normal differentiation of  $\text{CD34}^+$  cells sub-lineages (monocytes, erythrocytes, and granulocytes), Figure 4-11B. These findings are consistent with previously reported data that S100A4 knock out mice do not show any obvious phenotype at birth and develop normally (Naaman *et al.* 2004).

In summary,  $\text{CD34}^+$  cells can tolerate the loss of S100A4 and differentiate normally in the myeloid lineage. However,  $\text{CD34}^+$  cells may proliferate at lower growth rates as a result of S100A4 KD.

#### 4.4.4 Nuclear S100A4 over-expression does not affect AML cell lines proliferation

Previous studies have reported that S100A4 mislocalisation to the nucleus drives tumour growth and migration capacity in solid tumours such as breast cancers (Egeland *et al.* 2017). Following unsuccessful attempt to demonstrate nuclear overexpression of S100A4 in normal CD34<sup>+</sup> HSPCs to study the effects on proliferation and differentiation, S100A4 was overexpressed in the nucleus of leukaemia cell lines. Initially, a panel of lines (KG-1, Mv4;11, K562, and TF-1) that have little or no expression of S100A4 in the nucleus were chosen for this purpose. Overexpression of S100A4 in the nucleus was initially achieved in KG-1 and Mv4;11 cell lines. However, KG-1 and Mv4;11 cell lines showed rapid degradation of S100A4 while cells in culture (data not shown). KG-1 cells are less mature cell line classified as FAB M1 and derived from bone marrow of 59 years old male patient with erythroleukemia (Koeffler and Golde 1978). Interestingly, KG-1 cells' behaviour was consistent with primary CD34<sup>+</sup> HSPCs in degrading nuclear S100A4 post transduction. Conversely, K562, and TF-1 had sustained the overexpression of S100A4 in the nucleus during sub-culturing. Although K562 is not an AML cell line, both cell lines can still be used to study the effect on myeloid malignancy due to lack of S100A4 expression in the nucleus and ease of lentiviral infectability. Moreover, S100A4 is equally important in cytoplasm versus nucleus in K562 and TF-1 cells since these cell lines have no cytoplasmic expression of S100A4.

K562 and TF-1 cell lines were transduced with control GFP and 3xNLS-S100A4 pHIV vectors and cultured under serum deprivation conditions. Results show that nuclear expression of S100A4 does not affect the proliferation of K562 and TF-1 cells. In contrast, several studies in solid tumours have shown association between nuclear mislocalised S100A4 expression and increase of tumour growth, (Kikuchi *et al.* 2006), invasion (Wang *et al.* 2012), and metastasis (Mishra *et al.* 2012). One possible mechanism by which S100A4 promotes tumour proliferation is inhibiting the apoptotic cascade preventing cells from undergoing apoptosis. Interestingly, preferentially expressed antigen of melanoma (PRAME) which has previously been shown to reduce tumorigenicity of leukemic cells *in vivo*, has also been shown to reduce expression of S100A4, particularly in those leukaemia associated with favourable outcome (e.g., in leukaemia's harbouring RUNX1-ETO and PML-RAR $\alpha$ ) (Tajeddine *et al.* 2005). More recently, Xu *et al.* demonstrated that PRAME promotes apoptotic death of leukaemia cells

by regulating S100A4/p53 signalling (Xu *et al.* 2016). Indeed, it has been reported that S100A4 binds to C-terminal part of p53 in the nucleus of lung adenocarcinoma cells which in turn tags p53 for MDM2-dependent ubiquitination (Orre *et al.* 2013). Further, Orre *et al.* demonstrate that knockdown of S100A4 increased the transcriptional activity of p53 and led to p53-dependent cell cycle arrest.

Taken together, overexpressing S100A4 in the nucleus of KG-1 and Mv4;11 cell lines induced rapid degradation of S100A4 protein consistent to what was observed in normal CD34+ cells. Nevertheless, overexpressing S100A4 in the nucleus of K562 and TF-1 cell lines did not affect their proliferation.

#### 4.4.5 AML cell lines are dependent on S100A4 for their growth and survival

Several studies have reported that knocking down S100A4 in solid tumours reduces tumour growth, suppresses migration, and induces apoptosis (Zhang *et al.* 2016, Huang *et al.* 2012). However, the effect of S100A4 KD on growth and survival capacities in haematological malignancies has not been reported. In order to study these effects on leukemic cells, a panel of AML cell lines (OCI AML-2, NOMO-1, THP-1, and TF-1) were chosen. As shown previously in Fig. Ch. 3, OCI AML-2, NOMO-1, and THP-1 cell lines exhibited nuclear expression of S100A4. TF-1 cell line, however; expresses S100A4 in the cytoplasm but have no expression of S100A4 in the nucleus. Nevertheless, TF-1 can still be a useful model to study the effect of S100A4 KD on leukemic cells expressing cytoplasmic S100A4. As shown in Figure 4-12A, Figure 4-15A, and Figure 4-16A, S100A4 knockdown has significantly impaired the growth in OCI AML-2, NOMO-1, and TF-1 compared to control KD, respectively. Whereas, S100A4 KD in THP-1 cells slowed growth but was not statistically significant ( $p= 0.1$ ), Figure 4-17A. These findings are consistent with Zhang *et al.* study findings, where S100A4 KD through RNA interference (S100A4 siRNA) significantly decreased proliferation and increased apoptosis in anaplastic thyroid cancer (ATC) cell lines *in vitro*. Similar findings were achieved in an *in vivo* model when S100A4 was knocked down with shRNA, significant reduction in tumour growth and inhibition of metastasis in the abdominal cavity were observed (Zhang *et al.* 2016). Therefore, it can be hypothesised that S100A4 plays a crucial role in leukemic cells' proliferation and survival.

Furthermore, knocking down the expression of S100A4 in AML lines resulted in cell death through induction of apoptosis. As shown in Figure 4-14A, S100A4 KD has

significantly promoted apoptosis in OCI AML-2, TF-1, and THP-1 cell lines. OCI AML-2 cells is the most affected AML cell line by the absence of S100A4 expression as cell death was observed within the first 48 hours of S100A4 shRNA treatment (Figure 4-12B). However, in NOMO-1 line, cell death was delayed as NOMO-1 cells maintained their viability post knockdown of S100A4. Overall, these results of apoptosis assay clearly indicate that leukemic cells are dependent on S100A4 for their survival.

In conclusion, CD34<sup>+</sup> cells can tolerate the loss of S100A4 and differentiate normally in the myeloid lineage. However, knocking down expression of S100A4 in AML lines results in significant growth reduction of leukemic cells and cell death through induction of apoptosis. These findings provide an evidence that supports a novel role for S100A4 as a pro-survival factor in AML cell lines. Hence, these findings suggest that therapeutically targeting S100A4 would be an effective strategy in AML while sparing normal hematopoietic cells.

# Chapter 5

## Identification of S100A4 nuclear binding proteins in AML cells

## 5 Chapter 5 - Introduction

### 5.1 Overview

In the previous Chapter, I showed that knock down of S100A4 in AML cells impaired their ability to grow and survive, suggesting that AML cells are dependent on expression of this protein for their survival. Whilst it is well established that S100A4 has a role in cancer progression, the underlying mechanism in which S100A4 exerts its function is yet to be understood, especially in AML.

The biological functions of S100A4 are mediated by interactions with binding partner proteins through a  $\text{Ca}^{2+}$ -dependant manner (Donato *et al.* 2013). Interaction of S100A4 with  $\text{Ca}^{2+}$  ions is a critical process which leads to exposure of the hydrophobic surface of S100A4 to attract binding proteins (Bresnick *et al.* 2015); *in vitro* studies have shown that C-terminal EF hand (calcium binding domains) in S100 proteins has higher binding affinity to  $\text{Ca}^{2+}$  in the presence of its target ( $K_d \sim 10\text{-}100 \mu\text{M}$ ) (Gifford *et al.* 2007) (Figure 1-2). Although S100A4 protein binding partners such as RAGE and myosin have been identified in solid tumours (Bowers *et al.* 2012, Dahlmann *et al.* 2014), proteins that bind to S100A4 in the nucleus of haematological malignancies including AML remain unknown. One of the most widely used techniques to identify and validate physiologically relevant protein–protein interactions is co-immunoprecipitation (co-IP) (Turriziani *et al.* 2016). Co-IP uses specific antibodies to the target protein to indirectly pull down binding partners which can be subsequently identified by liquid chromatography mass spectrometry (LC/MS) (Lin and Lai 2017).

In this Chapter, a co-IP / MS approach was conducted to identify the binding partners of S100A4 in the cytoplasm and nucleus of AML cells. To replicate the biological conditions in which S100A4 binds proteins, S100A4 from both cytoplasmic and nuclear subcellular fractions had been pulled down under  $\text{Ca}^{2+}$  enriched conditions. These samples were assayed at the Bristol Proteomics Facility to identify a candidate list of potential S100A4 binding proteins. The co-IP/LCMS analysis of S100A4 binding partner in ME-1 cells and data analysis pipelines are detailed in Figure 2-8 and Figure 2-9 respectively.

## 5.2 Aims

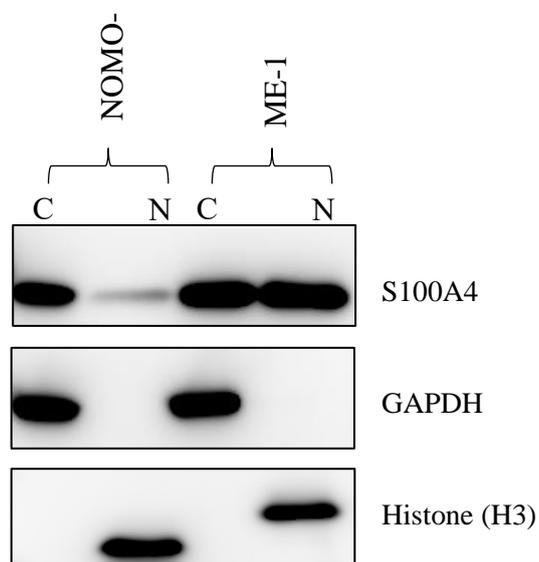
To identify the protein interaction partners of S100A4 in the ME-1 AML cell line. To achieve this objective, I aim to:

- Establish a co-IP protocol to pull down S100A4 binding partners under  $\text{Ca}^{2+}$  enriched conditions;
- Identify potential S100A4 binding proteins in the cytoplasm and nucleus of ME-1 using the optimised protocol established above coupled with LC/MS;
- Validate S100A4 binding proteins using a reciprocal co-IP approach coupled with western blot;
- Identify the over-represented gene ontology (GO) biological processes (BP) and molecular functions (MF) for S100A4 candidate binding proteins;
- Identify the protein-protein interaction (PPI) networks between S100A4 validated binding proteins and AML related proteins.

## 5.3 Results

### 5.3.1 Generating cytosolic and nuclear protein lysates from AML cell lines

In order to retain  $\text{Ca}^{2+}$ -based interactions between S100A4 and its protein binding partners, a buffer (TAEB) was used which was free of metal chelator (i.e. EDTA and EGTA) in order to retain calcium-dependent interactions. NOMO-1 and ME-1 cells were fractionated into cytosolic and nuclear sub-cellular fractions and GAPDH and Histone H3 protein expression were determined by western blot to verify fractionation efficiency and purity under these conditions (Figure 5-1) As before, S100A4 was expressed in both cytosol and nuclear sub-cellular fractions. Since NOMO-1 cells grow much faster than ME-1 cells, extracts from these lines were used for optimisation of S100A4 co-IP. However, given the increased nuclear expression of S100A4 in ME-1 cells, this line was used for subsequent LC-MS analyses using the optimal co-IP protocol.



**Figure 5-1: S100A4 is expressed in cytosolic and nuclear sub-cellular compartments of NOMO-1 and ME-1 cells.**

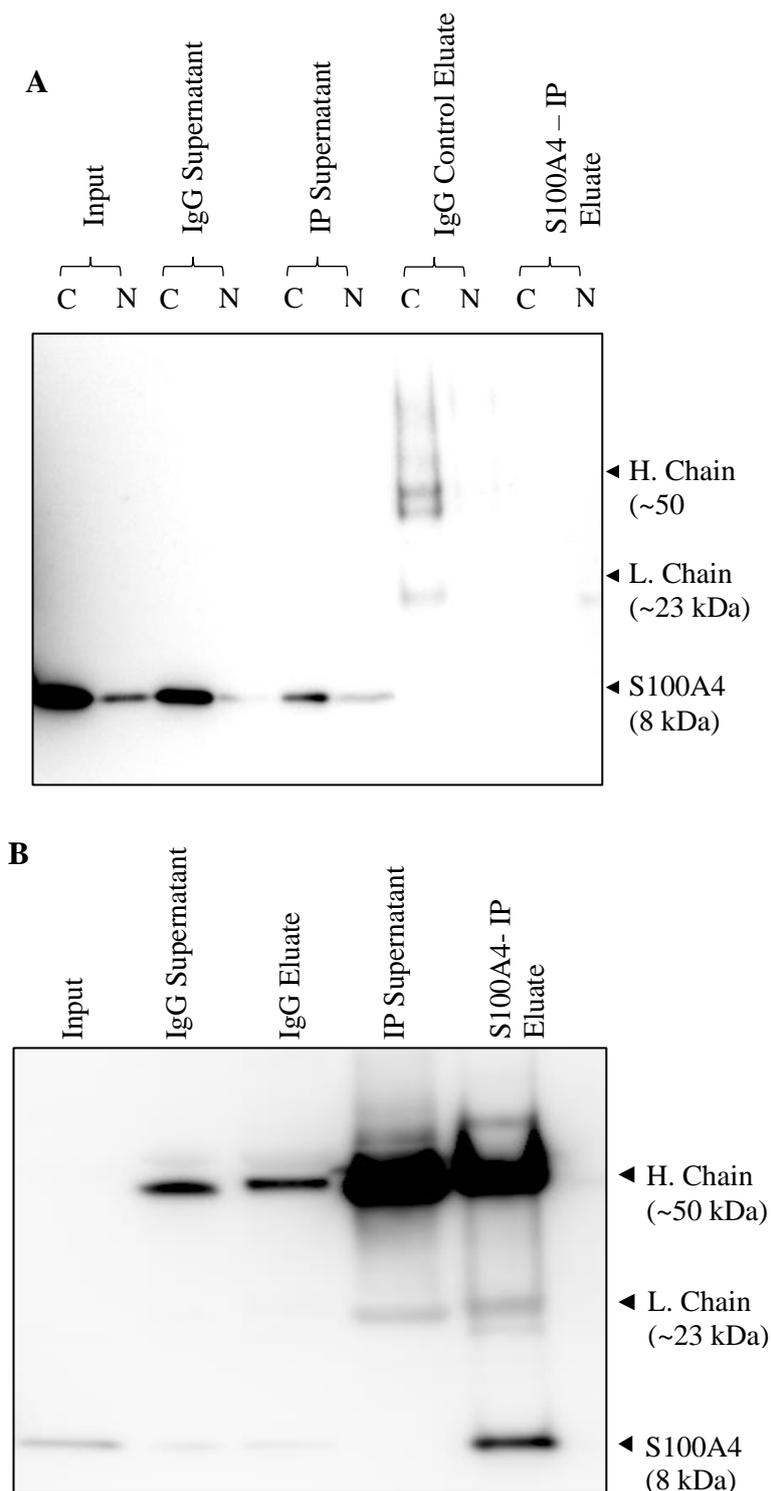
Example western blot showing S100A4 expression in NOMO-1 and ME-1 cells using non-ionic chelators containing extraction buffers (Cytoplasmic buffer and TEAB nuclear extraction buffer). (C) Cytoplasmic and (N) Nuclear subcellular compartments of NOMO-1 and ME-1 cells. GAPDH and Histone H3 were used to demonstrate fractionation efficiency.

### 5.3.2 Optimisation of co-IP of S100A4 and its binding partners complex in calcium enriched conditions for LC-MS analysis

Initially, a direct co-IP approach was used to pull down S100A4 binding proteins. The direct co-IP is based on pre-immobilization of the antibody to the magnetic beads followed by incubation with protein lysate mixture (Illustration diagrams are shown in Figure 2-6). As shown in Figure 5-2A, protein expression of S100A4 was determined by western blot following co-IP. Whilst S100A4 was detected in the original input lysates before conducting co-IP “input sample” and IgG co-IP (using NOMO-1 cells), S100A4 co-IP failed to show S100A4 protein expression in the IP eluted fraction.

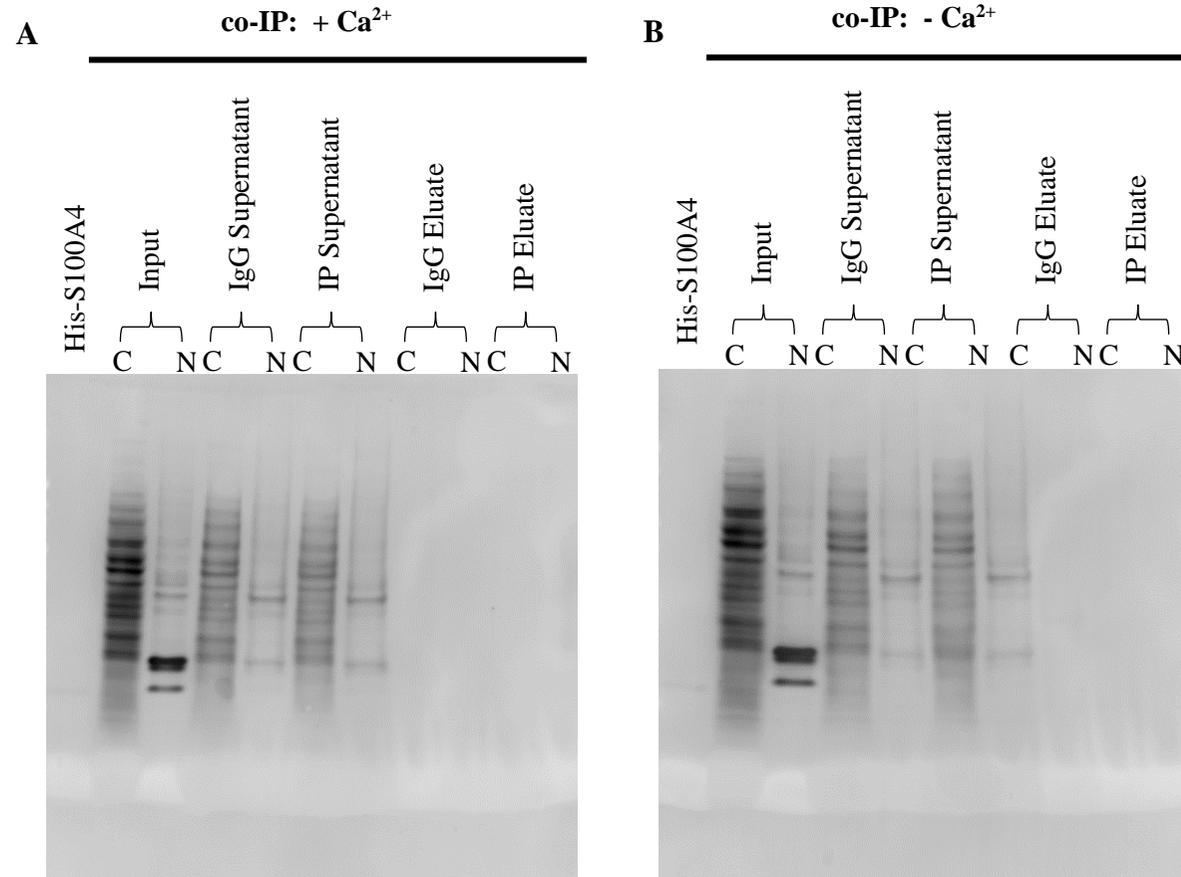
Alternatively, an indirect co-IP approach was undertaken where the antibody was incubated with a “protein mixture bait” to form an immune complex with the target protein antigen then retrieved by the magnetic beads. As shown in Figure 5-2B S100A4 protein expression was observed at 8 kDa in the input sample as well as the IP eluted sample. Further, upon eluting the co-IP fraction, the heavy (~50 kDa) and light (~23 kDa) chains of S100A4 mAb were visible suggesting elution was successful.

In order to retain S100A4 bound to protein partners in both cytosolic and nuclear lysates, co-IP was performed in the presence of 100  $\mu\text{M}$   $\text{CaCl}_2$ . A parallel co-IP using matched IgG control was performed under the same conditions to control for non-specific binding. Further, to control for  $\text{Ca}^{2+}$  independent interactions, an additional co-IP was performed in the presence of 10 mM  $\text{Na}_2\text{EDTA}$ . To quickly assess co-IP efficiency, protein complexes were subsequently separated using PAGE and proteins visualised within gel using SYPRO Ruby<sup>®</sup>. Immunoprecipitated S100A4 and partner proteins under both  $\text{Ca}^{2+}$  enriched and  $\text{Ca}^{2+}$  depleted conditions were not visible in the eluted fraction, due to the lower sensitivity of this method compared to immunoblotting (Figure 5-3A & B respectively). Taken together these data suggest that the indirect co-IP approach is more effective than the direct co-IP method in pulling down proteins.



**Figure 5-2: Immunoblots of direct and indirect S100A4 antibody binding co-IP of S100A4 binding partners.**

Immunoblots showing the efficiency of S100A4 co-IP using two different co-IP methods. A) Western blot of direct co-IP on both nuclear and cytosolic extracts of NOMO-1 cell line. B) Western blot of indirect co-IP on a cytosolic extract of NOMO-1 cell line. Input: original lysate before co-IP, IgG/IP Supernatant; unbound protein lysate, IgG/IP Eluate; eluted bound proteins. Each lane represents total lysate of NOMO-1 cell line. (H. Chain); Heavy IgG Chain (~50 kDa), (L. Chain); Light IgG Chain (~23 kDa).



**Figure 5-3: Quality Control check of co-IP on Tricine SDS-PAGE gel stained with SYPRO Ruby Fluorescent Stain.**

Quality control immunoblots stained with SYPRO Ruby fluorescent stain showing the co-IP efficiency under different enrichment conditions. A) PAGE gel of S100A4 co-IP under Ca<sup>+</sup> enriched conditions. B) Immunoblots of S100A4 co-IP under EDTA enriched conditions. (C) Cytoplasmic, (N) Nuclear. Input: original lysate before co-IP, IgG/IP supernatant; unbound protein lysate, IgG/IP Eluate; eluted bound proteins.

### 5.3.3 LC-MS Proteomic analysis and data filtration of S100A4 binding partners

#### 5.3.3.1 LC-MS Proteomic analysis of S100A4 binding partners in cytoplasmic and nuclear compartment of ME-1 cells

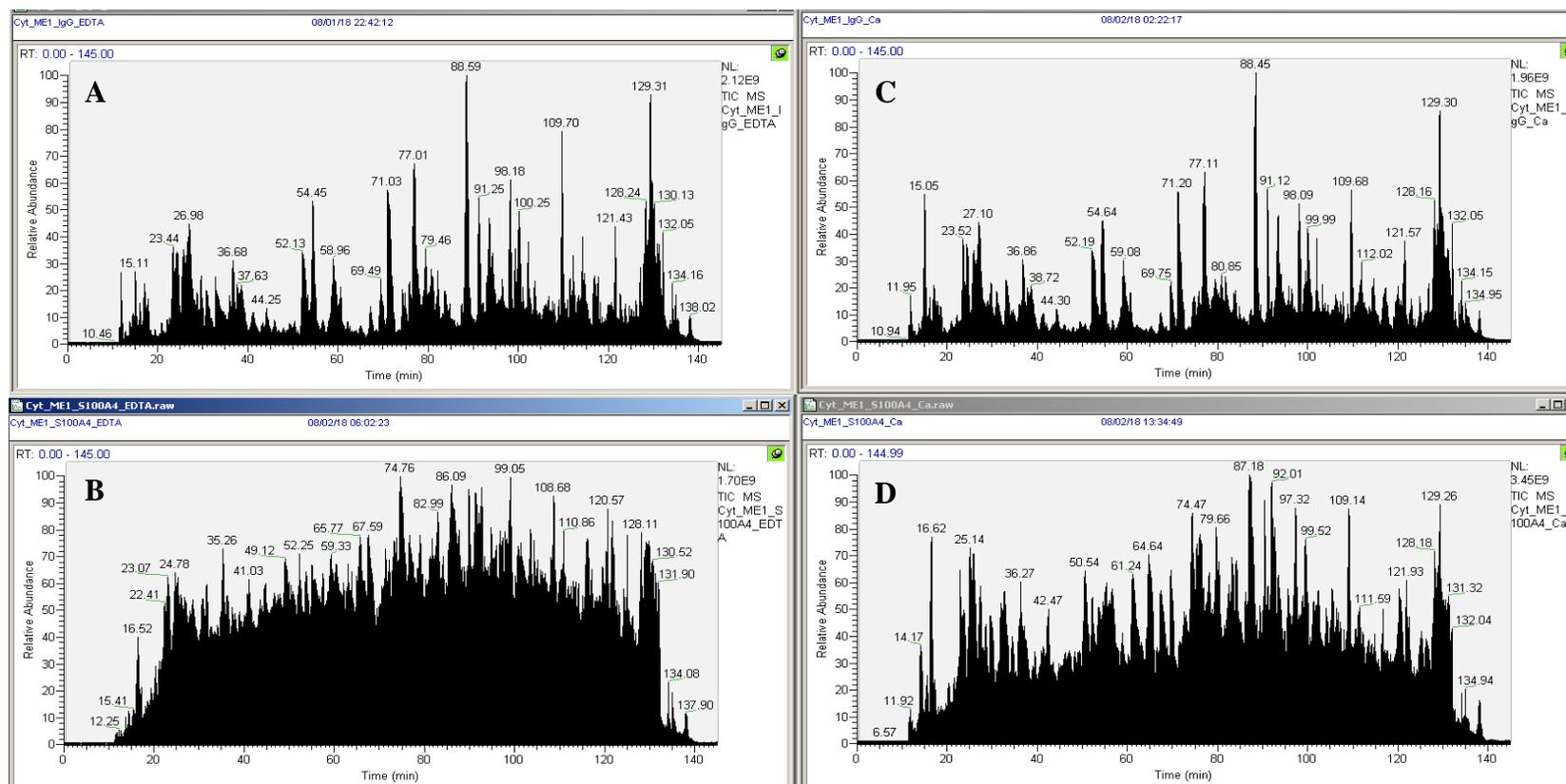
The indirect co-IP protocol established above was used to ‘pulldown’ proteins bound to S100A4 in the cytoplasm and nucleus of ME-1 AML cells. Each sample was subsequently analysed by LC-MS (*See full raw and analysed LCMS data in attached electronic CD*). In the cytoplasmic fractions, LC-MS analysis detected lower number of peptides as compared to nuclear counterparts (Table 5-1). LC-MS chromatograms show lower relative peptide abundance in the cytoplasmic IgG control co-IP samples under  $\text{Ca}^{2+}$  and EDTA enriched conditions (Figure 5-4A & C) compared to their S100A4 co-IP samples (Figure 5-4B & D) However, when spectra generated from these two samples (S100A4 co-IP/ $\text{Ca}^{2+}$  and S100A4 co-IP/EDTA) were searched against UniProt Human database, less peptides IDs were identified at 1% FDR. Further, matching the cytoplasmic peptides sequences against curated NCBI/taxonomy database resulted in more hits for BSA (Area Score:  $1.722 \times 10^{10}/\text{Ca}^{2+}$  and  $2.1 \times 10^9/\text{EDTA}$  and rabbit IgG (Area Score:  $1.71 \times 10^9/\text{EDTA}$  only) compared to other cytoplasmic peptides identified by MS analysis. On the contrary, nuclear MS data showed that higher number of peptide IDs were identified as shown in Table 5-1.

In summary, LC-MS analysis identified peptides IDs pulled down with S100A4 binding partners in cytoplasmic and nuclear co-IP at 95% confidence level at 1% FDR cut-off. These data will be filtered further according to the data filtration strategy shown in Figure 5-5 for cytoplasmic peptides and in Figure 5-6 for nuclear peptides.

**Table 5-1: Number of Peptides Identified by LC-MS in Cytoplasmic and Nuclear Co-IP Samples**

This table compares the number of peptides identified in cytoplasmic and nuclear fractions under calcium enriched and deprived conditions co-IP in S100A4 and IgG control.

% FDR →	1% FDR			
Enrichment →	Ca <sup>2+</sup>		EDTA	
Lysate	IgG Control	S100A4	IgG Control	S100A4
Cytoplasmic	82	19	142	25
Nuclear	191	681	222	696



**Figure 5-4: LC-MS Chromatograms Showing the Relative Abundance of Peptides Detected in Cytoplasmic Co-IP.**

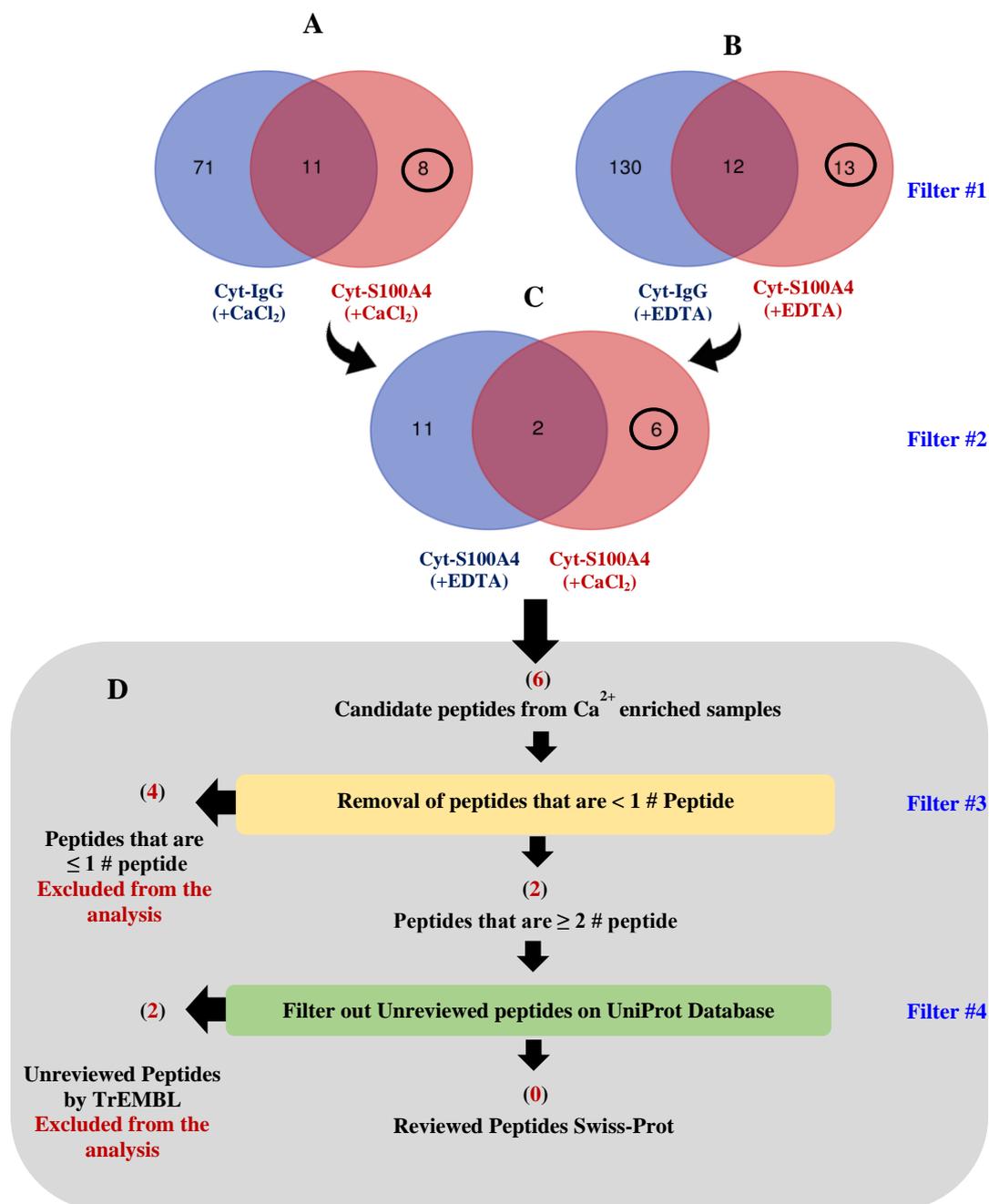
A) Example of LC-MS chromatogram of cytosolic IgG control co-IP in the presence of EDTA. B) Example of LC-MS chromatogram of cytosolic S100A4 co-IP in the presence of EDTA. C) Example of LC-MS chromatogram of cytosolic IgG co-IP under  $\text{Ca}^{2+}$  enriched condition. D) Example of LC-MS chromatogram of cytosolic S100A4 co-IP under  $\text{Ca}^{2+}$  enriched condition. Y-axis represents the relative abundance. X-axis; represents the run time in minutes.

### 5.3.3.2 LC-MS data filtration for cytoplasmic peptides

Following LC-MS analysis of potential S100A4 binding proteins in the cytoplasm of ME-1 cells, identified peptides were filtered and refined based on criteria outlined in (Figure 2-9). At a 1% FDR cut-off, a total of 19 peptides were detected in the S100A4 co-IP and 82 peptides were detected in the IgG control co-IP enriched with 100  $\mu$ M of  $\text{CaCl}_2$ . Of the 19 peptides detected in the S100A4 co-IP, 11 non-specific peptides were detected (present in the IgG control co-IP) and 8 specific (exclusive) proteins detected in S100A4 co-IP (Figure 5-5A).

To control for  $\text{Ca}^{2+}$ -independent interactions with S100A4, a similar co-IP was performed on cytoplasmic and nuclear lysates of ME-1 using calcium chelators  $\text{Na}_2\text{EDTA}$  enriched conditions. A total of 25 peptides were detected in the S100A4 co-IP samples while 142 peptides were detected in the IgG control sample. Thirteen specific peptides were ‘pulled down’ when non-specific binding proteins were filtered out (Figure 5-5B). Common to both approaches were two proteins (Figure 5-5C). Six proteins were specific to the  $\text{Ca}^{2+}$  enriched conditions. The candidate list of proteins was further refined by removing peptides that have only 1 distinct peptide sequence in the protein group from the list, leaving 2 candidate proteins only (Figure 5-5D, *Filter #3*). These proteins were matched against reviewed protein sequences in UniProt Knowledgebase (UniProtKB). Both proteins were identified as unreviewed proteins sequences on TrEMBL (Figure 5-5D, *Filter #4*). Therefore, these two proteins were excluded from the analysis.

Taken together, the above approach identified very few proteins in the cytosolic compartment of ME-1 as S100A4 binding partners (discussed further in 5.4.2).



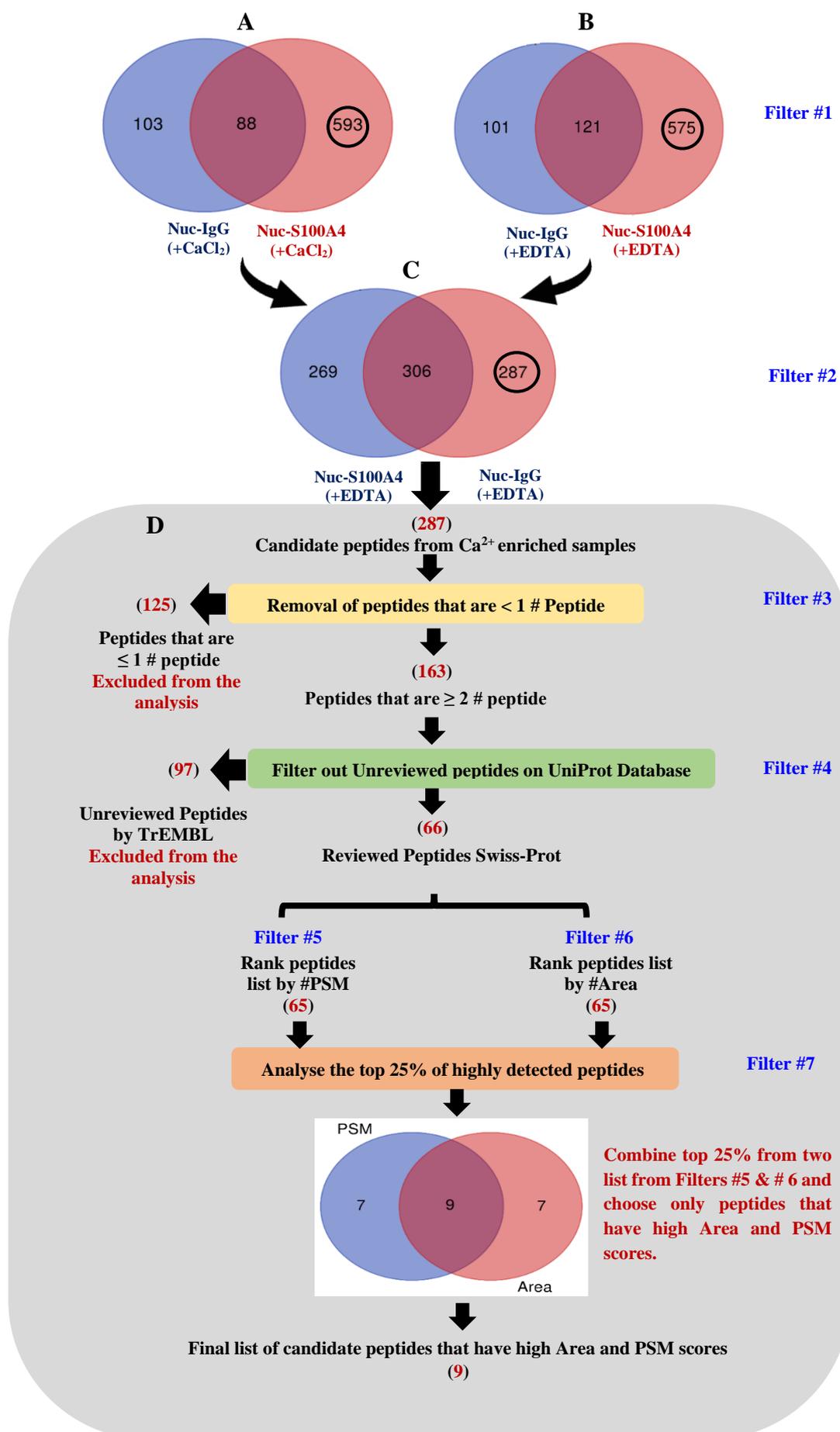
### Figure 5-5: LC-MS Preliminary Peptide filtration of cytoplasmic co-IP

A) Venn diagram showing filtering of non-specific peptides that appeared in IgG control from S100A4 calcium data set. B) Venn diagrams showing filtering of non-specific peptides that appeared in IgG control from S100A4 EDTA data set. C) Venn diagram showing filtering of peptides that were pulled down with S100A4 in the absence of calcium from S100A4 calcium data set. Circled numbers in red Venn diagram depict the number of peptides after filtration. D) Further filters of cytoplasmic peptides identified by LC-MS using spectra algorithmic scores including peptides score, PSM (displays the protein abundance in the sample), and Area (displays the highest detected peptide signal) as well as online peptide identification tools (UniProt-Human). Numbers in red in brackets depict the number of peptides resulting after each filter.

### 5.3.3.3 LC-MS data filtration for nuclear peptides

In the nuclear sub-cellular compartment of ME-1 cells, LC-MS analysis identified a total of 681 peptides in the S100A4 co-IP enriched with  $\text{Ca}^{2+}$ . Further, a total of 191 peptides were detected in the IgG control lysate under the same conditions. As shown in Figure 5-6A, non-specific binding peptides detected in the IgG control co-IP were removed from the nuclear calcium enriched fraction. Alternatively, a total of 696 peptides were detected in the nuclear lysate that had been treated with EDTA. In this approach, 222 peptides were detected in the IgG control co-IP sample under the same conditions. As shown in Figure 5-6B, non-specific binding peptides detected in the IgG control co-IP sample were removed from the nuclear EDTA enriched co-IP. To identify S100A4 binding peptides in a  $\text{Ca}^{2+}$ -dependent manner, peptides that are present in the EDTA data list were removed from the  $\text{Ca}^{2+}$  enriched sample remaining 287 peptides in  $\text{Ca}^{2+}$  enriched sample (Figure 5-6C). Subsequently, candidate peptides were refined based on their peptide score, in which peptides that have  $< 1$  peptide score were removed from the list leaving 163 peptides (Figure 5-6 D, *Filter #3*).

Using UniProtKB, the list of 163 candidate peptides were matched against reviewed peptides sequences. A total of 97 unreviewed peptides on TrEMBL/UniProt were removed from the list (Figure 5-6 D, *Filter #4*). The previous step resulted in identifying 66 manually reviewed peptides sequences on the Swiss-Prot/UniProt database and can be processed in the next step in the analysis (Table 5-2). However, Keratin, type II cytoskeletal 74 (KRT74) was removed as it has been recognised in the literature as major background contaminant (discussed in 5.4.2). Candidate list of 65 peptides were ranked twice, first rank is based on highest PSM score (displays the protein abundance in the sample) and second rank is based on Area score (displays the highest detected peptide signal). Subsequently, analysed peptides that are in the highest quartile of both data sets. This filtration step resulted in two sets of 16 proteins with highest PSM score and 16 peptides with highest area score (Figure 5-6 D, *Filter #5* and *#6*). In the final filtration step, ranked data sets from previous filtration step were combined and only 9 peptides which have both the highest PSM and area scores (Table 5-3) were chosen for further experimental validation, (Figure 5-6 D, *Filter #7*).



**Figure 5-6: LC-MS preliminary peptide filtration of nuclear co-IP.**

**A)** Venn Diagrams showing filtering non-specific peptides appeared in IgG control from S100A4 calcium data set. **B)** Venn Diagrams showing filtering non-specific peptides appeared in IgG control from S100A4 EDTA data set. **C)** Venn Diagrams showing filtering peptides that are pulled down with S100A4 in the absence of calcium from S100A4 calcium data set. Circled number in red Venn diagram depicted the number of peptides after filtration. **D)** Further refinements of Cytoplasmic peptides identified by LC-MS using spectra algorithmic scores including peptides score, PSM, (displays the protein abundance in the sample), and Area (displays the highest detected peptide signal) as well as online peptide identification tools (UniProt-Human). Numbers in red colour in brackets depicted the number of peptides resulted after each filter.

**Table 5-2: Identification of reviewed proteins on UniProt KB database in the S100A4 co-IP sample.**

List of 65 nuclear proteins identified as reviewed proteins in Swiss-Prot/UniProt KB database in the S100A4 co-IP sample enriched with calcium. These proteins were further refined as indicated in **Figure 5-6**.

<b>UniProt Accession No</b>	<b>HUGO Gene Name (Symbol)</b>	<b>Candidate Protein Name</b>
<b>O75533</b>	SF3B1	Splicing Factor 3B subunit 1
<b>P52272</b>	hnRNPM	heterogeneous nuclear Ribonucleoprotein M
<b>P33993</b>	MCM7	Minichromosome Maintenance complex component 7
<b>Q9NR30</b>	DDX21	DExD-box helicase 21
<b>P42167</b>	TMPO	Thymopoietin
<b>Q8WUM0</b>	NUP133	Nucleoporin 133
<b>Q13428</b>	TCOF1	Treacle Ribosome Biogenesis Factor 1
<b>Q9BW27</b>	NUP85	Nucleoporin 85
<b>Q9Y5B9</b>	SUPT16H	SPT16 homolog, Facilitates Chromatin Remodeling Subunit
<b>P29590</b>	PML	Promyelocytic Leukemia
<b>Q14684</b>	RRP1B	Ribosomal RNA Processing 1B
<b>Q8TAQ2</b>	SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2
<b>Q8IY81</b>	FTSJ3	FtsJ RNA 2'-O-methyltransferase 3

<b>Q6KC79</b>	NIPBL	NIPBL cohesin loading Factor
<b>Q96QV6</b>	HIST1H2AA	Histone Cluster 1 H2A Family Member A
<b>P55265</b>	ADAR	Adenosine Deaminase RNA Specific
<b>Q86WJ1</b>	CHD1L	Chromodomain Helicase DNA Binding protein 1 Like
<b>P46087</b>	NOP2	NOP2 Nucleolar Protein
<b>Q9BXY0</b>	MAK16	MAK16 homolog
<b>P46781</b>	RPS9	Ribosomal Protein S9
<b>P12270</b>	TPR	Translocated Promoter Region, Nuclear Basket Protein
<b>Q9ULK4</b>	MED23	Mediator Complex Subunit 23
<b>Q9NY61</b>	AATF	Apoptosis Antagonizing Transcription Factor
<b>P46060</b>	RANGAP1	Ran GTPase Activating Protein 1
<b>Q9BTA9</b>	WAC	WW domain Containing Adaptor with Coiled- Coil
<b>P07437</b>	TUBB	Tubulin Beta Class I
<b>P68371</b>	TUBB4B	Tubulin Beta 4B Class IVB
<b>Q13242</b>	SRSF9	Serine and Arginine Rich Splicing Factor 9
<b>P39023</b>	RPL3	Ribosomal Protein L3
<b>Q9UQR0</b>	SCML2	Scm Polycomb Group Protein-Like 2
<b>P22626</b>	hnRNPA2B1	heterogeneous nuclear Ribonucleoprotein A2/B1
<b>Q86UE4</b>	MTDH	Metadherin
<b>Q99848</b>	EBNA1BP2	EBNA1 Binding Protein 2
<b>P55081</b>	MFAP1	Microfibril Associated Protein 1
<b>P35659</b>	DEK	DEK proto-oncogene
<b>P62277</b>	RPS13	Ribosomal Protein S13
<b>P38159</b>	RBMX	RNA Binding Motif Protein X-linked
<b>Q92769</b>	HDAC2	Histone Deacetylase 2
<b>Q14669</b>	TRIP12	Thyroid Hormone Receptor Interactor 12
<b>Q96DI7</b>	SNRNP40	Small Nuclear Ribonucleoprotein U5 Subunit 40
<b>P17026</b>	ZNF22	zinc finger protein 22

<b>Q5QJE6</b>	DNTTIP2	Deoxynucleotidyl Transferase Terminal Interacting Protein 2
<b>Q14692</b>	BMS1	BMS1 Ribosome Biogenesis Factor
<b>P53999</b>	SUB1	SUB1 homolog, transcriptional regulator
<b>P55735</b>	SEC13	SEC13 homolog, nuclear pore and COPII coat complex component
<b>O60216</b>	RAD21	RAD21 cohesin complex component
<b>Q96PK6</b>	RBM14	RNA Binding Motif Protein 14
<b>Q8TEM1</b>	NUP210	Nucleoporin 210
<b>Q13595</b>	TRA2A	Transformer 2Alpha homolog
<b>Q92785</b>	DPF2	Double PHD Fingers 2
<b>Q8IY67</b>	RAVER1	Ribonucleoprotein, PTB binding 1
<b>Q13619</b>	CUL4A	Cullin 4A
<b>Q9UBW7</b>	ZMYM2	Zinc finger MYM-type containing 2
<b>Q6UN15</b>	FIP1L1	Factor Interacting with PAPOLA and CPSF1
<b>P25705</b>	ATP5F1A	ATP synthase F1 subunit alpha
<b>Q9NRH3</b>	TUBG2	Tubulin Gamma 2
<b>Q14344</b>	GNA13	G Protein subunit Alpha 13
<b>Q96MU7</b>	YTHDC1	YTH Domain Containing 1
<b>Q8NEJ9</b>	NGDN	Neuroguidin
<b>Q9Y232</b>	CDYL	Chromodomain Y Like
<b>P62263</b>	RPS14	Ribosomal Protein S14
<b>Q8N7H5</b>	PAF1	PAF1 homolog, Paf1/RNA polymerase II complex component
<b>Q13315</b>	ATM	ATM serine/threonine kinase
<b>Q92556</b>	ELMO1	Engulfment and cell Motility 1
<b>P49756</b>	RBM25	RNA Binding Motif Protein 25

**Table 5-3: S100A4 nuclear candidate binding partners identified by LC-MS**

A total of 9 proteins were concluded as S100A4 candidate binding partners (Figure 5-6) with highest abundance (High PSM scores) and highly detected peptide signal (high Area score) in MS list of S100A4 candidate binding partners under calcium enriched conditions. PSM score (displays the protein abundance in the sample), Area score (displays the highest detected peptide signal).

UniProt Accession No	HUGO Gene Name (Symbol)	Candidate Protein Name	Function	MS Area Score	MS PSM Score	Reference
P52272	hnRNPM	heterogeneous nuclear ribonucleoprotein M	Involved in pre-mRNA binding, Splicing, mRNA Transcription	$1.34 \times 10^9$	65	Geuens <i>et al.</i> 2016
Q96QV6	HIST1H2AA	histone cluster 1 H2A family member A	Involved in transcription regulation, DNA repair, DNA replication and chromosomal stability	$1.34 \times 10^9$	33	McGinty and Tan 2015
P42167	TMPO	Thymopoietin	Involved in the structural organization of the nucleus and in the post-mitotic nuclear assembly	$4.14 \times 10^8$	29	Naetar <i>et al.</i> 2017
O75533	SF3B1	Splicing Factor 3B subunit 1	pre-mRNA splicing Factor	$1.81 \times 10^8$	40	Dolatshad <i>et al.</i> 2015
P33993	MCM7	MiniChromosome Maintenance complex component 7	putative replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells	$1.4 \times 10^8$	29	Ishimi 2018
Q14684	RRP1B	Ribosomal RNA processing protein 1 homolog B	Positively regulates DNA damage-induced apoptosis by acting as a transcriptional coactivator of proapoptotic target genes of the transcriptional activator E2F1	$1.26 \times 10^8$	16	Lee <i>et al.</i> 2014
Q9BXY0	MAK16	MAK16 homolog	Part of the apparatus concerned with the nuclear events of the cell cycle	$1.22 \times 10^8$	9	Pellett and Tracy 2006
Q9Y5B9	SUPT16H	SPT16 homolog facilitates chromatin remodelling subunit (FACT complex)	Component of the FACT complex and acts to reorganize nucleosomes.	$1.19 \times 10^8$	17	Formosa 2013
Q9NR30	DDX21	DExD-box helicase 21	RNA helicase that acts as a sensor of the transcriptional status of both RNA polymerase (Pol) I and II	$9.78 \times 10^7$	22	Calo <i>et al.</i> 2015

Among the list of 65 proteins in Table 5-2, there were several proteins that have been implicated to have a role in AML pathogenesis including HDAC2, PML, DEK, RBM25, RAD21, and ATM (Table 5-4). Four of these proteins DEK, RBM25, RAD21, and ATM, were excluded from subsequent analysis due to their low abundance (PSM, *Filter #5*) and low signal intensity (Area, *Filter #6*) scores, Figure 5-6. The remaining proteins (HDAC2 and PML) were initially prioritised into the top quartile of proteins list for either high abundance or high signal intensity scores. However, after applying *Filter #7* (Figure 5-6) HDAC2 and PML were excluded from the prioritized list due to having a low score in either PSM or Area scores. Despite the systematic data filtration exclusion of these AML related proteins, pulling down these proteins with S100A4 in Ca<sup>2+</sup> enriched conditions from nuclear lysate may be of biological relevance.

In summary, 9 nuclear candidate proteins were identified as S100A4 binding partners candidates. Subsequently, these proteins' interaction with S100A4 will be validated experimentally. In this study, proteins pulled down under Ca<sup>2+</sup> enriched conditions but identified as known abnormalities, will be considered as S100A4 binding partners in AML and will be included in the subsequent functional analysis with S100A4 binding partners in 5.3.6.

#### **5.3.4 Validation of S100A4 nuclear binding proteins using two-step co-IP and western blot**

The data above identified 9 proteins that were potentially bound to S100A4. To confirm this, a two-step validation approach using ME-1 cells was applied. Firstly, a co-IP was performed where S100A4 was 'pulled down' in the presence of Ca<sup>2+</sup> (controlled using an IgG control co-IP) followed by western blot of the candidate binding partner proteins. Secondly, co-IP using the reciprocal candidate proteins and a subsequent western blot of S100A4 was then performed. Of the 9 proteins identified, only 6 commercial antibodies that are suitable for immunoprecipitation were available to test; hnRNPM, HIST1H2AA, MCM7, DDX21, SUPT16H, and MAK16.

**Table 5-4: AML proteins have been detected in S100A4 in Ca<sup>2+</sup> enriched co-IP .**

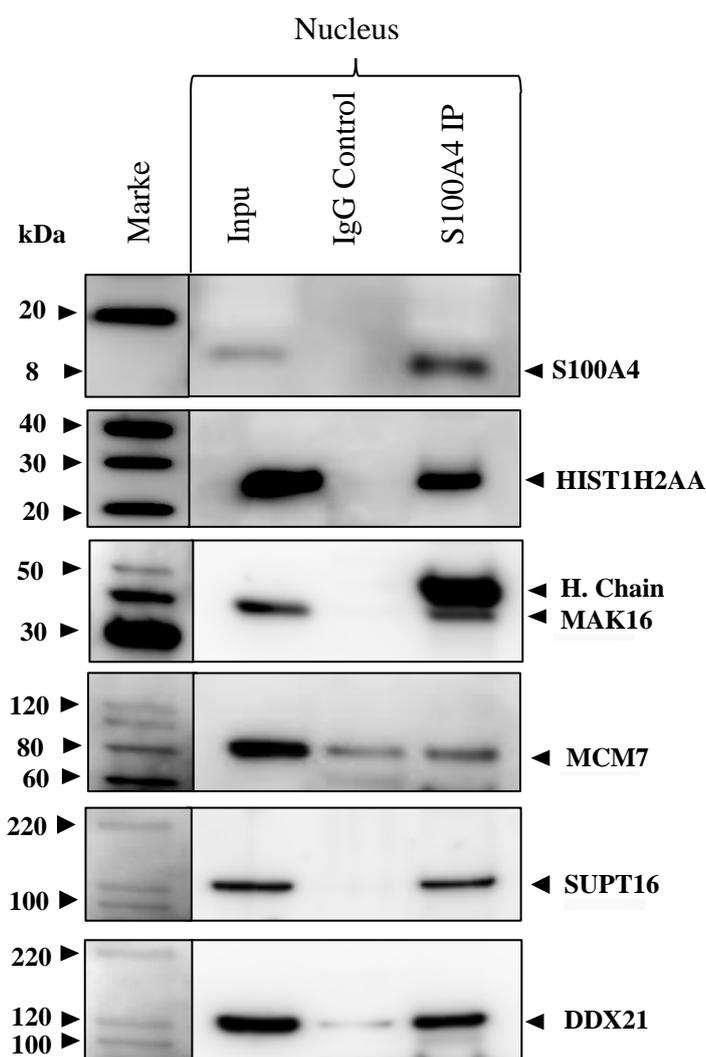
A total of 6 proteins known in AML bound to S100A4 in the calcium enriched co-IP. PSM score (displays the protein abundance in the sample), Area score (displays the highest detected peptide signal). ↑ Upward green arrow denotes that this protein has a high PSM/Area score ↓ downward red arrow denotes that this protein has a low PSM/Area score.

UniProt Accession No	HUGO Gene Symbol	Protein Name	Function	MS Area Score	MS PSM Score	Reference
<b>Q92769</b>	HDAC2	Histone Deacetylase 2	Involved in histone deacetylation	↑ 1.01×10 <sup>8</sup>	↓ 5	Yang <i>et al.</i> 2015
<b>P29590</b>	PML	Promyelocytic Leukemia	Involved in tumour suppression, transcriptional regulation, apoptosis, and DNA damage response	↓ 7.67×10 <sup>7</sup>	↑ 15	Hsu and Kao 2018
<b>P35659</b>	DEK	DEK proto-oncogene	Involved in chromatin organization, epigenetic and transcriptional regulation	↓ 4.87×10 <sup>7</sup>	↓ 6	Sanden and Gullberg 2015
<b>P49756</b>	RBM25	RNA-Binding Motif protein 25	Involved in regulation of alternative pre-mRNA splicing and in regulation of the apoptotic factor BCL2L1 isoform expression	↓ 3.1×10 <sup>7</sup>	↓ 4	Carlson <i>et al.</i> 2017
<b>O60216</b>	RAD21	Double-strand-break repair protein RAD21	Involved in sister chromatid cohesion and post-replicative DNA repair.	↓ 2.72×10 <sup>7</sup>	↓ 4	Rocquain <i>et al.</i> 2010;
<b>Q13315</b>	ATM	Ataxia Telangiectasia Mutated (Serine-protein kinase)	Acts as a DNA damage sensor activating checkpoint signalling upon double strand breaks (DSBs)	↓ 6.93×10 <sup>6</sup>	↓ 2	Paull 2015; Stein <i>et al.</i> 2018

Using the former approach, as expected S100A4 was ‘pulled down’ (Figure 5-7). Subsequent western blot reprobing demonstrated protein expression of HIST1H2AA (23 kDa), Mak16 (48 kDa), SUPT16H (120 kDa) and DDX21 (115 kDa). However, in case of MCM7 (80 kDa) and hnRNPM (65kDa) I was unable to confirm that these proteins were pulled down with S100A4 as no difference between the IgG and S100 lanes was observed (Figure 5-7 and Figure 5-8 respectively). To provide further evidence that the 6 candidate proteins (MCM7, DDX21, MAK16, SUPT16H, and hnRNPM, HIST1H2AA) are bound to S100A4, a reciprocal co-IP was performed in ME-1 nuclear protein lysates. Candidate proteins were pulled down individually and lysates analysed for S100A4 protein expression by western blot.

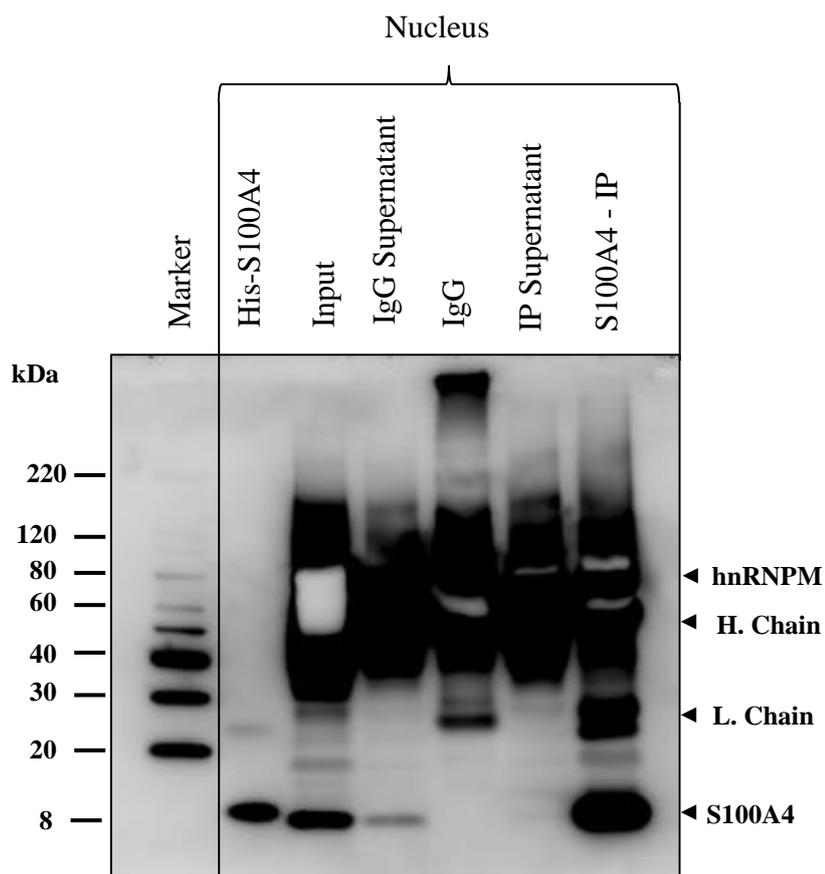
In the input sample for “MCM7” (Figure 5-9A), protein expression of S100A4 and MCM7 (80 kDa) was observed. Following co-IP using anti-MCM7 (IP eluted), there was an enrichment of MCM7 protein expression compared to IgG co-IP. However, there was no detectable protein expression for S100A4. A similar observation was observed for the reciprocal co-IP of DDX21 and MAK16 at their expected molecular weights (115 and 48 kDa respectively) (Figure 5-9B & C). In contrast the co-IP for “SUPT16H” failed to detect SUPT16H expression, nor was there detectable protein expression of S100A4 when assayed by western blot (Figure 5-9D).

When pulling down HIST1H2AA, protein expression of S100A4 and HIST1H2AA was detected at their expected molecular weights of 8 kDa and 23 kDa respectively in the eluted co-IP sample (Figure 5-10A). A similar observation was noted for the reciprocal co-IP of hnRNPM; with hnRNPM protein expression observed at 65 kDa (Figure 5-10B). Although the immunoblots of HIST1H2AA and hnRNPM show that the bands are oversaturated at 30 seconds of exposure, S100A4 can be easily visible on the blot. In summary, S100A4 was pulled down reciprocally with only two binding partners (HIST1H2AA and hnRNPM).



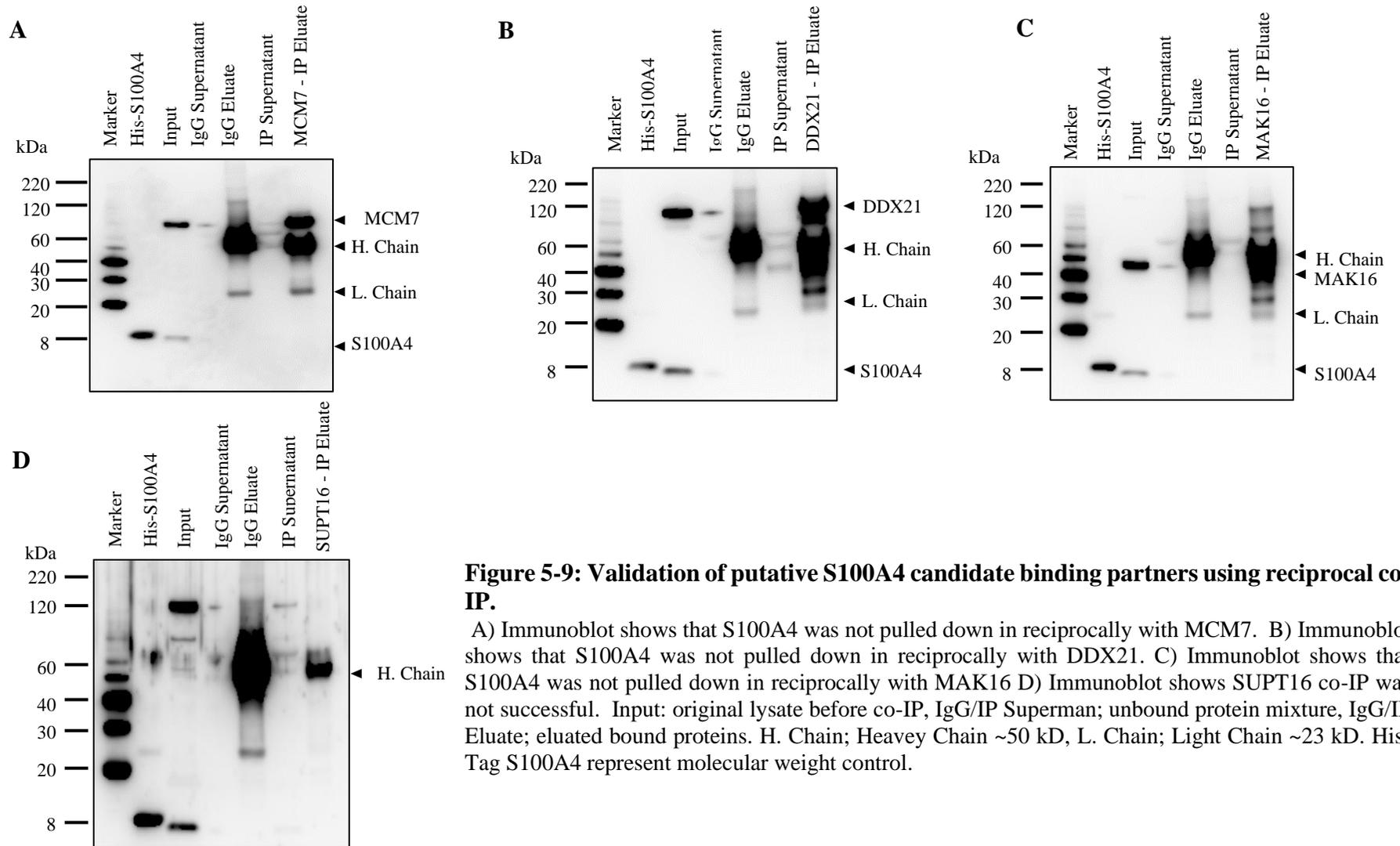
**Figure 5-7: S100A4 candidate binding partners are validated using co-IP.**

Western blots of S100A4 co-IP from the nuclear lysate of ME-1. All immunoblots were assayed for S100A4 expression and the candidate binding partner simultaneously; S100A4 (8 kDa); HIST1H2AA (23 kDa); MAK16 (48 kDa); MCM7 (80 kDa); SUPT16 (120 kDa); DDX21 (115 kDa). Input: original lysate before co-IP, IgG control; negative co-IP control, S100A4 co-IP; eluted bound proteins.



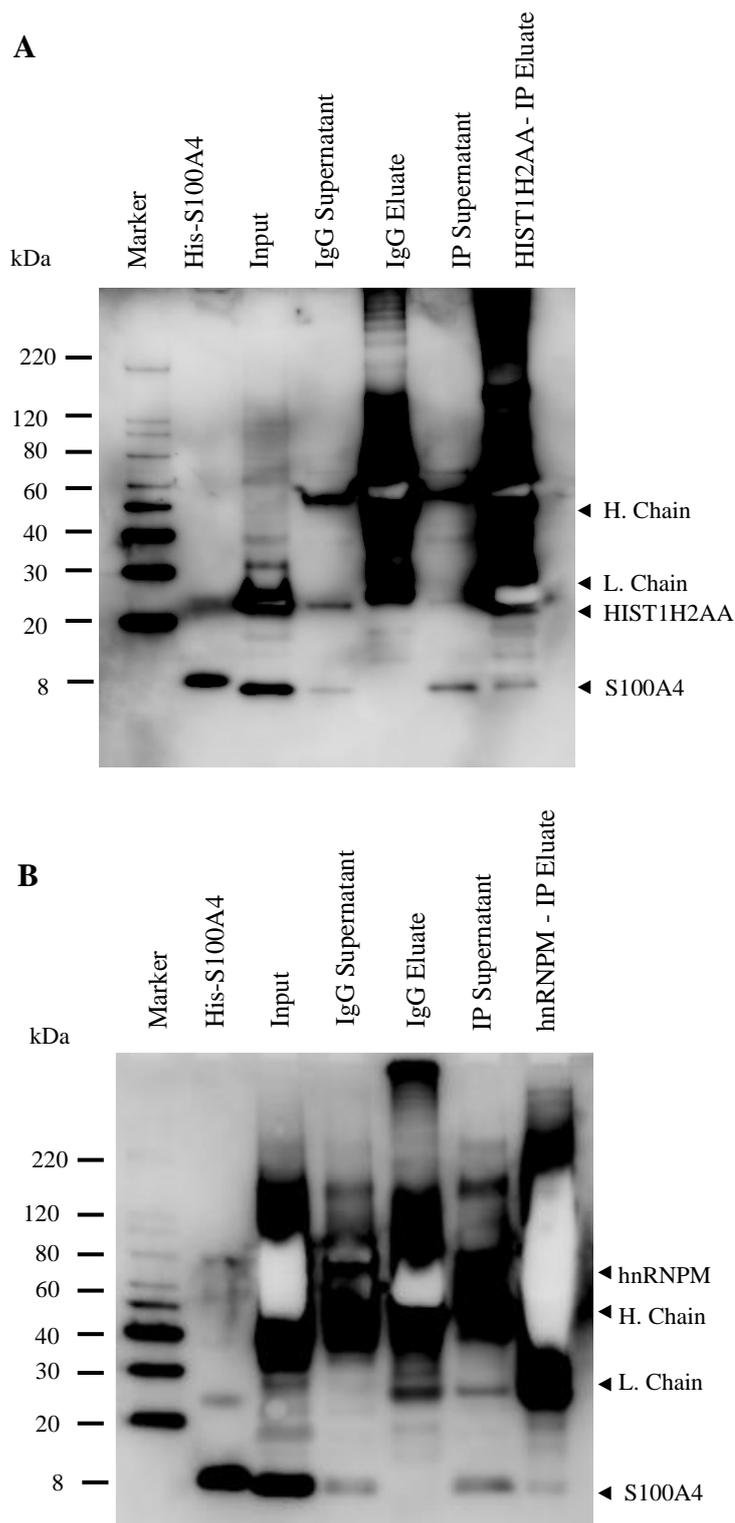
**Figure 5-8:hnRNPM is validated as a candidate binding partner of S100A4 using co-IP.**

Western blot of S100A4 co-IP from the nuclear lysate of ME-1 probed with S100A4 Ab and hnRNP Ab. S100A4 is detected at 8 kDa. Input: original lysate before co-IP, IgG/IP Supernatant; unbound protein mixture, IgG/IP Eluate; eluted bound proteins. H. Chain; Heavy Chain ~50 kDa, L. Chain; Light Chain ~23 kDa.



**Figure 5-9: Validation of putative S100A4 candidate binding partners using reciprocal co-IP.**

A) Immunoblot shows that S100A4 was not pulled down in reciprocally with MCM7. B) Immunoblot shows that S100A4 was not pulled down in reciprocally with DDX21. C) Immunoblot shows that S100A4 was not pulled down in reciprocally with MAK16 D) Immunoblot shows SUPT16 co-IP was not successful. Input: original lysate before co-IP, IgG/IP Superman; unbound protein mixture, IgG/IP Eluate; eluated bound proteins. H. Chain; Heavey Chain ~50 kD, L. Chain; Light Chain ~23 kD. His-Tag S100A4 represent molecular weight control.



**Figure 5-10: HIST1H2AA and hnRNPM are validated as S100A4 binding partners using reciprocal co-IP.**

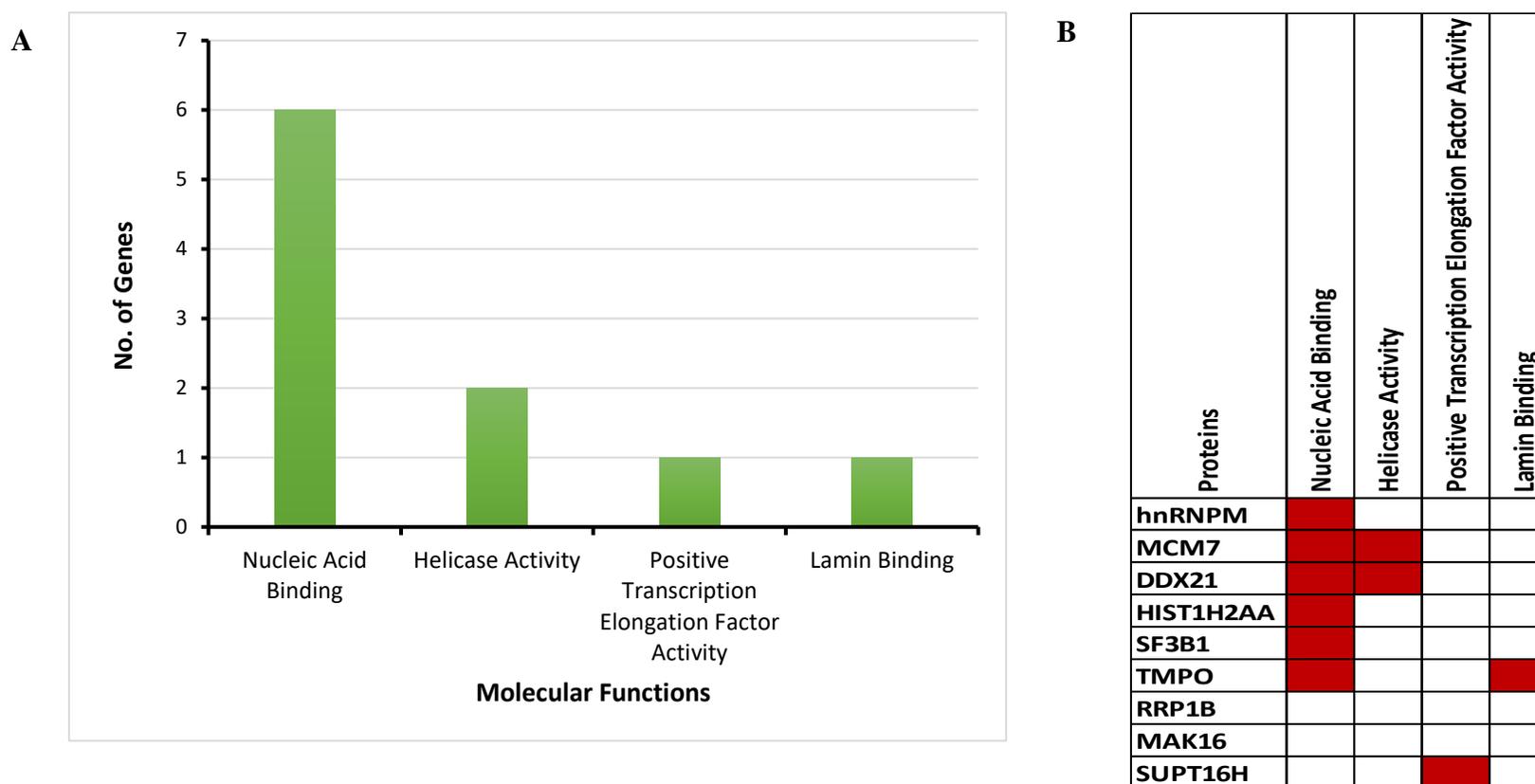
S100A4 is reciprocally detected in binding partner co-IP. A) Immunoblot shows that S100A4 was pulled down reciprocally with HIST1H2AA. B) Immunoblot shows that S100A4 was pulled down reciprocally with hnRNPM. Input: original lysate before co-IP, IgG/IP Supernatant; unbound protein mixture, IgG/IP Eluate; eluted bound proteins. H. Chain; Heavy Chain ~50 kD, L. Chain; Light Chain ~23 kD. His-Tag S100A4 represent molecular weight control. Image exposure time: 30 seconds

### 5.3.5 Functional analysis of S100A4 final binding partners candidates

To gain more insight on how S100A4 and its binding proteins could be linked AML pathophysiology, Gene Ontology (GO) enrichment analysis for “Biological Process” (BP) and “Molecular Function” (MF) were applied. The GO analysis was performed using PANTHER Over-representation Test Gene Ontology Consortium online tool. Initially, the nine candidate binding partner proteins of S100A4 were searched against the reference list of 65 proteins pulled down with S100A4 in presence of  $Ca^{2+}$  (Table 5-2) for BP and MF enrichment analysis.

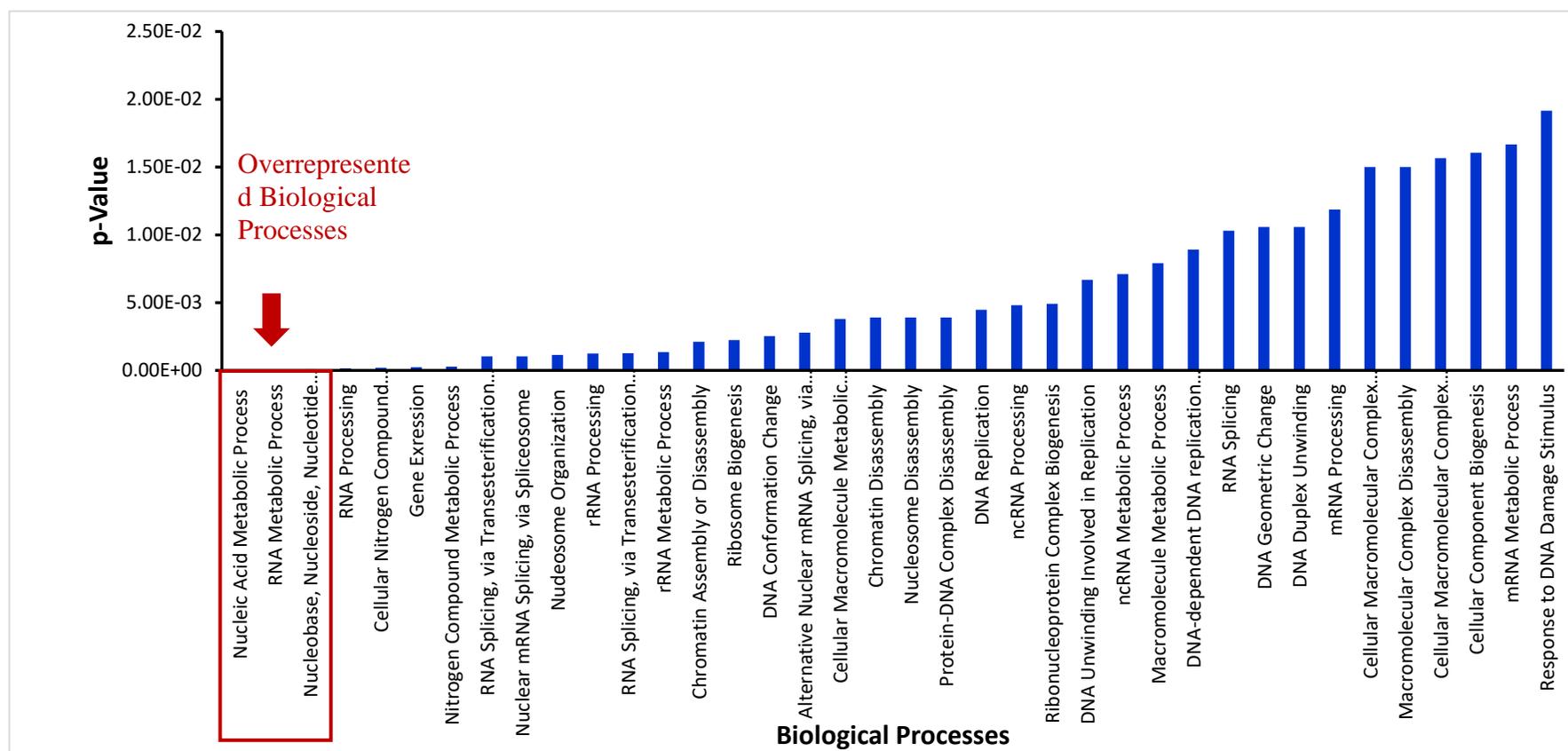
Using the MF enrichment analysis, “Nucleic Acid Binding” (in which hnRNPM, HIST1H2AA, SF3B, MCM7, TMPO, and DDX21 were identified) was the most over-represented BP (Figure 5-12). Alternatively, GO analysis identified several processes with the three most over-represented BPs involving “Nucleic Acid Metabolic Process”, “RNA Metabolic Process”, and “Nucleobase, Nucleoside, Nucleotide and Nucleic Acid Metabolic Process”; in which SUPT16H, MCM7, RRP1B, DDX21, hnRNPM, and SF3B1 were involved (Figure 5-12). Interestingly, hnRNPM and SF3B1 were observed in all nucleic acids related BPs and RNA related BPs reported in Figure 5-12 such as RNA splicing, RNA processing suggesting strong protein-protein interaction (PPI) between them.

Taken together, GO functional enrichment analysis demonstrated the involvement of validated S100A4 binding partners’ hnRNPM and HIST1H2AA in the overrepresented molecular functions and biological processes analysed.



**Figure 5-11: Functional analysis using molecular function enrichment for the final list of A100A4 binding partners' candidates.**

**A)** Gene Ontology (GO) Enrichment analysis using Molecular Function (MF) shows that “Nucleic Acid Binding” is the most overrepresented molecular function for S100A4 binding partners final candidate list. Functional representation is shown in the number of genes identified for each molecular function on the Y-axis of the graph. **B)** MF/Protein set overlap matrix showing the number of proteins observed in each MF. The GO/MF analysis was performed using PANTHER14.1 and searched against the reference list of 65 proteins pulled down with S100A4 in presence of  $\text{Ca}^{2+}$  (Table 5-2). PANTHER14.1 is generated from the 2019\_04 release of [Reference Proteome dataset](#) (based on UniProt Release 2019\_04, Ensembl release 95 and Ensembl Genome release 42).



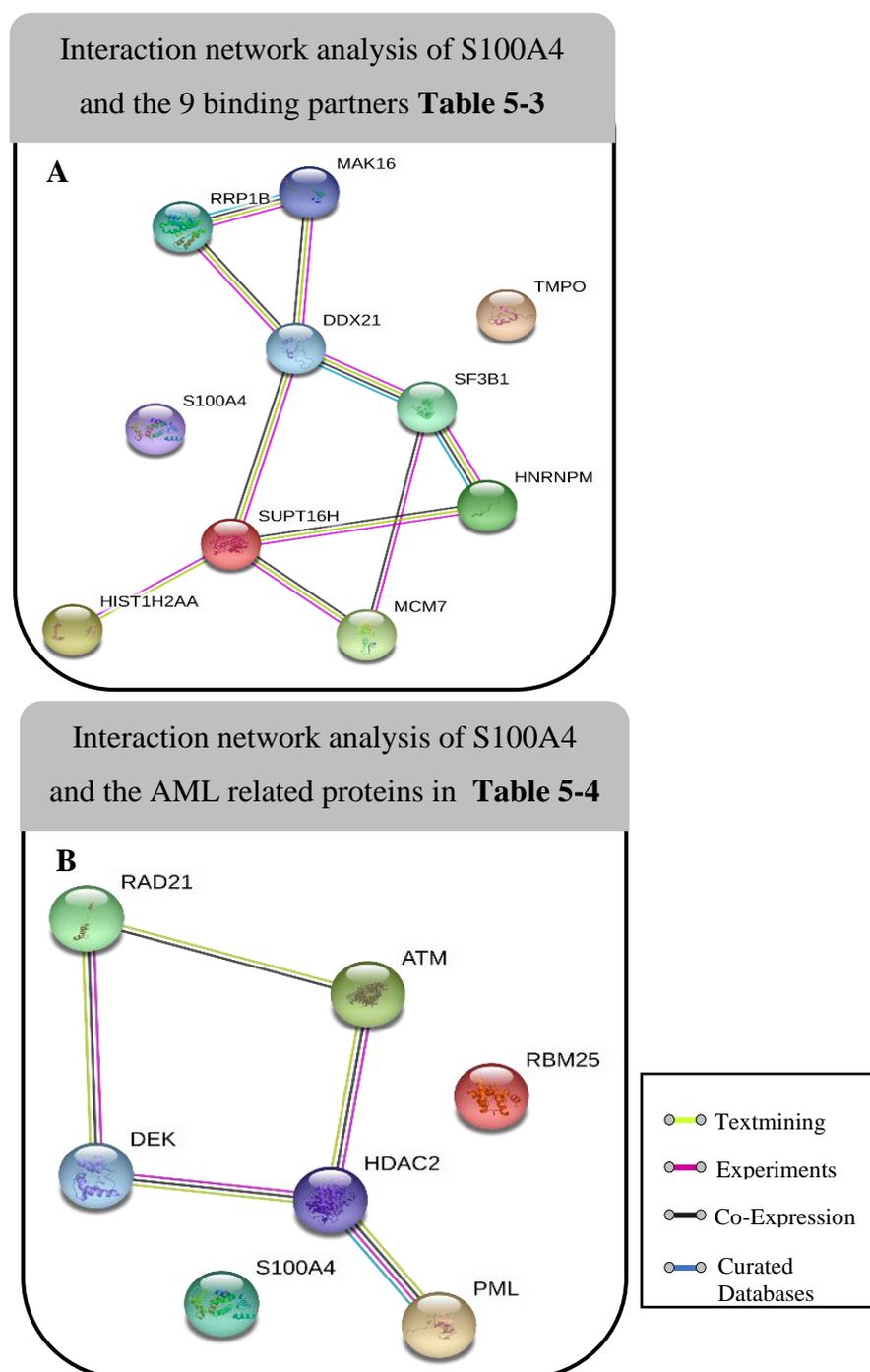
**Figure 5-12: Functional analysis using biological process enrichment for the final list of S100A4 binding partners candidates.**

Enrichment analysis using Gene Ontology (GO) using Biological Process (BP) using shows that “Nucleic Acid Metabolic Process”, “RNA Metabolic Process”, and “Nucleobase, Nucleoside, Nucleotide and Nucleic Acid Metabolic Process” are the most overrepresented biological processes for S100A4 binding partners final candidate list. Functional representation is shown in *p*-value on the left panel of the graph. Red rectangle indicates the overrepresented BPs in this analysis. PANTHER Overrepresentation Test (Released 20190711). The GO/BP analysis was performed using PANTHER14.1 and searched against the reference list of 65 proteins pulled down with S100A4 in presence of Ca<sup>2+</sup> (Table 5-2). PANTHER14.1 is generated from the 2019\_04 release of [Reference Proteome dataset](#) (based on UniProt Release 2019\_04, Ensembl release 95 and Ensembl Genome release 42).

### 5.3.6 Protein-protein interaction analysis between S100A4 and its binding partners

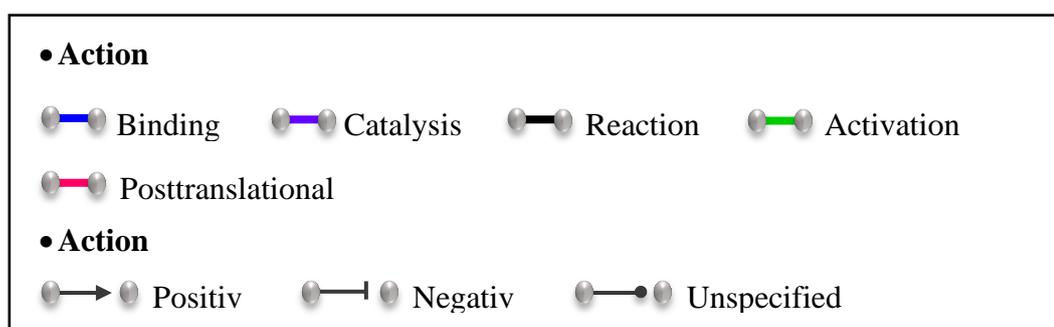
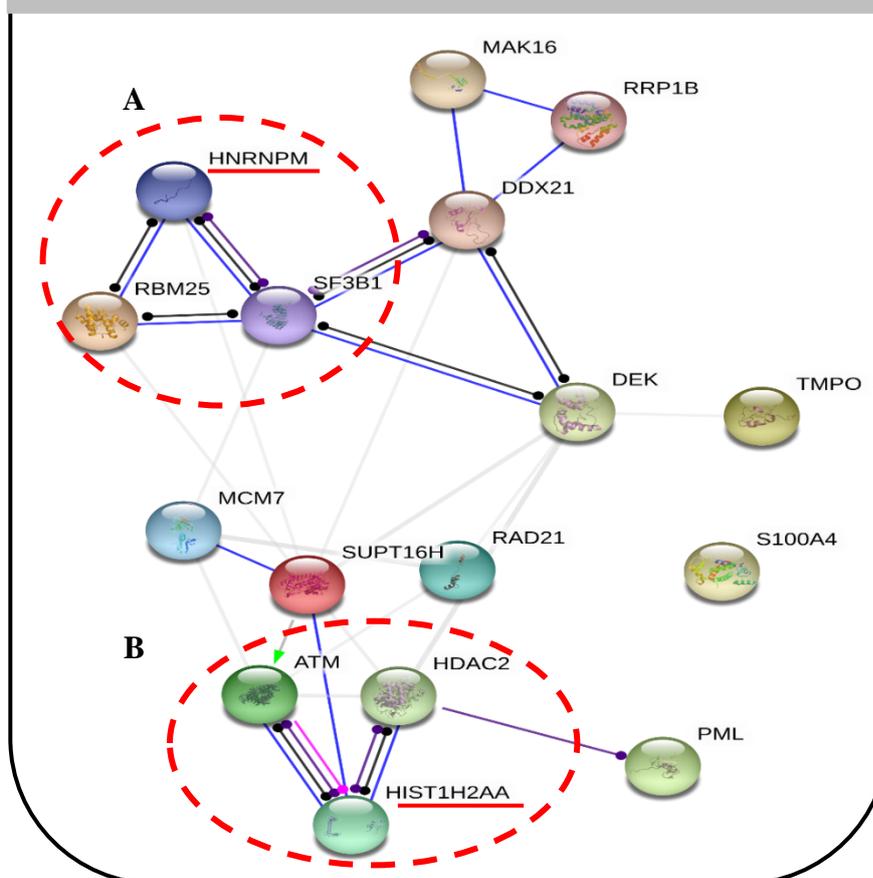
To analyse the PPI of S100A4 with the nine binding partners identified (Table 5-3), a functional protein association network analysis using STRING database was performed. The analysis was extended to include the AML related binding partners in Table 5-4. This analysis suggested no interaction were previously reported between S100A4 and the 9 binding partners (Figure 5-13A). Likewise, no interaction was reported between S100A4 and the AML related binding partners (Figure 5-13B). However, STRING combined analysis of S100A4 nine binding partners and AML related proteins shows significant interaction between each other ( $p\text{-value} = 5.8 \times 10^{-10}$ ), Figure 5-14. Interestingly, when the interaction between the S100A4 candidate binding partners and AML-related proteins was analysed, hnRNPM predicted to bind to SF3B1 and RBM25. Moreover, hnRNPM is predicted to have a catalytic and reaction functions with SF3B1 and only reaction function with RBM25 (Figure 5-14A). Likewise, HIST1H2AA is predicted to bind to HDAC2 and ATM and have a catalytic and reaction functions with them (Figure 5-14B). However, ATM is predicted to induce posttranslational modification (PTM) on HIST1H2AA (Figure 5-14B).

In summary, PPI network analysis elucidated the interaction of hnRNPM and HIST1H2AA with proteins that are implicated in known molecular pathways of AML development such as dysregulated splicing machinery, gene expression, and epigenetics regulation. Thus S100A4 could mediate its function through its validated binding partners' hnRNPM and HIST1H2AA in these pathways.



**Figure 5-13: String interaction network analysis of S100A4 and binding partners.** Protein interactions network analysis was performed on S100A4 and the 9 candidate binding partners and with AML-related proteins by STRING version 11.0 using interaction based on “evidence” **A**) PPI analysis using STRING network analysis of S100A4 with final MS candidate list of S100A4 binding partners. Average local clustering coefficient: 0.317; PPI enrichment  $p$ -value =  $1.98 \times 10^{-5}$  **B**) Network analysis of S100A4 with AML related proteins from MS dataset. Average local clustering coefficient: 0.143; PPI enrichment  $p$ -value = 0.0142. Coloured lines between proteins indicate various types of interactions evidence. PPI network analysis performed using STRING (version 11.0).

Interaction network analysis of combined S100A4, 9 binding partners **Table 5-3** and the AML related proteins in **Table 5-4**



**Figure 5-14: S100A4 nuclear candidate binding partners and AML-related proteins are highly interactive.**

Protein interactions network analysis was performed on the 9 candidate binding partners of S100A4 and AML-related proteins by STRING version 11.0 using interaction based on “molecular action” **A)** HnRNPM binds to SF3B1 and RBM25 and share catalytic and reactions functions. **B)** HIST1H2AA binds to HDAC2 and ATM AML-related proteins and share catalytic, reaction and PTM functions. Underlined proteins are validated S100A4 binding partners, red dotted circles highlight the interaction complexes in which hnRNPM and HIST1H2AA are involved. Interaction line thickness indicates the strength of data support. Average local clustering coefficient: 0.553; PPI enrichment  $p$ -value =  $5.8 \times 10^{-10}$ . PPI network analysis performed using STRING (version 11.0).

## 5.4 Discussion

The main objective of this Chapter was to identify proteins that bound S100A4 in the cytoplasm or nuclear subcellular compartment in AML cells. To achieve this, a co-IP protocol was established to ‘pull down’ S100A4 and its binding partners in the presence of calcium. As shown in (Figure 3-7), four cell lines (ME-1, HEL, NOMO-1, and THP-1) demonstrated expression of S100A4 in both cytoplasmic and nuclear subcellular fractions. However, ME-1 AML cell line was chosen to determine the binding proteins of S100A4 due to the high abundance of expression of S100A4 in the cytoplasmic and nuclear compartment compared to other cell lines. Given the ease of culturing NOMO-1 cells (as opposed to ME-1) and the generation of sufficient nuclear protein lysate to use in several co-IP assays (approximately doubling every 35 hours), this line was chosen to establish the co-IP protocol. This study shows that S100A4 is bound to hnRNPM and HIST1H2AA when co-immunoprecipitated under  $\text{Ca}^{2+}$  enriched conditions. Further functional and PPI analysis identified hnRNPM and HIST1H2AA in overrepresented BP and MF as well as actively interacting with proteins known to be changed in AML.

### 5.4.1 Optimal protein extraction and co-IP conditions were achieved in ME-1 cell line

For protein extractions in co-IP experiments, it is recommended to solubilise the cell membrane by using non-ionic non-denaturing buffers that will disrupt the integral membrane proteins and preserve endogenous protein interactions (Lin and Lai 2017). Further, given S100A4 is a protein which has a conserved EF-hand motif and binds  $\text{Ca}^{2+}$  to change its conformation, there is a need to retain this interaction complex in the protein extraction buffer. This buffer should not contain any high-affinity metal ion chelators such as EDTA or EGTA. EDTA has a higher affinity to bind to magnesium and low affinity towards  $\text{Ca}^{2+}$  ions whereas EGTA is known for its high affinity for binding with  $\text{Ca}^{2+}$  ions (Auld 1995). This study used a non-ion chelating buffer CEB-A (BioVision, Oxford, UK) for cytoplasmic extraction and triethylammonium bicarbonate buffer (TEAB) for nuclear extraction.

One of the most commonly used protocols for co-IP is the direct co-IP approach where the antibody is pre-immobilised to magnetic beads (Lin and Lai 2017). This study compared the direct co-IP method to an in house optimised indirect co-IP protocol and

found that direct co-IP was ineffective in ‘pulling down’ S100A4 binding complexes (Figure 5-2). The likely reason for this poor efficiency is that free antibody in the lysate can form stable immune complexes with the target protein (present in low concentrations) before it has been bound by the magnetic beads. Moreover, indirect co-IP is beneficial when the binding kinetics of the antibody to the antigen is slow or the antibody has a weak binding affinity for the antigen (Kaboord *et al.* 2015).

Most target-free S100 proteins, at normal physiological concentrations of  $\text{Ca}^{2+}$ , have low  $\text{Ca}^{2+}$  affinity when they are in an inactive form. However, S100 proteins  $\text{Ca}^{2+}$  binding affinity rise significantly when they bind to their functional targets (Liriano *et al.* 2012). Thus, the stability of S100A4 with its targets depends on  $\text{Ca}^{2+}$  ion concentrations which *in vivo* ranges from 0.1–10  $\mu\text{M}$  (Scott and Kekenus-Huskey 2016). To replicate the naturally occurring mechanism in S100A4 interaction with its binding partners in a  $\text{Ca}^{2+}$  dependent manner, all co-IP buffers were enriched with excess 100  $\mu\text{M}$   $\text{CaCl}_2$  to preserve the reversible interaction throughout the co-IP process. To control for non- $\text{Ca}^{2+}$  protein interactors with S100A4, a further set of co-IP lysates were generated where  $\text{Ca}^{2+}$  ions were depleted from the co-IP buffers using 10 mM  $\text{Na}_2\text{EDTA}$ . Stripping out  $\text{Ca}^{2+}$  ion using metal chelator such as EDTA during co-IP acts as a negative control in this experiment and served as a useful comparator in the LC-MS data analysis to rule out non- $\text{Ca}^{2+}$  binders as well as non-specific binding proteins from  $\text{Ca}^{2+}$ -enriched S100A4 co-IP.

Co-IP buffers are carefully chosen to ensure high antibody-antigen binding affinity as well as retaining S100A4 interaction complex throughout the co-IP process. In this co-IP experiment, Triton X-100 was chosen for its mild and non-denaturing property on the antibody-antigen reaction which ranges between 8% to 10% *v/v* (Qualtiere *et al.* 1977). Moreover, Triton X-100 has rigid and large nonpolar heads that do not disturb water-soluble proteins and does not disturb indigenous interactions complex or structures of proteins (Lee *et al.* 2018). Co-IP quality control for  $\text{Ca}^{2+}$  enriched conditions and  $\text{Ca}^{2+}$  depleted conditions co-IP experiments were verified by western blot. These blots show that both co-IP conditions have low non-specific background noise in the IP eluted lanes (Figure 5-3).

In summary, indirect co-IP was more efficient than the commonly used direct co-IP approach. Further, enriching extraction buffers with  $\text{Ca}^{2+}$  ions is essential to retain  $\text{Ca}^{2+}$  dependent interactions between S100A4 and its binding partners. Additionally, using a mild detergent that does not affect antibody-antigen affinity to ensure high co-IP efficiency.

### 5.4.2 Mass Spectrometry proteomic analysis and data filtration strategy

Following the ‘pull-down’ S100A4 and its binding proteins in the presence and absence of  $\text{Ca}^{2+}$  in ME-1 cells, cytoplasmic and nuclear lysates samples were analysed by LC-MS. The taxonomy analysis for cytoplasmic co-IP identified hits for BSA and rabbit IgG. This discrepancy between the observed high relative abundance signal on MS spectra outputs and number of cytoplasmic peptides IDs identified could be attributed to contamination occurred during either co-IP of Ms preparation. Another possible explanation is that this discrepancy could be attributed to heavy post-transcriptional modification (PTM) that could not be identified in a standard spectral search (Doll and Burlingame 2015). To illustrate, the MS analysis could detect the backbone of cytoplasmic peptides without detecting PTM at standard enrichment conditions. When peptides’ spectra matched to target-decoy database, the confidence (measured by percentage of FDR) of these identified peptides is very low and thereby they will be excluded from the final peptides list (Wang *et al.* 2017) resulting in a fewer number of peptides than observed spectral outputs. However, this is highly unlikely due to the sensitivity of LC-MS that could detect basic isoforms of proteins analysed. In most cases of MS data analysis, the analysis strategy is often designed to identify PPI, PTM, or quantification of protein(s) expression levels under certain stimuli (Matthiesen *et al.* 2011). Thus, in this study, the MS proteomic data filtration strategy targeted primarily identification of  $\text{Ca}^{2+}$  based PPI between S100A4 and its binding proteins which are represented in peptides identified in  $\text{Ca}^{2+}$  enriched co-IPs (cytoplasmic/nuclear).

The proteomic data filtration strategy applied in this study is based on removing non-specific / non  $\text{Ca}^{2+}$  dependent binders, filtering out underrepresented and unreviewed peptides’ sequences, and prioritising peptides for experimental validation based on their MS abundance and signal intensity scores. As a result, non-specific (IgG control co-IP) and non  $\text{Ca}^{2+}$  dependent binders ( $\text{Ca}^{2+}$  depleted co-IP) were removed from the  $\text{Ca}^{2+}$  enriched co-IP samples [cytoplasmic (Figure 5-5) and nuclear (Figure 5-6)].

Further refinement has resulted in removing underrepresented peptides’ sequences that have peptide score  $\leq 1$  and “Unreviewed” peptides reported by TrEMBL/UniProt KB database. The reason for excluding non-curated peptide sequences is to limit the analysis to true peptides matches. UniProt KB is a public database composed of Swiss-Prot (protein sequences source), PIR (protein information source),

and TrEMBL (translated nucleotide sequences from EMBL). PIR and Swiss-Prot databases contain manually curated protein sequences with several cross-references to other resources, including literature references (UniProtConsortium 2017). Thus, after applying this filter, the cytoplasmic data set analysis was terminated as the concluded peptides after removing the unreviewed peptides were zero. These findings are surprising given that S100A4 is primarily expressed in cytoplasm and therefore would expect to detect more interactions in the cytoplasmic compartment.

On the contrary, nuclear peptide data set were reduced to 66 peptides that are reported as reviewed peptides on Swiss-Prot database Table 5-2. However, Keratin (KRT74) was removed from this pre-final candidate list. Keratin is a fibrous structural protein abundant in skin and hair and is considered as one of the most frequently reported contaminants in MS (Hodge *et al.* 2013). Nuclear peptides are classified based on PSM score and based on Area score and analysed the highest quartile of both data sets. This step narrowed down the list of candidate proteins to more logical and manageable number of proteins to validate their interactions with S100A4. The final filtration step resulted in nine proteins (Table 5-3) that are highly abundant and have highest detected signal in MS analysis. These 9 proteins could be potential S100A4 binding partners and needed to be validated using reciprocal co-IP described in section 5.3.4.

Taken together, MS data filtration strategy has removed all non-specific, non-Ca<sup>2+</sup> based binding proteins of S100A4, and non-reviewed peptide sequences in UniProt database. This filtration strategy resulted in prioritising only 9 proteins based on their MS peptide abundance and highly detected peptide signals.

### 5.4.3 S100A4 is bound to proteins that are implicated in AML

LC-MS analysis is a highly sensitive technique that can pick up significantly low peptides signals in the mixture (Di Falco 2018). In many cases, during raw data processing many of these peptides get excluded from the data set due to their low abundance or signal intensity scores. However, those peptides may be biologically relevant interactors with S100A4 either in a less frequent or indirect manner of interactions. In this study, I decided not only to focus my functional analysis on novel S100A4 binding proteins in AML but also include those proteins that have a meaningful biological role and already implicated AML pathophysiology such as Histone deacetylase2 (HDAC2) and RNA binding motif protein 25 (RBM25) among other

proteins reported in (Table 5-4). It may not be possible at this stage to experimentally validate whether these proteins bind with S100A4, however, it could shed some lights on relevance of S100A4 in AML interaction network.

One of the known molecular mechanisms of AML is dysregulation of gene expression which involves many genes such as HDAC2, DEK. For example, HDAC2 is involved in repression of gene transcription by deacetylating lysine residues on the N-terminus of the core histones (H2A, H2B, H3 and H4) resulting in condensation of chromatin (Yang and Seto 2007). Interestingly, one the downstream targets of HDAC2 “HIST1H2AA” was ‘pulled down’ with S100A4 and validated by reciprocal co-IP as binding partner of S100A4. This may suggest indirect interaction of S100A4 with HDAC2 via its binding partner HIST1H2AA.

The role of other RNA binding proteins such as RBM25 which regulates pre-mRNA splicing are yet to be described in AML. For example, RBM25 functions as a regulator of the alternative pre-mRNA splicing through regulating the expression ratio of the pro-apoptotic factor BCL2L1 isoform S to antiapoptotic BCL2L1 isoform L mRNA expression. In AML, normal expression levels of RBM25 in leukemic cells controls the activity of growth oncogene MYC. However, downregulated RBM25 confers a proliferative advantage on pre-malignant cells by activating MYC and escaping apoptosis via shifting the ratio of pro-apoptotic BCL2L1 isoform S towards the anti-apoptotic BCL2L1 isoform L (Ge *et al.* 2019). Moreover, a recent study showed that approximately 51% (247/484) of RBP genes to be differentially expressed in AML patients (Wang *et al.* 2019). In this study, 107 RBP genes have been identified to be significantly upregulated in AML patients compared to normal CD34+ hematopoietic stem/progenitor cells. In Wang, E. *et al* study, RBM39 have been identified to play an essential role in AML survival and maintenance. CRISPR-mediated deletion of RBM39 caused altered splicing of HOXA9 target genes, known as an essential transcriptional network required in AML. Thus, RBM25 as well as other RNA binding proteins play an essential role in AML maintenance and pro-survival of leukemic cells through regulation of transcriptional networks and pro-apoptotic factors.

In summary, both epigenetic modifications and splicing machinery are commonly dysregulated molecular pathways in AML. The AML-related proteins that are ‘pulled down’ with S100A4 under  $\text{Ca}^{2+}$  enriched conditions may interact indirectly with S100A4 via its validated binding partners hnRNPM and HIST1H2AA. Functional and protein-

protein interaction analysis of the AML-related proteins with S100A4 and its binding partners may reveal how S100A4 mediates its function (s) in AML development.

#### 5.4.4 Validation of S100A4 candidate binding partners

In order to experimentally validate S100A4 binding to the 9 candidate proteins (Table 5-3), a two steps validation co-IP were performed. Only six candidate proteins (HIST1H2AA, MAK16, SUPT16H, and DDX21 ) (Figure 5-8 and Figure 5-8) were successfully pulled down with S100A4 in the first co-IP. However, initial results of pulling down MCM7 and hnRNPM with S100A4 were questionable as no difference have been observed between IgG and S100A4 co-IPs but the latter was validated in reciprocal IP. Subsequently, using a reciprocal co-IP approach, S100A4 was successfully pulled down with only two candidate proteins HIST1H2AA (Figure 5-10A) and hnRNPM (Figure 5-10B).

HIST1H2AA is a highly conserved core component protein of the nucleosome complex, highly abundant in the nucleus, and heavily positively charged protein (Chakraborty *et al.* 2018). Thus, HIST1H2AA protein can act as a magnet to small negatively charged proteins such as S100A4 when pulled down. Although, S100A4 was pulled down reciprocally with HIST1H2AA, I have to be cautious in naming HIST1H2AA as binding partner of S100A4 as it may be a non-specific binding arising during nuclear extraction. Nevertheless, the interaction was dependent on the presence of calcium which would make it more likely that this was a specific interaction. In addition, it is well reported that PTMs of HIST1H2AA may be linked to regulation of gene expression and other cellular activities such as transcription, DNA repair, replication, and epigenetics (reviewed in Corujo and Buschbeck 2018).

Heterogeneous nuclear ribonucleoprotein M (hnRNPM) belongs to a large RNA-binding protein (RBP) family localised mainly in the nucleus where it functions as a component of the spliceosome complex (reviewed in Geuens *et al.* 2016). Previous studies linked overexpression of hnRNPM with tumour progression, aggressiveness, and metastasis. For example, Yang *et al* reported that overexpressed hnRNP in breast cancer induced cell growth, colony formation, and inhibited apoptosis (Yang *et al.* 2018). Another study reported that hnRNPM promoted tumour metastasis through mediating alternative splicing programme (Xu *et al.* 2014). As mentioned previously in section 5.4.3, Wang, E. *et al* study identified 107 upregulated RBM genes in AML. Interestingly, among the dysregulated RBPs, hnRNPM was one of the upregulated genes ( $p < 0.05$ ) and

have been implicated as a member of the as pro-survival spliceosome machinery of AML interacting with RBM39 (Wang *et al.* 2019).

S100A4 was not detected in the other three candidate protein pull downs, MCM7, DDX21, and MAK16 (Figure 5-9A, B, and C respectively). A possible explanation is that the interaction between these proteins and S100A4 may not be frequent. The number of these proteins that are bound to S100A4 could be an underrepresented fraction of the ‘pulled down’ protein and therefore undetectable. An alternative explanation is that S100A4 is relatively less abundant in the nuclear compartment as compared to the candidate binding partners.

Finally, co-IP was not achieved in the case of SUPT16H (Figure 5-9D). The antibody may need further optimization in terms of antibody dilution or testing of alternative clones which was not possible given the time limitation. Alternatively, I may try another SUPT16H antibody that is validated to work in co-IP. In summary, the data above suggest that HIST1H2AA and hnRNPM are bound to S100A4 in the nucleus of ME-1 and could be the basis of future studies (6.2).

#### **5.4.5 Validated S100A4 binding partners (hnRNPM and HIST1H2AA) are identified in overrepresented MF and BP**

S100A4 mediates its function indirectly via interacting with target proteins. Thus, it is very important to understand what the molecular functions (MF) of S100A4 binding proteins are and the biological processes they are involved in. As shown in Figure 5-11, the GO enrichment analysis for MF of the final binding partners of S100A4 in Figure 5-11 revealed that five binding partners were involved in the highest overrepresented MF of “Nucleic Acid Binding”. The validated binding partners of S100A4 (HIST1H2AA and hnRNPM) were observed in this MF.

Gene ontology enrichment analysis assigns general MF terms to proteins of interest; thus, it is important to couple the MF enrichment analysis with biological process to identify the BP these proteins involved in. Therefore, GO enrichment analysis for BP in Figure 5-12 shows that hnRNPM observed in all over-represented BPs together with SF3B1. In addition to the over-represented BPs, hnRNPM and SF3B1 were observed together in all RNA related biological processes such as mRNA processing and RNA splicing suggesting strong protein-protein interaction between them. Further, both hnRNPM and SF3B1 were among the pulled down list of proteins in the nuclear-Ca<sup>2+</sup>

enriched S100A4 co-IP and detected by LC-MS analysis as highly abundant proteins in the group. These findings may suggest a physical interaction between hnRNPM and SF3B1. Although SF3B1 has not been experimentally validated as S100A4 binding partner in this study, the co-appearance with hnRNPM in RNA related BPs indicates its importance as a potential indirect binding partner with S100A4. Mutated SF3B1 is frequently detected in several haematological malignancies such as MDS, CLL and AML and associated with poor patients' outcomes (Ganguly and Kadam 2016, Wang *et al.* 2011, Larsson *et al.* 2013 respectively). For instance, Hou *et al* reported that SF3B1 mutation is associated with poor OS and DFS in *de novo* AML. In this study, AML patients who undergone conventional intensive chemotherapy achieved CR 79.7%. However; SF3B1 mutation was associated with lower CR (22.2% vs. 79.7%,  $P = 0.0005$ ). Furthermore, patients who have SF3B1 mutation had significantly shorter OS (2 months vs. 29.5 months,  $P < 0.001$ ) and DFS (0 month vs. 9 months,  $P < 0.001$ ) (Hou *et al.* 2016).

In summary, hnRNPM and HIST1H2AA have been identified in over-represented MF and BP suggesting that these proteins have an important role in the cell.

#### **5.4.6 Validated S100A4 binding partners (hnRNPM and HIST1H2AA) are highly interactive with AML-related proteins**

It has been reported in AML that key biological processes such as gene expression regulation, epigenetics modifications, and cell survival are mediated by proteins assembly via coordinated mechanisms of action called PPI which may play a vital role in disease development and progression (Shi *et al.* 2014). Thereby, to identify the PPI between S100A4 and its binding partners reported in Table 5-3, a protein association analysis (STRING) was used.

As shown in Figure 5-13A, no interaction has been predicted between S100A4 and the candidate list of binding partners. Similarly, STRING analysis has shown no predicted interaction between S100A4 and AML-related proteins that are pulled down with nuclear S100A4  $Ca^{2+}$  enriched co-IP, Figure 5-13B. However, STRING's protein-protein network analysis is based on importing data from experimentally derived PPI through literature-curated interactions. Further, STRING also uses computational PPI prediction tools that calculate the interaction confidence score based on text mining, experimentally determined, co-expression, and interactions that are imported from other curated databases (Szklarczyk *et al.* 2015). Thus, novel and under-reported interactions

such as the interaction between S100A4 and the pulled down proteins in literature and other curated databases may not be imported in STRING network analysis (Wodak *et al.* 2009). Therefore, S100A4 could be a novel interactor with hnRNPM and HIST1H2AA but under-reported in literature.

Interestingly, two of the nine binding proteins of S100A4, hnRNPM and HIST1H2AA showed an interaction with AML-related proteins; SF3B1, RBM25 known splicing factors in AML and HDAC2 and ATM known epigenetics and cell cycle regulators in AML respectively as shown in Figure 5-13. This protein-protein network analysis has significantly more interaction than predicted by STRING ( $p\text{-value} = 5.8 \times 10^{-10}$ ). To illustrate, the group of proteins included in this analysed have more interactions among each other than what would be expected from a random set of proteins of similar size drawn from the genome.

In conclusion, S100A4 may interact indirectly with AML driver proteins such as SF3B1 and RBM25 to mediate splicing events through its binding partner hnRNPM. Likewise, S100A4 may interact indirectly with HDAC2 and ATM to regulate gene expression and cell cycle through binding with HIST1H2AA. The PPI analysis was consistent with the observed MF and BP of the validated binding partners of S100A4 hnRNPM and HIST1H2AA. In future studies, we may extend this project to show how S100A4 can modulate the activities of hnRNPM and HIST1H2AA in AML as discussed in section (6.1).

# Chapter 6

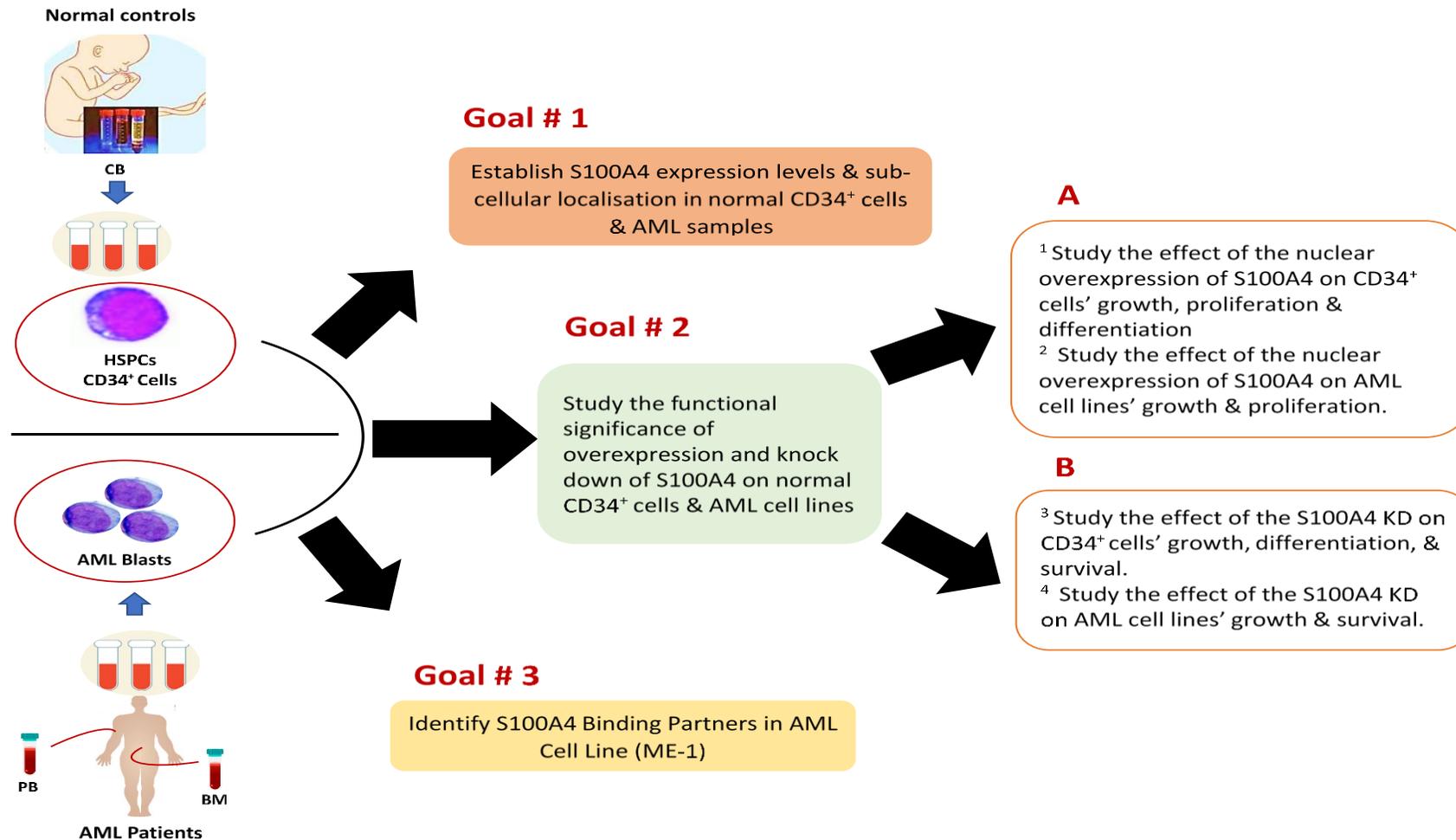
## General Discussion and Future Directions

## 6 Chapter 6 - General discussion

### 6.1 Summary and conclusion

Acute myeloid leukaemia is very heterogeneous disease where multiple factors contribute to AML development, progression, and relapse. Given that AML is much more prevalent in older people with a median age of diagnosis (>68 years of age), it has been reported that 70% of patients in this age group die of AML within the first year of diagnosis (Short *et al.* 2018). Further, the overall 5 year survival rate is still below 50% for younger patients despite the fact that they are able to receive intensive treatment (Khwaja *et al.* 2016). Older AML patients generally have lower tolerance towards therapeutic cytotoxic agents which limits their treatment options. Consequently, this group of patients are good candidates for less intensive therapies such as low-intensity chemotherapy or targeted therapeutic agents (Burnett *et al.* 2011). In the last decade, targeted therapeutic agents in AML have become the centre of focus in many clinical trials as solo treatment, in combination of chemotherapy regimen, or in combination of other target agents (Kayser and Levis 2018). Recent advances in transcriptome contributed to better understanding of aberrant genes expression profiles in leukemia (Ye *et al.* 2019). However, transcriptomic data provide insufficient information about the protein expression profiles and does not identify altered subcellular localization of proteins. Often, protein mislocalisation between cytoplasm and nucleus may interfere with normal cellular functions by interacting with key oncoproteins or tumour suppressors which cooperatively may lead to tumour development and metastasis (Wang and Li 2014). Thus, analysing the AML proteome in tandem with transcriptomics with focus on protein changes within the nuclear compartment led to identification of S100A4 as the most significant and fold changing protein in AML blasts that had not been previously associated with AML (Alanazi *et al.* 2019). S100A4 belongs to the S100 multigene family of calcium-binding proteins of the EF-hand type. These proteins have diverse roles in a variety of cellular processes including regulation of proliferation, cell cycle progression, apoptosis, differentiation, Ca<sup>2+</sup> homeostasis, migration, adhesion and transcription (Bresnick *et al.* 2015, Donato *et al.* 2013). The functions of S100A4 and other family members may vary in normal and cancer cells based on the expression levels and subcellular localisation (Chen *et al.* 2014).

In cancer, S100A4 is implicated in metastasis, apoptosis regulation, cytoskeletal rearrangements, angiogenesis regulation, transcription factors regulation which are highly dysregulated processes in cancer (Chen *et al.* 2014). Overexpression of S100A4 is widely reported in several solid tumour and often associated poor prognosis (Xu *et al.* 2019), tumour progression (Zhou *et al.* 2018), chemoresistance (Mencia *et al.* 2010), and metastasis (Li *et al.* 2013). Moreover, S100A4 is predominantly expressed in the cytoplasm under normal physiological conditions (Kikuchi *et al.* 2006) and secreted to the extracellular space upon  $\text{Ca}^{2+}$  homeostasis (Donato *et al.* 2013). However, several studies have identified that S100A4 is mislocalised to the nucleus of transformed cells (Boye *et al.* 2016, Boye *et al.* 2010). In this study, the levels and subcellular localisation of S100A4 were established in normal  $\text{CD34}^+$  cells and AML patients' samples. Following, the functional significance of nuclear overexpressed and knocked down S100A4 on  $\text{CD34}^+$  cells and AML cell lines were studied as summarised in the working model (Figure 6-1). Thus, this study demonstrated that S100A4 expression is restricted to the cytoplasmic compartment in normal  $\text{CD34}^+$  HSPCs (Figure 3-10A) and differentiated progenitors (Figure 3-10B). Further, restriction of cytoplasmic S100A4 expression was also confirmed in mixed progenitors' population in normal BM samples (Figure 3-11). However, in AML, S100A4 is mislocalised to the nucleus in 83% of AML (20/24) of FAB M1 (Figure 3-13B) and 44% (4/9) of FAB M4 (Figure 3-14B) compared normal controls. Indeed, nuclear mislocalisation of S100A4 in many solid tumours including colorectal, ovarian, and breast cancers was linked to aggressiveness, progression, metastasis phenotypes (Kikuchi *et al.* 2006, Egeland *et al.* 2017, Flatmark *et al.* 2003). The pathological link between the presence of S100A4 in the nucleus and cancer phenotypes could be explained by its direct interactions with key regulatory proteins in the nucleus. For example, S100A4 regulates cell cycle in HeLa cells through negatively regulating p53 in the nucleus leading to its degradation (Orre *et al.* 2013).



**Figure 6-1: Integrated working model summarising the investigation plan of role of S100A4 in AML carried out in this study.**

This study also demonstrated that *S100A4* mRNA is significantly overexpressed across all AML subtypes as compared to normal HSCs (Figure 3-12) and was associated with poor OS of AML patients ( $p=0.01$ ) (Figure 3-15). However, it would more informative to perform the survival analysis on nuclear protein samples to see whether the subcellular distribution of S100A4 influences patient outcomes.

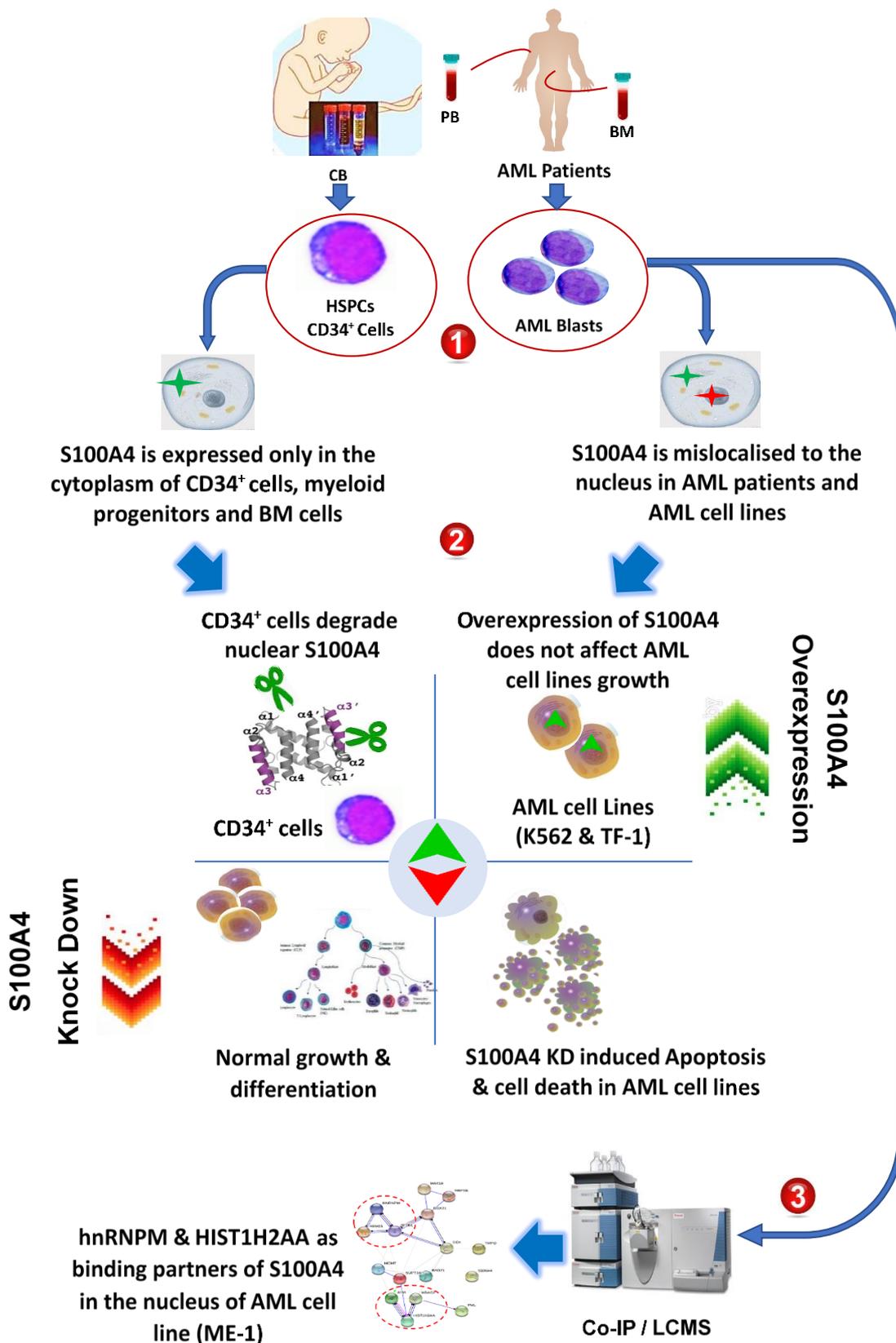
Nuclear expression of S100A4 in solid tumours promoted proliferation, and tumour survival; however, little is known about its role in leukemia. Thus, I attempted to be ectopically overexpress S100A4 in the nucleus of CD34<sup>+</sup> HSPCs and leukaemia cell lines (K562 and TF-1) to study the effect on proliferation, differentiation and survival. Although, the viral vectors used in this study contained a nuclear localisation signal to force nuclear expression of S100A4 (that was successfully validated in HEK cells) primary CD34<sup>+</sup> HSPCs did not overexpress of S100A4 in the nucleus (Figure 4-4). This suggests that CD34<sup>+</sup> HSPCs recognise the nuclear expression of S100A4 as abnormal and therefore rapidly degrade it. In contrast, I was able to overexpress S100A4 in the nucleus of leukaemia cell lines but this had little or no effect on proliferation even under serum deprivation conditions (Figure 4-7).

Knock-down of S100A4 in CD34<sup>+</sup> HSPCs had little impact on their proliferation, differentiation or survival (Figure 4-11). In contrast, knocking down the expression of S100A4 in AML cell lines showed significant reduction in growth and resulted in cell death through induction of apoptosis (Figure 4-14A). Several studies have reported similar results in solid tumours. For instance, S100A4 KD in colorectal tumours suppresses cell growth (Huang *et al.* 2012), inhibits tumour angiogenesis in endothelial cells (Ochiya *et al.* 2014), and blocks metastasis in anaplastic thyroid cancer (Zhang *et al.* 2016). Knock down of S100A4 has also been shown to reduce the self-renewal capability and tumorigenic properties of solid tumour cancer initiating cells (Lo *et al.* 2011).

S100 family members in general have no known enzymatic activity (Donato *et al.* 2013), instead; S100 proteins mediate their functions through interacting with other proteins and regulate their functions (Bresnick *et al.* 2015). The majority of these interactions occur in Ca<sup>2+</sup> dependent manner (Badyal *et al.* 2011). Several binding partners of S100A4 have shown a role cancer development and metastasis in solid tumours such as non-muscle myosin heavy chain IIA (NM-MHC IIA), p53, MMP9, among others. In such heterogeneous and complex disease such as AML, proteins interactions play a vital role in disease pathogenesis. This study identified

novel binding partners of S100A4 through mass spectrometry analysis. S100A4 was pulled-down with hnRNPM and HIST1H2AA in Ca<sup>2+</sup> dependent manner. HnRNPs comprise a family of RNA binding proteins involved in processing heterogeneous nuclear RNAs into mature mRNAs (Geuens *et al.* 2016). Interestingly, hnRNPM has been reported to be one of the upregulated genes ( $p < 0.05$ ) among RNA-binding proteins (RBM) and have been implicated as a member of the as pro-survival spliceosome machinery of AML interacting with RBM39 (Wang *et al.* 2019). Moreover, this study demonstrated through protein network analysis that hnRNPM interacts with SF3B1 known to be frequently mutated MDS and linked poor prognosis (Malcovati *et al.* 2011). Further, HIST1H2AA interacts with HDAC2 and ATM known epigenetics and cell cycle regulators in AML.

In conclusion, this study demonstrated for the first time the subcellular localisation of S100A4 in normal haematopoiesis and AML. Normal CD34<sup>+</sup> cells and myeloid differentiated lineages express detectable S100A4 in the cytosol only while, S100A4 is mislocalised to the nucleus of AML patients' samples. Further, CD34<sup>+</sup> cells can tolerate the loss of S100A4 and differentiate normally in the myeloid lineage. In contrast, knocking down expression of S100A4 in AML lines results in significant growth reduction and induction of apoptosis. These findings provide an evidence that supports a novel role for S100A4 as a pro-survival factor in AML cell lines and suggest that therapeutically targeting S100A4 would be an effective strategy in AML while sparing normal hematopoietic cells. Mechanistically, S100A4 may interact indirectly with proteins such as SF3B1 and RBM25 to dysregulate splicing through its binding partner, hnRNPM. Likewise, S100A4 may interact indirectly with HDAC2 and ATM to regulate gene expression and cell cycle through binding with HIST1H2AA. All finding of this study regarding the role of S100A4 in AML are summarised in a graphical abstract in Figure 6-2.



**Figure 6-2: Graphical abstract summarising the findings of this study about the role of S100A4 in Acute Myeloid Leukaemia.**

## 6.2 Future directions

The effect of subcellular distribution of S100A4 on patient outcome, could be performed using nuclear fractionated S100A4 patient samples. Further exploration of the mechanistic role of S100A4 in modulating the activities of hnRNPM and HIST1H2AA in AML could also be performed. This could be done by knocking down S100A4 and analysing its binding partners' activity. Another exciting avenue to explore, would be to test the efficacy of S100A4 inhibitors in cell lines and in patient derived xenograft mouse models (PDX) expressing nuclear S100A4. Saleem *et al.*, have developed novel small molecules (SMI1 and SMI2) which have shown high binding efficiency with S100A4. When SMI1 and SMI2 have been tested in vitro on neuroendocrine prostate (NE-CaP), both inhibitors inhibited tumour growth, proliferation, migration and invasiveness (Saleem *et al.* 2017). Further, the role of S100A4 in the regulation of HSC in the niche is yet to be described. However, HSC self-renewal and differentiation fates are regulated through Wnt/ $\beta$ -Catenin canonical signalling pathway (Katayama 2019). Some studies showed that the expression of S100A4 is positively regulated by  $\beta$ -Catenin activity in colorectal and lung cancers (Stein *et al.* 2006, Hou *et al.* 2018). Reporter assays confirmed that  $\beta$ -Catenin regulates the expression of S100A4 through binding to TCF which in turn induces S100A4 promoter activity (Stein *et al.* 2006). One study showed that inhibiting the expression of S100A4 in colon cancer cell lines by an inhibitor called calcimycin reduced the activity of Wnt/ $\beta$ -Catenin pathway (Sack *et al.* 2011). Thus, inhibiting the expression of S100A4 could reduce the activity of overexpressed  $\beta$ -Catenin in Wnt/ $\beta$ -Catenin pathway and consequently keep the balance of HSCs self-renewal capacity. To do this, first, we have to confirm that S100A4 expression is regulated by  $\beta$ -Catenin by knocking down  $\beta$ -Catenin and check the expression of S100A4. Alternatively, we could induce the expression of  $\beta$ -Catenin in AML cell line that is known to have less expression of  $\beta$ -Catenin and analyse the expression of S100A4.

# References

- Abelson, S., G. Collord, S. W. K. Ng, *et al.* 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* 559(7714): 400-404.
- Abreu, T. R., N. A. Fonseca, N. Gonçalves and J. N. Moreira. 2020. Current challenges and emerging opportunities of CAR-T cell therapies. *Journal of Controlled Release* 319: 246-261.
- Adams, G. B., K. T. Chabner, I. R. Alley, *et al.* 2006. Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* 439(7076): 599-603.
- Alanazi, B., C. R. Munje, N. Rastogi, *et al.* 2019. Integrated nuclear proteomics and transcriptomics identifies S100A4 as a therapeutic target in acute myeloid leukemia. *Leukemia*.
- Arber, D. A., A. Orazi, R. Hasserjian, *et al.* 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127(20): 2391-405.
- Asada, N., S. Takeishi and P. S. Frenette. 2017. Complexity of bone marrow hematopoietic stem cell niche. *Int J Hematol* 106(1): 45-54.
- Atlasi, Y., R. Noori, I. Marolin, *et al.* 2016. The role of S100a4 (Mts1) in Apc- and Smad4-driven tumour onset and progression. *Eur J Cancer* 68: 114-124.
- Auld, D. S. 1995. Removal and replacement of metal ions in metallopeptidases. *Methods Enzymol* 248: 228-42.
- Badar, T. and N. N. Shah. 2020. Chimeric Antigen Receptor T Cell Therapy for Acute Lymphoblastic Leukemia. *Curr Treat Options Oncol* 21(2): 16.
- Badyal, S. K., J. Basran, N. Bhanji, *et al.* 2011. Mechanism of the Ca(2)+-dependent interaction between S100A4 and tail fragments of nonmuscle myosin heavy chain IIA. *J Mol Biol* 405(4): 1004-26.
- Bagger, F. O., D. Sasivarevic, S. H. Sohi, *et al.* 2016. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res* 44(D1): D917-24.
- Beerman, I. 2017. Accumulation of DNA damage in the aged hematopoietic stem cell compartment. *Semin Hematol* 54(1): 12-18.
- Bennett, J. M., D. Catovsky, M. T. Daniel, *et al.* 1976. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* 33(4): 451-8.
- Beyar-Katz, O. and S. Gill. 2018. Novel Approaches to Acute Myeloid Leukemia Immunotherapy. *Clin Cancer Res* 24(22): 5502-5515.
- Binz, P. A., M. Muller, C. Hoogland, *et al.* 2004. The molecular scanner: concept and developments. *Curr Opin Biotechnol* 15(1): 17-23.

- Bonnet, D.2002. Haematopoietic stem cells. *J Pathol* 197(4): 430-40.
- Bose, P., P. Vachhani and J. E. Cortes.2017. Treatment of Relapsed/Refractory Acute Myeloid Leukemia. *Curr Treat Options Oncol* 18(3): 17.
- Bowers, R. R., Y. Manevich, D. M. Townsend and K. D. Tew.2012. Sulfiredoxin redox-sensitive interaction with S100A4 and non-muscle myosin IIA regulates cancer cell motility. *Biochemistry* 51(39): 7740-54.
- Boye, K., H. Jacob, K. A. Frikstad, *et al.*2016. Prognostic significance of S100A4 expression in stage II and III colorectal cancer: results from a population-based series and a randomized phase III study on adjuvant chemotherapy. *Cancer Med* 5(8): 1840-9.
- Boye, K. and G. M. Maelandsmo.2010. S100A4 and metastasis: a small actor playing many roles. *Am J Pathol* 176(2): 528-35.
- Boye, K., J. M. Nesland, B. Sandstad, G. M. Maelandsmo and K. Flatmark.2010. Nuclear S100A4 is a novel prognostic marker in colorectal cancer. *Eur J Cancer* 46(16): 2919-25.
- Bresnick, A. R., D. J. Weber and D. B. Zimmer.2015. S100 proteins in cancer. *Nat Rev Cancer* 15(2): 96-109.
- Burnett, A., M. Wetzler and B. Lowenberg.2011. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* 29(5): 487-94.
- Burnett, A. K.2018. Treatment of Older Patients With Newly Diagnosed AML Unfit for Traditional Therapy. *Clin Lymphoma Myeloma Leuk* 18(9): 553-557.
- Cabezón, T., J. E. Celis, I. Skibshøj, *et al.*2007. Expression of S100A4 by a variety of cell types present in the tumor microenvironment of human breast cancer. *Int J Cancer* 121(7): 1433-44.
- Cai, Y. and J. G. Mikkelsen.2016. Lentiviral Delivery of Proteins for Genome Engineering. *Curr Gene Ther* 16(3): 194-206.
- Calo, E., R. A. Flynn, L. Martin, *et al.*2015. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* 518(7538): 249-53.
- CancerResearchUK ( 2014). "Cancer mortality for common cancers." Retrieved 6 MAY 2016, 2016.
- Carlson, S. M., C. M. Soulette, Z. Yang, *et al.*2017. RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine monomethylation. *J Biol Chem* 292(32): 13381-13390.
- Chakraborty, K., M. Kang and S. M. Loverde.2018. Molecular Mechanism for the Role of the H2A and H2B Histone Tails in Nucleosome Repositioning. *J Phys Chem B*.
- Challen, G. A., D. Sun, M. Jeong, *et al.*2011. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* 44(1): 23-31.

- Chen, H., C. Xu, Q. Jin and Z. Liu.2014. S100 protein family in human cancer. *Am J Cancer Res* 4(2): 89-115.
- Chen, J., O. Odenike and J. D. Rowley.2010. Leukaemogenesis: more than mutant genes. *Nat Rev Cancer* 10(1): 23-36.
- Christophe, D., C. Christophe-Hobertus and B. Pichon.2000. Nuclear targeting of proteins: how many different signals?, *Cell Signal* 12(5): 337-41.
- Cornelissen, J. J., A. Gratwohl, R. F. Schlenk, *et al.*2012. The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. *Nat Rev Clin Oncol* 9(10): 579-90.
- Corujo, D. and M. Buschbeck.2018. Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. *Cancers (Basel)* 10(3).
- Crane, G. M., E. Jeffery and S. J. Morrison.2017. Adult haematopoietic stem cell niches. *Nat Rev Immunol* 17(9): 573-590.
- Dahlmann, M., A. Okhrimenko, P. Marcinkowski, *et al.*2014. RAGE mediates S100A4-induced cell motility via MAPK/ERK and hypoxia signaling and is a prognostic biomarker for human colorectal cancer metastasis. *Oncotarget* 5(10): 3220-33.
- Di Falco, M. R.2018. Mass Spectrometry-Based Proteomics. *Methods Mol Biol* 1775: 93-106.
- Dohner, H., E. Estey, D. Grimwade, *et al.*2017. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 129(4): 424-447.
- Dohner, H., D. J. Weisdorf and C. D. Bloomfield.2015. Acute Myeloid Leukemia. *N Engl J Med* 373(12): 1136-52.
- Dolatshad, H., A. Pellagatti, M. Fernandez-Mercado, *et al.*2015. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* 29(5): 1092-103.
- Doll, S. and A. L. Burlingame.2015. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol* 10(1): 63-71.
- Dombret, H. and C. Gardin.2016. An update of current treatments for adult acute myeloid leukemia. *Blood* 127(1): 53-61.
- Donato, R.1986. S-100 proteins. *Cell Calcium* 7(3): 123-45.
- Donato, R.2007. RAGE: a single receptor for several ligands and different cellular responses: the case of certain S100 proteins. *Curr Mol Med* 7(8): 711-24.

- Donato, R., B. R. Cannon, G. Sorci, *et al.* 2013. Functions of S100 proteins. *Curr Mol Med* 13(1): 24-57.
- Drissen, R., N. Buza-Vidas, P. Woll, *et al.* 2016. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol* 17(6): 666-676.
- Ebraldize, A., E. Tulchinsky, M. Grigorian, *et al.* 1989. Isolation and characterization of a gene specifically expressed in different metastatic cells and whose deduced gene product has a high degree of homology to a Ca<sup>2+</sup>-binding protein family. *Genes Dev* 3(7): 1086-93.
- Eckert, R. L., A. M. Broome, M. Ruse, *et al.* 2004. S100 proteins in the epidermis. *J Invest Dermatol* 123(1): 23-33.
- Egeland, E. V., K. Boye, D. Park, *et al.* 2017. Prognostic significance of S100A4-expression and subcellular localization in early-stage breast cancer. *Breast Cancer Res Treat* 162(1): 127-137.
- Ema, H. and H. Nakauchi. 2003. Self-renewal and lineage restriction of hematopoietic stem cells. *Curr Opin Genet Dev* 13(5): 508-12.
- Endele, M., M. Etzrodt and T. Schroeder. 2014. Instruction of hematopoietic lineage choice by cytokine signaling. *Exp Cell Res* 329(2): 207-13.
- Estey, E. H. 2018. Acute myeloid leukemia: 2019 update on risk-stratification and management. *Am J Hematol* 93(10): 1267-1291.
- Fabris, L., M. Cadamuro, L. Moserle, *et al.* 2011. Nuclear expression of S100A4 calcium-binding protein increases cholangiocarcinoma invasiveness and metastasization. *Hepatology* 54(3): 890-9.
- FDA, U. S. ( 2018). "FDA approves treatment for adult patients who have relapsed or refractory AML with a certain genetic mutation.". Retrieved 25/10/2019, 2019.
- Feuring-Buske, M., B. Gerhard, J. Cashman, *et al.* 2003. Improved engraftment of human acute myeloid leukemia progenitor cells in beta 2-microglobulin-deficient NOD/SCID mice and in NOD/SCID mice transgenic for human growth factors. *Leukemia* 17(4): 760-3.
- Fischer-Fantuzzi, L. and C. Vesco. 1988. Cell-dependent efficiency of reiterated nuclear signals in a mutant simian virus 40 oncoprotein targeted to the nucleus. *Mol Cell Biol* 8(12): 5495-503.
- Flatmark, K., K. B. Pedersen, J. M. Nesland, *et al.* 2003. Nuclear localization of the metastasis-related protein S100A4 correlates with tumour stage in colorectal cancer. *J Pathol* 200(5): 589-95.
- Formosa, T. 2013. The role of FACT in making and breaking nucleosomes. *Biochim Biophys Acta* 1819(3-4): 247-55.

- Gale, R. E., C. Green, C. Allen, *et al.* 2008. The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood* 111(5): 2776-84.
- Ganguly, B. B. and N. N. Kadam. 2016. Mutations of myelodysplastic syndromes (MDS): An update. *Mutat Res Rev Mutat Res* 769: 47-62.
- Garrett, S. C., K. M. Varney, D. J. Weber and A. R. Bresnick. 2006. S100A4, a mediator of metastasis. *J Biol Chem* 281(2): 677-80.
- Ge, Y., M. B. Schuster, S. Pundhir, *et al.* 2019. The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nat Commun* 10(1): 172.
- Geuens, T., D. Bouhy and V. Timmerman. 2016. The hnRNP family: insights into their role in health and disease. *Hum Genet* 135(8): 851-67.
- Gifford, J. L., M. P. Walsh and H. J. Vogel. 2007. Structures and metal-ion-binding properties of the Ca<sup>2+</sup>-binding helix-loop-helix EF-hand motifs. *Biochem J* 405(2): 199-221.
- Giles, F. J., G. Borthakur, F. Ravandi, *et al.* 2007. The haematopoietic cell transplantation comorbidity index score is predictive of early death and survival in patients over 60 years of age receiving induction therapy for acute myeloid leukaemia. *Br J Haematol* 136(4): 624-7.
- Gingras, A. R., J. Basran, A. Prescott, *et al.* 2008. Crystal structure of the Ca(2+)-form and Ca(2+)-binding kinetics of metastasis-associated protein, S100A4. *FEBS Lett* 582(12): 1651-6.
- Goto, K., H. Endo and T. Fujiyoshi. 1988. Cloning of the sequences expressed abundantly in established cell lines: identification of a cDNA clone highly homologous to S-100, a calcium binding protein. *J Biochem* 103(1): 48-53.
- Greenbaum, A., Y. M. Hsu, R. B. Day, *et al.* 2013. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* 495(7440): 227-30.
- Grignani, F., T. Kinsella, A. Mencarelli, *et al.* 1998. High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. *Cancer Res* 58(1): 14-9.
- Grigorian, M. S., E. M. Tulchinsky, S. Zain, *et al.* 1993. The mts1 gene and control of tumor metastasis. *Gene* 135(1-2): 229-38.
- Gross, S. R., C. G. Sin, R. Barraclough and P. S. Rudland. 2014. Joining S100 proteins and migration: for better or for worse, in sickness and in health. *Cell Mol Life Sci* 71(9): 1551-79.
- Haferlach, C., C. Mecucci, S. Schnittger, *et al.* 2009. AML with mutated NPM1 carrying a normal or aberrant karyotype show overlapping biologic, pathologic, immunophenotypic, and prognostic features. *Blood* 114(14): 3024-32.

- Haferlach, T., A. Kohlmann, L. Wiczorek, *et al.*2010. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J Clin Oncol* 28(15): 2529-37.
- Hanack, K., K. Messerschmidt and M. Listek.2016. Antibodies and Selection of Monoclonal Antibodies. *Adv Exp Med Biol* 917: 11-22.
- He, X., Y. Deng and W. Yue.2017. Investigating critical genes and gene interaction networks that mediate cyclophosphamide sensitivity in chronic myelogenous leukemia. *Mol Med Rep* 16(1): 523-532.
- Heizmann, C. W., G. Fritz and B. Schafer.2002. S100 proteins: structure, functions and pathology. *Front Biosci* 7(1): 1356-1368.
- Heizmann, C. W., G. Fritz and B. W. Schafer.2002. S100 proteins: structure, functions and pathology. *Front Biosci* 7: d1356-68.
- Helfman, D. M., E. J. Kim, E. Lukanidin and M. Grigorian.2005. The metastasis associated protein S100A4: role in tumour progression and metastasis. *Br J Cancer* 92(11): 1955-8.
- Hilpert, J., L. Grosse-Hovest, F. Grunebach, *et al.*2012. Comprehensive analysis of NKG2D ligand expression and release in leukemia: implications for NKG2D-mediated NK cell responses. *J Immunol* 189(3): 1360-71.
- Hodge, K., S. T. Have, L. Hutton and A. I. Lamond.2013. Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics* 88: 92-103.
- Hofmann, S., M. L. Schubert, L. Wang, *et al.*2019. Chimeric Antigen Receptor (CAR) T Cell Therapy in Acute Myeloid Leukemia (AML). *J Clin Med* 8(2).
- Hoggatt, J. and L. M. Pelus.2011. Mobilization of hematopoietic stem cells from the bone marrow niche to the blood compartment. *Stem Cell Res Ther* 2(2): 13.
- Holzinger, A., M. Barden and H. Abken.2016. The growing world of CAR T cell trials: a systematic review. *Cancer Immunol Immunother* 65(12): 1433-1450.
- Hou, H. A., C. Y. Liu, Y. Y. Kuo, *et al.*2016. Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia. *Oncotarget* 7(8): 9084-101.
- Hou, S., T. Tian, D. Qi, *et al.*2018. S100A4 promotes lung tumor development through beta-catenin pathway-mediated autophagy inhibition. *Cell Death Dis* 9(3): 277.
- Hsu, K. S. and H. Y. Kao.2018. PML: Regulation and multifaceted function beyond tumor suppression. *Cell Biosci* 8: 5.

- Huang, L., Y. Xu, G. Cai, Z. Guan and S. Cai.2012. Downregulation of S100A4 expression by RNA interference suppresses cell growth and invasion in human colorectal cancer cells. *Oncol Rep* 27(4): 917-22.
- Ishimi, Y.2018. Regulation of MCM2-7 function. *Genes Genet Syst* 93(4): 125-133.
- Ismail, N. I., G. Kaur, H. Hashim and M. S. Hassan.2008. S100A4 overexpression proves to be independent marker for breast cancer progression. *Cancer Cell Int* 8: 12.
- Ismail, T. M., D. G. Fernig, P. S. Rudland, *et al.*2008. The basic C-terminal amino acids of calcium-binding protein S100A4 promote metastasis. *Carcinogenesis* 29(12): 2259-66.
- Iwasaki, H.2013. [Roles of transcription factors in normal and leukemic hematopoiesis]. *Rinsho Ketsueki* 54(10): 1566-72.
- Jaiswal, S., P. Fontanillas, J. Flannick, *et al.*2014. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 371(26): 2488-98.
- Jan, M., T. M. Snyder, M. R. Corces-Zimmerman, *et al.*2012. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* 4(149): 149ra118.
- Kaboord, B., S. Smith, B. Patel and S. Meier.2015. Enrichment of low-abundant protein targets by immunoprecipitation upstream of mass spectrometry. *Methods Mol Biol* 1295: 135-51.
- Kanji, S., V. J. Pompili and H. Das.2011. Plasticity and maintenance of hematopoietic stem cells during development. *Recent Pat Biotechnol* 5(1): 40-53.
- Katayama, Y.2019. [Fate decision of hematopoietic stem cells by Wnt signaling.]. *Clin Calcium* 29(3): 343-347.
- Kayser, S. and M. J. Levis.2018. Advances in targeted therapy for acute myeloid leukaemia. *Br J Haematol* 180(4): 484-500.
- Kho, P. S., L. Jankova, C. L. Fung, *et al.*2012. Overexpression of protein S100A4 is independently associated with overall survival in stage C colonic cancer but only in cytoplasm at the advancing tumour front. *Int J Colorectal Dis* 27(11): 1409-17.
- Khwaja, A., M. Bjorkholm, R. E. Gale, *et al.*2016. Acute myeloid leukaemia. *Nat Rev Dis Primers* 2: 16010.
- Kikuchi, N., A. Horiuchi, R. Osada, *et al.*2006. Nuclear expression of S100A4 is associated with aggressive behavior of epithelial ovarian carcinoma: an important autocrine/paracrine factor in tumor progression. *Cancer Sci* 97(10): 1061-9.
- Koeffler, H. P. and D. W. Golde.1978. Acute myelogenous leukemia: a human cell line responsive to colony-stimulating activity. *Science* 200(4346): 1153-4.

- Kohlmann, A., T. J. Kipps, L. Z. Rassenti, *et al.* 2008. An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase. *Br J Haematol* 142(5): 802-7.
- Koshelev Iu, A., G. P. Georgiev and A. V. Kibardin. 2008. [Functions of protein MTS1 (S100A4) in normal and tumor cells]. *Genetika* 44(2): 149-62.
- Kronenwett, R. and R. Haas. 2006. [Differentiation potential of stem cells from bone marrow]. *Med Klin (Munich)* 101 Suppl 1: 182-5.
- Krug, U., C. Rollig, A. Koschmieder, *et al.* 2010. Complete remission and early death after intensive chemotherapy in patients aged 60 years or older with acute myeloid leukaemia: a web-based application for prediction of outcomes. *Lancet* 376(9757): 2000-8.
- Kumar, S. and R. Agrawal. 2014. Next generation tyrosine kinase inhibitor (TKI): afatinib. *Recent Pat Anticancer Drug Discov* 9(3): 382-93.
- Kung Sutherland, M. S., R. B. Walter, S. C. Jeffrey, *et al.* 2013. SGN-CD33A: a novel CD33-targeting antibody-drug conjugate using a pyrrolobenzodiazepine dimer is active in models of drug-resistant AML. *Blood* 122(8): 1455-63.
- Laouedj, M., M. R. Tardif, L. Gil, *et al.* 2017. S100A9 induces differentiation of acute myeloid leukemia cells through TLR4. *Blood* 129(14): 1980-1990.
- Larsson, C. A., G. Cote and A. Quintas-Cardama. 2013. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol Cancer Res* 11(8): 815-27.
- Lee, M., A. M. Dworkin, J. Lichtenberg, *et al.* 2014. Metastasis-associated protein ribosomal RNA processing 1 homolog B (RRP1B) modulates metastasis through regulation of histone methylation. *Mol Cancer Res* 12(12): 1818-28.
- Lee, Y. C., J. A. Baath, R. M. Bastle, *et al.* 2018. Impact of Detergents on Membrane Protein Complex Isolation. *J Proteome Res* 17(1): 348-358.
- Lewit-Bentley, A. and S. Rety. 2000. EF-hand calcium-binding proteins. *Curr Opin Struct Biol* 10(6): 637-43.
- Ley, T. J., C. Miller, L. Ding, *et al.* 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22): 2059-74.
- Li, H., Z. Liu, C. Xu, *et al.* 2013. Overexpression of S100A4 is closely associated with the progression and prognosis of gastric cancer in young patients. *Oncol Lett* 5(5): 1485-1490.
- Li, Z. H. and A. R. Bresnick. 2006. The S100A4 metastasis factor regulates cellular motility via a direct interaction with myosin-IIA. *Cancer Res* 66(10): 5173-80.

- Lin, J. S. and E. M. Lai.2017. Protein-Protein Interactions: Co-Immunoprecipitation. *Methods Mol Biol* 1615: 211-219.
- Linch, D. C., R. K. Hills, A. K. Burnett, A. Khwaja and R. E. Gale.2014. Impact of FLT3(ITD) mutant allele level on relapse risk in intermediate-risk acute myeloid leukemia. *Blood* 124(2): 273-6.
- Liriano, M. A., K. M. Varney, N. T. Wright, *et al.*2012. Target binding to S100B reduces dynamic properties and increases Ca(2+)-binding affinity for wild type and EF-hand mutant proteins. *J Mol Biol* 423(3): 365-85.
- Liu, J., Z. M. Xu, G. B. Qiu, *et al.*2014. S100A4 is upregulated via the binding of c-Myb in methylation-free laryngeal cancer cells. *Oncol Rep* 31(1): 442-9.
- Liu, Y. P. and B. Berkhout.2014. HIV-1-based lentiviral vectors. *Methods Mol Biol* 1087: 273-84.
- Lo, J. F., C. C. Yu, S. H. Chiou, *et al.*2011. The epithelial-mesenchymal transition mediator S100A4 maintains cancer-initiating cells in head and neck cancers. *Cancer Res* 71(5): 1912-23.
- Lucas, D.2017. The Bone Marrow Microenvironment for Hematopoietic Stem Cells. *Adv Exp Med Biol* 1041: 5-18.
- Luskin, M. R., J. W. Lee, H. F. Fernandez, *et al.*2016. Benefit of high-dose daunorubicin in AML induction extends across cytogenetic and molecular groups. *Blood* 127(12): 1551-8.
- Malashkevich, V. N., K. M. Varney, S. C. Garrett, *et al.*2008. Structure of Ca<sup>2+</sup>-bound S100A4 and its interaction with peptides derived from nonmuscle myosin-IIA. *Biochemistry* 47(18): 5111-26.
- Malcovati, L., E. Papaemmanuil, D. T. Bowen, *et al.*2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* 118(24): 6239-46.
- Mandinova, A., D. Atar, B. W. Schafer, *et al.*1998. Distinct subcellular localization of calcium binding S100 proteins in human smooth muscle cells and their relocation in response to rises in intracellular calcium. *J Cell Sci* 111 ( Pt 14): 2043-54.
- Marcon, E., H. Jain, A. Bhattacharya, *et al.*2015. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat Methods* 12(8): 725-31.
- Marenholz, I., C. W. Heizmann and G. Fritz.2004. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem Biophys Res Commun* 322(4): 1111-22.
- Matsuzaki, Y. and Y. Sato.2015. Hematopoietic stem cells. *Nihon Rinsho* 73 Suppl 5: 215-22.

- Matthiesen, R., L. Azevedo, A. Amorim and A. S. Carvalho.2011. Discussion on common data analysis strategies used in MS-based proteomics. *Proteomics* 11(4): 604-19.
- McGinty, R. K. and S. Tan.2015. Nucleosome structure and function. *Chem Rev* 115(6): 2255-73.
- Mencia, N., E. Selga, I. Rico, *et al.*2010. Overexpression of S100A4 in human cancer cell lines resistant to methotrexate. *BMC Cancer* 10: 250.
- Mendelson, A. and P. S. Frenette.2014. Hematopoietic stem cell niche maintenance during homeostasis and regeneration. *Nat Med* 20(8): 833-46.
- Miliotou, A. N. and L. C. Papadopoulou.2018. CAR T-cell Therapy: A New Era in Cancer Immunotherapy. *Curr Pharm Biotechnol* 19(1): 5-18.
- Miller, A. D.2014. Retroviral vector production. *Curr Protoc Hum Genet* 80: Unit 12.5.
- Miranda, K. J., R. F. Loeser and R. R. Yammani.2010. Sumoylation and nuclear translocation of S100A4 regulate IL-1beta-mediated production of matrix metalloproteinase-13. *J Biol Chem* 285(41): 31517-24.
- Mishra, S. K., H. R. Siddique and M. Saleem.2012. S100A4 calcium-binding protein is key player in tumor progression and metastasis: preclinical and clinical evidence. *Cancer Metastasis Rev* 31(1-2): 163-72.
- Moore, B. W.1965. A soluble protein characteristic of the nervous system. *Biochem Biophys Res Commun* 19(6): 739-44.
- Moran-Crusio, K., L. Reavie, A. Shih, *et al.*2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20(1): 11-24.
- Moravkova, P., D. Kohoutova, S. Rejchrt, J. Cyrany and J. Bures.2016. Role of S100 Proteins in Colorectal Carcinogenesis. *Gastroenterol Res Pract* 2016: 2632703.
- Mosaad, Y. M.2014. Hematopoietic stem cells: an overview. *Transfus Apher Sci* 51(3): 68-82.
- Munje, C., R. K. Hills, A. Whetton, *et al.*2015. Cord Blood-Derived Quiescent CD34+ Cells Are More Transcriptionally Matched to AML Blasts Than Cytokine-Induced Normal Human Hematopoietic CD34+ Cells. *Gene Expr* 16(4): 169-75.
- Naaman, C. E. L., B. Grum-Schwensen, A. Mansouri, *et al.*2004. Cancer predisposition in mice deficient for the metastasis-associated Mts1(S100A4) gene. *Oncogene* 23(20): 3670-80.
- Naetar, N., S. Ferraioli and R. Foisner.2017. Lamins in the nuclear interior - life outside the lamina. *J Cell Sci* 130(13): 2087-2096.
- Naoe, T. and H. Kiyoi.2013. Gene mutations of acute myeloid leukemia in the genome era. *Int J Hematol* 97(2): 165-74.

- Nicolas, E., C. Ramus, S. Berthier, *et al.* 2011. Expression of S100A8 in leukemic cells predicts poor survival in de novo AML patients. *Leukemia* 25(1): 57-65.
- Nombela-Arrieta, C., G. Pivarnik, B. Winkel, *et al.* 2013. Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nat Cell Biol* 15(5): 533-43.
- Ochiya, T., K. Takenaga and H. Endo. 2014. Silencing of S100A4, a metastasis-associated protein, in endothelial cells inhibits tumor angiogenesis and growth. *Angiogenesis* 17(1): 17-26.
- Oppenheim, J. J. 2001. Cytokines: past, present, and future. *Int J Hematol* 74(1): 3-8.
- Orre, L. M., E. Panizza, V. O. Kaminsky, *et al.* 2013. S100A4 interacts with p53 in the nucleus and promotes p53 degradation. *Oncogene* 32(49): 5531-40.
- Paczulla, A. M., K. Rothfelder, S. Raffel, *et al.* 2019. Absence of NKG2D ligands defines leukaemia stem cells and mediates their immune evasion. *Nature* 572(7768): 254-259.
- Park, S. W., C. W. Pyo and S. Y. Choi. 2015. High-efficiency lentiviral transduction of primary human CD34(+) hematopoietic cells with low-dose viral inocula. *Biotechnol Lett* 37(2): 281-8.
- Pathuri, P., L. Vogeley and H. Luecke. 2008. Crystal structure of metastasis-associated protein S100A4 in the active calcium-bound form. *J Mol Biol* 383(1): 62-77.
- Paull, T. T. 2015. Mechanisms of ATM Activation. *Annu Rev Biochem* 84: 711-38.
- Pellett, S. and J. W. Tracy. 2006. Mak16p is required for the maturation of 25S and 5.8S rRNAs in the yeast *Saccharomyces cerevisiae*. *Yeast* 23(7): 495-506.
- Pouzolles, M., L. Oburoglu, N. Taylor and V. S. Zimmermann. 2016. Hematopoietic stem cell lineage specification. *Curr Opin Hematol* 23(4): 311-7.
- Qualtiere, L. F., A. G. Anderson and P. Meyers. 1977. Effects of ionic and nonionic detergents on antigen-antibody reactions. *J Immunol* 119(5): 1645-51.
- Reinisch, A., S. M. Chan, D. Thomas and R. Majeti. 2015. Biology and Clinical Relevance of Acute Myeloid Leukemia Stem Cells. *Semin Hematol* 52(3): 150-64.
- Robb, L. 2007. Cytokine receptors and hematopoietic differentiation. *Oncogene* 26(47): 6715-23.
- Rocquain, J., V. Gelsi-Boyer, J. Adelaide, *et al.* 2010. Alteration of cohesin genes in myeloid diseases. *Am J Hematol* 85(9): 717-9.
- Rudland, P. S., A. Platt-Higgins, C. Renshaw, *et al.* 2000. Prognostic significance of the metastasis-inducing protein S100A4 (p9Ka) in human breast cancer. *Cancer Res* 60(6): 1595-603.

- Sack, U., W. Walther, D. Scudiero, *et al.* 2011. S100A4-induced cell motility and metastasis is restricted by the Wnt/beta-catenin pathway inhibitor calcimycin in colon cancer cells. *Mol Biol Cell* 22(18): 3344-54.
- Sakuma, T., M. A. Barry and Y. Ikeda. 2012. Lentiviral vectors: basic to translational. *Biochem J* 443(3): 603-18.
- Saleem, M., A. A. Ganaie, R. Maqbool, *et al.* 2017. Abstract 1586: Developing novel inhibitors of S100A4 for neuroendocrine (NE) and metastatic prostate cancer: systematic testing using relevant models and drug development techniques. *Cancer Res* 77(13 Supplement): 1586-1586.
- Saleem, M., M. H. Kweon, J. J. Johnson, *et al.* 2006. S100A4 accelerates tumorigenesis and invasion of human prostate cancer through the transcriptional regulation of matrix metalloproteinase 9. *Proc Natl Acad Sci U S A* 103(40): 14825-30.
- Sanchez-Aguilera, A. and S. Mendez-Ferrer. 2017. The hematopoietic stem-cell niche in health and leukemia. *Cell Mol Life Sci* 74(4): 579-590.
- Sanden, C. and U. Gullberg. 2015. The DEK oncoprotein and its emerging roles in gene regulation. *Leukemia* 29(8): 1632-6.
- Sanjuan-Pla, A., I. C. Macaulay, C. T. Jensen, *et al.* 2013. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* 502(7470): 232-6.
- Santamaria-Kisiel, L., A. C. Rintala-Dempsey and G. S. Shaw. 2006. Calcium-dependent and -independent interactions of the S100 protein family. *Biochem J* 396(2): 201-14.
- Schlenk, R. F., E. Taskesen, Y. van Norden, *et al.* 2013. The value of allogeneic and autologous hematopoietic stem cell transplantation in prognostically favorable acute myeloid leukemia with double mutant CEBPA. *Blood* 122(9): 1576-82.
- Schmidt-Hansen, B., J. Klingelhofer, B. Grum-Schwensen, *et al.* 2004. Functional significance of metastasis-inducing S100A4(Mts1) in tumor-stroma interplay. *J Biol Chem* 279(23): 24498-504.
- Scott, C. E. and P. M. Kekenes-Huskey. 2016. Molecular Basis of S100A1 Activation at Saturating and Subsaturating Calcium Concentrations. *Biophys J* 110(5): 1052-63.
- Seif, A. E. 2011. Pediatric leukemia predisposition syndromes: clues to understanding leukemogenesis. *Cancer Genet* 204(5): 227-44.
- Semov, A., M. J. Moreno, A. Onichtchenko, *et al.* 2005. Metastasis-associated protein S100A4 induces angiogenesis through interaction with Annexin II and accelerated plasmin formation. *J Biol Chem* 280(21): 20833-41.
- Seval, G. C. and M. Ozcan. 2015. Treatment of Acute Myeloid Leukemia in Adolescent and Young Adult Patients. *J Clin Med* 4(3): 441-59.
- Shi, M., M. Wu, P. Pan and R. Zhao. 2014. Network-based sub-network signatures unveil the potential for acute myeloid leukemia therapy. *Mol Biosyst* 10(12): 3290-7.

- Shlush, L. I., S. Zandi, A. Mitchell, *et al.* 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 506(7488): 328-33.
- Short, N. J., M. E. Rytting and J. E. Cortes. 2018. Acute myeloid leukaemia. *Lancet* 392(10147): 593-606.
- Signore, M., V. Manganelli and A. Hodge. 2017. Antibody Validation by Western Blotting. *Methods Mol Biol* 1606: 51-70.
- Song, K., L. Li, Y. Wang and T. Liu. 2016. Hematopoietic stem cells: multiparameter regulation. *Hum Cell* 29(2): 53-7.
- Stein, M. K., M. Crawley, E. Vick and M. G. Martin. 2018. Concomitant ATM Mutations Identified by Next Generation Sequencing in a Patient With New-Onset Acute Myeloid Leukemia Following Imatinib Treatment for Chronic Myeloid Leukemia. *World J Oncol* 9(2): 66-67.
- Stein, U., F. Arlt, J. Smith, *et al.* 2011. Intervening in beta-catenin signaling by sulindac inhibits S100A4-dependent colon cancer metastasis. *Neoplasia* 13(2): 131-44.
- Stein, U., F. Arlt, W. Walther, *et al.* 2006. The metastasis-associated gene S100A4 is a novel target of beta-catenin/T-cell factor signaling in colon cancer. *Gastroenterology* 131(5): 1486-500.
- Steinbach, D., N. Pfaffendorf, S. Wittig and B. Gruhn. 2007. PRAME expression is not associated with down-regulation of retinoic acid signaling in primary acute myeloid leukemia. *Cancer Genet Cytogenet* 177(1): 51-4.
- Strynadka, N. C. and M. N. James. 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu Rev Biochem* 58: 951-98.
- Szade, K., G. S. Gulati, C. K. F. Chan, *et al.* 2018. Where Hematopoietic Stem Cells Live: The Bone Marrow Niche. *Antioxid Redox Signal* 29(2): 191-204.
- Szklarczyk, D., A. Franceschini, S. Wyder, *et al.* 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43(Database issue): D447-52.
- Tajeddine, N., J. L. Gala, M. Louis, *et al.* 2005. Tumor-associated antigen preferentially expressed antigen of melanoma (PRAME) induces caspase-independent cell death in vitro and reduces tumorigenicity in vivo. *Cancer Res* 65(16): 7348-55.
- Takenaga, K., Y. Nakamura and S. Sakiyama. 1994. Cellular localization of pEL98 protein, an S100-related calcium binding protein, in fibroblasts and its tissue distribution analyzed by monoclonal antibodies. *Cell Struct Funct* 19(3): 133-41.
- Tamma, R. and D. Ribatti. 2017. Bone Niches, Hematopoietic Stem Cells, and Vessel Formation. *Int J Mol Sci* 18(1).

- Taylor, J., W. Xiao and O. Abdel-Wahab.2017. Diagnosis and classification of hematologic malignancies on the basis of genetics. *Blood* 130(4): 410-423.
- Taylor, S., S. Herrington, W. Prime, P. S. Rudland and R. Barraclough.2002. S100A4 (p9Ka) protein in colon carcinoma and liver metastases: association with carcinoma cells and T-lymphocytes. *Br J Cancer* 86(3): 409-16.
- Tomczak, K., P. Czerwinska and M. Wiznerowicz.2015. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 19(1a): A68-77.
- Tonks, A., L. Pearn, M. Musson, *et al.*2007. Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. *Leukemia* 21(12): 2495-505.
- Tonks, A., A. J. Tonks, L. Pearn, *et al.*2005. Optimized retroviral transduction protocol which preserves the primitive subpopulation of human hematopoietic cells. *Biotechnol Prog* 21(3): 953-8.
- Turriziani, B., A. von Kriegsheim and S. R. Pennington.2016. Protein-Protein Interaction Detection Via Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol* 919: 383-396.
- UniProtConsortium, T.2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45(D1): D158-d169.
- van Dieck, J., D. P. Teufel, A. M. Jaulent, *et al.*2009. Posttranslational modifications affect the interaction of S100 proteins with tumor suppressor p53. *J Mol Biol* 394(5): 922-30.
- Vardiman, J. W., N. L. Harris and R. D. Brunning.2002. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* 100(7): 2292-302.
- Vogel, C. and E. M. Marcotte.2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13(4): 227-32.
- Wang, E., S. X. Lu, A. Pastore, *et al.*2019. Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell* 35(3): 369-384.e7.
- Wang, L., M. S. Lawrence, Y. Wan, *et al.*2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 365(26): 2497-506.
- Wang, L., X. Wang, Y. Liang, X. Diao and Q. Chen.2012. S100A4 promotes invasion and angiogenesis in breast cancer MDA-MB-231 cells by upregulating matrix metalloproteinase-13. *Acta Biochim Pol* 59(4): 593-8.
- Wang, Q., K. Wang and M. Ye.2017. Strategies for large-scale analysis of non-histone protein methylation by LC-MS/MS. *Analyst* 142(19): 3536-3548.
- Wang, T. Y., S. J. Chang, M. D. Chang and H. W. Wang.2009. Unique biological properties and application potentials of CD34+ CD38- stem cells from various sources. *Taiwan J Obstet Gynecol* 48(4): 356-69.

- Wang, X. and S. Li.2014. Protein mislocalization: mechanisms, functions and clinical applications in cancer. *Biochim Biophys Acta* 1846(1): 13-25.
- Wei, Q. and P. S. Frenette.2018. Niches for Hematopoietic Stem Cells and Their Progeny. *Immunity* 48(4): 632-648.
- Welch, J. S., T. J. Ley, D. C. Link, *et al.*2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150(2): 264-78.
- Wilson, V. G.2017. Introduction to Sumoylation. *Adv Exp Med Biol* 963: 1-12.
- Wodak, S. J., S. Pu, J. Vlasblom and B. Seraphin.2009. Challenges and rewards of interaction proteomics. *Mol Cell Proteomics* 8(1): 3-18.
- Xu, J., N. Gross, Y. Zang, *et al.*2019. Overexpression of S100A4 Predicts Migration, Invasion, and Poor Prognosis of Hypopharyngeal Squamous Cell Carcinoma. *Mol Diagn Ther* 23(3): 407-417.
- Xu, Y., X. D. Gao, J. H. Lee, *et al.*2014. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev* 28(11): 1191-203.
- Xu, Y., L. J. Rong, S. L. Meng, *et al.*2016. PRAME promotes in vitro leukemia cells death by regulating S100A4/p53 signaling. *Eur Rev Med Pharmacol Sci* 20(6): 1057-63.
- Yamamoto, R., Y. Morita, J. Oechara, *et al.*2013. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* 154(5): 1112-1126.
- Yang, H., S. Maddipoti, A. Quesada, *et al.*2015. Analysis of class I and II histone deacetylase gene expression in human leukemia. *Leuk Lymphoma* 56(12): 3426-33.
- Yang, W. H., M. J. Ding, G. Z. Cui, M. Yang and D. L. Dai.2018. Heterogeneous nuclear ribonucleoprotein M promotes the progression of breast cancer by regulating the axin/beta-catenin signaling pathway. *Biomed Pharmacother* 105: 848-855.
- Yang, X. J. and E. Seto.2007. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* 26(37): 5310-8.
- Ye, J., D. Luo, J. Yu and S. Zhu.2019. Transcriptome analysis identifies key regulators and networks in Acute myeloid leukemia. *Hematology* 24(1): 487-491.
- Yokota, S., H. Kiyoi, M. Nakao, *et al.*1997. Internal tandem duplication of the FLT3 gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia* 11(10): 1605-9.
- Yu, V. W. and D. T. Scadden.2016. Hematopoietic Stem Cell and Its Bone Marrow Niche. *Curr Top Dev Biol* 118: 21-44.

Zhang, K., M. Yu, F. Hao, A. Dong and D. Chen.2016. Knockdown of S100A4 blocks growth and metastasis of anaplastic thyroid cancer cells in vitro and in vivo. *Cancer Biomark* 17(3): 281-291.

Zhang, S., G. Wang, D. Liu, *et al.*2005. The C-terminal region of S100A4 is important for its metastasis-inducing properties. *Oncogene* 24(27): 4401-11.

Zhao, Y., T. Zhang and Q. Wang.2013. S100 calcium-binding protein A4 is a novel independent prognostic factor for the poor prognosis of gastric carcinomas. *Oncol Rep* 30(1): 111-8.

Zhou, Y., Z. Li, Y. Ding, *et al.*2018. Overexpression of S100A4 protein may be associated with the development and progression of pancreatic cancer. *J Cancer Res Ther* 14(Supplement): S159-s166.

# Appendices

## 7 Appendices

### **Appendix 1**

*To view the complete AML patient's data, please refer to the attached electronic document on Google Drive Link Below:*

<https://drive.google.com/open?id=1Y-1WNC2qklBduY2DNL9MuzjpROfmQ2Jv>

### **Appendix 2**

*To view the complete LC/MS raw data, please refer to the attached electronic document on Google Drive Link Below:*

<https://drive.google.com/open?id=1PfOup6dqB9tpuVHUU51KlZtYWnnCXcQG>

### **Appendix 3**

*To view the complete LC/MS raw and analysed data of Cytoplasmic and Nuclear proteomic, please refer to the attached electronic document on Google Drive Link Below:*

[https://drive.google.com/open?id=17dLkYZZz\\_XjZpt-2ceVvJj-A9x02XuDf](https://drive.google.com/open?id=17dLkYZZz_XjZpt-2ceVvJj-A9x02XuDf)