

**Characterisation of a novel protein, ANKRD18A,
implicated in a severe form of thrombocytopenia**

by

Jane Wong Fütterer, MD

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Institute of Cardiovascular Sciences,
College of Medical and Dental Sciences,
University of Birmingham
September, 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Platelets (thrombocytes) mediate clot formation following vessel injury. Inherited thrombocytopenias, characterised by platelet counts below $150 \times 10^9 \text{ L}^{-1}$, have variable bleeding severity.

Whole exome sequencing of two cousins with severe thrombocytopenia identified a shared single-codon deletion in ANKRD18A. ANKRD18A's function is unknown. We hypothesise that this mutation in ANKRD18A is the aetiology of these children's thrombocytopenia.

Despite limited patients' blood, quantitative real-time polymerase chain reaction demonstrated reduced ANKRD18A mRNA levels in patient leukocytes compared to controls. Patients' platelets demonstrated decreased activity when stimulated with adenosine diphosphate, collagen related peptide and the synthetic peptide SFLLRN, which mimics the N-terminal sequence of protease activated receptors. Electron microscopy demonstrated macrothrombocytopenia.

To assist characterisation studies, a recombinant fragment of ANKRD18A, the ankyrin repeat domain, was purified to generate a polyclonal antibody, α -25277. Experiments demonstrated possible specificity of α -25277. Cellular functions of ANKRD18A remain unclear, yet the availability of α -25277, once better characterised, will be instrumental in demonstrating the potential link between ANKRD18A and the observed thrombocytopenia. Another in-house antibody, α -21234, may also be effective. Antibody specificity can be proven by immunoblotting lysates of cells, where ANKRD18A is silenced by RNAi. Should both antibodies not be specific, generating a monoclonal antibody is requisite.

Acknowledgements

I would like to thank Dr Neil Morgan and Prof Steve Watson for providing me the opportunity to embark on a PhD research project in a fascinating area at an age where others attend university graduation ceremonies of their sons or daughters.

I owe thanks to members of the Oxford Protein Facility: Ray Owens, Louise Bird and Heather Rada, who were invaluable in helping me come up with the 32 expressing constructs which is the basis of this project, and provided me the opportunity to experience what OPPF has to offer.

Many thanks also to members the Willcox group, Carrie Willcox, Mahb Salim, and Fiyaz Mohammed. Their open door policy allowed me to barge in whenever I needed help, not to mention their generous sharing of reagents, and finally, many thanks for all your support.

I need to recognise colleagues of the IBR first floor: Yotis Senis for three years of good laughs, Saturday philosophical conversations, and on the side, answers to scientific questions; Steve Thomas and Natalie Poulter for letting me pester them with confocal problems; Tanya Brill and Julie Reyes, for being such pleasant office mates and many, many thanks for all your help with the Accuri B6; Steph Watson, thanks for being the SUPER TECHNICIAN; Ben Johnson, for being cheerful everyday and for your help with genetics, Abdullah Khan, for his expertise on ImageJ and the confocal, and finally, Sarah Fletcher, many, many thanks for her infinite patience when on the confocal with me. Thanks also to the remaining members of the lab for making it such a friendly environment to work in. Special thanks to Kabir Khan, formerly from the Bicknell lab, for keeping me company during late night and weekend sessions.

My work was funded through a PhD stipend by the British Heart Foundation.

Finally, I could not have completed the three years of lab work nor written this thesis without the understanding and encouragement of my family.

Table of Contents

Table of Contents	4
List of Figures.....	9
List of Tables	16
List of Supplementary Figures	17
List of Abbreviations	18
1. Introduction	20
1.1. Platelet physiology.....	20
1.1.1. Platelet Activation.....	21
1.1.2. Platelet inactivation	31
1.2. DNA Sequencing.....	35
1.2.1. Next Generation Sequencing.....	38
1.2.2. Exome Sequencing.....	46
1.3. Platelet Disorders	53
1.4. Platelet Biogenesis and Inherited Thrombocytopenias.....	57
1.4.1. Inherited Thrombocytopenias caused by genes affecting Mk production.....	62
1.4.2. Inherited Thrombocytopenias caused by genes affecting Mk maturation	66
1.4.3. Inherited Thrombocytopenias caused by genes affecting platelet production and development.....	68
1.5. Identification of genes responsible for inherited thrombocytopenias.....	77
1.6. A family with severe thrombocytopenia	80

1.6.1. A genetic variant in <i>ANKRD18A</i> correlates with severe thrombocytopenia	81
1.7. Aims of this project.....	92
2. Materials and methods	93
2.1. Cell Culture	93
2.2. Transient Transfections.....	93
2.2.1. HEK 293T cells	93
2.2.2. Dami cells	94
2.3. Preparation of cellular lysates.....	95
2.3.1. Washed Platelets	95
2.3.2. Platelet lysates.....	96
2.3.3. Lysates of HEK293T (adherent cells) and Dami (suspended) cells	96
2.4. Construct design utilizing high-throughput cloning and expression at the Oxford Protein Production Facility (OPPF).....	97
2.4.1. ANKRD18A Functional Domains Used.....	97
2.4.2. Vectors used	97
2.4.3. Cloning.....	98
2.4.4. HEK293T transient transfections	99
2.4.5. <i>E. coli</i> transformations and protein expression.....	99
2.5. Plasmid DNA Preparation.....	102
2.6. SDS-PAGE (sodium dodecyl sulphate polyacrylamide gels) and Western Blotting	103
2.6.1. Immunoprecipitation (IP)	104
2.7. Protein Purification	105
2.7.1. Preparation of cleared lysate for Ni ²⁺ -NTA column chromatography	105
2.7.2. Preparation of Ni ²⁺ -NTA Column.....	105

2.7.3.	Preparation of Anion Exchange Column	107
2.7.4.	Concentration of eluted protein	109
2.7.5.	Passivation of Amicon ultra-filtration membranes.....	110
2.8.	Coomassie stain.....	110
2.9.	Lentiviral production.....	110
2.9.1.	Concentrating Lentivirus.....	111
2.10.	Isolating cDNA for RT-PCR and quantitative real time PCR	111
2.10.1.	Preparing quantitative real-time reaction plates.....	112
2.11.	Immunofluorescence confocal microscopy	113
2.11.1.	Plasmids transfected into HEK293T cells and HeLa cells for confocal microscopy	115
2.12.	Platelet spreading.....	115
2.13.	Flow Cytometry.....	115
2.14.	Isolation of CD34⁺ cells from human umbilical cord blood by density gradient centrifugation	116
2.14.1.	Isolation of Mononuclear Cells	116
2.14.2.	Isolation of CD34 ⁺ cells from mononuclear cells.....	117
2.14.3.	Magnetic Separation	118
2.14.4.	Evaluation of hematopoietic progenitor cell purity.....	119
2.14.5.	Expansion and differentiation of CD34 ⁺ cells.....	119
2.14.6.	Evaluation of ploidy in megakaryocytes	119
2.15.	Transmission electron microscopy	120
2.16.	Statistical analysis of α-granules in electron micrographs	120
3.	Initial characterisation of the ANKRD18A variant.....	121
3.1.	Introduction	121

3.2. Quantitation of <i>ANKRD18A</i> mRNA expression in patient platelets	121
3.3. Functional Studies using patient platelets.....	124
3.4. Electron microscopy of <i>ANKRD18A</i> patient platelets.....	129
3.5. Utilising full-length wild-type and mutant <i>ANKRD18A</i> plasmids for transfections and protein expression	131
3.5.1. Determining transfection efficiencies of full-length GFP-tagged <i>ANKRD18A</i> constructs	131
3.6. Summary	139
4. Characterisation of <i>ANKRD18A</i> peptide-derived antibodies	142
4.1. Introduction	142
4.2. Comparison of <i>ANKRD18B</i> and <i>ANKRD18A</i>	143
4.3. Initial characterisation of existing <i>ANKRD18A</i> antibodies.....	148
4.4. <i>ANKRD18A</i> and <i>ANKRD18B</i> in megakaryocyte development.....	155
4.5. Probing for <i>ANKRD18A</i> and <i>ANKRD18B</i> protein expression.....	157
4.6. Summary	159
5. Generation of <i>ANKRD18A</i> recombinant protein and antibody production 163	
5.1. Introduction	163
5.2. Generation of <i>ANKRD18A</i> recombinant protein.....	164
5.2.1. Establishment and selection of <i>ANKRD18A</i> constructs to provide suitable expression levels.....	164
5.2.2. Testing expression of <i>ANKRD18A</i> constructs	166
5.2.3. Testing of transfection agents	174
5.2.4. Purification of selected <i>ANKRD18A</i> recombinant proteins.....	176

5.2.5. Proteolytic cleavage of the MBP tag from the ANKRD18A recombinant protein (clone F2).....	180
5.2.6. Production and purification of the ANKRD18A recombinant protein (clone G2)	187
5.3. Characterisation of ANKRD18A polyclonal antibody 25277	190
5.3.1. Assessing specificity of α -25277	191
5.3.2. Immunoblotting to assess ANKRD18A antibody 25277	196
5.4. Summary	201
6. Discussion	205
6.1. Conclusions from work of this thesis.....	205
6.1.1. Evolutionary context of ANKRD18A and its link to platelet disorders.....	205
6.1.2. Generation of ANKRD18A recombinant protein and subsequent antibodies	207
6.1.3. Characterisation of α -25277	212
6.1.4. Comparison of experimental findings with predictions by the Blueprint progenitors database	215
6.2. Future research.....	219
6.3. Final conclusion	221
7. Supplementary Data	225
References	231

List of Figures

Figure 1.1 - Platelet activation at sites of vascular injury. ECM, extracellular....	22
Figure 1.2 - Signalling pathways of 3 major platelet adhesion receptors: GPIb-IX-V, GPVI and GPIIb/IIIa (integrin α Ib β 3). sGC, soluble guanylyl cyclase; eNOS, endothelial nitric oxide synthase (Li, Delaney et al. 2010).....	23
Figure 1.3 - Platelet G protein coupled receptors' signalling. PGI ₂ , prostacyclin; PI, prostacyclin receptor; 5HT, serotonin; SFK, Src family kinases; PI3K, phosphoinositide 3-kinase; PDE3, phosphodiesterase-3; NOS, nitric oxide synthase; NO, nitric oxide; sGC, soluble guanylate cyclase; PKG, phosphokinase G; MAPK, mitogen-activated protein kinases; PLA2, phospholipase A2; IP3, inositol triphosphate; DAG, diacyl glycerol; CalDAGGEF1, calcium and DAG-regulated guanine nucleotide exchange factor 1; PKC, phosphokinase C; RIAM, Rap1-GTP-interacting adapter molecule; GEF, guanine nucleotide exchange factor; AC, adenylate cyclase; TXA ₂ , thromboxane; ADP, adenosine diphosphate (Li, Delaney et al. 2010).	25
Figure 1.4 - Schematic diagram of the coagulation cascade (Illustration from epomedicine.com)	28
Figure 1.5 - Simplified schematic diagram of the fibrinolytic system. t-PA, tissue plasminogen activator; u-PA, urokinase plasminogen activator; PAI-1, plasminogen activator inhibitor-1; α ₂ -AP, α ₂ -antiplasmin; TAF1, thrombin-activatable fibrinolysis inhibitor. Blue denotes stimulation and red denotes inhibition (Rijken and Lijnen 2009).....	29
Figure 1.6 - Schematic diagram of eicosanoid (prostaglandins and related compounds) synthesis (Illustration from www.tankonyvtar.hu).....	33
Figure 1.7 - Nucleic acid structures and DNA chain elongation. (A) Molecular structures of dideoxyribonucleotide and deoxyribonucleotide (B) Example of DNA chain elongation. (Illustrations from www.ubooks.pub – panel A, and binf.snipcademy.com – panel B).....	36
Figure 1.8 - Changes in the various sequencing instrument read outputs, timeline of newly introduced sequencing instruments and major sequencing projects over the past decade. Top, data output per run plotted on logarithmic scale. Middle, introduction year of various NGS sequencing platforms. Bottom, timeline of the major sequencing projects which have been completed. (Mardis 2011).....	39
Figure 1.9 - Comparison of fragment library construction for (a) mate-end vs. (b) mate-pair sequencing. “A” signifies adaptor and “SP” signifies specific primer (Mardis 2013).	45
Figure 1.10 - Schematic diagram of endomitosis (Geddis 2010).	58

Figure 1.11 – Schematic diagram of platelet development (Geddis 2010).	60
Figure 1.12 - Schematic diagram of the different stages of platelet development associated with the genes which can mutate at each stage. Above each stage are the different genes which can be mutated to cause an inherited thrombocytopenia. <i>THPO</i> is the gene for TPO (Savoia 2016).	62
Figure 1.13 - Structural domain organisation of Wiskott-Aldrich Syndrome protein (WASp). WH2(WASP-homology-2; also called verprolin-homology), C(cofilin-homology or connector), A(acidic region), PP(polyproline domain), GBD(GTPase binding domain), B(basic domain), WH1(WASP-homology-1 domain) (Goley and Welch 2006).	70
Figure 1.14 - Pedigree of patients affected by <i>ANKRD18A</i> -linked severe thrombocytopenia. Shaded square is the male patient (II:5) with the homozygous non-frameshift deletion of <i>ANKRD18A</i> (p.801-801). Shaded circle (II:3) is his cousin. Double lines linking I:1 with I:2 and I:3 with I:4 signify a first cousin union (analysis performed by N. Morgan).	81
Figure 1.15 - Segregation analysis of the exome candidates in family members. The 3 variants (in the genes <i>GNE</i> , <i>FRMPD1</i> and <i>ANKRD18A</i>) were shared by both children and were located within a region of homozygosity on chromosome 9p13.3. Double lines linking parents signify first cousin unions. (analysis performed by N. Morgan)	83
Figure 1.16 - Linear domain organisation of <i>ANKRD18A</i> derived from bioinformatical sequence analysis. Domains (or regions of homology) have been identified with the fold prediction server HHpred using the sequence of <i>ANKRD18A</i> (isoform 2). <i>ANKRD18A</i> isoform 1 has a 63-amino acid insertion, between the ankyrin repeat domain and the tropomyosin homology region. Threading suggests that, compared to isoform 2, isoform 1 (1054 residues) has a more extended ankyrin repeat domain, comprising 11 repeats instead of 9 in the shorter isoform 2 (992 residues). The approximate position of the in-frame deletion mutation (pGlu801del) is indicated.	86
Figure 1.17 - Dendrogram of the <i>ANKRD</i> gene family. <i>ANKRD18A</i> is only expressed in primates whereas <i>ANKRD26</i> expression is not limited to primates.	90
Figure 3.1 - A homozygous <i>ANKRD18A</i> mutation results in decreased expression of <i>ANKRD18A</i> mRNA in 2 patients with a severe thrombocytopenia. <i>ANKRD18A</i> mRNA production, as assessed by real-time quantitative PCR (qRT-PCR) on total RNA isolated from PBMCs from the 2 patients (II:5 and II:3) and 2 healthy controls (C1 and C2). <i>ANKRD18A</i> mRNA levels were expressed as relative expression compared to $\beta 2$ Microglobulin mRNA levels. Values are mean \pm SEM for 3 experiments for all samples. * $P < 0.05$ (N. Morgan, unpublished data).	124
Figure 3.2 – Patients' platelets [(A) II:5 and (B) II:3] have similar surface receptor expression levels compared to control when surface area of patient platelets	

were normalised to control. However, CD42b levels for patient II:5 is reduced and GPVI levels are slightly reduced for patient II:3, both due to unclear reasons (N. Morgan, unpublished data).....	125
Figure 3.3 – Patients’ platelets have decreased responses to α -CD62P (A) and decreased fibrinogen binding to α IIb β 3 (B) after stimulation with low and high dose agonists: ADP, CRP and PAR-1 peptide. Healthy volunteers 1 in 10 = PRP from control individuals diluted 1 in 10 with PBS. Data for healthy volunteers shown as mean \pm 1 s.d. (n=9). Isotype control = IgGk1. (N. Morgan, unpublished data).	128
Figure 3.4 – Transmission electron microscopy image demonstrating patient platelets to be enlarged but number of α granules in patient platelets not significantly different from number of α granules in healthy control platelets. Patient with macrothrombocytopenia but the number of α granules (arrow) remain comparable to control α granule numbers when surface area of platelets were taken into account (N=1). TEM images courtesy of Denai Bem.	130
Figure 3.5 - Lower transfection efficiencies using GFP-full length wild-type and mutant ANKRD18A plasmids compared to GFP-empty vector. (A) Flow cytometry analysis of HEK293T cells untransfected and transfected with plasmids encoding GFP-ANKRD18A (WT), GFP-ANKRD18A (Δ 801), GFP vector control. (B) Immunoblot of HEK293T WCLs treated as in panel A and immunoblotted with α -GFP. Untransfected (lane 1), GFP vector (lane 2) GFP-ANKRD18A (WT) (lane 3), GFP-ANKRD18A (Δ 801) (lane 4). Arrow in left panel identifies the GFP (MW 27 kDa) protein and the arrowhead in the right panel demonstrates where the GFP-tagged full length ANKRD18A protein should be located, 143 kDa. (N=20)	133
Figure 3.6 – Confocal microscopy images of HEK293T cells demonstrating cytoplasmic localisation of WT and MUT ANKRD18A following transient transfection with plasmids encoding full length (B) wild-type GFP-tagged ANKRD18A and (C) mutant GFP-tagged ANKRD18A (Δ 801). Row A demonstrate images of untransfected HEK293T cells. Presence of cells is verified using the nuclear stain TO-PRO3. Scale bar = 10 μ m (N=2).....	136
Figure 3.7 – Immunofluorescence confocal microscopy images of HeLa cells transiently transfected with plasmids encoding (A) myc-tag empty vector (B) myc-tagged WT ANKRD18A and (C) myc-tagged mutant ANKRD18A (Δ 801) demonstrating cytoplasmic localisation of WT and MUT ANKRD18A. Primary antibody used was a monoclonal anti-myc antibody and secondary antibody was α -mouse Alexa Fluor 488. Nuclear staining was with TO-PRO-3 iodide. Scale bar = 8 μ m (N=2).....	138
Figure 4.1 - <i>ANKRD18B</i> mRNA is not expressed in platelets or leukocytes as determined by RT-PCR. Expected size of <i>ANKRD18B</i> is 371 base pairs (bp). Expected size of β -actin is 223 bp. Bands seen in left gel are primer dimers. Lower bands seen in both gels are primer dimers. (N=1)	145

- Figure 4.2 – *ANKRD18A* mRNA is expressed in platelets and leukocytes as determined by RT-PCR. Lanes 1, 5, 9, 13, 16 use *ANKRD18A* primers. Lanes 2, 6,10,14,17 use β -actin primers and Lanes 3, 7,11,15,18 use *GAPDH* primers. P = platelets. B = buffy coat. Expected size of *ANKRD18A* is 394 bp. Expected size of β -actin is 223 bp. Expected size of *GAPDH* is 225 bp. (N = 3)..... 146
- Figure 4.3 – *ANKRD18A* mRNA is expressed in lymphoid cells lines, Damis, HeLAs and HEK293T cells and its derivatives. *ANKRD18B* mRNA is expressed in Damis, HeLAs and HEK293T cells and its derivatives. Lanes 1=H₂O, lane 2= 1654, a B-cell line, lane 3= Lgn, a lymphocyte cell line, lane 4=D+, a HEK293 cell line derivative, lane 5=YTW, a HEK293 derivative, 6= RY, a lymphocyte cell line, 7= HeLa, 8= HK, a lymphocyte cell line, 9=Dami. Upper band is *ANKRD18B* or *ANKRD18A* mRNA. Expected size of *ANKRD18B* is 371 bp. Expected size of *ANKRD18A* is 394 bp. Expected size of *GAPDH* is 225 bp. Lower bands are primer dimers. (N=1) 148
- Figure 4.4 – Sequence alignment of *ANKRD18A* and *ANKRD18B* for residues 1-329 (of *ANKRD18A* amino acid sequence). *ANKRD18B* exists in two isoforms (1011 and 1401 amino acids, respectively). The peptide sequence used as antigens to raise polyclonal antibodies are highlighted by horizontal bars..... 151
- Figure 4.5 – Western blots demonstrating Abgent antibody detecting (A) *ANKRD18A* expression in HEK293T and Dami cells at correct molecular weight of 116 kDa whereas (B) α -21234 detects a dominant band at the incorrect molecular weight. WCLs were analysed by SDS-PAGE and immunoblotted with two different α -*ANKRD18A*, (A) Abgent antibody and (B) α -21234. Lane 1- HEK293T cells, lane 2-adherent Dami cells, lane 3- Dami cells in suspension, lane 4- purified recombinant protein, clone G2 (Mr = 65kDa). Bottom panel: loading control, blotted with α -*GAPDH* (Mr = 37kDa). The expected MW of *ANKRD18A* is 116kDa. Arrows indicate the expected positions of antigens (N=5). 152
- Figure 4.6 – Western blots demonstrating Abgent antibody detecting (A) no *ANKRD18A* expression in platelets whereas (B) α -21234 detects *ANKRD18A* at correct molecular weight. Platelet WCLs analysed by SDS-PAGE immunoblotted with (A) Abgent antibody and (B) α -21234. Lanes 1-5 different donor platelets, lane 6-purified recombinant protein, clone G2. Bottom panels, loading controls. Left bottom panel, Coomassie stained blot of above blot. Right bottom panel, immunoblot for *GAPDH*. Arrow in top right panel indicates *ANKRD18A*. Arrowhead in top right panel indicates clone G2. (N=5). 153
- Figure 4.7 – *ANKRD18A* mRNA expression increases with megakaryocytic development as determined by qRT-PCR. (A) Determining the fold difference in *ANKRD18A* mRNA expression compared to the *GAPDH* reference during 3 stages of megakaryocyte differentiation. Samples are run in triplicate from 2 different cord blood (CB) preparations and retinoblastoma

(RB) cell line as a control. (B) Agarose gel confirming expression of ANKRD18A mRNA from the samples that were run on the qRT-PCR plate. 1/5=1:5 dilution (N=3).	156
Figure 4.8 – No <i>ANKRD18B</i> mRNA expression in developing megakaryocytes. Dami cDNA was used as positive control. Dami 1/5 is cDNA at a dilution of 1:5, CB1 = cord blood sample 1 (N = 1).	157
Figure 4.9 – Western blots demonstrating (A) ANKRD18B expression in HEK293T and Dami cells but (B) no ANKRD18B expression in platelets. HEK293T, Dami and platelet WCLs are immunoblotted with α -ANKRD18B (top panel) and for loading control, α -GAPDH (left bottom panel). Right bottom panel is Coomassie stain for actin, which is the loading control. (N=5).	158
Figure 5.1 - Coomassie-stained SDS-PAGE gel of lysed <i>E. coli</i> (B834) cell cultures expressing plasmids encoding ANKRD18A or fragments thereof. Molecular weight markers are in units of kDa. For clone designations, refer to Table 5.2. (N=1).	169
Figure 5.2 - Coomassie-stained SDS-PAGE gel of lysed <i>E. coli</i> (B834 and Rosetta) cell cultures expressing plasmids encoding ANKRD18A or fragments thereof. Molecular weight markers are in units of kDa. For clone designations, refer to Table 5.2. (N=1).	170
Figure 5.3 - Coomassie-stained SDS-PAGE gel of lysed <i>E. coli</i> (Rosetta) cell cultures expressing plasmids encoding ANKRD18A (or fragments thereof). Molecular weight markers are in units of 10^3 Dalton. For clone designations, refer to Table 5.2. (N=1).	171
Figure 5.4 – Western blot analysis of whole cell lysates from HEK293T cells transiently transfected with expression plasmids encoding ANKRD18A (and fragments thereof) and immunoblotted with anti-His ₆ . Refer to Table 5.2 for clone designations. Molecular weight markers are given in units of kDa. (N=1).	173
Figure 5.5 - Anti-His ₆ blot of whole cell lysates from HEK293T cells transiently transfected with expression plasmids encoding ANKRD18A (and fragments thereof). Refer to Table 5.2 for clone designations. Molecular weight markers are given in units of kDa. (N=1).	174
Figure 5.6 – Anti-His ₆ immunoblot of HEK293T cells transiently transfected with plasmids encoding ANKRD18A fragments and its respective tags using PEI. Refer to Table 5.2 for clone designations and sizes. The highest expression clones (F2 and G2) are encircled in red. (N=2).	175
Figure 5.7 – Coomassie stained SDS-PAGE gel demonstrating purification of clone F2 using Ni ²⁺ -NTA chromatography. Elution fractions using 100-500mM imidazole were collected for dialysis. Arrow points to recombinant protein, MBP-ankyrin repeat protein, clone F2. Molecular markers on the left are in units of kDa. (N=10).	177

Figure 5.8 – Coomassie stained SDS-PAGE gel demonstrating purification of clone F2 using anion exchange chromatography. Fractions eluted using increasing concentrations of NaCl (250-500mM) are pooled for concentration. The arrow points to purified clone F2. (N=10).....	179
Figure 5.9 – Removing MBP tag from recombinant protein MBP-ANKRD18A (clone F2) with HRV3C protease. Lane 1- Uncut F2, lane 2-10 µg F2, 2 µl HRV3C, 8 hours incubation at 4°C, lane 3- 50 µg F2, 6 µl HRV3C, 8 hours at 4°C, lane 4- Control protein, 2 µl HRV3C, 8 hours at 4°C, lane 5- 10 µg F2, 3 µl HRV3C, overnight incubation at 4°C, lane 6- 50 µg F2, 11 µl HRV3C, overnight incubation at 4°C, lane 7- Control, 3 µl HRV3C, overnight incubation at 4°C. Each lane contains 10 µl of sample plus 30 µl of Laemmli sample buffer. (N=5).....	181
Figure 5.10 – Cleaving MBP tag from clone F2 (recombinant protein, MBP-ankyrin repeat domain). Each reaction mixture contains 200 µg of F2 with 40 µl (80 units) of HRV3C protease. Lane 1- unconcentrated F2, lane 2- filtrate from concentration of F2, lanes 3-7 cleaved F2. Each lane contains 10 µl of sample and 30 µl of Laemmli sample buffer. (N=1).....	182
Figure 5.11 – Coomassie stained SDS-PAGE demonstrating purified protein of clone F2 sent to Biogenes for generation of antibody. Lane 1- pre-concentrated protein, lane 2-concentrated protein, lane 3- flow through from concentration. Each lane contains 5 µl of sample and 30 µl of Laemmli sample buffer. Total concentration of protein after concentration is 2.16 mg/ml in total volume of 1 ml. (N=1).....	187
Figure 5.12 – Coomassie stained SDS-PAGE demonstrating purification of precleared lysates from thirty 15cm plates of HEK293T cells transiently transfected with cDNA from clone G2 (E3C Halo-ankyrin repeat domain). Elutions using 300-500mM imidazole are collected for dialysis and further purification using the anion exchange column. Each lane contains 40 µl of sample and 10 µl of Laemmli sample buffer (N=20).....	189
Figure 5.13 – Coomassie stained SDS-PAGE demonstrating purification of pooled Ni ²⁺ -NTA eluants of clone G2 on an anion exchange column. Elutions from lanes 5-7 using 350 to 500mM NaCl are pooled for dialysis and then concentration. Each lane contains 40 µl of sample and 10 µl of Laemmli sample buffer. Arrow points to purified elutions (N=20).	190
Figure 5.14 - Flow cytometry experiment probing specificity of the polyclonal α-ANKRD18A, α-25277, using HEK293T cells. A) and B) Concentrations of antibody stocks were as follows: Non-specific IgG (1 mg/ml), α-rabbit IgG conjugated to FITC (2 mg/ml), α-25277 (0.24 mg/ml). Clone G2 = purified ANKRD18A (1-260) (N=2).	193
Figure 5.15 – Immunofluorescence confocal microscopy images of Dami cells stained with α-His and TO-PRO-3 demonstrating intranuclear localization of fusion protein (clone G2). Dami cells are transiently transfected with cDNA from clone G2 (E3C-Halo ankyrin repeat domain). A) primary antibody: α-	

His monoclonal (1:750), secondary antibody: α -mouse Alexa Fluor 488 (1:300) B) TO-PRO-3 (1:2000) C) Overlay of panels A and B.....	194
Figure 5.16 – Immunofluorescence confocal microscopy images of Dami cells stained with α -His and α -25277 demonstrating intranuclear localisation of ANKRD18A. Dami cells are transiently transfected with cDNA of plasmid for Clone G2. A) primary: α -His6, secondary: α -mouse Alexa Fluor 488, B) primary: α -25277, secondary: α -rabbit Alexa Fluor 647, C) overlay of A and B (N=2).	195
Figure 5.17 – Dot blot demonstrating the antibody dilution of α -25277 to be used for experiments. 1 = WCLs from HEK293T cells, 2 = WCLs from Dami cells differentiated for 24 hours, 3 = WCLs of Dami cells differentiated at 48 hours, 4 = WCLs of Dami cells differentiated at 72 hours. N=2.....	197
Figure 5.18 – Immunoblot of Dami and HEK293T WCLs using α -25277. Blotting antibody dilution is 1:1500 (N=3).	198
Figure 5.19 – Immunoblot of healthy donor platelet WCLs using α -25277. Blotting antibody dilution is 1:1500 (N= 3).	198
Figure 5.20 – Immunoblot of ANKRD18A immunoprecipitated from whole cell lysates of HEK293T and Dami cells. IP antibodies are (A) IgG and (B) α -ANKRD18A, α -25277. Blotting antibodies used are monoclonal (A) IgG and (B) α -ANKRD18A, α -25277 at 1:1500 dilution. Lane 1 - Clone G2, lane 2 - Dami cells, lane 3 - HEK293T cells, lane 4 - BSA (N=2).	200
Figure 5.21 – Immunoblot of ANKRD18A immunoprecipitated from whole cell lysates of healthy donor platelets (D1 and D2). IP antibodies are (A) IgG and (B) α -ANKRD18A, α -25277. Blotting antibodies used are (A) monoclonal IgG and (B) α -ANKRD18A, α -25277 at 1:1500 dilution. Lane 1- Clone G2, lane 2- D1, lane 3- D2, lane 4-BSA (N=1).	201
Figure 6.1 – River plot representing relative mRNA expression of <i>ANKRD26</i> , <i>ANKRD18B</i> and <i>ANKRD18A</i> across the various cell lineages of hematopoiesis. Line thickness increases with more mRNA expression. HSC = hematopoietic stem cell, MPP = multipotent progenitor, LMPP = lymphoid primed multipotent progenitor, CLP = common lymphoid progenitor, CMP = common myeloid progenitor, GMP = granulocyte-monocyte progenitor, MEP = megakaryocyte-erythrocyte progenitor, EB = erythroblast, MK = megakaryocyte (blueprint.haem.cam.ac.uk).....	218
Figure 6.2 – Riverplot representing <i>GNE</i> mRNA expression across the various cell lineages of hematopoiesis (blueprint.haem.cam.ac.uk).	224

List of Tables

Table 1.1 - Comparison of NGS platforms (Mardis 2011).....	39
Table 1.2 – Summary of differences between various exome sequencing platforms (Warr, Robert et al. 2015).	51
Table 1.3 - Overview of inherited thrombocytopenias (confer section 1.3). Frequencies are indicated as follows: *++++, >100 families reported; +++, >50 families reported; ++, > 10 families reported; +, < 10 families reported. AD, autosomal dominant; AR, autosomal recessive; XL, X-linked. (Balduini and Savoia 2012, Balduini, Savoia et al. 2013).	76
Table 2.1 - Protocol for HEK293T transient transfections using PEI.	94
Table 2.2 - Protocol for Dami cells transient transfections using PEI.	95
Table 2.3 - DNA plasmid constructs generated at OPPF. Abbreviations used: 3C = cleavage site for picornavirus endopeptidase 3C; SUMO = small ubiquitin-like modifier; MBP = maltose binding protein; GST = glutathione-S-transferase; HALO = Promega HaloTag; Phyre2 = protein fold prediction server Phyre2 (www.sbg.bio.ic.ac.uk/phyre2); HHpred = homology detection and structure prediction by HMM-HMM comparison (toolkit.tuebingen.mpg.de/hhpred).	101
Table 2.4 - Volumes of Binding and Elution Buffers that are mixed to arrive at the different concentrations of imidazole needed to elute protein from the Ni ⁺² -NTA column.	106
Table 2.5 - Volumes of Start and Elution Buffers mixed to arrive at the concentrations of NaCl needed to elute protein from the anion exchange column.	108
Table 2.6 - Primer sequences used in RT-PCR and Sanger sequencing.....	112
Table 2.7 - Choosing the appropriate MACS column and separator based on total and CD34+ cell numbers	118
Table 5.1 – Functional domains of ANKRD18A. Phyre2 = protein fold prediction server Phyre2 (www.sbg.bio.ic.ac.uk/phyre2); HHpred = homology detection and structure prediction by HMM-HMM comparison (toolkit.tuebingen.mpg.de/hhpred)	165
Table 5.2 - DNA plasmid constructs generated at OPPF. Abbreviations used: 3C = cleavage site for human rhinovirus 3C protease; SUMO = small ubiquitin-like modifier; MBP = maltose binding protein; GST = glutathione-S-transferase; HALO = Promega HaloTag.....	168
Table 5.3 – Summary of ANKRD18A recombinant proteins with detectable expression when transiently transfected into HEK293T cells. The clones highlighted in green were chosen for purification. MBP = maltose binding	

protein, Sumo = small ubiquitin-like modifier, Halo = modified haloalkane dehalogenase; all constructs, except for E3, include a cleavage site for human rhinovirus 3C protease.....	176
Table 5.4 - Flow cytometry experiment probing specificity of the polyclonal α -ANKRD18A, α -25277. Values indicated are the mean FL1-A fluorescence intensities for the conditions shown in the panels in Figure 5.18 above (N=2).	193

List of Supplementary Figures

Figure S 1 – Utilising flow cytometry to probe specificity of α -ANKRD18A-25277 in Dami cells. Dami cells are undifferentiated and impermeabilised (panels A-D) or differentiated and permeabilised (panels E-H). Dami cells are differentiated using TPO (3 ng/ml) and PMA (6.2 ng/ml). Antibody stocks: IgG (1 mg/ml), α -25277 (0.24 mg/ml), and α -rabbit Alexa Fluor 488 (2 mg/ml). IgG used at a dilution of 1:100, α -25277 used at dilution of 1:100 and α -rabbit Alexa Fluor 488 used at dilution 1:500. N=1	225
Figure S 2 – Bar graph demonstrating the mean co-localisation coefficients of confocal microscopy overlay images 1-3. M1 and M2 = Manders split coefficient. Images are differentiated Dami cells transiently transfected with cDNA of Clone G2. Cells are stained with primary antibodies: α -His and α -25277 and secondary antibodies: α -mouse Alexa Fluor 488 and α -rabbit Alexa 647. N=3	226
Figure S 3 – Sequence alignment of ANKRD18A with two isoforms of ANRKD18B. The sequences were aligned using Clustal Omega (McWilliam, Li et al. 2013) and the alignment formatted using Esript (Robert and Gouet 2014). Identical amino acids are indicated by white type on red background, similar residues by red type on white background. Numbers above the aligned sequences indicate residue numbers for ANKRD18A. ANKRD18B_1011 and ANKRD18B_1401 indicate the 1011-residue and 1401-residue alternative splice forms of ANRKD18, respectively.	228
Figure S 4 – Alignment of ANKRD18A sequences from different primates. Hs = human, Gg = gorilla, Pg = chimpanzee, NI = gibbon. The sequences were aligned using Clustal Omega (McWilliam, Li et al. 2013) and the alignment formatted using Esript (Robert and Gouet 2014). The EEE motif mutated in the thrombocytopenic patients is at residues 799 to 801 (human sequence numbers).	230

List of Abbreviations

ACD - acid citrate dextrose	GAPP - genotyping and platelet phenotyping study
ADP - adenosine diphosphate	GEF – guanine nucleotide exchange factor
AEBSF - 4-(2-amino ethyl) benzene sulfonyl fluoride hydrochloride	GFP - green fluorescent protein
ALL - acute lymphoblastic leukemia	GPVI - glycoprotein VI
AML - acute myelogenous leukemia	GP1b/V/IX - glycoprotein 1b/V/IX
ARD - ankyrin repeat domain	GPS - gray platelet syndrome
ATP - adenosine triphosphate	GT - Glanzmann thrombasthenia
ANKRD - ankyrin repeat domain	HEK - human endothelial kidney
BSA - bovine serum albumin	HRP - horseradish peroxidase
BSS - Bernard Soulier syndrome	HSC - haematopoietic stem cell
CAMT - congenital amegakaryocytic thrombocytopenia	IP – immunoprecipitation
CMP - common myeloid progenitor	IP ₃ – inositol triphosphate
CRP- collagen related peptide	IT - inherited thrombocytopenia
DAG – diacyl glycerol	ITP - immune thrombocytopenia purpura
DMEM - Dulbecco's modified eagle medium	LTA-light transmission aggregometry
ECL- enhanced chemiluminescence	MAPK-mitogen-activated protein kinase
EDTA - Ethylene diamine tetra-acetic acid	MCV - mean corpuscular volume
EGTA - Ethylene glycol-bis (β amino ethyl ether)-N'N'N'N' tetra-acetic acid	MEP - megakaryocyte-erythroid progenitor
ER - endoplasmic reticulum	MK - megakaryocyte
FBS - fetal bovine serum	MPV - mean platelet volume
FT – flow through	MUT - mutant
GA - glutaldehyde	MW - molecular weight
	NGS - next generation sequencing

NO – nitric oxide	RT-PCR- reverse transcription
NOS – nitric oxide synthase	polymerase chain reaction
NP-40/IGEPAL CA-630 –	RUSAT - radioulnar synostosis with
octylphenoxypolyethoxyethanol	amegakaryocytic thrombocytopenia
OPPF - Oxford Protein Production	SFK – Src family kinases
Facility	SDS-PAGE - sodium dodecyl
PAR - protease activated receptor	sulphate polyacrylamide gel
PBS – phosphate-buffered saline	electrophoresis
PCR - polymerase chain reaction	SOC - super optimal broth
PDE3 – phosphodiesterase 3	TBS-T – Tris-buffered saline-Tween
PEI – polyethylenimine	TPO – thrombopoietin
PGI ₂ – prostacyclin	TXA ₂ - thromboxane
PI3K – phosphoinositide 3-kinase	UTR - untranslated region
PKC – phosphokinase C	vWF - von Willebrand Factor
PKG –phosphokinase G	WAS - Wiskott-Aldrich syndrome
PLA ₂ – phospholipase A ₂	WASp - Wiskott-Aldrich syndrome
PLCβ – phospholipase C beta	protein
PMA - phorbol 12-myristate 13-	WCL – whole cell lysates
acetate	WES - whole exome sequencing
POTE - prostate, ovary, testis,	WT - wild-type
placenta	XLT - X-linked thrombocytopenia
PRP - platelet rich plasma	5HT - serotonin
P/S - penicillin/streptomycin	
PVDF - polyvinylidene difluoride	
qRT-PCR - quantitative reverse	
transcription polymerase chain	
reaction	
RIPA - radioimmunoprecipitation	
assay buffer	
RPMI - Roswell Park Memorial	
Institute	

1. Introduction

1.1. Platelet physiology

Human platelets, derived from its precursor, the megakaryocyte (Mk), are discoid shaped hematopoietic anucleated cellular fragments that are 1 to 3 μm in diameter. Platelets play important roles in hemostasis, wound healing, angiogenesis, inflammation and innate immunity (Machlus and Italiano 2013). The platelet's role as the initial mediator in hemostasis following vascular injury was first elucidated by Giulio Bizzozero in 1881 (Coller 2011). Because of this initial discovery of the platelet's function by Bizzozero and because patients have presented with bleeding diatheses secondary to thrombocytopenia (low platelet counts), diseases have been identified and ascribed to platelet abnormalities. In healthy individuals the physiological concentration range for platelets is $(150 - 400) \times 10^9 /\text{L}$. As the sentinels of vascular integrity, platelets normally circulate in the vasculature not interacting with unimpaired inner surfaces of vessels (Ruggeri and Mendolicchio 2007). However, when vessels are injured, a series of events occur to form a hemostatic plug: platelet adhesion, platelet aggregation and thrombus formation (Figure 1.1).

1.1.1. Platelet Activation

Following vessel injury, endothelial cells are damaged, exposing extracellular matrices which allow platelets to quickly adhere to exposed adhesive macromolecules [von Willebrand factor (vWF) and collagen] of the extracellular matrix. This critical adhesion step of hemostasis occurs under conditions of high shear rates such that this interaction needs to be strengthened using platelet receptors GPIb/V/IX, which binds vWF and GPVI, which binds collagen. The copy numbers of GPIb/V/IX are estimated at ~50,000 per platelet and the copy numbers of GPVI are ~4000-6000 per platelet. GPVI, which is non-covalently coupled with the Fc Receptor γ (FcR γ) chain, when cross-linked by its ligand, initiates signalling utilising the immunoreceptor tyrosine-based activation motif (ITAM) located in the cytoplasmic region of FcR γ . The ITAMs are tyrosine phosphorylated by Src kinases Fyn and Lyn resulting in the recruitment of the tyrosine kinase Syk to bind the ITAMS via its SH2 domains, leading to Syk activation which initiates the phosphorylation and activation of downstream proteins such as the transmembrane adapter linker protein, LAT, (linker for activated T-cells) and SLP-76 (Src homology 2 domain-containing leukocyte phosphoprotein of 76 kDa). This signalling cascade is further activated involving proteins such as PLC γ 2, PKC and the release of Ca^{+2} stored in the endoplasmic reticulum (Stegner, Haining et al. 2014), (Li, Delaney et al. 2010). Elevated levels of Ca^{+2} lead to the Ca^{+2} dependent association of GPIIb and GPIIIa

resulting in the activated platelet membrane receptor complex, GPIIb/IIIa, which binds fibrinogen (Figure 1.2).

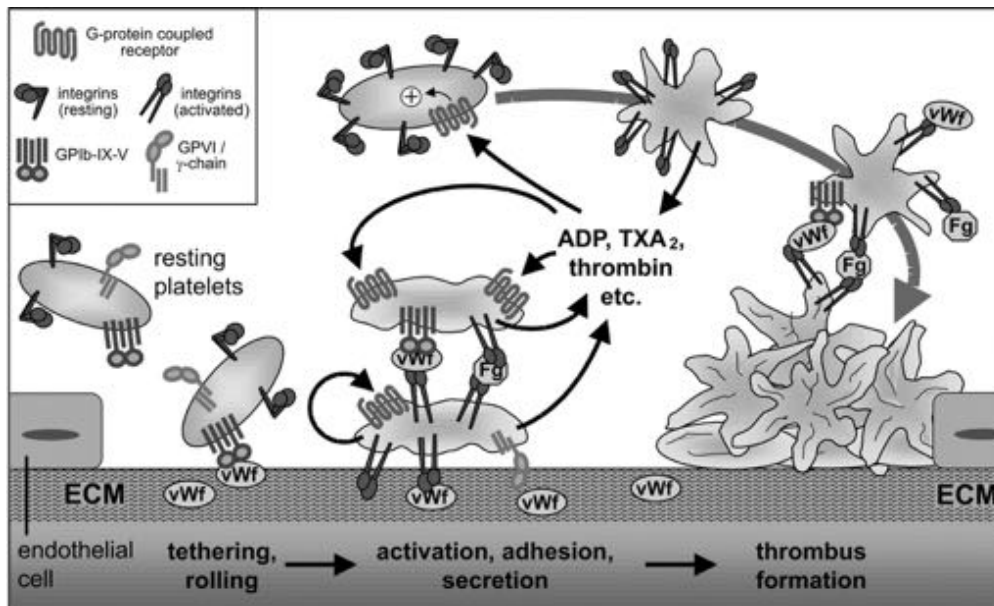


Figure 1.1 - Platelet activation at sites of vascular injury. ECM, extracellular matrix; fg, fibrinogen. (Offermanns 2006)

The copy numbers of GPIIb/IIIa per platelet is estimated at 80,000 to 100,000. Induction of the signalling cascade also activates GPIa/IIa (copy numbers estimated at 2,000-4,000 per platelet), which also binds collagen. The adhesion of platelets to the damaged vessel wall thus brings about the formation of a platelet monolayer. Following adhesion, platelet aggregation occurs, whereby more platelets are recruited from the circulation and utilising GPIIb/IIIa mediated aggregation, a platelet plug forms.

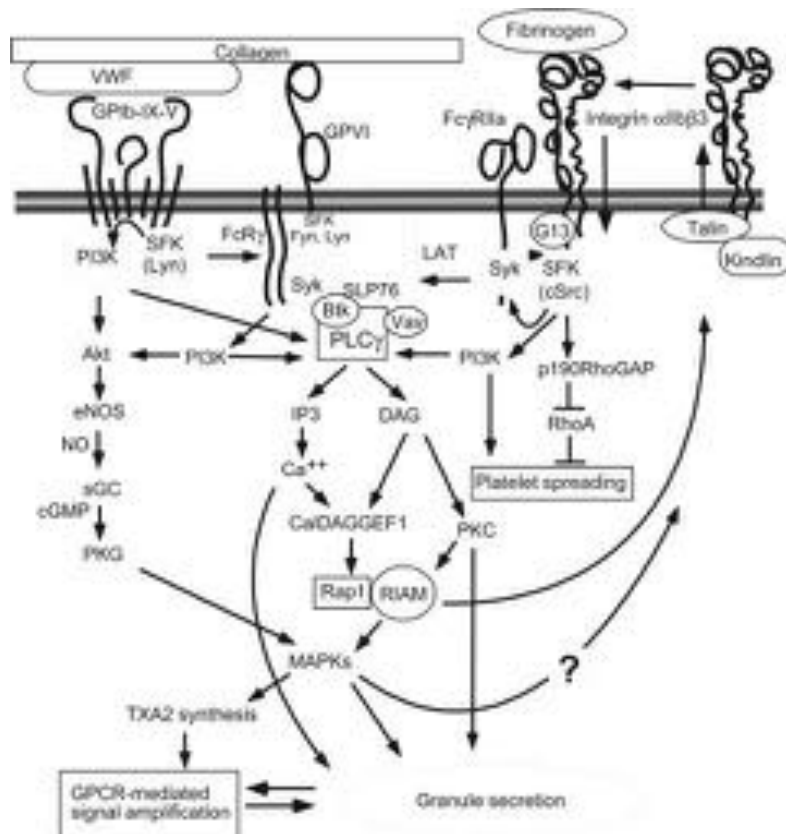


Figure 1.2 - Signalling pathways of 3 major platelet adhesion receptors: GPIb-IX-V, GPVI and GPIIb/IIIa (integrin α IIb β 3). sGC, soluble guanylyl cyclase; eNOS, endothelial nitric oxide synthase (Li, Delaney et al. 2010).

The recruitment of more platelets is brought about by a number of locally generated mediators such as thrombin (which is produced on the surface of activated platelets), ADP (released from dense bodies of activated platelets) and thromboxane A₂ (TXA₂), which is formed by a series of metabolic reactions:

platelet membrane phospholipids are converted to arachidonic acid (AA) followed by AA conversion to prostaglandin H_2 (PGH_2), using the COX-1 enzyme; and finally PGH_2 , is converted to TXA_2 using thromboxane synthase (Figure 1.5). All three mediators activate platelets via G-protein coupled receptors (ADP receptors P_2Y_{12} and P_2Y_1 coupled to G proteins G_i and G_q , respectively; thrombin receptors PAR1 and PAR4 coupled to G proteins G_{13} and G_q and thromboxane receptors coupled to G proteins G_{13} and G_q), which act through their signalling pathways to facilitate the formation and release of more of the same mediators. This positive-feedback mechanism amplifies the initial adhesion step so that there is quick activation and recruitment of platelets to an area of injury to form a thrombus (Figure 1.3) (Li, Delaney et al. 2010), (Offermanns 2006).

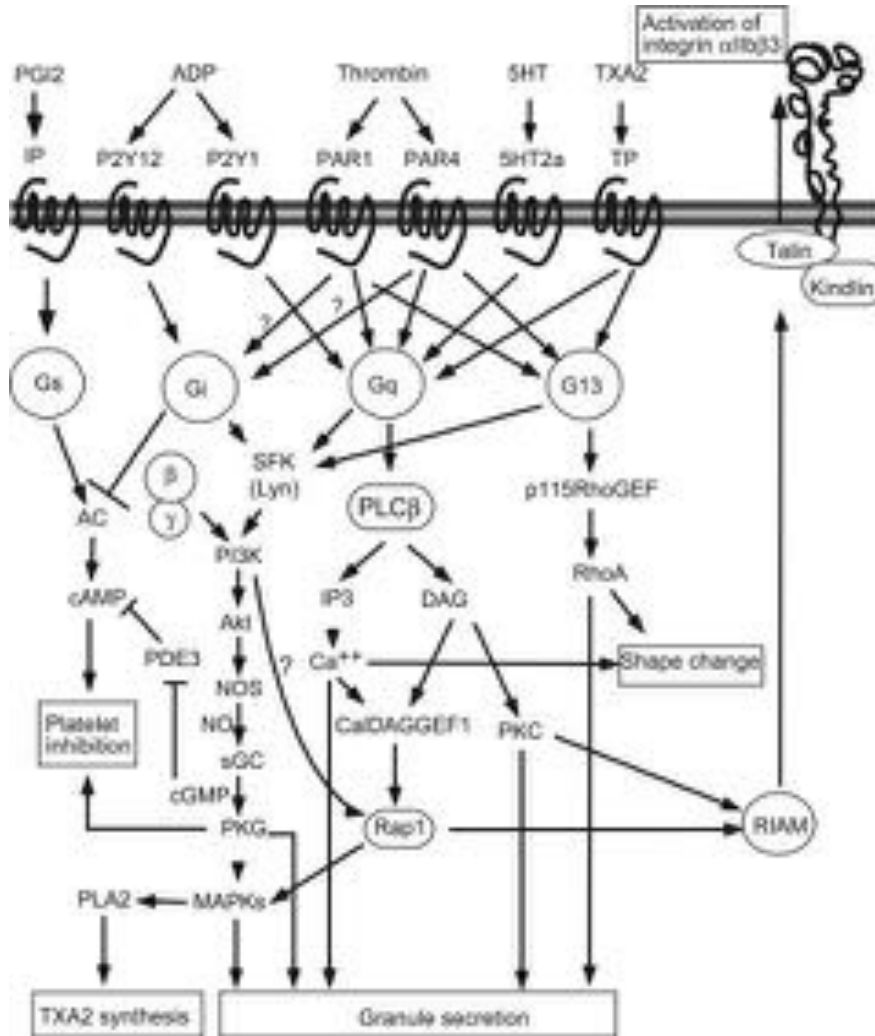


Figure 1.3 - Platelet G protein coupled receptors' signalling. PGI₂, prostacyclin; PI, prostacyclin receptor; 5HT, serotonin; SFK, Src family kinases; PI3K, phosphoinositide 3-kinase; PDE3, phosphodiesterase-3; NOS, nitric oxide synthase; NO, nitric oxide; sGC, soluble guanylate cyclase; PKG, phosphokinase G; MAPK, mitogen-activated protein kinases; PLA2, phospholipase A2; IP3, inositol triphosphate; DAG, diacyl glycerol; CalDAGGEF1, calcium and DAG-regulated guanine nucleotide exchange factor 1; PKC, phosphokinase C; RIAM, Rap1-GTP-interacting adapter molecule; GEF, guanine nucleotide exchange factor; AC, adenylyl cyclase; TXA₂, thromboxane; ADP, adenosine diphosphate (Li, Delaney et al. 2010).

The initial platelet response to activators such as ADP, thromboxane (TxA₂) and thrombin is a change of its shape from discoid to spherical. The concentrations of agonist required for this shape change are much lower than that needed for degranulation and aggregation. Shape change is a rapid process whereby the platelet cytoskeleton is reorganised: new actin filaments are laid down to form the submembranous actin filament network; filopodia are extended; actomyosin-based contractile processes are stimulated, resulting in the centralisation of dense and α granules and finally, the circumferential microtubule coil depolymerizes (Offermanns 2006).

Following shape change, platelets accumulate to form multiple platelet/platelet interactions (platelet aggregation) to form a platelet plug. Platelet aggregation is carried out by GPIIb/IIIa, which binds various extracellular ligands including dimeric fibrinogen and multimeric vWF. The structures of fibrinogen and vWF allow them to cross-bridge platelets to form platelet aggregates (Offermanns 2006).

Although the signalling cascade of ADP, TxA₂ and thrombin has not yet been definitely clarified, what is currently known is that signalling following platelet activation with these three ligands induces the intracellular domains of GPIIb/IIIa to rapidly convert into an active conformation (inside-out signalling), which lead to the activation of the ligand binding function of GPIIb/IIIa and the initiation of the “outside-in” signalling. Results of “outside-in” signalling lead to platelet

spreading, more granule secretion, stabilisation of platelet adhesion and aggregation and clot retraction. For aggregation to be rapid and efficient, the activation of all three of the heterotrimeric G protein coupled receptors, G_q , G_{13} and G_i is required (Offermanns 2006), (Li, Delaney et al. 2010). Clot retraction is the decrease in size of a blood clot after several days; its purpose is to bring together the edges of the blood vessel wall, which has been damaged. Clot retraction is dependent on the presence of platelets because clot retraction is regulated by signalling pathways utilised by receptors found on platelets: the thrombin receptor and GPIIb/IIIa (Kasahara, Kaneda et al. 2013).

To amplify platelet activation, platelets need to secrete its α granule and dense body contents as well as lysosomal contents, which serve to act on the vessel wall as well as other circulating hematopoietic cells. Contents of α granules include growth factors, chemokines, adhesive molecules and coagulation factors. Dense body granules include ATP, ADP and serotonin (Offermanns 2006).

A physiological process that complements as well as influences platelet activation is the activation of the coagulation cascade (Figure 1.4). Secretion of platelet α -granules contributes to the pro-coagulant activity of platelets by providing additional coagulant factors such as Factor V, VIII and I (fibrinogen). “Strong” platelet activation (prolonged increases in intracellular $[Ca^{++}]$) leads to phosphatidylserine exposure on the outer membrane of platelets (Offermanns 2006). Phosphatidylserine provides binding sites for plasma protein complexes

such as factor VIII_a-IX_a (tenase) and factor V_a-X_a (prothrombinase), which are needed to convert prothrombin to thrombin (factor II_a) (Zwaal, Comfurius et al. 2004). The coagulation cascade assists in stabilising the clot since thrombin converts soluble fibrinogen to insoluble fibrin strands, which are cross-linked by factor XIII to form a stabilised blood clot (Undas and Ariens 2011).

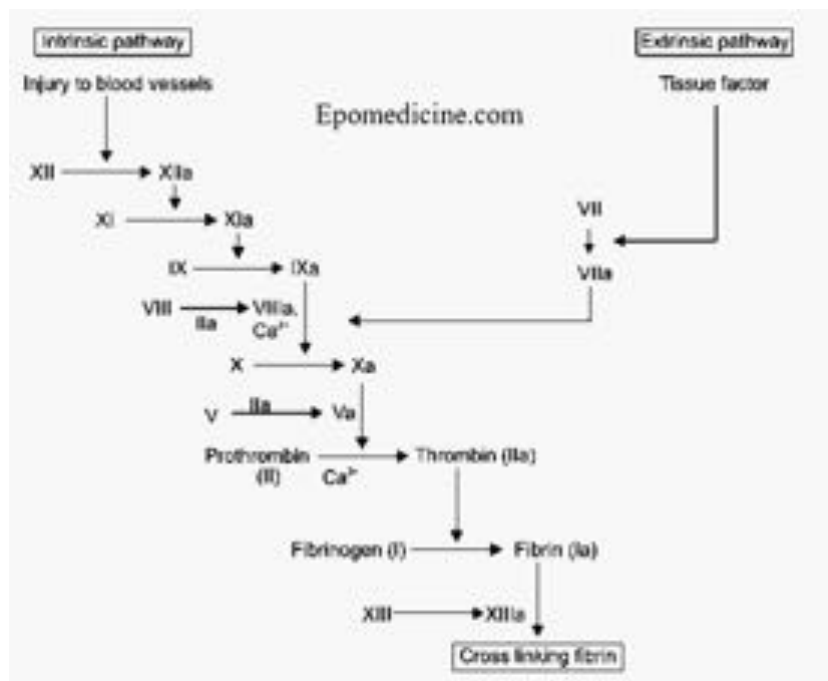


Figure 1.4 - Schematic diagram of the coagulation cascade (Illustration from epomedicine.com)

Dissolution of blot clots to maintain a patent vasculature is dependent on fibrinolysis (Figure 1.5). Under normal physiological conditions, coagulation and fibrinolysis are finely regulated. The major fibrinolytic protease is plasmin, which is formed by the conversion of the circulating plasma zymogen, plasminogen (PLG), by both activators: tPA (tissue PLG activator, produced by endothelial

cells) and uPA (urokinase, produced by renal epithelial cells, monocytes/macrophages and endothelial cells). It appears that tPA is the major intravascular activator of PLG. There are currently no human examples of complete deficiencies of tPA or/and uPA. However, evaluations of tPA and uPA deficient (single and doubly deficient) mice indicate that both are not required for normal fertility and embryologic development but are vital for clot lysis in the adult mice (Cesarman-Maus and Hajjar 2005, Rijken and Lijnen 2009).

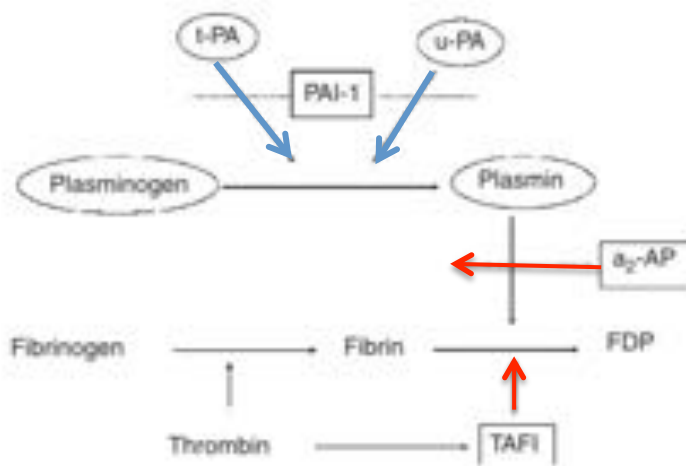


Figure 1.5 - Simplified schematic diagram of the fibrinolytic system. t-PA, tissue plasminogen activator; u-PA, urokinase plasminogen activator; PAI-1, plasminogen activator inhibitor-1; α_2 -AP, α_2 -antiplasmin; TAFI, thrombin-activatable fibrinolysis inhibitor. Blue denotes stimulation and red denotes inhibition (Rijken and Lijnen 2009).

Utilising a positive feedback mechanism, plasmin cleaves both tPA and uPA, transforming them from single chains to active two-chain polypeptides. Plasmin's major substrate, fibrin, regulates its own degradation by binding both PLG and tPA on its surface to localise and enhance plasmin generation. tPA is dependent

on the presence of fibrin to increase its efficiency in generating plasmin. Once plasmin is formed, it is rapidly inhibited by its major inhibitor, α_2 -plasmin inhibitor and its lesser inhibitor, α_2 -macroglobulin, unless it remains bound to fibrin or to its cell surface receptors (Cesarman-Maus and Hajjar 2005).

Once it's formed, plasmin cleaves fibrin to generate soluble degradation products. It also exposes carboxy-terminal lysine residues on fibrin, which provide further binding sites for tPA and PLG, leading to enhanced plasmin generation and fibrin removal. Binding of the lysine residues can be blocked by lysine analogues as well as by TAF1 (thrombin-activatable fibrinolysis inhibitor). TAF1 is activated by thrombin to remove carboxy-terminal lysine residues, thus decreasing plasmin generation to stabilise a fibrin thrombin and providing a counter regulatory mechanism to fibrinolysis. Fibrin dissolution is also regulated by PLG inhibitors such as PAI-1 (PLG activator inhibitor-1) and plasmin inhibitors such as α_2 -PI (α_2 -plasmin inhibitor). Plasmin when bound to fibrin has its lysine binding sites occupied and is protected from α_2 -PI. TAFI, however, can decrease this protection by deleting the lysine residues on fibrin to prevent plasmin binding (Rijken and Lijnen 2009).

Plasmin degradation of fibrin produces a number of molecular products. When fibrin, cross-linked by factor XIII, is degraded by plasmin, fragments called D-dimers are released. Other fibrin breakdown products are also formed which

may inhibit platelet function, potentiate bradykinin's hypotensive effects and affect chemotaxis and immune modulation (Hajjar 2003).

1.1.2. Platelet inactivation

Platelet activation is inhibited by the vasoactive hormones nitric oxide (NO) (Palmer, Ashton et al. 1988) and prostacyclin (PGI₂) (Moncada, Gryglewski et al. 1976), both of which are produced by the endothelium. The endothelium covers the inner surface of the vasculature, serving not only as a structural barrier against circulating elements in the blood, but is also an endocrine organ, releasing various hormones such as NO and PGI₂, which act on the circulatory system. In healthy blood vessels, the endothelium constitutively produces the enzymes nitric oxide synthase (NOSIII) and cyclo-oxygenase (COX-1), which catalyse the formation of NO and PGI₂. Both NO and PGI₂ function to vasodilate vessels and inhibit platelet aggregation. On blood vessels, the actions of NO and PGI₂ are additive whereas in platelets, they are synergistic. NO mediates its actions mainly by activating intracellular guanylyl cyclase, leading to the formation of cGMP. Prostacyclin, in contrast, mediates its actions via cell surface receptors (IP) and/or intracellular receptors peroxisome proliferator-activated receptors (PPAR)β. In diseased (i.e. occluded) vessels, vascular smooth muscle cells produce inducible forms of NOSII and COX-2, which produce substantial amounts of NO, PGI₂ and prostaglandin E₂ (Mitchell, Ali et al. 2008).

Nitric oxide synthesis arises from the catalysis of L-arginine, a non-essential amino acid, and oxygen, using the enzyme NOSIII. Byproducts of the reaction are L-Citrulline and H₂O. NOS exists as three isoforms, NOSI, NOSII and NOSIII

(Fedorov, Hartmann et al. 2003). It is NOSIII which is mainly found in endothelial cells, where it is a membrane-bound, constitutively expressed homodimer requiring calcium to activate it (Forstermann, Schmidt et al. 1991). Calcium is required for calmodulin binding, which together associate with NOSIII, to regulate its activity. Other modifications of NOSIII include post-translational lipid modifications, protein-protein interactions, and protein phosphorylation (Sessa 2004). Once synthesised in the endothelial cell, NO diffuses out of the cell via the luminal side to affect platelet and other vascular cellular functions and via the abluminal side to affect smooth muscle function. For platelets, NO inhibits adhesion and aggregation; for smooth muscle cells, NO inhibits vasoconstriction and vascular remodelling and smooth muscle cell proliferation. The mechanism by which platelets and smooth muscle cells are principally affected is through the activation of cytosolic soluble guanylyl cyclase by NO, which leads to an increase in intracellular cGMP, which activates protein kinase G, resulting in reduced intracellular calcium and reduction in cell activation (Hanafy, Krumenacker et al. 2001). Prostacyclin synthesis initiates at the endothelial membrane where phospholipase A₂ (PLA₂) releases arachadonic acid (AA) from membrane-bound lipids with the assistance of calcium (Mitchell and Warner 1999) (Figure 1.6). AA, once liberated, can be metabolised further with cyclo-oxygenase (COX), which exists as two isoforms, COX-1 and COX-2. COX-1 is the major isoform in healthy endothelial cells and is constitutively expressed (Mitchell, Lucas et al. 2006). The second step in prostacyclin synthesis involves two enzymatic reactions utilising cyclo-oxygenase: an oxygenase step converting AA to

prostaglandin G_2 (PGG_2) and a peroxidase step converting PGG_2 to prostaglandin H_2 (PGH_2). PGH_2 serves as a substrate for a number of subsequent prostaglandin synthase enzymes, including prostacyclin synthetase (PGIS) and thromboxane synthase. In endothelial cell, COX-1 and PGIS predominate, resulting in prostacyclin being the major product made. However, in platelets, COX-1 also predominates, but thromboxane synthase is the major prostaglandin synthase enzyme present, resulting in thromboxane being the predominant product made (Needleman, Moncada et al. 1976).

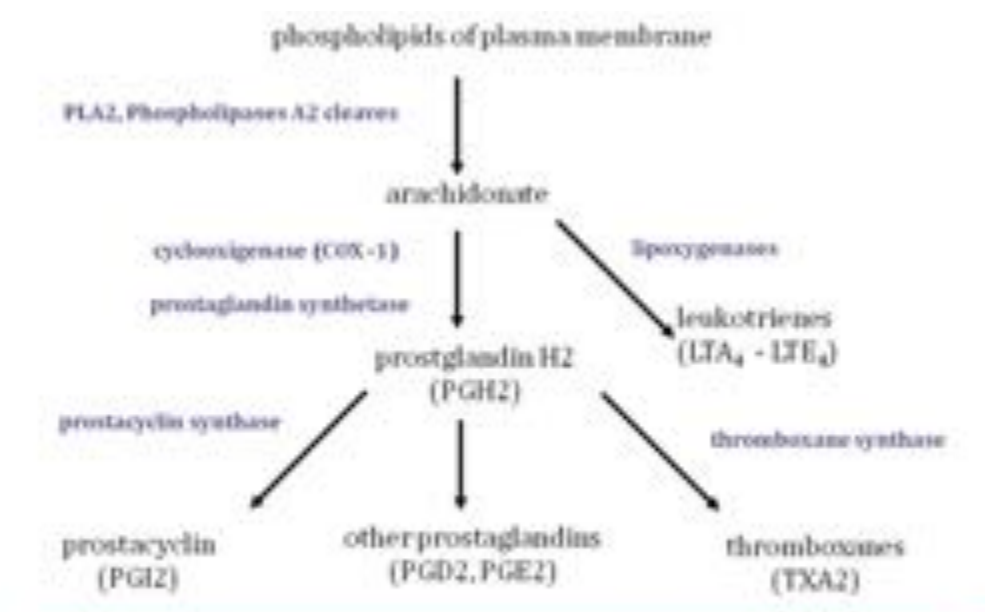


Figure 1.6 - Schematic diagram of eicosanoid (prostaglandins and related compounds) synthesis (Illustration from www.tankonyvtar.hu).

Prostacyclin binds to two receptor types: the cell surface prostacyclin (IP) receptor and the nuclear receptor PPAR β (peroxisome proliferator-activated receptor β). IP receptors, thought to be the predominant surface receptors for

prostacyclins, are seven trans-membrane spanning G-protein coupled receptors, which are ubiquitous throughout the body but are especially enriched on platelets and smooth muscle cells (Stitham, Arehart et al. 2007), (Mitchell, Ali et al. 2008). Activation of IP receptors by prostacyclin leads to G-protein mediated activation of adenylate cyclase, leading to formation of cAMP, which phosphorylate PKA (Figure 1.3). Phosphorylated PKA leads to a decrease in calcium in cells such as platelets and smooth muscle cells (Schwarz, Walter et al. 2001).

It is still unclear how PPAR β plays a role in preventing platelet aggregation since PPAR receptors affect function at the level of gene transcription and platelets have no nuclei. Studies have indicated that specific PPAR β agonists inhibit platelet aggregation independent of IP receptors; and, PPAR β agonists can also act in synergy with IP receptor agonists to inhibit platelet aggregation, suggesting that IP and PPAR β receptors have separate but complementary signalling pathways (Ali, Davidson et al. 2006).

In conclusion, following vessel injury, exposure of extracellular matrices allows platelet receptors to bind adhesive macromolecules collagen, (vWF and fibrinogen), thus laying down the initial monolayer of platelets. Further activation of platelets from local mediators such as thrombin, ADP and thromboxane lead to platelet aggregation and clot formation, which is stabilised by certain members of the coagulation cascade (I, V, VIII, and XIII). To counteract the activation of platelets are the vasoactive hormones NO and prostacyclin, which are co-

released by the endothelium, acting in synergy to inhibit platelet aggregation and vasoconstriction of the vasculature. They mediate these effects via the second messengers cGMP and cAMP respectively (Mitchell, Ali et al. 2008).

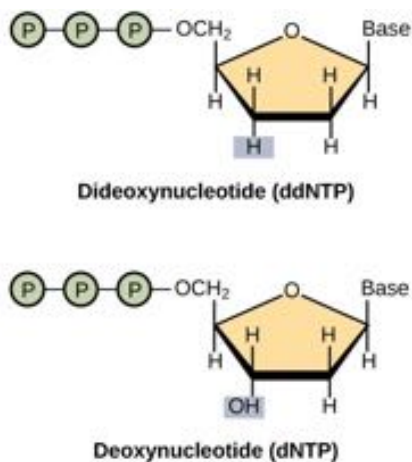
1.2. DNA Sequencing

Modern day DNA sequencing, termed dideoxy DNA sequencing, is based on the sequencing method developed by Dr. Fred Sanger in the 1970's. This enzymatic sequencing method depends on random inhibition of DNA chain elongation.

The DNA to be sequenced needs to be single-stranded thus serving as a template for the construction of a new complementary DNA strand *in vitro* using a DNA polymerase. The newly synthesized DNA strands are of various lengths, which are then separated by size on a slab polyacrylamide gel. The sequencing reaction occurs via four parallel reactions, each of which contains the four dNTPs (dATP, dCTP, dGTP, and dTTP) plus a small amount of one of the four analogous dideoxynucleotides (ddNTPs), which are a base-specific chain terminators. Structurally, ddNTPs are nearly identical to dNTPs except that they lack a hydroxyl group at the 3' carbon position (Figure 1.7A). Dideoxynucleotides can be incorporated into a DNA chain, but cannot join in the phosphodiester bonding at its 3' carbon with other dNTPs because its 3' carbon lacks a hydroxyl group. Therefore, once a ddNTP has been added to a DNA chain, it causes the termination of DNA chain elongation (Figure 1.7B). In order to visualise the elongated DNA chains, one of the four dNTPs or the primer is labelled with ^{35}S or

^{32}P , thus labelling the DNA. The dNTPs' concentrations are in excess compared to the concentrations of the ddNTPs, creating a competition between a specific ddNTP and its dNTP counterpart for inclusion in the lengthening DNA chain.

A.



B.

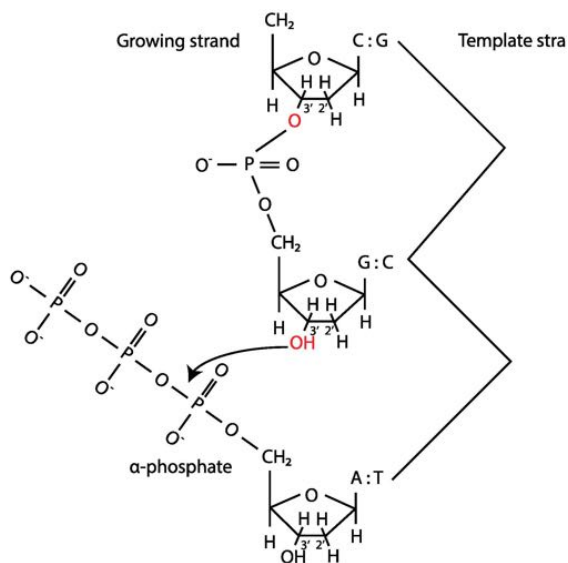


Figure 1.7 - Nucleic acid structures and DNA chain elongation. (A) Molecular structures of dideoxyribonucleotide and deoxyribonucleotide (B) Example of DNA chain elongation. (Illustrations from www.ubooks.pub— panel A, and binf.snipcademy.com— panel B).

DNA chain termination will occur randomly at any one of the four nucleotide bases. Since the sequencing sample is a population of identical DNA molecules, each of the four base-specific reactions will be a collection of labelled DNA

fragments of different lengths. In each base-specific reaction, DNA fragments will have a common 5' end and variable 3' ends because of the random insertion of ddNTPs. Each base-specific reaction will be loaded onto a high urea concentration denaturing polyacrylamide gel to size-fractionate the DNA fragments. Even fragments differing by only one nucleotide can be separated. Once electrophoresis is completed, the gel is dried and exposed to X-ray film to allow visualisation of the DNA fragments. DNA sequence is determined by using the naked eye to read from the shortest fragment (bottom of film) to the longest fragment (top of film) (Strachan 2011).

To improve the efficiency of dideoxy DNA sequencing, automated DNA sequencing machines, which sequence fluorescently labelled DNA, were introduced in the 1980's. Four different fluorescent dyes, each with a different emission wavelength, were used; thus, permitting all four reactions to be loaded into a single sample well on a polyacrylamide gel. During the running of the gel, DNA fragments passed by a laser, which excited the fluorescent dyes and a monitor detected and recorded the fluorescence signals. This generated an intensity profile for each of the four different fluorophores as well as electronically storing the information. Automated sequencing was subsequently made more efficient by introducing capillary sequencing instruments in 1999. For these instruments, rather than having DNA samples on a slab gel, samples were instead injected into long and very thin (0.1 mm diameter) capillary tubes containing polyacrylamide gel (Strachan 2011). The use of capillary tubes

eliminated many rate-limiting steps in the sequencing protocol. Also, run times could be accelerated because heat is more rapidly dissipated using capillary tubes vs. glass plates which were used for slab gels. Each capillary sequencing instrument was able to detect 500-600 bases from each of 96 reactions in about 10 hours, producing around 115 kbp per day (24 hour unattended operation) (Mardis 2011). These instruments ultimately became the principal data-generating workhorses used for the human and mouse genome projects in 2001/2002 (Mardis 2013).

1.2.1. Next Generation Sequencing

The automated Sanger method of DNA sequencing is considered a “first-generation” technology, and subsequent newer methods of sequencing, arriving in 2005, came to be referred to as “next-generation sequencing” (NGS) or “massively parallel” sequencing (Metzker 2010, Mardis 2011). The major advantages offered by NGS are the enormous amount of data which can be generated at a much cheaper cost (Metzker 2010). Unlike the past, where there was only one company which manufactured the capillary sequencing instruments (Applied Biosystems), there are now many commercially available NGS instruments, resulting in fierce competition among the manufactures to provide better sequencing platforms. These platforms aim to increase the amount of sequence output per run, to increase read lengths (the number of nucleotides per sequence read), to improve base-calling accuracy and to decrease costs (Mardis 2011) (Figure 1.8).

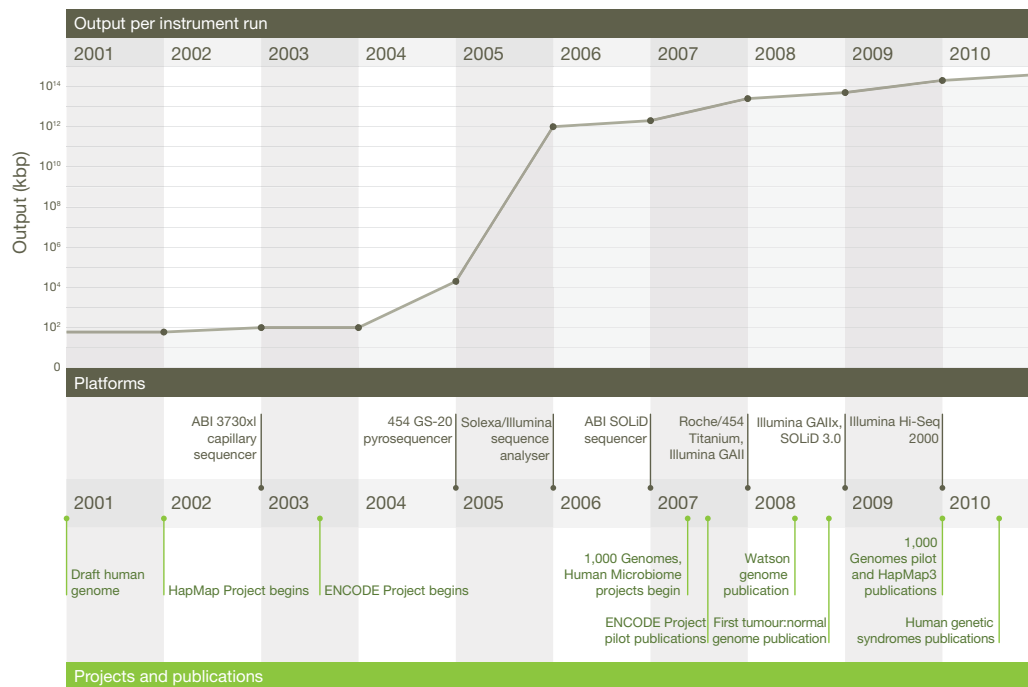


Figure 1.8 - Changes in the various sequencing instrument read outputs, timeline of newly introduced sequencing instruments and major sequencing projects over the past decade. Top, data output per run plotted on logarithmic scale. Middle, introduction year of various NGS sequencing platforms. Bottom, timeline of the major sequencing projects which have been completed. (Mardis 2011)

Table 1 | Sequencing platform comparison

	Roche/454	Life Technologies SOLiD	Illumina Hi-Seq 2000	Pacific Biosciences RS
Library amplification method	emPCR* on bead surface	emPCR* on bead surface	Enzymatic amplification on glass surface	NA (single molecule detection)
Sequencing method	Polymerase-mediated incorporation of unlabelled nucleotides	Ligase-mediated addition of 2-base encoded fluorescent oligonucleotides	Polymerase-mediated incorporation of end-blocked fluorescent nucleotides	Polymerase-mediated incorporation of terminal phosphate labelled fluorescent nucleotides
Detection method	Light emitted from secondary reactions initiated by release of PPi	Fluorescent emission from ligated dye-labelled oligonucleotides	Fluorescent emission from incorporated dye-labelled nucleotides	Real time detection of fluorescent dye in polymerase active site during incorporation
Post incorporation method	NA (unlabelled nucleotides are added in base-specific fashion, followed by detection)	Chemical cleavage removes fluorescent dye and 3' end of oligonucleotide	Chemical cleavage of fluorescent dye and 3' blocking group	NA (fluorescent dyes are removed as part of PPi release on nucleotide incorporation)
Error model	Substitution errors rare, insertion/deletion errors at homopolymers	End of read substitution errors	End of read substitution errors	Random insertion/deletion errors
Read length (fragment/paired end)	400 bp/variable length mate pairs	75 bp/50+25 bp	150 bp/100+100 bp	>1,000 bp

Comparison of commercially available next generation platforms (Roche/454, Life Technologies and Illumina) and a single molecule platform (Pacific Biosciences), illustrating the similarities and differences in these technologies, according to several metrics. NA, not applicable; PPi, pyrophosphate.

* emPCR (emulsion PCR) is a bulk amplification process whereby library fragments are combined with beads and PCR reactants in an oil emulsion that allows *en masse* amplification of millions of bead-DNA combinations in a single tube.

Table 1.1 - Comparison of NGS platforms (Mardis 2011)

Although each of the NGS instruments is different in certain specific ways (Table 1.1), they also share certain aspects, which are characteristic of massively parallel sequencing, making it different from Sanger sequencing. Firstly, for NGS, there are fewer initial DNA preparatory steps. Rather than performing a bacterial cloning step followed by DNA purification, which is required for Sanger sequencing, for NGS, DNA is fragmented and a library is produced by ligating specific synthetic DNAs (adapters) to the ends of the DNA fragments to be sequenced. Secondly, the library fragments are amplified on a solid surface (glass slide or microbead) by a DNA polymerase reaction. Amplification by PCR, forming template clusters, is required so that subsequent sequencing reactions will produce enough signal for detection by the optical system. However, the downside to this step is that it introduces sequencing errors, which will be carried to the analysis step. Immobilisation of the template onto a solid surface allows the template to be spatially separated so that millions of sequencing reactions can be performed simultaneously (Metzker 2010). Thirdly, the sequencing reactions are a series of repeating steps that are performed and detected automatically. Each different NGS instrument has its own DNA sequencing reaction (Table 1.1). Regardless of the specifics, these sequencing reactions are notable for the fact that they occur in a nucleotide-by-nucleotide stepwise fashion rather than by separation and detection of previously produced sequencing reaction products, as seen in Sanger sequencing. Therefore, in NGS, the sequencing instruments perform sequencing and detection simultaneously rather than as distinct processes. Also, during each sequencing

run, thousands to millions of fragments are sequenced, creating enormous data sets. Fourthly, for NGS, sequence information can be obtained from both ends of the DNA fragments being sequenced. One can either sequence both ends of linear fragments (paired end sequencing) or from both ends of previously circularised fragments (mate pair sequencing) (Figure 1.9). Paired end sequencing libraries, requiring little DNA, consist of DNA fragments less than 1 kb, are not complicated to make, making it the method most often used to prepare the template for human genome sequencing. For paired-end sequencing, the linear fragment has adapter sequences with different priming sites for each adapter. Sequence begins from an adapter priming site of a fragment followed by the sequencing from the opposite adapter priming site. These sequences are then paired during the alignment step of data analysis. Having two sequences of the same region provides more certainty of placement than having only one end read of the same length. Mate pair libraries, made from DNA fragments greater than 1 kb, require much more DNA because of the low yield of circularisation of large DNA molecules. The library for mate-pair sequencing is constructed by circularising a fragment around a single adapter so that both fragment ends ligate to the adapter ends (Figure 1.9). The fragment is then linearized by type IIS endonuclease digestion or by nick translation to produce a linear fragment with both ends of the original fragment and a central adapter. The central adapter is biotinylated, allowing for capture and purification using streptavidin magnetic beads. The linear fragments again have unique adapters ligated to their ends and sequencing begins from one adapter site

followed by sequencing from the adapter of the other pair. Ultimately, the resulting sequences are longer than the sequences from the paired-end sequencing. Often, both sequencing approaches are used together to sequence genomes with difficult sequences (repetitive regions) or when attempting *de novo* sequencing. Mate-pair sequences provide longer reads (up to 20 kb is possible) while paired-end sequences allow, through its shorter sequences, the assembly of difficult to sequence regions; paired-end sequences provide additional detail to the sequencing scaffold provided by the mate-pair reads. Finally, read lengths generated from NGS are much shorter than those generated from Sanger sequencing. Unfortunately, these short reads will also contain sequencing errors, which is indistinguishable from a variant (Sims, Sudbery et al. 2014). For Sanger sequencing, read length is determined by a combination of gel-related factors: percentage of the polyacrylamide gel, the length and thickness of the gel, time spent running the gel and the electrophoresis conditions. In NGS, read length is a function of the signal to noise ratio. Each NGS instrument has its own source of noise, which creates certain types of sequencing errors; this interplay between noise and error types gives rise to the term “error model,” which is specific to each instrument (Table 1.1). Noise adds up during the sequencing process and ultimately leads to decreased read length because the signal from nucleotide base incorporation is outcompeted by incorrect signals resulting from noise such as incorrect or out-of-phase incorporation events, residual signals from prior reactions or reactants and other sources of noise (Mardis 2013). To determine what the read lengths and error types are for an

instrument, one needs to sequence a reference set of genes or a genome and compare the results with a high quality reference gene set or genome. Doing this will allow one to determine the different types of errors (substitution, insertion or deletion errors) as well as the error model (random vs. systemic errors). Overall, the more a sequence is examined following alignment with a high-quality reference sequence, the better the chance of identifying the error models and coverage biases and their contributing sources (Mardis 2013). Difficulties raised by short reads arise during the analysis step. Although short reads can be assembled like Sanger sequencing reads (which is based on shared sequences), the fact that there are fewer numbers of shared sequences due to the shorter read lengths makes it difficult to assemble the overall length of contiguous sequences. This limitation is compounded when the genome to be sequenced is large and complex. Thus, because high quality reference genomes exist for many organisms, using NGS reads for sequence read alignment is a better approach for data analysis (Mardis 2013). Specific algorithms have been devised for short read alignments and there is a score- based metric that indicates a sequence's best fit in the genome. Sequences with mostly repetitive content score lowest because of the uncertainty of their validity (Li, Ruan et al. 2008). To overcome error problems, the number of sequencing reads (depth of coverage) can be increased (Sims, Sudbery et al. 2014). Also, more accurate reads can be obtained by sequencing from both ends of each library fragment using paired-end sequencing and/or mate-pair sequencing (as outlined above) (Metzker 2010, Mardis 2013).

In conclusion, NGS has revolutionised DNA sequencing by massively increasing data production while significantly lowering the cost of sequencing; however, there are drawbacks to NGS. These include the presence of polymerase errors during library construction early on in the sequencing process, resulting in nucleotide bases wrongly appearing as variants in the sequenced genome; the occurrence of preferential amplification of certain fragments in the library population, making them appear to be more abundant relative to other fragments; and finally, the dilution of DNA modifications, such as different types of methylation, during the amplification process. Also, because of the massive amount of data generated by NGS, burdens are placed on research centres for the storage of the sequences. And, there are data analysis challenges as well, due to the increased quantity of data, the extreme short reads and the different error profiles of each read type (Mardis 2011).

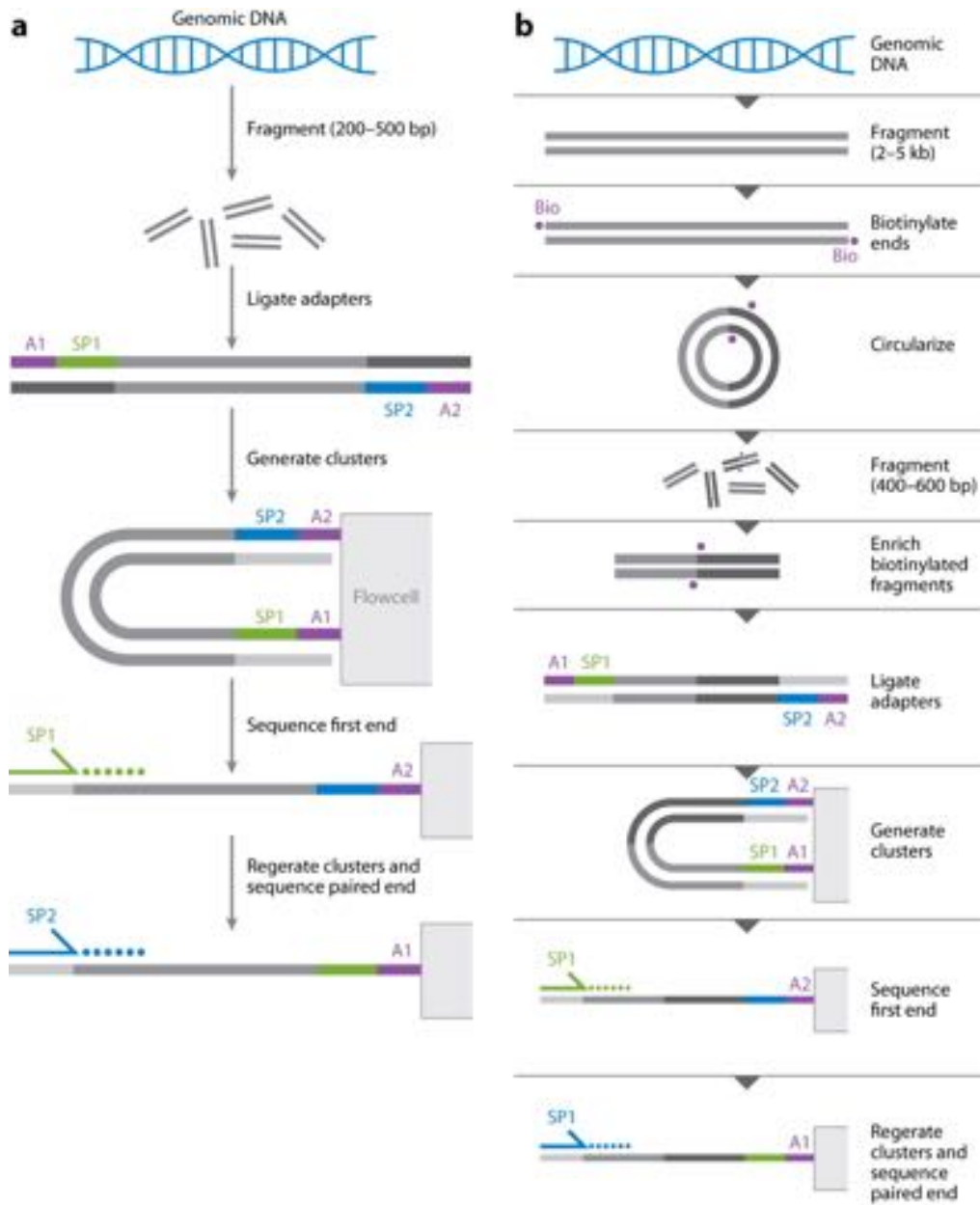


Figure 1.9 - Comparison of fragment library construction for (a) mate-end vs. (b) mate-pair sequencing. “A” signifies adaptor and “SP” signifies specific primer (Mardis 2013).

In order to overcome some of the above shortcomings of NGS, instrument detectors were subsequently devised which could, as accurately as possible, sense very low levels of signal produced from unamplified individual DNA molecules as they are being sequenced. This form of sequencing, termed single-molecule sequencing, has been labelled “third-generation” sequencing. This newest form of sequencing produces sequences faster than NGS (20,000 times faster) and at lower costs as well. Read lengths are quite long (>1000 bp) but there is an overall higher error rate. An example of this platform is the instrument by Pacific Biosciences, Inc., which involves sequencing-by-synthesis whereby synthesis occurs in real-time, thus making the sequencing reactions extremely fast. This platform has also been shown to successfully sequence as well as detect specific methyl residues known to be on DNA residues serving a regulatory role. The error model for single-molecule sequencing consist mainly of insertion and deletion errors plus a small percentage of substitution errors. To offset the high error rate seen with single-molecule sequencing of an individual read, multiple reads of the same molecule are required so that there is an overall higher consensus read accuracy for the molecule (Mardis 2013).

1.2.2. Exome Sequencing

Despite the huge technological progress that has been made in DNA sequencing, it is still too expensive to sequence a whole genome to enough depth to identify variants that affect phenotypic expression (Warr, Robert et al. 2015). Most read lengths from NGS are short and contain errors, which look no different from a variant. To overcome this conundrum, the number of sequencing

reads can be increased: a read containing a 1% error rate when sequenced identically 8 times, will result in a strongly supported identification of a variant associated with an error rate of 10^{-16} (Schatz, Delcher et al. 2010). Thus, increased depth (the average number of times a nucleotide is sequenced in a collection of random raw sequences) of short reads can be used to overcome the inadequacies of NGS but incurs greater costs secondary to increased number of sequencing runs, to increased number of analytical inspections of cumbersome captured image files, and for the storage of the large number of sequences (Mamanova, Coffey et al. 2010, Sims, Sudbery et al. 2014).

One method to bypass the continued costliness of NGS is to perform targeted sequencing. This “target-enrichment” strategy is based on the capture of specific regions of the genome from a DNA sample followed by massively parallel sequencing of the targets. Exome sequencing targets all the exons of the protein coding genes in the genome (encompassing ~1% of the human genome), and may also target functional non-protein coding elements such as microRNA, long intergenic noncoding RNA and other RNA molecules as well as specific candidate loci (Warr, Robert et al. 2015); (Mamanova, Coffey et al. 2010).

Exome sequencing has proven to be an efficient way to identify variants and genes responsible for Mendelian disorders; it is faster than conventional approaches and it can surplanted conventional approaches which have failed to identify the culprit gene (Bamshad, Ng et al. 2011). Traditional methods of identifying rare Mendelian disorders can be hampered by having only small

number of cases or families to study, reduced penetrance (the proportion of persons with a specific phenotype among people with the genotype), locus heterogeneity (having phenotypically similar characteristics despite having mutations at different genetic loci) and reduced reproductive fitness. However, there are drawbacks with exome sequencing as well. One major concern is that it is still unclear exactly which sequences in the human genome are protein coding. Initial exome sequencing efforts used as its targeted DNA, genes that were identified by the Consensus Coding Sequence (CCDS) Project, a highly reputable compilation, although not complete. Currently, all commercial kits target at a minimum, genes identified in the RefSeq collection (an open-access compilation of annotated and curated DNA and RNA nucleotide sequences) and even hypothetical proteins as well. Other limitations of exome sequencing include capture probe efficiency – some sequences are not even targeted by the probes offered; sequencing of all templates are not equally efficient thus sequences can not all be aligned to the reference genome, and it is still unclear whether other sequences such as promoters, microRNA and ultra-conserved elements (genome subsequences, which seem to have high levels of sequence constraint) should be sequenced as well (Bamshad, Ng et al. 2011).

There are mainly two different methods of exome capture: solution-based and array-based. The initial preparation steps for solution-based and array-based whole exome sequencing (WES) are similar: the DNA is fragmented and oligonucleotide probes (baits) are used to hybridize to selective target areas of

the genome. In solution-based methods, use of biotinylated oligonucleotide probes to selectively bind targeted regions of the genome followed by “pull-down” with streptavidin metal beads is used. Non-targeted regions of the genome, which do not bind to the probes, are washed away and targeted regions bound to the metal beads are amplified, enriching the sample for the targeted region of the genome (Warr, Robert et al. 2015). For the array-based method, oligonucleotide probes, rather than being biotinylated are bound to a high-density microarray. The array-based method was the original method used in WES but has since been replaced by the solution-based method. Benefits of the solution-based method include less DNA needed for the process, the number of probes it can use is not limited by what can be accommodated on the microarray, and no additional equipment or time is required to process the microarray (Warr, Robert et al. 2015).

Similar to NGS platforms, exome capture platforms evolve quickly. Basically, each version differs mainly in the region of the genome which it targets; however, 84% of the targeted region still overlap (Chilamakuri, Lorenz et al. 2014). There are three major companies that provide exome capture platforms: NimbleGen, Agilent, and Illumina (Table 1.2). We chose to use the Agilent’s SureSelect Human All Exon Kit, which of the three kits, is the only one to use RNA probes; the probes used are longer than those found in other kits (114-126 bp) and the targeted sequences are adjacent to one another rather than overlapping (Clark, Chen et al. 2011). Advantages of the Agilent exome capture kits are that it is

good at identifying insertions and deletion (indels) because longer probes allow for larger mismatches (Clark, Chen et al. 2011, Bodi, Perera et al. 2013, Chilamakuri, Lorenz et al. 2014); it produces fewer duplicate reads when compared to NimbleGen (Sulonen, Ellonen et al. 2011) and it has a greater alignment rate and fewer PCR duplicates than NimbleGen (Bodi, Perera et al. 2013). Disadvantages with the Agilent kit include fewer high-quality reads (Sulonen, Ellonen et al. 2011) and less uniform coverage (Bodi, Perera et al. 2013). However, the Agilent kit can detect the most SNPs and small indels with more depth of coverage (Clark, Chen et al. 2011). Overall, all three kits have high levels of targeting efficiency and cover a very large part of what we know of the exome. Which kit to chose for an experiment depends on the question one wishes to answer because the kits differ somewhat in their target choice, bait (oligonucleotide) lengths, bait density and molecule used for capture (Warr, Robert et al. 2015).

	NimbleGen's SeqCap EZ Exome Library	Agilent's Sure Select Human All Exon Kit	Illumina's TruSeq Exome Enrichment Kit	Illumina's Nextera Rapid Capture Exome Kit
Probe size, bp ^a	55–105	114–126	95	95
Probe type	DNA	RNA	DNA	DNA
Coverage strategy	High-density, overlapping probes	Adjacent probes	Gaps between probes	Gaps between probes
Fragmentation method	Ultrasonication	Ultrasonication	Ultrasonication	Transposomes
Target region size (human), Mb ^b	64	50	62	62
Reads remaining after filtering ^c	66%	71.7%	54.8%	40.1%
Major strengths	(i) High sensitivity and specificity (ii) Most uniform coverage in difficult regions	(i) Better coverage of indels (ii) High alignment rate (iii) Fewer duplicate reads than other platforms	(i) Good coverage of UTRs and miRNAs	(i) Good coverage of UTRs and miRNAs
Major weaknesses	(i) More duplicate reads than Agilent (ii) Lower alignment rate than Agilent	(i) Fewer high-quality reads than NimbleGen	(i) High off-target enrichment	(i) High off-target enrichment (ii) Coverage bias for high GC content areas reducing uniformity
Nonhuman supported	Yes	Yes	No	No

UTR, untranslated region; miRNA, microRNA.

^a As described in Clark *et al.* (2011), NimbleGen SeqCap EZ v2.0, Agilent SureSelect Human All Exon 50Mb, Illumina TruSeq Exome Enrichment.

^b For NimbleGen SeqCap EZ v3.0, Agilent SureSelect V5, and Illumina TruSeq and Illumina Nextera original versions.

^c Filtering for duplicates, multiple mappers, improper pairs, and off-target reads, data from Chilamakuri *et al.* (2014).

Table 1.2 – Summary of differences between various exome sequencing platforms (Warr, Robert *et al.* 2015).

Clinical application of WES was first successfully used to identify a rare disease in a child presenting with severe symptoms suggestive of inflammatory bowel disease; however, definite diagnosis could not be made despite comprehensive clinical evaluation. Following WES, analysis identified a novel hemizygous missense mutation in the gene, X-linked inhibitor of apoptosis, whereby a cysteine residue had been replaced with a tyrosine. Further functional assays confirmed deficiency of the protein, X-linked inhibitor of apoptosis, and the diagnosis of X-linked lymphoproliferative disease 2 was made, although the child exhibited inflammatory bowel-like symptoms, which had never been associated with X-linked lymphoproliferative disease 2. Based on this diagnosis, the child received an allogenic bone marrow transplant because the recommended treatment for X-linked lymphoproliferative disease 2 is a bone marrow transplant

to prevent the development of a lethal ailment, hemophagocytic lymphohistiocytosis. Following his bone marrow transplant, the child's gastrointestinal symptoms resolved. The child had inflammatory bowel-like symptoms due to loss of tolerance to the bacteria in his gastrointestinal tract (Worthey, Mayer et al. 2011, Warr, Robert et al. 2015).

Since the initial success of using WES to diagnose the above rare Mendelian disorder, exome sequencing has been used to diagnose novel Mendelian diseases as well as identify novel variants for diseases with known phenotypes. Furthermore, exome sequencing can assist physicians in diagnosing patients who have not yet exhibited the full spectrum of symptoms and it can be used for prenatal diagnoses (Sassi, Guerreiro et al. 2014, Warr, Robert et al. 2015).

Benefits of early diagnosis and finding the causal variant include correct treatment, alleviation of further invasive testing, better prognostic advice and eligibility for benefits and clinical trials (Iglesias, Anyane-Yeboa et al. 2014, Warr, Robert et al. 2015).

It is assumed that with time, as the cost of sequencing decreases, the enthusiasm for WES will wane because scientists want more comprehensive sequences. It is thought that currently, WES is used more often than whole genome sequencing (WGS) because it enriches for the coding region of the genome, which is more easily interpretable, and is less costly, allowing for greater numbers of samples to be sequenced and analysed. WGS

encompasses most of the genome, but it requires a larger amount of sequencing to achieve the same amount of coverage as WES. However, when comparing the performances of WES to WGS, one sees that there are certain advantages to WES. Because WES can achieve higher coverage due to target-enriched sequencing over specific regions of the genome, WES is able to pick up regions of genome missed by WGS. Thus, exome sequencing, because of greater base coverage due to enrichment, can identify variants not identified using WGS. Also, WES can be used to clarify or validate reads from WGS where there has been low depth of coverage (Bamshad, Ng et al. 2011).

1.3. Platelet Disorders

Among the diseases which are manifested by platelet-related bleeding are inherited thrombocytopenias (IT). ITs result from genetic defects affecting the biogenesis of platelets from its precursor, the megakaryocyte (Mk). In ITs, defects in platelet structure or function results in thrombocytopenia which can then lead to haemorrhaging, which can range from mild to life-threatening, regardless of platelet count; however, the severity of thrombocytopenia (defined as platelet counts of less than $150 \times 10^9/L$) does not positively correlate with the severity of bleeding (Nurden and Nurden 2011). Also, the clinical expression of the same bleeding disorder can vary from one patient to the next. This lack of correlation and consistency and the lack of understanding of why this is so, further underscores the fact that we still have much to learn concerning platelets:

the biology underlying platelet biogenesis, the structures and interactions of platelet proteins, and the receptors and signalling pathways used by platelets for activation/aggregation.

Until the end of the century, only a few ITs were known and they were identified because of severe haemorrhaging episodes. Thus, diseases labelled as inherited thrombocytopenias were few in number and not clearly characterized until around 15 years ago (Noris and Balduini 2015). Since then, knowledge concerning ITs has burgeoned because of the application of next generation sequencing (NGS) to identify new genes that cause ITs. NGS, which parallelizes the DNA sequencing process, produces thousands of sequence variants concurrently, allowing sequencing to be faster, more efficient, and less expensive. This has facilitated the sequencing of large numbers of patients' DNA to identify new genes as well as new variants of known genes that are associated with diseases resulting in thrombocytopenia. The recruitment and characterisation of large series of patients equipped with genetic information along with large coordinated international collaborations have assisted in identifying and characterising the increased number of novel disorders in recent years. Fourteen newly identified genes have been identified since 2010 - increasing the total number of well-characterised genes attributed to IT to 25. Currently, hereditary thrombocytopenia is thought to affect at least 2.7 per 100,000 individuals (Balduini and Noris 2016). However, it is believed that these

genes account for only around 50% of patients with IT, suggesting that there are still many more genes that need to be identified (Balduini, Pecci et al. 2012).

The best known IT prior to the usage of NGS was the biallelic form of Bernard-Soulier syndrome (BSS), which is characterised at birth with recurrent life-threatening haemorrhaging (Balduini and Noris 2016). Other ITs with severe bleeding tendencies include Wiscott-Aldrich syndrome (WAS), congenital amegakaryocytic thrombocytopenia (CAMT), and gray platelet syndrome (GPS). In the past, spontaneous haemorrhaging and prevention of hemorrhages and attainment of hemostasis after a large bleed following injury was always a major concern (Noris and Balduini 2015). Nowadays, most patients diagnosed with these “newer” ITs have only mild or moderate thromobocytopenia with little to no bleeding diathesis (Balduini and Noris 2016). Excessive bleeding may only become evident when the patient is appropriately challenged with surgery, childbirth, menstruation, or injury (Watson, Lowe et al. 2013). Platelet abnormalities are only suspected and diagnosed when the patient presents with a disproportionate amount of bruising for the severity of injury, with bruising not associated with trauma, or excessive bleeding at mucocutaneous membranes (Cox, Price et al. 2011). This “pattern” of bleeding i.e. petechiae, bruising at non-contact areas of the body, or mucosal bleeding such as gingival bleeding, nosebleeds and menorrhagia, allows the clinician to differentiate platelet disorders from other coagulation defects such as haemophilia and type III von Willebrand disease (vWD), where hemorrhaging are more severe and are

located in joint spaces, soft tissues, and muscles (van Ommen and Peters 2012). Although patients with “newer” ITs do not have as severe bleeding risks, they nonetheless are predisposed to other risks because such ITs are associated with systemic diseases: renal failure due to glomerulopathies, hematologic malignancies, skeletal defects, immunodeficiency, etc. which can be fatal (Noris and Balduini 2015). Now, because the vast majority of patients with hereditary thrombocytopenia have mild to no bleeding risks, the diagnosis of thrombocytopenia is often made incidentally in adult life and at times with the misdiagnosis of immune-mediated thrombocytopenia, known as Immune thrombocytopenic purpura (ITP). ITP is an autoimmune disorder caused by auto-antibodies (IgG) in ~60% of cases to GPIIb/IIIa or GPIb/IX. Patients are thrombocytopenic but have normal bone marrow morphology and no other known cause for the decreased platelet counts. Misdiagnosis can result in futile and dangerous therapies such as splenectomy or intravenous steroids (Balduini, Savoia et al. 2013).

Studying inherited platelet disorders is important because it can assist in identifying and providing the correct diagnosis. It will additionally assist in the further understanding and clarification of platelet development and biology. Such studies can ultimately lead to discovery of treatments for ITs, treatments for thrombocytopenias due to causes other than genetic mutations as well as for treatments of diseases caused by coagulopathies. Also, because many of the newly identified ITs are part of a syndrome with other associated pathologies,

identification of the genetic defect will assist in appropriate medical follow-up and care when the other diseases develop (Balduini and Noris 2016).

1.4. Platelet Biogenesis and Inherited Thrombocytopenias

The production of platelets involves two phases. The first phase involves Mk differentiation, which takes days to be completed – there is an enormous amount of nuclear proliferation and enlargement of the Mk cytoplasm so that cytoskeletal proteins, platelet-specific granules and sufficient membrane are present for platelet formation. The second phase lasts relatively quickly – completion is within hours. In the second phase, Mks remodel their cytoplasm to form proplatelets and then preplatelets, which subsequently are shed into the blood stream as platelets (Machlus and Italiano 2013). In humans, it takes about 5 days for Mks to undergo polyploidization, mature and release platelets (Odell, Jackson et al. 1970). Once in the blood stream, human platelets survive for 7 to 10 days (Jackson and Edwards 1977).

Platelets, despite being anuclear, have unique receptors, a highly organized cytoskeleton and specialized secretory granules (Thon and Italiano 2010). They originate from Mks, which are the largest (50 to 100 μm) yet one of the rarest of cells in the bone marrow (Nakeff and Maat 1974). For the assembly and release of platelets into the blood stream, Mks need to become polyploid by the process of endomitosis, which involves cycles of DNA replication without cell division (Figure 1.10). Endomitosis in Mks is similar to other cell cycles: G1, S and G2 phases are followed by mitosis during which anaphase chromosomes separate

and cleavage furrow formation begins; however, in Mks the nuclei do not completely separate and cleavage furrows regress before cytokinesis is completed. Because there is no cleavage, nuclei of Mks are multilobulated with a polyploid nucleus (Geddis 2010).

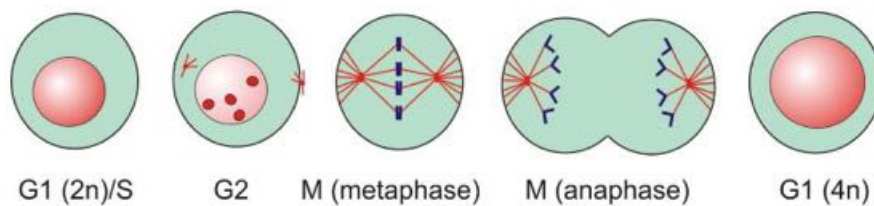


Figure 1.10 - Schematic diagram of endomitosis (Geddis 2010).

After nuclear polyploidization, Mk maturation commence whereby they migrate from their osteoblastic niches to their vascular niches in the bone marrow to form pseudopodia, which contain the majority of the Mk's cytoplasm in long narrow processes called proplatelets. A Mk can have between 10 to 20 proplatelets. Mks extrude their nucleus (Machlus and Italiano 2013) and proplatelets develop tandemly placed platelet-like swellings that are connected by thin cytoplasmic bridges (Italiano, Lecine et al. 1999). In the sinusoidal blood vessels of the bone marrow, bar-bell shaped proplatelets form, and are released. Proplatelets interconvert to preplatelets and eventually, platelets form when bar-bell shaped proplatelets undergo fission (Machlus and Italiano 2013).

Polyploidy is a distinctive feature of mature Mks. It has been proposed that Mk polyploidization serves to amplify functional genes in order to simplify the massive production of lipids and proteins needed to form the extensive invaginated membrane system (IMS) (Zimmet and Ravid 2000). The IMS, the hallmark of a mature Mk cytoplasm, is a large complicated membrane network of cisternae and tubules that is continuous with the Mk plasma membrane. It occupies the entire megakaryocytic cytoplasm except for a narrow area at the periphery. The IMS is thought to serve as the source of membrane needed to form proplatelets (Schulze, Korpai et al. 2006). During the development of platelets, their granule and organelle contents are transported from the Mk cell body as streams of individual particles (Italiano, Lecine et al. 1999). Alpha granules, the most abundant organelle in platelets (Machlus, Thon et al. 2014), are made in the Golgi apparatus of Mks and consist of endogenous proteins as well as proteins from the extracellular environment. Alpha granule contents include von Willebrand factor, fibrinogen, fibronectin, thrombospondin, platelet factor 4, transforming growth factor β , platelet derived growth factor, and insulin-like growth factor. Dense granule contents include ADP/ATP, Ca^{2+} , histamine, and serotonin. These granules will eventually be transferred to the newly made platelets (Whiteheart 2011).

Mks are derived from hematopoietic stem cells (HSC) (Figure 1.11), which differentiate in the osteoblast to more committed progenitors: the common myeloid progenitor (CMP) and the megakaryocyte-erythroid progenitor (MEP).

MEPs develop into cells that belong to the megakaryocyte and erythrocyte lineages (Kaushansky 2008).

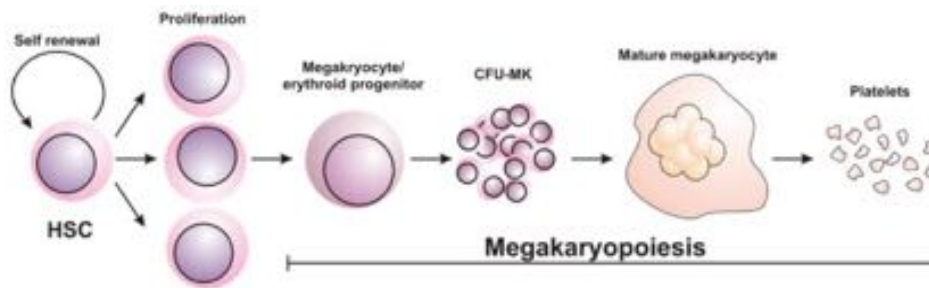


Figure 1.11 – Schematic diagram of platelet development (Geddis 2010).

Thrombopoietin is a glycoprotein produced by the liver and kidney and it functions to regulate the production of platelets. The commitment of HSCs to Mks is initiated by the binding of thrombopoietin (TPO) to its Mk specific receptor, c-Mpl, a homologue of the oncogene, v-Mpl. Both HSCs and Mks express c-Mpl. TPO and c-Mpl are essential for HSC and megakaryocyte growth and differentiation. Ligation of TPO with c-Mpl results in the cytoplasmic protein tyrosine kinase, Janus kinase 2 (Jak2), to bind to c-Mpl and become activated. Based on crystallographic data of the erythropoietin receptor, it is believed that c-Mpl, in the unbound state, is a homodimer. However, upon binding to TPO, the C-terminal ends of c-Mpl are brought closer together, allowing Jak2 molecules, which are bound to the receptor, to activate each other through trans-autophosphorylation (Witthuhn, Quelle et al. 1993). Jak2 phosphorylates itself at multiple tyrosines as well as phosphorylating c-Mpl, which provides docking sites

for SH2 domain containing proteins. Jak2 activation leads to activation of multiple downstream proteins: signal transducer and activator of transcription (STAT), mitogen-activated protein kinase (MAPK) and phosphoinositol-3 kinase (PI3K). There is also negative regulation of TPO signalling to prevent thrombocythemia or leukemia. Proteins involved in the negative signalling pathways include the suppressor of cytokine signalling (SOCS) family and Lnk, an adaptor protein, both of which inhibit Jak signalling (Geddis 2010). Mk growth and differentiation requires multiple growth factors, the most important being TPO. This is exemplified in *c-Mpl* and *Tpo*-null mice, which are thrombocytopenic and have reduced HSCs and progenitors of all lineages (Geddis 2010). Also, transcription factors such as RUNX1, GATA1, FLI-1 and ETV6 form complex networks regulating the differentiation of Mks (Dore and Crispino 2011).

Many different genes control each of the phases of megakaryocyte and platelet development. A causal variant occurring in any of these genes may result in thrombocytopenia. When a disease-causing variant is present in the genes involved in the proliferative phase of megakaryopoiesis, where proliferation of promegakaryoblasts is affected, there is absent or reduced number of Mks in the bone marrow. When genes involved in the maturation phase are affected, the number of Mks can be normal or increased but Mks are immature i.e. small with hypolobulated nucleus and sometimes with reduced alpha granules. When Mk production is normal, thrombocytopenia is due to defective proplatelet formation

or abnormally increased platelet removal (Johnson, Fletcher et al. 2016, Savoia 2016). Some examples of genes with disease-causing variants which affect each of the stages of platelet development include *THPO/MPL* and *HOXA11* affecting Mk production; *RUNX1*, *FLI1*, *ETV6*, *GATA1* affecting Mk maturation; and *MYH9*, *WAS*, *GP1BA*, *GP1BB*, *GP9*, *ITGA2B* and *ITGB3* affecting platelet production and development (Figure 1.12 and Table 1.3).

1.4.1. Inherited Thrombocytopenias caused by genes affecting Mk production

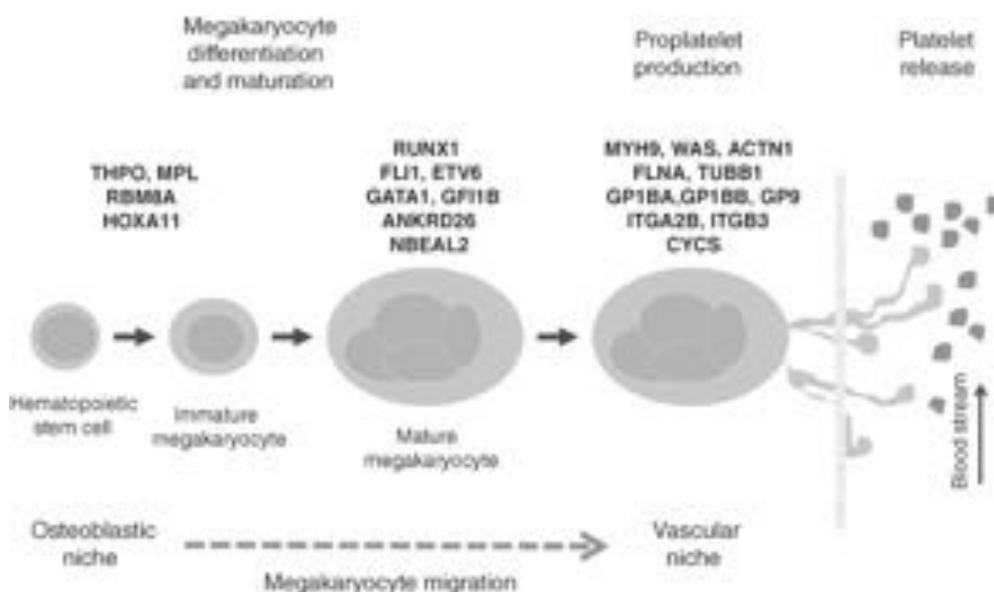


Figure 1.12 - Schematic diagram of the different stages of platelet development associated with the genes which can mutate at each stage. Above each stage are the different genes which can be mutated to cause an inherited thrombocytopenia. *THPO* is the gene for TPO (Savoia 2016).

THPO and *MPL* are essential genes in the production of platelets. Many different causal variants inherited in a homozygous or heterozygous manner have been identified in these two genes. Loss of function variants of the *MPL* gene result in congenital amegakaryocytic thrombocytopenia (CAMT). Patients with CAMT, an autosomal recessive disease, have a high risk of life-threatening hemorrhages and progress within the first years of life to develop bone marrow aplasia which can result in death unless bone marrow transplants are performed (Balduini, Pecci et al. 2012). Forty-one different disease-causing variants, found along the entire *MPL* gene, have been identified (Balduini, Pecci et al. 2012). The functional consequences resulting from the variants have not all been completely identified. However, it appears that patients with complete loss of function of the *MPL* gene have permanently low platelet counts whereas presence of any residual *MPL* function will result in improvement of thrombocytopenia after the first year of life (Ballmaier and Germeshausen 2011). However, the currently reported *MPL* variants explain only 60% of the CAMT cases, suggesting that CAMT might be caused by other genes not yet identified (Ballmaier and Germeshausen 2011).

In regards to the *THPO* gene, there is a gain of function variant resulting in increased platelet production (Wiestner, Schlemper et al. 1998), as well as a loss of function variant of the *THPO* gene (p.R38C) which reduces binding affinity of TPO to MPL. An example of this gain of function variant is the disease hereditary thrombocythemia, which is transmitted in an autosomal dominant

fashion. TPO is persistently elevated in the afflicted individuals studied. The mutation that was identified was a G to C transversion at the splice donor site of intron 3 of *THPO*. This mutation leads to mRNAs with shortened 5'UTR resulting in more efficiently translated mRNA and thus systemic overproduction of TPO (Wiestner, Schlemper et al. 1998). Homozygous patients with the loss of function variant develop aplastic anemia, whereas heterozygous patients have mild thrombocytopenia (Dasouki, Rafi et al. 2013). Another example of causal gene variants that affect megakaryocyte production involves the gene *HOXA11*, a transcription factor with an unknown role in hematopoiesis but which encodes for a homeobox protein that regulates gene expression, morphogenesis and differentiation. Pathological variants of the *HOXA11* gene result in an autosomal dominant IT named radio-ulnar synostosis with amegakaryocytic thrombocytopenia (RUSAT). Patients with RUSAT present with proximal fusion of the radius and ulna (resulting in difficulty with supination and pronation of forearm) plus thrombocytopenia caused by absent or reduced number of megakaryocytes in the bone marrow which also may become aplastic. The initial patients which presented with RUSAT were two unrelated families which were nonconsanguineous and of different ethnic backgrounds. Both fathers and affected children had radio-ulnar synostosis but only three of the four affected children had symptomatic thrombocytopenia since birth. Utilising Sanger sequencing, the mutation was identified as a single base-pair deletion in the highly conserved homeodomain, located at exon 2 of *HOXA11*. The point mutation is located at a critical DNA binding site whereby the mutation causes a

nonsense mutation because there is a translational frameshift resulting in a premature stop. There appears to be no *HOXA11* mRNA expression as determined by RT-PCR in platelets, megakaryocytic cell lines or TPO-induced CD34⁺ umbilical cord stem cells. It is postulated that the patients that were studied developed amegakaryocytic thrombocytopenia secondary to alterations in the interaction of *HOXA11* or its protein with other genes that play an important role in megakaryocytic differentiation during the specific embryonic stage important for hematopoiesis (Thompson and Nguyen 2000). *HOX* genes' function in normal hematopoiesis has been extensively studied using gene expression analysis, targeted genetic insertion and deletion studies in hematopoietic stem cells (HSC) and early hematopoietic progenitors of murine models and cell lines. However, it should be realised that targeted genetic deletion studies are difficult to interpret given the functional redundancy of *HOX* genes. Highly expressed in most primitive HSCs and progenitors, *HOX* genes' expression is virtually absent in CD34⁻ cells, which are differentiated bone marrow cells (Alharbi, Pettengell et al. 2013). Examples of "knock-down" studies include those for *HOXA9*, the most widely expressed *HOX* family member, *HOXA5* and *HOXA7*: *HOXA9*^{-/-} mice are very deficient in myeloid and lymphoid cells (Magnusson, Brun et al. 2007); *HOXA5*^{-/-} mice have an increase in erythroid progenitors and a decrease in myelomonocytic cells (Crooks, Fuller et al. 1999) and *HOXA7*^{-/-} mice demonstrate a decrease in megakaryocytic/erythroid progenitors resulting in reticulocytosis and thrombocytopenia (So, Karsunky et al. 2004). To date, *HOXA11*^{-/-} mice have been generated whereby skeletal

deformities were analysed but haematological abnormalities were not mentioned (Small and Potter 1993).

1.4.2. Inherited Thrombocytopenias caused by genes affecting Mk maturation

The next stage of megakaryopoiesis where genetic causal variants can affect the production of platelets is during megakaryocyte maturation. Most of the genes which are currently known and which are affected at this stage, encode for transcription factors such as GATA1, FLI1, RUNX1 and ETV6. *GATA1* encodes for a zinc finger transcription factor found in megakaryocytic/erythroid progenitors and controls megakaryocytic and erythroid development. GATA1, which binds cofactor FOG1, has two zinc fingers, a C-terminal one which binds DNA avidly and an N-terminal one which stabilizes DNA binding. *GATA1* variants can cause an X-linked macrothrombocytopenia with decreased α -granule content, whereby Mks are immature. *GATA1* variants may also cause dyserythropoietic anemia or β -thalassemia. GATA1 also interacts with other transcription factors such as FLI1 (Millikan, Balamohan et al. 2011). FLI1 and ETV6 both belong to the same ETS family of transcription factors, which is characterized by the following structure: at the N-terminus is a pointed (PNT) domain, which dimerizes and interacts with other transcription factors; a central inhibitory domain; and, at the C-terminus, the DNA-binding domain (ETS). Disease-causing variants in the *ETV6* gene result in autosomal dominant thrombocytopenia, an elevated erythrocyte mean corpuscular volume (MCV) and B cell precursor acute lymphoblastic leukemia (ALL). Mks are immature i.e. small and hypolobulated with abnormal red cell precursors (Noetzli, Lo et al. 2015). FLI1 interacts with GATA1 as well as with

RUNX1 to regulate many genes needed for terminal megakaryocytic maturation (Huang, Yu et al. 2009). *FLI1* when deleted at its chromosomal region, 11q24.3, results in Jacobsen syndrome, which, when thrombocytopenia is also present, is named Paris-Trousseau syndrome. The Mks of these patients are small with low ploidy but platelets are large i.e. macrothrombocytopenic with giant α -granules (Bluteau, Glembotsky et al. 2012). *RUNX1*, which belongs to the family of genes called core binding factor $\alpha 2$ (CBF- α), is heterodimeric because it complexes with core binding factor β (CBF- β). CBF- β doesn't directly bind DNA but increases the DNA binding ability of the RUNX1-CBF- β complex and makes it more stable. Mutations of *RUNX1* result in the disease called familial platelet disorder with propensity to acute myelogenous leukemia (FPD-AML) (Savoia 2016). Patients with FPD-AML have moderate thrombocytopenia with decreased α -granules, small, increased number of Mks and a predisposition to develop myelodysplastic syndromes or acute myeloid leukemia (AML) (Savoia 2016). One of the genes to which RUNX1 binds is the 5'UTR of *ANKRD26*, thereby inhibiting *ANKRD26* transcription. Causal variants that occur at the 5'UTR of *ANKRD26* prevent RUNX1 binding and result in the autosomal dominant disease, ANKRD26-related thrombocytopenia (ANKRD-RT or THC2). Patients with ANKRD-RT have an increased number of megakaryocytes, which are small with hypolobulated nuclei and have in their cytoplasm, particulate structures (PaCS) which are made of proteasome complexes and polyubiquinated proteins (Necchi, Balduini et al. 2013). Patients with ANKRD26-RT are also predisposed to developing haematological malignancies, especially AML (Noris, Favier et al. 2013).

Targetted disruption of the *ANKRD26* gene in mice results in hyperphagia, obesity and gigantism, all of which are caused by defects in the primary cilia found in regions of the brain which control appetite and energy homeostasis (Bera, Liu et al. 2008, Acs, Bauer et al. 2015). The molecular mechanism for adipogenesis is attributed to increased activation of ERK and mTOR signalling pathways (Fei, Bera et al. 2011). Analyses of *ANKRD26*^{-/-} mice did not demonstrate altered haematological cell counts. However, a study of 11 Italian and French pedigrees with THC2 revealed no dramatic changes in megakaryocyte differentiation but a defect in proplatelet development, as seen in electron micrographs of THC2 patient megakaryocytes derived from purified blood CD34⁺ cells treated with TPO and SCF (Stem cell factor). Correction of this defect in proplatelet development was accomplished by using a MEK inhibitor. The mechanism postulated to explain this defect is the persistant activation of the MAPK/ERK1/2 pathway, which was seen using THC2 patient-derived megakaryocytes when compared to controls. (Bluteau, Balduini et al. 2014).

1.4.3. Inherited Thrombocytopenias caused by genes affecting platelet production and development

The next group of ITs result from defective proplatelet formation or release. Examples of diseases from this group include MYH9-RD, Wiskott-Aldrich syndrome (WAS), Bernard Soulier Syndrome (BSS), and Glanzmann thrombasthenia (GT). MYH9-RD, one of the most commonly identified ITs, is caused by causal variants in the *MYH9* gene, which encodes for the heavy chain of non-muscle myosin IIA in non-muscle cells (Balduini, Pecci et al. 2011). Non-

muscle myosin IIA's protein structure consists of an amino terminal globular head followed by a coiled coil domain and a non-helical tail at the carboxy terminus, which dimerizes and associates with two pairs of light chains. The heavy chain of non-muscle myosin IIA is expressed throughout hematopoiesis and normally contributes to cytokinesis, membrane rigidity, and contraction of matrix. The proposed mechanistic cause for macrothrombocytopenia is the lack of heavy chain non-muscle myosin IIA reactivation in terminal platelet formation. When proplatelets squeeze through the sinusoidal blood vessels, it is shear, which divides the proplatelets to form preplatelets. Fluid stress activates heavy chain non-muscle myosin IIA, which accumulates at stressed sites to cause fragmentation of preplatelets to form platelets. When heavy chain non-muscle myosin IIA mutates, it doesn't aggregate, but is diffuse, preventing preplatelets from fragmenting and thus forming normal sized platelets (Spinler, Shin et al. 2015). Patients with MYH9-RD are macrothrombocytopenic at birth and have neutrophils, which contain cytoplasmic aggregates of wild-type and mutated myosin-9 (Savoia, De Rocco et al. 2010). Later in life, the majority of patients with MYH9-RD develop sensorineural hearing loss, presenile cataracts, chronic or intermittent elevated liver enzymes, and glomerulonephritis which can progress to endstage renal failure (Pecci, Biino et al. 2012).

WAS and X-linked thrombocytopenia (XLT) are both caused by causal variants located on the gene for Wiskott-Aldrich syndrome protein (WASp). WASp is located on the X chromosome, is expressed in hematopoietic cells and functions

to regulate the actin cytoskeleton. Males with WAS have multiple medical problems: moderate to severe congenital microthrombocytopenia; eczema; severe immune compromise with risks for developing lymphoproliferative disorders. Thrombocytopenia is due to inefficient platelet release into the blood stream as well as increased clearance of platelets from the circulation (Balduini, Pecci et al. 2012). Almost 150 different mutated alleles have been described for WAS: the most common being missense and splice-site mutations, short deletions and nonsense mutations (Pai and Notarangelo 2010). Loss-of-function variants usually result in WAS whereas variants which allow expression of the WASp, however small, results in XLT. There is a strong correlation between severity of WAS variants and the phenotypic expression of WAS/XLT (Massaad, Ramesh et al. 2013).

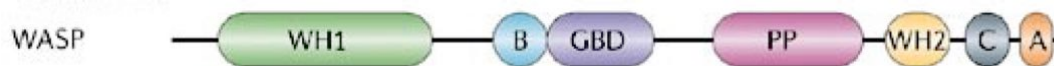


Figure 1.13 - Structural domain organisation of Wiskott-Aldrich Syndrome protein (WASp). WH2(WASP-homology-2; also called verprolin-homology), C(cofilin-homology or connector), A(acidic region), PP(polyproline domain), GBD(GTPase binding domain), B(basic domain), WH1(WASP-homology-1 domain) (Goley and Welch 2006).

WASp (Figure 1.13) is a class I nucleation promoting factor. It is needed for actin-related protein 2/3 (ARP2/3) complex activation because the ARP2/3 complex (the nucleator) itself has no catalytic activity. The ARP2/3 complex is a

major participant in cellular actin polymerization, forming branched-actin-filament networks in the actin cytoskeleton (Goley and Welch 2006).

The key member of the actin cytoskeleton is the monomeric globular (G) actin, an ATPase, which can assemble itself to form filamentous (F) actin. Actin polymerisation is tightly coupled to ATP hydrolysis, which controls the kinetics in the assembly and disassembly of actin and its binding with interacting proteins (Pollard and Borisy 2003). The formation of a new actin filament is termed nucleation. Spontaneous nucleation is possible but kinetically laborious so for a cell to be efficient it has factors, which can accelerate or bypass nucleation to allow for more effective assembly of actin. There are three classes of proteins called nucleators, which can initiate actin polymerisation: the actin-related protein 2/3 (ARP2/3) complex, the formins and spire. Each nucleator initiates nucleation by a particular method. The ARP2/3 complex is believed to mimic an actin dimer or trimer and serves as a template for the initiation of a new actin filament which branches off an existing filament to generate γ -branched actin networks (Goley and Welch 2006). The ARP2/3 complex consists of seven polypeptides of which two, ARP2 and ARP3, are actin-related proteins of the ARP2 and ALP3 subfamilies. Other subunits of the complex include ARPC1, ARPC2, ARPC3, ARPC4 and ARPC5. By itself, the ARP2/3 complex has no biochemical activity; but, when bound to a nucleation promoting factor (NPF), it will be activated to begin formation of a new filament (daughter) that branches off from an existing filament (mother) at a 70° branch angle (Mullins, Heuser et al. 1998, Amann,

2001 #152). Termed autocatalytic branching or dendritic branching, this coupling of nucleation and branching is the functional essence of the ARP2/3 complex in the cell.

Nucleation promoting factor (NPF) proteins are made up of two groups, class I and class II, based on how they activate the ARP2/3 complex and their effect on the γ -branching reaction (Goley and Welch 2006). There are a diverse array of class I and II NPF proteins with different overall domain structures. For class I members, of which WASp (Figure 1.12) is one, the only common domain shared by all, is the WCA domain, which consists of a WASP-homology-2 (WH2 or verprolin-homology) domain that binds G-actin and a central (C or cofilin-homology) domain and an acidic (A) region, which together (CA) binds the ARP2/3 complex (Marchand, Kaiser et al. 2001, Chereau, Kerff et al. 2005). A model for how class I NPF functions involves the acidic region binding to the ARP2/3 complex which allows the central region to initiate activating conformational changes in the complex and the WH2 and central regions to bind and present actin monomers to the ARP2/3 complex to facilitate the formation of a nucleus for the polymerization of the daughter filament (Panchal, Kaiser et al. 2003). Class I NPFs are activated by a number of factors, the most studied being the Rho-family GTPases, CDC42 and Rac, which bind the GBD domain (Bompard and Caron 2004).

Actin polymerization is required for multiple vital cellular processes such as cell migration, phagocytosis, endocytosis, vesicle trafficking and cytokinesis. However, the nucleation promoting factor required for each of the cellular processes vary. Phagocytosis is the process whereby large (greater than 0.5 μm in diameter) items such as bacteria are engulfed. Cells such as phagocytes utilise actin polymerization to negotiate the plasma membrane around the bacteria's surface. The only NFP which is used for phagocytosis is WASp, which is expressed in hematopoietic cells in mammals (Castellano, Le Clainche et al. 2001). The internalisation of small particles (endocytosis) also requires actin polymerisation; however, the precise role that the ARP2/3 complex and WASp play in this process which includes invagination of the membrane, pinching off vesicles, or driving vesicles away from the plasma membrane, remains to be clarified (Kaksonen, Sun et al. 2003). WASp may also play a role in adhesion and in podosome formation (Goley and Welch 2006).

The GPIb/IX/V complex consists of glycoproteins (GP) Ib α , GPIb β , GPIX, and GPV, which associate in a ratio of 2:4:2:1 in the endoplasmic reticulum (ER) of Mks and platelets before migrating to the membrane (Luo, Mo et al. 2007). Platelets require this complex to bind and aggregate on the exposed endothelium of damaged blood vessels. GPIb α binds von Willebrand factor (vWF). GPIb α , GPIb β and GPIX are needed for the correct assembly of the complex in the ER. More than 50 different causal variants have been identified in *GPIBA*, *GPIBB* and *GP9* (Lopez, Andrews et al. 1998). Variants that decrease the expression of the

GPIb/IX/V complex or its ability to bind vWF result in the biallelic (recessive form) or the monoallelic (dominant variant) form of Bernard-Soulier syndrome (BSS). Patients with biallelic BSS have moderate macrothrombocytopenia but a severe bleeding diathesis. The GPIb/IX/V complex on platelets is severely reduced to undetectable and when patient platelets are tested for aggregation using ristocetin, there is none to barely any aggregation. Generally, the monoallelic form of BSS is clinically less severe: patient platelet counts are only slightly decreased, bleeding risk is mild or absent and ristocetin-induced aggregation is normal or slightly abnormal. Ristocetin is an antibiotic, which was previously clinically used to treat Staphylococcal infections but was discontinued because of its side effects in patients: thrombocytopenia and platelet aggregation. Why ristocetin causes platelet aggregation is hypothesized to be due to its phenolic groups, which bind platelets, reducing the platelets negative surface charge. Normally, the platelets' net negative charge leads to repulsion and prevents bridging by vWF via GPIb. When ristocetin, which is positively charged, is added, it binds via its phenolic groups to platelets (probably at protein-associated free carboxy groups on the platelets), reducing the platelets' net negative charge, allowing for closer platelet interaction, which then allows for the vWF to bridge between platelets, causing agglutination (Coller and Gralnick 1977).

There are also variants on *GP1BA* which increases the affinity of GPIb α for VWF resulting in macrothrombocytopenia - this is called platelet type von Willebrand disease (Savoia, Kunishima et al. 2014). Another important receptor found on

platelet and MK cell surfaces is the GPIIb-IIIa complex, which when exposed, has a binding site for fibrinogen. When its genes, *ITGA2B* and *ITGB3*, which encode for the two integrin subunits of the receptor, α IIb and β 3, have biallelic pathological variants, Glanzmann thrombasthenia (GT) develops. GT is an autosomal recessive disease whereby platelets are functionally abnormal but phenotypically normal: platelets cannot aggregate but are of normal size and number (Gresele, Falcinelli et al. 2009). However, there is also an autosomal dominant form of macrothrombocytopenia whereby gain of function variants are found in either the α IIb or β 3 subunit resulting in constitutive activation of the GPIIb-IIIa complex; however, there is no platelet aggregation defect (Kashiwagi, Kunishima et al. 2013).

Disease (abbreviation)	Frequency* /spontaneous bleeding	Inheri- tance	Gene (chromosome location)	Other features
Normally sized platelets				
Congenital amegakaryocytic thrombocytopenia (CAMT)	+++/yes	AR	<i>MPL</i> (1p34)	Always evolves into bone marrow aplasia in infancy.
Amegakaryocytic thrombocytopenia with radio-ulnar synostosis (RUSAT or CTRUS)	+/yes	AD	<i>HOXA11</i> (7p15-14)	Radio-ulnar synostosis ± other defects. Possible evolution into Aplastic anemia.
Familial platelet disorder and predisposition to acute myelogenous leukemia (FPD/AML)	++/no	AD	<i>RUNX1</i> (21q22)	High risk of developing leukemia or myelodysplastic syndrome.
ANKRD26-related thrombocytopenia (THC2 or ANKRD26-RD)	++/no	AD	<i>ANKRD26</i> (10p2)	Risk of developing acute myelogenous leukemia
Large platelets				
GATA1-related disease (GATA-RD)	++/yes	XL	<i>GATA1</i> (Xp11)	Hemolytic anemia, possible unbalanced globin chain synthesis, possible congenital erythropoietic porphyria
Paris-Trousseau thrombocytopenia (TCPT), Jacobsen syndrome (JBS)	++++/yes	AD	<i>FLI1</i> (11q24.3)	Cardiac and facial defects, developmental delay ± other defects
<i>MYH9</i> -related disease (<i>MYH9</i> -RD)	++++/no	AD	<i>MYH9</i> (22q12-13)	Cataracts, nephropathy, and/or deafness
Platelet-type vWD (VWDP)	++++/yes	AD	<i>GP1BA</i> (17p13)	Platelet count goes down under stress
Bernard-Soulier syndrome (BSS) Biallelic Monoallelic	+++/no	AR AD	<i>GP1BA</i> (17p13) <i>GP1BB</i> (22q11), <i>GP9</i> (3q21)	Giant platelets Large platelets
Glanzmann thrombasthenia (GT)	+/no	AD	<i>ITGA2B</i> (17q21.31) <i>ITGB3</i> (17q21.32)	Platelet anisotropy
Small platelets				
Wiskott-Aldrich syndrome (WAS)	++++/yes	XL	<i>WAS</i> (Xp11)	Severe immunodeficiency leading to death in infancy.
X-linked thrombocytopenia (XLT)				Mild immunodeficiency

Table 1.3 - Overview of inherited thrombocytopenias (confer section 1.3). Frequencies are indicated as follows: *++++, >100 families reported; +++, >50 families reported; ++, > 10 families reported; +, < 10 families reported. AD, autosomal dominant; AR, autosomal recessive; XL, X-linked. (Balduini and Savoia 2012, Balduini, Savoia et al. 2013).

1.5. Identification of genes responsible for inherited thrombocytopenias

The genetic causal variants identified in ITs usually involve a single gene that follows a Mendelian pattern of inheritance. Of the Mendelian diseases that have been studied, the majority are caused by rare mutations that affect the function of the protein; around 85% of disease-causing mutations are found in exomes (the coding regions of the genome) or in canonical splice sites (Choi, Scholl et al. 2009). The human genome has around 180,000 exons so exomes constitute only approximately 1.5% of the human genome. As a result, using whole exome sequencing (WES) to identify rare Mendelian mutations is a more efficient and cost-effective approach to studying Mendelian disorders than whole genome sequencing. Exome sequencing allows targeted sequencing of predefined regions at high depth, thus reducing the complexity of the genomic analyses as well as reducing storage costs. Reduced costs allow more samples to be sequenced and thus enable larger population based comparison (Schneeberger 2014). Other benefits of exome sequencing include the ability to identify genes in certain instances when conventional approaches have failed. For instance, linkage studies cannot be done on extremely rare diseases because there are not enough participants for a statistically powerful study. Linkage studies also cannot identify sporadic mutations, genetic heterogeneity (causal variants that are located on different genes) nor phenotypic heterogeneity (diverse presentation of the disease leading to uncertain diagnosis) (Ku, Naidoo et al. 2011). However, WES has limitations as well (Schneeberger 2014): if causal

variants are located in regulatory regions (introns), they will be missed by WES; genetic markers can only be genotyped if they are in the targeted regions and WES requires enrichment kits and information about targeted regions.

Similar to other genetic analyses, once candidate genes and variants are identified by WES, further functional studies are required to confirm that the coded protein does what is hypothesised and that the causal variants identified are deleterious.

Although the number of patients which have been identified with ITs have increased, absolute numbers remain small, making it difficult to find patients to study. In the past, genetic studies of IT focused on candidate gene sequencing with whole exome sequencing of individual cases when a causative gene wasn't obvious (Zhang, Churpek et al. 2015). For the past 8 years our laboratory has undertaken a collaborative and more large scale approach to identifying and studying ITs. Called the genotyping and platelet phenotyping (GAPP) study, this has involved recruiting patients with suspected IT or sustained thrombocytopenia from over 25 Haemophilia Comprehensive Care Centres across the UK. Before patients enter the study, they've had clinical and genetic studies to rule out known platelet disorders: GT, BSS, WAS, MYH9-RD, and ITP as well as known non-platelet disorders such as vWD and inherited coagulation factor deficiencies. Once patients have entered the study, their platelets undergo several functional assays to characterize the platelets before having their genomic DNA sequenced

using WES. The diagnosis of platelet disorders therefore involves a two-prong approach - platelet phenotyping plus WES.

Normally, the definitive diagnosis of a platelet disorder is determined by using light transmission aggregometry (LTA), the gold standard method (Born 1962), to analyze platelet rich plasma (PRP) for any defects in platelet aggregation. This is accomplished by using a streamlined panel of platelet agonists at specified concentrations (Dawood, Lowe et al. 2012). Assessment of platelet dense granule ATP release, using lumi-aggregometry provides additional information regarding platelet secretion disorders. However, in thrombocytopenic patients, because lumi-aggregometry is unreliable for platelet counts below $150 \times 10^9/L$, a flow cytometry assay, developed in our laboratory, is performed (Dawood, Wilde et al. 2007). Using the flow cytometer, thrombocytopenic patients' PRP are stimulated with the following agonists to check for the presence and functionality of the following receptors as well as their signalling pathways: ADP (3 and $30\mu M$) - $P2Y_{12}$ receptor, G_i signalling; collagen related peptide (CRP, 0.3 and $3\mu g/ml$) - GPVI receptor and signalling and PAR-1 peptide (10 and $100\mu M$) - protease-activated receptor, G_q signalling. Also checked by flow cytometry is membrane expression for P-selectin, a marker of α and dense granule release and fluorescent fibrinogen binding, a marker of integrin activation.

1.6. A family with severe thrombocytopenia

This project involves the characterization of a novel protein, ANKRD18A (Ankyrin Repeat Domain 18A). The corresponding gene, *ANKRD18A* was identified following the sequencing of the exomes of two cousins from a consanguineous family (Figure 1.14). The proband (patient who is the starting point for a genetic study), patient II:5, was entered in this study at age 3 because of a platelet count of $3 \times 10^9/L$. Phenotypically, he had developmental delay and skull abnormalities because at birth he developed bilateral intraventricular hemorrhages requiring insertion and several revisions of a ventriculo-peritoneal shunt. He received HLA matched platelet transfusions every 1-2 weeks for the first six to twelve months of life. His baseline platelet count remained at approximately $10 \times 10^9/L$. There were no abnormalities in the rest of his full blood count. He had a bone marrow aspirate and trephine (a hole saw used in surgery to remove a circle of tissue) which was normocellular with normal megakaryocyte numbers and morphology and normal cytogenetics. His cousin, II:3, is the only sibling with a platelet abnormality in her immediate family. She has a baseline platelet count of $(15 - 20) \times 10^9/L$, but requires weekly HLA matched platelet transfusions to prevent epistaxis and hematomas, which have been severe enough to necessitate hospitalization. Both the proband and his cousin had normal coagulation parameters and there were no identifiable anti-platelet auto-antibodies or HLA antibodies (Initial clinical work-up of both patients were performed by Dr. Jayashree Motwani and Dr. Mike Williams at Birmingham Children's Hospital).

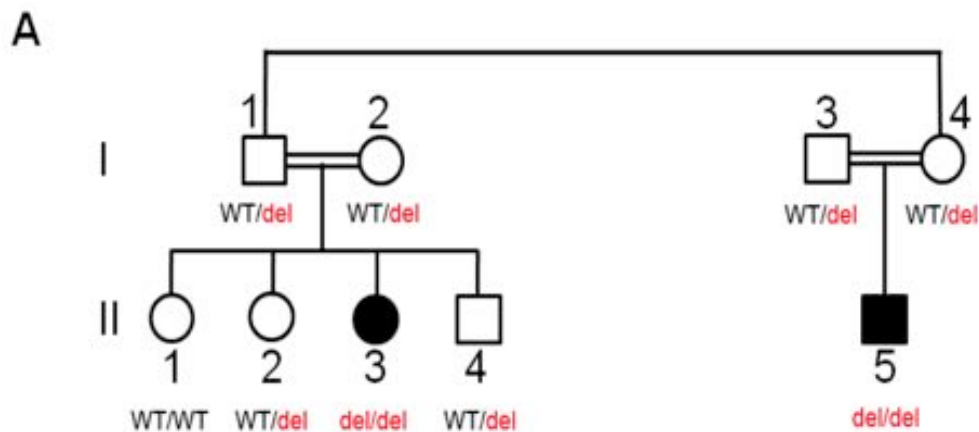


Figure 1.14 - Pedigree of patients affected by ANKRD18A-linked severe thrombocytopenia. Shaded square is the male patient (II:5) with the homozygous non-frameshift deletion of ANKRD18A (p.801-801). Shaded circle (II:3) is his cousin. Double lines linking I:1 with I:2 and I:3 with I:4 signify a first cousin union (analysis performed by N. Morgan).

1.6.1. A genetic variant in *ANKRD18A* correlates with severe thrombocytopenia

To determine the genetic defect underlying the thrombocytopenia seen in these two patients, exome sequencing was performed using the SureSelect human AllExon 50Mb kit (Agilent Technologies) and sequenced on the Hi Seq 2000 (Illumina) with 100 bp paired-end reads. Because Patient II:5 and II:3 were cousins born from consanguineous relationships derived from a single consanguineous kindred, analysis was focussed on identifying a shared homozygous variant secondary to the recessive segregation of the disease. Patient II:5 had 24,293 variations and his cousin, patient II:3, had 23,943. Comparing these variants with the 1000 Genomes project database and our in-house database, which was composed of >250 exomes at the time, 83

homozygous novel variants were identified for patient II:5 and 77 for patient II:3. Of these variants, only 2 homozygous non-synonymous variants [*FRMPD1* (p.A509V) and *GNE* (p.G447R)] and 1 non-frameshift deletion [*ANKRD18A* (p.E799-799 del)] were found in both patients on chromosome 9p13.3. All three variants are located within a tightly linked region of homozygosity on chromosome 9. Information from Sanger sequencing was used to verify these mutations as well as to perform segregation analysis among immediate family members (Figure 1.15). The 2 non-synonymous variants were found in the genes *FRMPD1* (A509V) and *GNE* (G447R). *FRMPD1* encodes for a FERM and PDZ domain containing protein which functions to regulate the subcellular localization of activator G-protein signalling 3 (An, Blumer et al. 2008). It is weakly expressed in hematopoietic cells; and, the alanine that is mutated in *FRMPD1* is not conserved in different species, suggesting that the alanine residue is insignificant. Also, most recently, the variant in *FRMPD1* has been observed to occur at a frequency of 0.0003708, implicating that this low-frequency variant is unlikely to be pathogenic. *GNE* encodes for UDP-N-acetylglucosamine 2-epimerase, a bifunctional enzyme which catalyses the initial two steps in the biosynthesis of N-acetylneuraminic acid (sialic acid). Its structural domain organisation reveals an epimerase domain at its amino terminus and a kinase domain at its carboxy terminus (Stasche, Hinderlich et al. 1997, Eisenberg, Avidan et al. 2001, Tong, Tempel et al. 2009). Mutagenesis and enzymatic activity studies have established that the activities of the epimerase and kinase domains are interrelated and that mutations in either one

of the domains will affect the activities of both domains (Penner, Mantey et al. 2006). Sialic acids are monosaccharides, which are expressed on the cell surfaces of all eukaryotic hematopoietic cells.

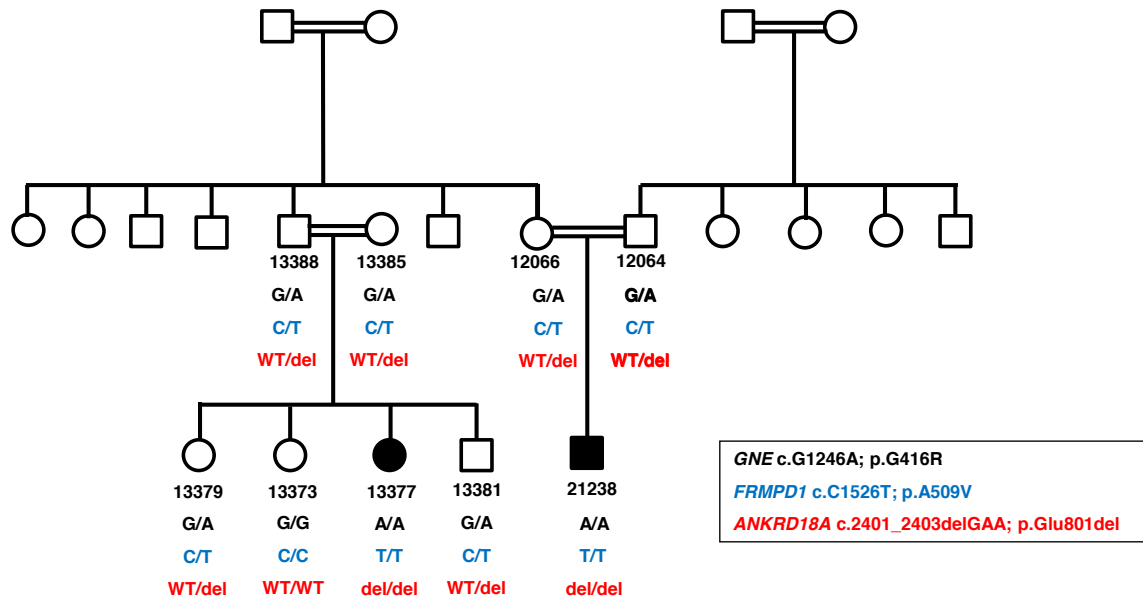


Figure 1.15 - Segregation analysis of the exome candidates in family members. The 3 variants (in the genes *GNE*, *FRMPD1* and *ANKRD18A*) were shared by both children and were located within a region of homozygosity on chromosome 9p13.3. Double lines linking parents signify first cousin unions. (analysis performed by N. Morgan)

Heterozygous variants (autosomal dominant) located in the epimerase domain of the *GNE* gene results in sialuria, which is an overexcretion of sialic acid in the urine. Homozygous and compound heterozygous (both alleles are mutated at different locations) variants located in either the epimerase or the kinase domain result in a group of neuromuscular disorders called hereditary inclusion body

myopathies (HIBM), which are characterized by adult-onset, slowly progressive, distal and proximal muscle weakness. Muscle biopsies from these patients demonstrate rimmed vacuoles and filamentous inclusions. Initial identification of the genetic aetiology of HIBM by Eisenberg et. al. (Eisenberg, Avidan et al. 2001) studied 47 Middle Eastern Jewish families and one family from the countries India, Georgia and the Bahamas. All the patients from the Jewish families had a homozygous M712T missense mutation in the kinase domain of the protein whereas the family from India had a V696M and C303X compound heterozygous missense mutation in the kinase and epimerase domain, respectively; the family from Georgia had a G576E and A631T compound heterozygous missense mutation, both in the kinase domain; and the family from the Bahamas had a R246Q and D225N compound heterozygous missense mutation, both in the epimerase domain. However, in this study no mention was made regarding the patients' hematopoietic profiles (Eisenberg, Avidan et al. 2001).

Since the initiation of this project, there have since appeared two separate reports of two sets of siblings, of nonconsanguineous parents, who developed myopathies during late adolescence. Interestingly, these patients also had thrombocytopenia since birth: platelet counts ranged from $1.1 - 40 \times 10^9/L$. Genetic analyses of one of the sets of siblings, who are Japanese, had a compound heterozygous V603L and G739 missense mutation, both in the kinase domain. For the other set of siblings, who are Chinese, genetic analyses demonstrated both siblings having a compound heterozygous T217H missense

and D515Q frameshift mutation, the former in the epimerase domain and the latter in the kinase domain (Izumi, Niihori et al. 2014, Zhen, Guo et al. 2014).

Variants in the *GNE* gene are suggested to have a role in causing thrombocytopenia. Sialic acid is known to play a role in platelet function: it is located in the platelet membrane and is involved in glycosylation. Several groups have shown that platelets, when lacking sialic acid, are removed from the circulation, resulting in shortened platelet lifespan and thrombocytopenia (Sorensen, Rumjantseva et al. 2009).

The homozygous non-synonymous variant in the *GNE* gene (p.G447R) of these two patients, II:5 and II:3, is located in the kinase domain. Despite these findings regarding the *GNE* gene, it is still possible that the genetic defect which is causing the thrombocytopenia in Patients II:5 and II:3 is due to the homozygous non-frameshift deletion of the amino acid glutamic acid 801 in the gene *ANKRD18A*. It is hypothesised that the mutation in *ANKRD18A* is the causal variant because neither patient suffered from sialuria nor any myopathies. *ANKRD18A* (Ankyrin Repeat Domain 18A) encodes a protein, which is expressed only in primates (Figure 1.17), and is weakly expressed in cells of hematopoietic lineage (Figure 6.1). Currently functionally uncharacterized, it encodes for a 992-residue, multi-domain protein of 116kD. Speculatively, *ANKRD18A* has three distinct regions of homology to proteins of known structure as determined by fold prediction using threading (Phyre2, HHpred). The amino

terminal region (residues 1-260 or 35-235, as determined by two different fold prediction algorithms) consists of 9 ankyrin-like repeats, followed by a coiled coil region that appears to resemble the structures of tropomyosin (residues 470-780) and at the carboxy terminus, the colicin IA (residues 529-942) homology domain (Figure 1.16). The homozygous non-frameshift deletion of the last of a trio of glutamic acid residues in *ANKRD18A* (p.E801del) is located in the colicin IA homology domain. The carboxy terminal domain of ANKRD18A has only a very limited homology to colicin IA.

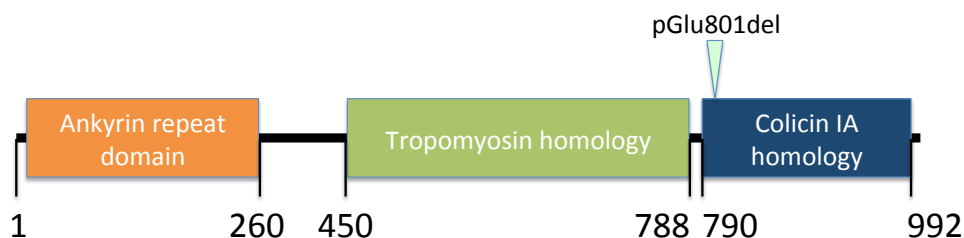


Figure 1.16 - Linear domain organisation of ANKRD18A derived from bioinformatical sequence analysis. Domains (or regions of homology) have been identified with the fold prediction server HHpred using the sequence of ANKRD18A (isoform 2). ANKRD18A isoform 1 has a 63-amino acid insertion, between the ankyrin repeat domain and the tropomyosin homology region. Threading suggests that, compared to isoform 2, isoform 1 (1054 residues) has a more extended ankyrin repeat domain, comprising 11 repeats instead of 9 in the shorter isoform 2 (992 residues). The approximate position of the in-frame deletion mutation (pGlu801del) is indicated.

Colicin IAs are *E. coli* bacteriocidal proteins formed by its plasmids. Colicins are transmembrane proteins that form channels to depolarize the bacterial cellular cytoplasmic membrane, resulting in the depletion of cellular energy and the killing of bacteria. Colicins are released by *E. coli* to reduce competition from other bacterial strains. Colicin IA has three domains: an amino-terminal translocation domains which functions to allow movement across the outer membrane and periplasmic space; the receptor-binding domain and the carboxy-terminal channel-forming domain which functions to form channels in the cytoplasmic domain (Gouaux 1997, Stroud, Reiling et al. 1998). The sequence identity over the homologous region between ANKRD18A and colicin IA is very low (9% identity) but the prediction has a high confidence level (98%). The homology extends over the coiled-coiled region of colicin IA, including the receptor binding region but excludes the channel-forming domain. In the fold prediction, the ANKRD18A Δ 801 deletion mutation maps to a small helix in the receptor-binding region of colicin IA. Because of the very low level of sequence identity between the two proteins, a prediction of what effect the analogous deletion would have on colicin IA is not possible. (Wiener, Freymann et al. 1997).

ANKRD18A is specific to primates (Figure 1.17) and is a member of the *POTE* (expressed in Prostate, Ovary, Testis and placenta) family of genes, of which *ANKRD26* is the ancestral gene (Figure 1.17). Causal variants located at multiple sites in the 5'-UTR of *ANKRD26* have been shown to cause an autosomal dominant form of thrombocytopenia (ANKRD-RT or THC2), which

have been identified in multiple families in Italy (Noris, Perrotta et al. 2011). ANKRD18A and ANRKD26 share 44% identity at the amino acid sequence level. There is yet another POTE protein named ANKRD18B (also located on chromosome 9), which is 93%-identical to ANKRD18A at the amino acid sequence level (Figure S3). The three glutamic acid residues, which are in ANKRD18A (residues 797-799) are also present in ANKRD18B (residues 870-873). The MLEEEVL sequence motif, which encompasses the three consecutive Glu residues ($\Delta 801$, third Glu residue mutated in patient ANKRD18A), is conserved in *Pan troglodytes* (chimpanzee) and partially conserved in *Nomascus leucogenys* (gibbon; third Glu residue missing) (Figure S4). It is absent in gorillas, because the gorilla orthologue of ANKRD18A (760 amino acids) is truncated at the C-terminus relative to the human sequence (Figure S 4).

ANKRD18A is specific to primates and is a member of the *POTE* gene family, which encompasses 13 known paralogs (Hahn, Bera et al. 2006). Several *POTE* paralogs have been implicated in apoptosis and cytoskeletal regulation. For instance, the *POTE* paralogs, *POTE2 α -actin* and *POTE2 γ -C*, which have an actin retrogene, were shown to induce apoptosis when transiently transfected into HeLA cells (Liu, Bera et al. 2009). POTE proteins RAI14 and UACA share the linear domain organisation of ANKRD26 and other POTE proteins. RAI14 is ubiquitously expressed and has thread-like projections in the cytoplasm that are POTE actin fusion proteins. The mouse orthologue of RAI14 is associated with cortical actin cytoskeletal structures in the terminal web and is also seen in

intestinal epithelial cell-cell adhesion sites. UACA is more widely expressed in skeletal muscles. The mouse orthologue of UACA localises to the cell cytoplasm and reportedly mediates apoptosis (Hahn, Bera et al. 2006). Based on these findings and invoking homology between POTE proteins, it is conceivable that ANKRD18A may too be involved in cytoskeletal regulation.

Tissue and cellular expression information for *ANKRD18A* mRNA is based on data obtained from the Genotype-Tissue Expression (GTEx) Project and the BLUEPRINT Epigenome Project. GTEx is a resource database and associated tissue bank launched by the NIH to facilitate the study of the relationship between genetic variation and gene expression in tissues (Consortium 2013) (www.genome.gov). Based on the database from GTEx, which RNA sequenced 53 human tissue samples, *ANKRD18A* mRNA expression was noted to be highest in testis (5 FPKM) followed by spinal cord (2 FPKM), thyroid (1 FPKM), pituitary gland (0.8 FPKM), pancreas (0.7 FPKM), prostate gland (0.5 FPKM) and various regions of the brain with expression levels ranging from 0.5 to 1.0 FPKM. The cut-off level for detection was 0.5 FPKM. (FPKM is the fragments per kilobase of transcript per million mapped reads. It is the relative expression of transcript in proportion to the number of cDNA fragments that originate from it.) Cellular expression of *ANKRD18A* mRNA is from the Blueprint haematopoietic database.

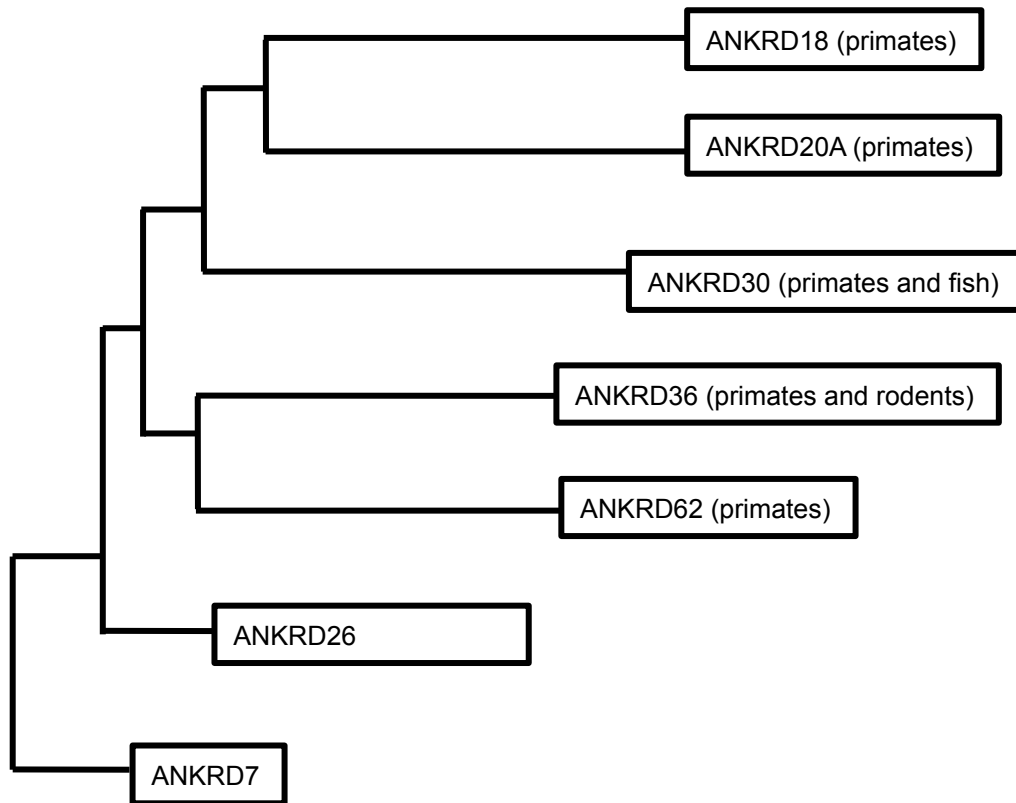


Figure 1.17 - Dendrogram of the ANKRD gene family. ANKRD18A is only expressed in primates whereas ANKRD26 expression is not limited to primates.

The BLUEPRINT epigenome project is a collaborative endeavour funded by the EU to study how genes are activated or repressed in healthy and diseased haematopoietic cells (www.blueprint-epigenome.eu). Based on the database from the Blueprint project, *ANKRD18A* mRNA expression in CD34⁺CD41⁺CD42⁺ MK cells was below the 0.5 FPKM cut-off. There was however, higher *ANKRD18A* mRNA expression in other hematopoietic cells: neutrophilic metamyelocyte (7 FPKM), mature eosinophil (5 FPKM), segmented neutrophil of bone marrow (2 FPKM), erythroblast (1 FPKM), conventional dendritic cell (0.8 FPKM), and various $\alpha\beta$ T and B cells (0.5-1 FPKM). Contrary to the Blueprint

data, the PlateletWeb (plateletweb.bioapps.biozentrum.uni-wuerzburg.de) indicates that there is protein expression of ANKRD18A in platelets (unquantified and source of information unclear). However, PlateletWeb does not indicate that ANKRD18B is expressed in platelets.

The speculative domain structure of ANKRD18A gives few clues as to the physiological role of this protein. In order to characterise the function of ANKRD18A, a molecular probe specific for this protein is a key requirement. Previously, this project used polyclonal antibodies directed against peptides derived from amino acid residues 53-66 and 260-289 of the ANKRD18A sequence. However, variable results in Western blotting experiments raised doubts as to the reliability and specificity of these antibodies.

1.7. Aims of this project

The overall aim of this project is to characterise ANKRD18A. Only then will it be possible to determine whether or not the genetic aetiology of Patient II:5 and II:3's severe thrombocytopenia is due to the mutation in *ANKRD18A*. An essential reagent for studying proteins is a specific antibody. Because of variable results obtained when using ANKRD18A peptide antibodies that were already present in the laboratory, as well as the fact that there exists another protein, ANKRD18B, which has a 93% amino acid sequence identity with ANKRD18A, the first objective of this project was to generate an antibody which would be specific to ANKRD18A. This specificity entailed an antibody which recognises the tertiary structure of ANKRD18A (or functional domains thereof). This antibody will permit subsequent *in vitro* experiments such as imaging studies for cellular localization and protein expression; flow cytometry assays to check for protein expression; and Westerns and immunoprecipitation experiments to confirm protein expression.

2. Materials and methods

2.1. Cell Culture

Human endothelial kidney (HEK) 293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (P/S), 1% glutamine. Dami (human megakaryoblastic) cells were grown in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% FBS, 1% P/S and 1% glutamine (From Watson Lab protocol manual).

2.2. Transient Transfections

2.2.1. HEK 293T cells

HEK293T cells were transfected with plasmid DNA using transfectant reagents transfectamine, lipofectamine or PEI. Protocols for commercial reagents transfectamine and lipofectamine were according to kit protocols. HEK293 cells transfected using PEI followed the protocol in Table 2.1. Cells were plated at the aforementioned densities 24 hours prior to transfection in media consisting of DMEM + 10% FCS + 1% glutamine but no P/S. A day later, media was changed to DMEM + 2% FCS + 1% glutamine and no P/S. A transfection cocktail cDNA + PEI + Opti-MEM was made, vortexed and incubated at room temperature for 10 minutes before adding dropwise onto cells. 72 hours later, cells were harvested

by discarding media and placed in -80 °C for more than 1 hour to assist in cell lysis (Protocol from Ben Willcox's group).

Plate size	Cells plated	Media volume	Opti-mem	DNA	PEI
24-well	1x10 ⁵	500 µl	25 µl	400 ng	2.5 µl
6-well	5x10 ⁵	2 ml	125 µl	2 µg	12.5 µl
15 cm	5x10 ⁶	20 ml	2.25 ml	36 µg	225 µl

Table 2.1 - Protocol for HEK293T transient transfections using PEI.

2.2.2. Dami cells

Dami cells were transiently transfected using only PEI. PEI was used because it is cheaper than the other proprietary transfection kits. Transfections were performed using the same method as HEK293T cells except that the media used for plating the Dami cells consisted of RPMI + 10% FCS + 1% glutamine and no P/S. Also, added to the media to differentiate Dami cells when plating the cells were 28.5 ng/ml hTPO + 6.2 ng/ml PMA per 10 ml of media (Protocol from S. Fletcher). TPO and PMA are added on the day of plating Dami cells (24 hours before transfection).

Plate size	Cells plated	Media volume	Opti-mem	DNA	PEI
24-well	8×10^4	500 μ l	30 μ l	500 ng	1.5 μ l
12-well	2×10^5	1 ml	60 μ l	1 μ g	3 μ l
6-well	4×10^5	2 ml	150 μ l	2.5 μ g	7.5 μ l

Table 2.2 - Protocol for Dami cells transient transfections using PEI.

2.3. Preparation of cellular lysates

2.3.1. Washed Platelets

Platelets were first obtained from whole blood by venipuncture from healthy, drug free volunteers with a syringe containing 10% sodium citrate. The blood was further anticoagulated by adding 10% acid citrate dextrose (ACD: 120 mM sodium citrate, 110 mM glucose, 80 mM citric acid). The blood was spun at 200g for 20 minutes to separate the blood into its cellular fractions. Platelets (platelet rich plasma, PRP), located in the top fraction, were retained, prostacyclin (10 μ g/ml) was added to inhibit platelet activation, and the PRP was again spun at 1000g for 10 minutes. The plasma supernatant was discarded and the platelet pellet washed with 25 ml modified Tyrodes-HEPES buffer (134 mM NaCl, 0.34 mM Na_2HPO_4 , 2.9 mM KCl, 12 mM NaHCO_3 , 20 mM HEPES, 5 mM glucose, 1 mM MgCl_2 , pH 7.3) and 3 ml ACD. Prostacyclin was added again for the final spin at 1000g x 10minutes. The platelet pellet was resuspended in Tyrodes-HEPES buffer at a concentration of $5\text{-}10 \times 10^8$ /ml for Westerns and IPs and

2×10^7 /ml for spreading experiments. Platelets rested for 30 minutes before being used for further experiments (From Watson lab protocol manual).

2.3.2. Platelet lysates

Platelet lysates were prepared by adding the same volume of 2X Lysis buffer (300 mM NaCl, 20 mM Tris Base, 2 mM EGTA, 2 mM EDTA, 2% NP-40) and protein inhibitors (0.5 mM Leupeptin, 0.5 mM AEBSF, 0.05 mM Apoptinin, and 0.02 mM pepstatin), as the volume of washed platelets and placed on ice for 10 minutes. Once lysis was completed, the tube of lysed platelets was centrifuged at 10,000g for 10 minutes at 4 °C (From Watson Lab protocol manual).

2.3.3. Lysates of HEK293T (adherent cells) and Dami (suspended) cells

Lysates of HEK293T and Dami cells were prepared by using RIPA buffer (50 mM Tris-HCl, pH 8, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 5 mM EDTA, 1% EDTA). Adherent cells at ~80% confluency were lysed with 700 μ l RIPA buffer plus inhibitors (as mentioned above) for a 10 cm dish, on ice. Suspension cultures at a density of $1 - 2 \times 10^6$ cells/ml were lysed with 100 μ l of RIPA buffer plus inhibitors, on ice (From Cold Spring Harbor Protocols).

2.4. Construct design utilizing high-throughput cloning and expression at the Oxford Protein Production Facility (OPPF)

2.4.1. ANKRD18A Functional Domains Used

Constructs listed in Table 2.3 were made using various forms (because of different fusion tags) of the expression vector pOPINF. Inserts were the different functional domains of ANKRD18A:

1. ankyrin repeat domain (amino acids 1-260, as determined by HHpred)
2. ankyrin repeat domain (amino acids 35-235, as determined by Phyre2)
3. tropomyosin homology domain (amino acids 450-790)
4. colicin IA homology domain (amino acids 790-992)

Cloning was done with a ligase-free technology called InFusion (Clontech). Inserts were generated by PCR so cloning was independent of restriction sites.

2.4.2. Vectors used

1. pOPINE- POI-C-HIS₆ (POI=protein of interest, C-HIS₆=C terminal His tag)
2. pOPINF- N-HIS₆-3C-POI (N-HIS₆=N terminal His tag, 3C= 3C protease site)
3. pOPINFS- N-HIS₆-3C-POI-C-STREP^{II} (STREP^{II}=streptavidin binding peptidell)
4. pOPINJ- N-HIS₆-GST-3C-POI
5. pOPINM- N-HIS₆-MBP-3C-POI (MBP=maltose binding protein)
6. pOPINS3C- N-HIS₆-SUMO-3C-POI (SUMO=small ubiquitin-like modifier)
7. pOPIN-3C-HALO7- POI-3C-HALO-C-HIS₆ (HALO= Halotag)

2.4.3. Cloning

Inserts were generated by PCR with primers designed for In-Fusion cloning by OPF. Primers had melting temperatures $T_m > 68$ °C on target gene. Template was purified plasmid DNA encoding GFP-tagged wild-type ANKRD18A. Once the PCR reaction cycle was completed, PCR products were purified using the AMPureXP Magnetic Bead Purification kit (AgenCourt/Beckman). Prior to purification, PCR reaction was treated with DpnI restriction enzyme (specific for methylated DNA) to digest the template but not the PCR product. This step was required for cases where the template had the same antibiotic resistance as the target pOPIN vector. Once purification was completed, the PCR product was eluted from magnetic beads with elution buffer (10 mM Tris-HCl, pH 8.0).

Ligation reaction was performed using 1 µl of linearized pOPIN vector with 10 - 250 ng of DNA encoding the purified insert, with added appropriate volume of sterile water for total volume of 10 µl. This was added to a dry-down In-Fusion plate (Clontech). The ligation reaction was allowed to proceed at 42 °C for 30 minutes. Following that, the reaction mixture was immediately placed on ice and diluted with 40 µl of TE buffer. For transformation, 5 µl of diluted mixture was immediately added to OmniMaxII competent cells, and incubated on ice for 30 minutes, followed by heat shock at 42 °C for 30 seconds, then back on ice for 2 minutes, then 300 µl of Power broth/SOC was added and the mixture was incubated at 37 °C for 1 hour. LB agar plates were prepared by mixing appropriate antibiotic (carbenicillin 50mg/ml, 1:1000)+ 20% X-gal (in DMF) + IPTG (500mM, 1:500). After an hour, competent cells were plated onto agar

plates and then incubated overnight at 37 °C. Next day, white colonies were picked. Blue colonies are secondary to inefficiently linearized parental plasmid. Picked colonies were placed in power broth supplemented with appropriate antibiotics. Cultures were grown overnight in shaker. Plasmids are then mini-prepped using a Bio-Robot 8000 to obtain cDNA.

2.4.4. HEK293T transient transfections

Transfection utilised the transfectant reagent GeneJuice (EMDMillipore), which was composed of a nontoxic cellular protein and a small amount of a proprietary polyamine. The transfection procedures followed manufacturers' recommendations.

2.4.5. *E. coli* transformations and protein expression

A volume of 3 µl of purified plasmid was added to competent cell *E. coli* strains B834 or Rosetta (DES) LysS aliquots in 96 well plates and the transformation was done as described above. Agar plates were prepared as mentioned above and supplemented with antibiotics at a dilution of 1:1000 of the standardised stocks. In order to avoid loss of the pRARELysS plasmid in Rosetta, chloramphenicol was added as well. Typically, five transformation reactions were added onto agar plates and incubated overnight at 37 °C. On the following day, single colonies were picked and used to inoculate growth cultures (power broth) supplemented with appropriate antibiotics and incubated overnight at 37 °C while shaking (200 rpm). The overnight cultures were diluted by inoculating 3 ml of power broth with either 150 µl of B834 or 250 µl of Rosetta

cultures, incubated in 37 °C under shaking for 3 - 5 hours. Subsequently, cultures were cooled (under shaking) in 20 °C shaker for 20 min, and then induced by adding IPTG to a final concentration of 1 mM per well, following by incubation overnight at 20 °C (under shaking).

Clone	OPTIC	Gene name	aa_N	aa_C	OPPF No.	Vector	Tag	MW of Tag (kDa)	Domains	Domain MW
A1	13915	ANKRD-FL	2	992	14229	pOPINE	(His) ₆	0.9	Full length	116 kD
B1	13915	ANKRD-FL	2	992	14230	pOPINF	3CProt+His ₆	2.2		
C1	13915	ANKRD-FL	2	992	14231	S3C	Sumo-3C	13.2		
D1	13915	ANKRD-FL	2	992	14232	M	MBP	43		
E1	13915	ANKRD-FL	2	992	14233	E-3C-HALO	E3CHalo	35.3		
F1	13915	ANKRD-35	35	235	14234	pOPINE	(His) ₆	0.9	ANK repeat	23 kD
G1	13915	ANKRD-35	35	235	14235	pOPINF	3CProt+His ₆	2.2		
H1	13915	ANKRD-35	35	235	14236	S3C	Sumo-3C	13.2		
A2	13915	ANKRD-35	35	235	14237	M	MBP	43		
B2	13915	ANKRD-35	35	235	14238	E-3C-HALO	E3CHalo	35.3		
C2	13915	ANKRD-1	2	260	14239	pOPINE	(His) ₆	0.9	ANK repeat	30 kD
D2	13915	ANKRD-1	2	260	14240	pOPINF	3CProt+His ₆	2.2		
E2	13915	ANKRD-1	2	260	14241	S3C	Sumo-3C	13.2		
F2	13915	ANKRD-1	2	260	14242	M	MBP	43		
G2	13915	ANKRD-1	2	260	14243	E-3C-HALO	E3CHalo	35.3		
H2	13915	ANKRD-450	450	790	14249	pOPINJ	GST	30	Tropomyosin	40 kD
A3	13915	ANKRD-450	450	790	14250	pOPINFs	Strep II	1.1		
B3	13915	ANKRD-450	450	790	14246	S3C	Sumo-3C	13.2		
C3	13915	ANKRD-450	450	790	14247	M	MBP	43		
D3	13915	ANKRD-450	450	790	14248	E-3C-HALO	E3CHalo	35.3		
E3	13915	ANKRD-450	450	790	14244	E	(His) ₆	0.9		
F3	13915	ANKRD-790	790	992	14251	pOPINJ	GST	30	Colicin 1A	24 kD
G3	13915	ANKRD-790	790	992	14252	pOPINFs	Strep II	1.1		
H3	13915	ANKRD-790	790	992	14253	pOPINS3C	Sumo-3C	13.2		
A4	13915	ANKRD-790	790	992	14254	pOPINM	MBP	43		
B4	13915	ANKRD-790	790	992	14255	E	(His) ₆	0.9		
C4	13915	ANKRD-790	790	992	14256	E-3C-HALO	E3CHalo	35.3		
D4	13915	ANKRD-	450	992	14257	pOPINJ	GST	30	Tropomyosin +	64 kD
E4	13915	ANKRD-	450	992	14258	pOPINFs	Strep II	1.1		
F4	13915	ANKRD-	450	992	14259	pOPINM	MBP	43		
G4	13915	ANKRD-	450	992	14260	pOPINS3C	Sumo-3C	13.2		
H4	13915	ANKRD-	450	992	14261	E-3C-HALO	E3CHalo	35.3		

Table 2.3 - DNA plasmid constructs generated at OPPF. Abbreviations used: 3C = cleavage site for picornavirus endopeptidase 3C; SUMO = small ubiquitin-like modifier; MBP = maltose binding protein; GST = glutathione-S-transferase; HALO = Promega HaloTag; Phyre2 = protein fold prediction server Phyre2 (www.sbg.bio.ic.ac.uk/phyre2); HHpred = homology detection and structure prediction by HMM-HMM comparison (toolkit.tuebingen.mpg.de/hhpred).

Next day, 1 ml of induced culture was centrifuged at $6000 \times g$ for 10 min. The supernatant was discarded and cell pellets were placed in a $-80\text{ }^{\circ}\text{C}$ freezer for at least 20 min to aid cell lysis. Frozen cells were resuspended in lysis buffer (NPI-Tween), supplemented with 1 mg/ml lysozyme and 400 units/ml DNAase type I, then stored on ice for 30 min., centrifuged at $5000 \times g$ for 30 min. at $4\text{ }^{\circ}\text{C}$. The lysate was kept and transferred to a 96 well plate for purification of proteins using Qiagen Magnetic Ni-NTA beads with the Qiagen Bio-Robot 8000.

2.5. Plasmid DNA Preparation

Validation of DNA construct correction was done by Sanger sequencing of all constructs that expressed protein when transfected into HEK293T cells: clones A4, B3, C3, E2, E3, F2, G2, H1, H3. Sequencing reaction included 200 ng of plasmid DNA, 3.2 pmol of primer plus molecular grade water for a total volume of 10 μl . Primers (Sigma) used for sequencing were the T7 primer (5' end): TAA TAC GAC TCA CTA TAG GG and the TriEX Down Primer (3' end): CAC AAA TAC CAC TGA GAT CGA and ANKRD18A_JFR (starting at bp 493): CTT ATT ATA CAC GGC ATA ATG (Table 2.6). Reaction mix was then sent to the Birmingham Genomics Facility for sequencing (sequencing mixture protocol per Birminham Genomics Facility). Plasmid DNA used for sequencing was prepared following the protocol from DNA Plasmid miniprep kit (Sigma PLN350). Following confirmation of the validity of construct sequence, the highest expressing clones were chosen for protein purification: clones F2 and G2. Large amounts of plasmid DNA were required for large scale transfection of HEK293T

cells in 15 cm plates. Plasmid DNA was prepped following the protocol from a MaxiPrep kit (Sigma NA0310).

2.6. SDS-PAGE (sodium dodecyl sulphate polyacrylamide gels) and Western Blotting

Denatured proteins were separated by electrophoresis using 8%, 9%, or 10% SDS-PAGE gels. Running buffer consisted of 25mM Tris Base, 190mM glycine 0.1% SDS, pH 8.3. Prestained molecular weight markers (New England Biolabs P7712) were run alongside samples. Using the wet transfer method, separated proteins in the gels were electro-transferred onto PVDF (polyvinylidene difluoride) membranes (Biorad 162-0177) at 20V overnight or 100V for 1 hour. Transfer buffer consisted of 25 mM Tris Base, 190 mM glycine, and 20% methanol. Following completion of transfer, membranes were blocked for 1 hour at room temperature in blocking buffer: 5% whole milk in TBS-T (Tris buffered saline with Tween-20): 200 mM Tris, pH 7.6, 1.37 M NaCl, 0.1% Tween 20. Following block, primary antibody was placed in block buffer and incubated at 4 °C overnight. The following day membranes were washed 5X, 10 minutes each in TBS-T with increased salt concentration (200 mM Tris, 5 M NaCl, 0.1% Tween 20) (TBS-T protocol per M. Tomlinson). Then membranes were placed in HRP (horseradish peroxidase) conjugated secondary antibodies (GE Healthcare NA931V) in TBS-T for 1 hour at room temperature. After incubation, membranes were again washed 5X with high salt TBS-T. Once washes were completed, the membranes were incubated with enhanced chemiluminescence (ECL,

Amersham Biosciences, RPN 2232) reagent, 1 ml from each bottle, for 1 minute at room temperature, then developed by exposure to autoradiographs. Primary antibodies used for Western blotting: rabbit polyclonal α -ANKRD18A (Abgent, AP18045a), rabbit polyclonal ANKRD18A: α -21233 and α -21234 (Biogenes), and rabbit polyclonal ANKRD18A: α -25277 (Biogenes) and rabbit polyclonal α -GAPDH (Abcam, ab9485) (Protocols for Westerns from Watson lab protocol manual, transfer protocol from N. Morgan).

2.6.1. Immunoprecipitation (IP)

Cellular lysates, prepared as described above, were pre-cleared with 25 μ l protein A sepharose (PAS) beads, rocking for 1 hour at 4 °C. Insoluble debris and beads then were pelleted by centrifugation at 10,000g at 4 °C. The cleared lysate was transferred to a new tube, 1 μ g of antibody was added and the eppendorf tube was rocked again for 15-60 minutes at 4 °C. Another 25 μ l of PAS was added and the mixture again rocked at 4 °C, overnight. Next day, PAS beads were pelleted at 9,000 rpm, 20 seconds, 4 °C. The beads were then washed with four changes of 1X lysis buffer without inhibitors. Proteins were eluted from the beads by adding 35-50ul 2X Laemmli sample buffer, boiled for 5 minutes at 100 °C. Antibody used for IP was the rabbit polyclonal α -ANKRD18A, α -25277 (Protocol from Y. Senis).

2.7. Protein Purification

2.7.1. Preparation of cleared lysate for Ni²⁺-NTA column chromatography

Cells from the -80 °C freezer were thawed on ice for 15 minutes or more then lysed on ice with 10 ml of lysis buffer containing 20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, 1% NP-40, DNase I, protease inhibitors, pH 7.4. Protease inhibitors (Sigma S8830), used at 1:10, were made by dissolving a tablet containing various inhibitors (EDTA free) into 10 ml of H₂O. DNase I (Sigma DN-25), used at 1:500, was prepared by dissolving 100 mg in 1 ml of H₂O. Following lysis, cells were scraped off the plate and the lysate was placed into ultracentrifuge tubes (Beckman 50 ml polyallomer tubes with caps) that fit the JA25.50 rotor. The lysate was spun for 45 min. at 70,000g using the Beckman Coulter J2 ultracentrifuge. The supernatant was then cleared through a 0.45 µm filter before it was loaded onto the Ni²⁺-NTA column.

2.7.2. Preparation of Ni²⁺-NTA Column

Gibson pump tubing was filled with distilled H₂O. The 1 ml Ni²⁺-NTA column (GE Healthcare 17-5247-01) was clamped onto a stand and the top stopper removed. The tubing was connected to the column by allowing the distilled H₂O to fill the column 'drop-to-drop' to avoid introducing air into the system and then the connector with the tubing was screwed onto the column and the snap-off end of the column removed. The column was then washed with 5 column volumes of distilled H₂O, then equilibrated with 10 column volumes of Binding Buffer (20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4). Elution Buffer

consisted of 20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4. Columns were run at a rate of around 1 ml/min using the Gibson pump. Lysate were loaded onto the column twice to maximize binding to the column. Once loading was completed, the column was washed with 10 column volumes of the Binding Buffer. Both clones F2 (MBP-Ank repeat domain) and G2 (E3C Halo-Ank repeat domain) were eluted using a buffer with increasing concentrations of imidazole. Elutions of 10 column volumes were used per step. Concentrations of imidazole used per step were 60 mM, 80 mM, 100 mM, 150 mM, 200 mM, 300 mM and 500 mM (Table 2). To make the buffers with these concentrations of imidazole, one mixed the Binding Buffer with the Elution Buffer at various ratios using the formula: $(10 \text{ mM})x + (500 \text{ mM})y = 60 \text{ mM}(x + y)$, x = Binding buffer volume, y = Elution buffer volume

Final imidazole concentration	Binding Buffer volume (ml)	Elution Buffer volume (ml)
60 mM	4.5	0.5
80 mM	4.4	0.6
100 mM	4.1	0.9
150 mM	3.35	1.65
200 mM	3.05	1.95
300 mM	2	3
500 mM	0	5

Table 2.4 - Volumes of Binding and Elution Buffers that are mixed to arrive at the different concentrations of imidazole needed to elute protein from the Ni^{+2} -NTA column.

Each different concentration of imidazole elution was collected in a different 15 ml conical tube. Once collection was completed, 40 μl was taken from each tube to be run on a Coomassie gel to determine at which fraction is the elution

cleanest. 40 μ l of elution was mixed with 10 μ l of 5x sample buffer, boiled for 5 min and run on a 10% acrylamide gel. Clone F2 eluted with the highest purity after the imidazole concentration of 100 mM. Elutions from using imidazole concentrations of 100 mM and greater (100 mM, 150 mM, 200 mM, 300 mM, and 500 mM imidazole) were combined and dialyzed thrice-overnight and then next day twice for 2 hours or greater, each in dialysis buffer consisting of 20 mM Tris, 150 mM NaCl, pH 7.8. Clone G2 eluted with imidazole concentrations of 300 mM and greater so the following fractions were pooled for dialysis: 300 mM and 500 mM imidazole. Dialysis also consisted of 3 exchanges lasting two hours or greater, each time. Dialysis buffer consisted of 20 mM Tris, 200 mM NaCl, pH 7.8. Dialysis tubing was boiled for 5 min in 10 mM EDTA and stored in 20% ethanol prior to use (Protocol for preparation and use of Ni²⁺-NTA column and buffers from GE Healthcare manual, and as described in Wang et. Al. (Wang, Sambandan et al. 2013).

2.7.3. Preparation of Anion Exchange Column

Following completion of dialysis, dialysate was loaded onto a 1 ml anion exchange column (HiTrap Q HP, GE Healthcare 17-1151-01), at a rate of 1 ml/min using the Gibson pump. The anion exchange column was prepared by first removing the top stopper and loading the column with the Start Buffer by first adding it dropwise so not to introduce air into the column. Once loaded, the bottom stopper was removed and the column was washed with 10 column volumes of Start Buffer: 20 mM Tris, 50 mM NaCl, pH 7.8 to remove the preservatives in the column. The column was then washed with 10 column

volumes of Regeneration Buffer (20 mM Tris, 1 M NaCl, pH 7.8) and then equilibrated with 10 column volumes of Start Buffer. Once the above column preparation was completed, the dialysate was loaded onto the column once followed by washing of the column once with Start Buffer and then elution commenced with stepwise ionic strength gradients (increasing concentrations of NaCl). Start Buffer: 20 mM Tris, 50 mM NaCl, pH 7.8, Elution Buffer: 20 mM Tris, 500 mM NaCl, pH 7.8. Formula used to determine the increasing NaCl concentrations: $x + y = 10$, $50x + 500y = 175(x+y)$. x = Start Buffer volume, y = Elution Buffer volume (Protocol for column preparation and buffers from GE Healthcare).

Final NaCl concentration	Start Buffer volume (ml)	Elution Buffer volume (ml)
175mM	7.222	2.778
200mM	6.667	3.333
250mM	5.556	4.444
300mM	4.444	5.556
350mM	3.333	6.667
400mM	2.222	7.778
500mM	0	10

Table 2.5 - Volumes of Start and Elution Buffers mixed to arrive at the concentrations of NaCl needed to elute protein from the anion exchange column.

Each elution fraction was collected in 15 ml conical tubes. From each fraction, 40 μ l was taken and mixed with 5x sample buffer, boiled, run on a SDS-PAGE, then stained with Coomassie stain to determine at which NaCl concentration is the protein eluted. For Clone F2, fractions eluted with 250-500 mM NaCl, inclusive, were pooled, then dialyzed in 20 mM Tris, 200 mM NaCl, pH 7.8 thrice,

2 hours or greater, each. Following dialysis, protein was concentrated with Amicon Ultra-15 centrifugal filters (Millipore UFC903024). For Clone G2, fractions eluted with 350-500 mM NaCl, inclusive, were pooled, then dialyzed thrice in 20 mM sodium phosphate, 500 mM NaCl, pH 7.5. Each dialysis lasted 2 hours or more, each time. Afterwards the dialysate is concentrated with the Amicon Ultra-15 filters (Protocol for Amicon use from Millipore instructions, protocols for dialysis buffer recipes from Biogene recommendation).

2.7.4. Concentration of eluted protein

Following completion of dialysis, the dialysate concentration was measured using a NanoDrop spectrophotometer. Determining protein concentration using the NanoDrop required the molecular weight (MW) of the protein and its coefficient of extinction. For Clone F2, its MW is 72.126 kilodaltons (g/mol) and extinction coefficient is $81165 \text{ M}^{-1}\text{cm}^{-1}$. For clone G2, its MW is 64.842 kD and the extinction coefficient is $73,380 \text{ M}^{-1}\text{cm}^{-1}$. The protein was concentrated using Amicon Ultra-15 filters. The filters were first pre-wetted with ddH₂O, spinning at 2300 rpm for 15 min. Dialysate was then placed into the column and spun cautiously so as not to dry out the filter. First spin was at 2500g for 10 min and repeated until volume decreased to about 2 ml. Speed was then decreased to 1000g for 10 min. When the volume reached 1 ml, the concentration was checked on the NanoDrop. This was done 3 times and the values averaged to arrive at the concentration value.

2.7.5. Passivation of Amicon ultra-filtration membranes

Membranes were overnight in 2 ml of 5% PEG 6000 (Stock, 20%). Next day, p PEG 6000 was poured out and column rinsed with sterilised double distilled water. 2 ml of sterilised double distilled water was placed into column again and spun at 2300 rpm for ~20 minutes so that only 500 µl of water remained. Previous step was repeated. Remaining 500 µl of water was removed and replaced with 2 ml of sterilised distilled water inside filter (column) and 5 ml outside filter (Protocol from K. Fütterer).

2.8. Coomassie stain

8-10% SDS-PAGE was stained with Coomassie stain to determine at which imidazole or NaCl fractions the protein was eluted. The gel was stained for at least 1 hour, then destained until bands are discernable and background blue stain is light blue. Following destain, the gel needs to be placed in a gel drying solution for at least 1 hour to delay cracking of the gel. Coomassie stain: 0.1% Coomassie R250, 10% glacial acetic acid, 40% methanol. Coomassie destain: 20% methanol, 10% glacial acetic acid. Gel drying solution: 10% methanol, 5% glycerol (Protocol from K. Fütterer).

2.9. Lentiviral production

On day prior to transfection, 3×10^6 HEK293T cells were plated onto 10 cm dishes in 10 ml of DMEM with 10% FCS plus glutamine. Next day, media was removed

and replaced with 3 ml of DMEM with 2% FCS + glutamine. 1 ml of Opti-mem plus 9 µg of DNA (4.39 µg of target gene + 3.29 µg of packaging vector, PAX2, +1.32 µg of envelope vector, PMD2G) plus 36 µg of PEI were pipetted into an eppendorf tube. Tube was vortexed then mixture sat at room temperature for 10 minutes then added dish drop-wise to dish followed by north-south and east-west movements of plate to mix its contents. Plates were put in 37°C incubator for 48 hours then the supernatant was harvested for lentivirus (Protocol from R. Bicknell Lab).

2.9.1. Concentrating Lentivirus

All the supernatant from the above lentiviral transfection was placed into 50 ml conical tubes. Cellular debris was removed from the supernatant by centrifuging at 300g for 5 minutes. The supernatant was placed into concentrators (Corning Spin X UF, 20 ml, Sigma 431487) once concentrators had been rinsed with 10 ml of sterile distilled water by centrifuging at 4000g for 10 minutes. The supernatant was centrifuged for 2 hours at 4000g. Once completed, what remained at the top of the concentrator (~2 ml) was kept and 8 µg/ml of polybrene (Sigma 107689) was added. The mixture was filtered through a 0.45 µm filter (Protocol from R. Bicknell Lab).

2.10. Isolating cDNA for RT-PCR and quantitative real time PCR

Total RNA was isolated and purified using the Machery-Nagel kit (cat. no. 740955.10). 10 µl was taken from RNA prep and mixed with 10 µl of premix from

the ABI RT kit (cat.no. 4368814) which converted total RNA to single-strand cDNA by using the following program on the PCR thermocycler: 25 °C for 10 min, 37 °C for 2 hours, 85 °C for 5 min, 4 °C forever. The RT-PCR program which used to determine mRNA expression was: Denaturation 94 °C for 3 min, Denaturation 94 °C for 1 min, Annealing 60 °C for 1 min, Extension 72 °C for 1 min, repeat 39 times, Extension 72 °C for 5 min, 4 °C forever. RT-PCR reaction mixture consisted of 12.5 µl Sigma Red Taq polymerase, 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 10.5 µl H₂O, and 1 µl cDNA (Protocol from N. Morgan).

Primer name	Primer sequence (5'-3')
ANKRD18A_CF (Forward)	GCCCGGTATTCGCAACAGCTT
ANKRD18A_BR (Reverse)	CGATGCTGTGCTTGCTTAGG
ANKRD18B_CF (Forward)	CTCAGTTTTGGGAGACGCCTG
ANKRD18B_CR (Reverse)	GCGCCACGTTTCAGGAGAATA
β-actin (Forward)	GCGGGAAATCGTGCGTGACATT
β-actin (Reverse)	GATGGAGTTGAAGGTAGTTTCGT
GAPDH (Forward)	GAAGGTGAAGGTCGGAGT
GAPDH (Reverse)	AAGATGGTGATGGGATTT
T7 (Forward)	TAATACGACTCACTATAGGG
TriEX Down (Reverse)	CACAAATACCACTGAGATCGA
ANKRD18A_JFR (Reverse)	CTTATTATACACGGCATAATG

Table 2.6 - Primer sequences used in RT-PCR and Sanger sequencing

2.10.1. Preparing quantitative real-time reaction plates

Reaction mixture consisted of SYBR green 12.4 µl, Forward Primer 0.1 µl (10 µM), Reverse Primer 0.1 µl (10 µM), H₂O 9.8 µl, cDNA 2.5 µl. Every reaction was done in triplicate. NTC is PCR H₂O (negative control). Control neat was the

positive control. In this case, Dami and the retinoblastoma cell lines were used. These dilutions of neat were used: the first set is 1:5, 2nd set is 1:25, 3rd set is 1:50. Sequences of ANKRD18A and 18B primers are as written in Table 2.6. Quantitation of *ANKRD18A* and *18B* mRNA expression were compared against two internal standards, β -*actin* and *GAPDH*. Sequence of primers for β -*actin* and *GAPDH* are in Table 2.6. The instrument used was Applied Biosystems real-time PCR 7500 Fast and the thermal cycling conditions were UP activation 95 °C 20 sec HOLD, Denature 95 °C 3 sec 40 cycles, Anneal/Extend 60 °C 30 sec 40 cycles. The Sequence Detection System software calculated the baseline and threshold for the amplification curves (Protocol from N.Morgan).

2.11. Immunofluorescence confocal microscopy

HEK293 and Dami cells were transfected as above in 24 well plates with coverslips placed in them. Coverslips were initially sterilized in 70% EtOH and dried prior to placing into wells of plate. 48 hours after transfection, media in the wells were removed and cells were washed with PBS 3 times. After final wash, cells were fixed in 200 μ l of 10% formalin for 5 min. Once fixation was completed, cells were washed again 3 times with PBS. After final wash, if cell permeabilization was required, 200 μ l of 0.1% TritonX-100 was added to each well and incubated for 5 minutes. After 5 minutes, cells were again washed with PBS 3 times before 200 μ l of the blocking solution (45 ml of 5% BSA + 5 ml of 10% goat serum) was added to each well and incubated for a minimum of 1 hour. Once blocking was completed, primary antibody was added directly into blocking

solution and cells were incubated in primary antibody for 1 hour. The amount of antibody used varied, depending on antibody concentration: monoclonal anti-myc antibody 1:200 (Cell Signalling Technology, 9B11), monoclonal anti-His antibody 1:750 (R&D Systems, MAB050), polyclonal anti-ANKRD18A 1:100. After incubation with primary antibody, cells were again washed 3 times with PBS. After washing, cells were incubated with 200 μ l of fluorochrome labelled secondary antibody for 1 hour in the dark. [Alexa Fluor 488 goat anti-mouse IgG H+L 2 mg/ml (A10680, Life Technology), Alexa Fluor 647 goat anti-rabbit IgG H+L 2 mg/ml (A21244, Life Technology)]. After incubating in secondary antibody, cells were again washed 3 times with PBS. If visualization of the nucleus was required, cells were then stained with TO-PRO-3 Iodide (Invitrogen T3605), 1:2000 for 20 min in the dark. Following TO-PRO, cells were again washed 3 times with PBS and then coverslips were taken out of the wells and mounted onto slides with Hydromount (National Diagnostics, HS-106). Slides were then examined on the Leica DMIRE2 confocal microscope. All cells were imaged with a 60X oil objective. GFP tagged proteins and Alexa Fluor 488 secondary antibody were imaged following excitation with the 488 nm line on an Argon-ion laser. TO-PRO-3 and Alexa Fluor 647 were imaged after excitation with a 633 nm line on an Argon-ion laser. Analyses of images were performed using ImageJ software (Fletcher, Iqbal et al. 2014).

2.11.1. Plasmids transfected into HEK293T cells and HeLa cells for confocal microscopy

1. pCMV-GFPWT ANKRD18A
2. pCMV-GFP MUT ANKRD18A
3. pCMV-GFP vector
4. pCMV-MycWT ANKRD18A
5. pCMV-MycMUT ANKRD18A
6. pCMV-Myc vector

2.12. Platelet spreading

Glass coverslips were coated overnight at 4 °C using 100 µg/ml fibrinogen followed by washing with three changes of PBS. The coverslips were then blocked with 5 mg/ml of heat denatured BSA for 1 hour at room temperature followed by washing with another three changes of PBS. Washed platelets were then added to the coverslips to allow for adherence and spreading for 45 minutes at 37 °C. Afterwards, non-adherent platelets were removed with three changes of PBS washes. Adherent platelets on the coverslips were fixed with 10% formalin at room temperature for 10 minutes. Coverslips were then mounted onto glass slides using Hydromount and imaged by confocal microscopy as above (Protocol from Watson lab).

2.13. Flow Cytometry

Cells transiently transfected in 6 well plates were trypsinised if adherent (HEK293T cells) or cell plate is tapped forcefully to dislodge the small amount of adherent cells (Dami cells). Media was collected and cells were spun down at 1000 rpm for 5 min to remove supernatant. Cells were then washed with PBS

and incubated with PBS/10% FBS for 20 min. Cells were washed with PBS again, then fixed with 500 µl of 10% formalin for 5 min. Following fixation, cells were permeabilised with 500 µl of 1% TritoninX-100/PBS for 5 min. Again, afterwards, cells were washed with PBS. Following wash, cells were incubated with primary antibody (anti-25277) at 1:100 for 20 min in PBS + 2% FBS. Control antibody used was a monoclonal anti-IgG (CalTag, MAB8406). After incubation, cells were washed with PBS/1% FBS. After wash, cells were then incubated with secondary antibody, Alexa Fluor 488 goat anti-rabbit IgG (Life Technology A11008) or Alexa Fluor 488 goat anti-mouse IgG at 1:500 in PBS/2% FBS for 20 min. Cells were again washed with PBS/1% FBS. Following wash, cell were resuspended in PBS and analysed on the AccuriC6 (BD Biosciences) flow cytometer (Protocol from J. Reyes).

2.14. Isolation of CD34⁺ cells from human umbilical cord blood by density gradient centrifugation

2.14.1. Isolation of Mononuclear Cells

All reagents were at room temperature. Cord blood was diluted with 15 ml of blood with 35 ml of Buffer 1 (Sterile Buffer 1= 5 ml of ACD in 45 ml of PBS, pH 7.2). Acid Citrate Dextrose (ACD): Sodium Citrate - 25 g/L, Glucose - 20 g/L, Citric Acid- 15 g/L). 35 ml of diluted sample was carefully laid over 15 ml of Ficoll-Paque (d=1077 g/ml, GE Healthcare Life Sciences, #17-1440-02) in 50 ml Falcon tube – the two layers must not be mixed. Low density mononuclear cells were obtained after centrifugation for 30 minutes at 400g at 20 °C with no brake.

The upper layer was aspirated and discarded. The mononuclear cell layer (lymphocytes, monocytes, thrombocytes) is at the interphase. The interphase was transferred to a new 50 ml Falcon tube. The volume of transferred interphase layer was estimated and added to at least 3 volumes of Buffer 1, mixed and centrifuged at 300g for 10 minutes at 20 °C. After spin, supernatant was removed completely. To isolate CD34⁺ cells, pellet (by flicking the tube) was resuspended in 50 ml of Buffer 1 and centrifuged at 200g for 10 minutes at 20 °C. Supernatant was removed completely. Pellet was resuspended in 500 µl of Buffer 2 (Sterile Buffer 2: ACD, PBS pH 7.2, 0.5% BSA) (Protocol from D. Bem).

2.14.2. Isolation of CD34+ cells from mononuclear cells

Subsequent steps required all reagents to be at 4°C. 30 µm pre-separation filter was moistened (Miltenyi Biotec, #130-041-407) with Buffer 2 and then cells passed from above through nylon mesh. Cells were washed once with 500 µl of Buffer 2 and then pellet resuspended in 500 µl Buffer 2 and cell suspension placed into MACS column (Miltenyi Biotec, #130-042-201) and centrifuged at 300g for 10 minutes at 4 °C. Cell pellet resuspended in 300 µl Buffer 2 for up to 10⁸ total cells. 100 µl FcR Blocking Reagent (MACS MicroBead Kit, Miltenyi Biotec, #130-046-702) was added for up to 10⁸ total cells. 100 µl CD34 MicroBeads was added for up to 10⁸ total cells. Mixture was mixed and incubated for 30 minutes at 4 °C. Afterwards, cell number was determined by counting [10 µl of trypan blue (Sigma) plus 90 µl of cell solution]. Cells were washed by adding 5-10 ml of Buffer 2 for up to 10⁸ total cells and centrifuged at

300g for 10 minutes at 4 °C. Supernatant was aspirated completely. Cells up to 10^8 cells were resuspended in 500 μ l Buffer 2 (Protocol from D. Bem).

2.14.3. Magnetic Separation

An appropriate MACS column and MACS separator (Miltenyi Biotec) were chosen according to the number of total cells and the number of CD34⁺ cells. The column was placed in the magnetic field of a suitable MACS Separator. The column was prepared by rinsing with Buffer 2. For MS column, 500 μ l of Buffer 2 and for LS column, 3 ml of Buffer 2. Cell suspension was placed onto column drop-wise. Flow-through, which contained unlabelled cells, (can be used as negative control) was collected. Column was washed three times with the appropriate amount of Buffer 2: MS-3x500 μ l and LS-3x3 ml. Unlabelled cells were collected from wash and combined with previous flow-through. Column removed from the separator and placed onto a suitable collection tube. An appropriate amount of Buffer 2 was pipetted onto the column and the magnetically labelled cells were immediately flushed out by firmly pushing the plug into the column (MS-1 ml and LS- 5 ml) (Protocol from D. Bem).

Column	Max # of labelled cells	Max # of total cells	Separator
MS	10^7	2×10^8	MiniMACS (Octo Vario Super)
LS	10^8	2×10^9	MidiMACS(Quadro Vario Super)

Table 2.7 - Choosing the appropriate MACS column and separator based on total and CD34+ cell numbers

2.14.4. Evaluation of hematopoietic progenitor cell purity

Purity of the isolated hematopoietic progenitor cells were evaluated by flow cytometry or fluorescence microscopy. Analysis of CD34⁺ cells were accomplished by direct immunofluorescent staining using an antibody recognising an epitope different from that recognised by the CD34 monoclonal antibody QBEND/10.

2.14.5. Expansion and differentiation of CD34⁺ cells

CD34 cells were expanded for 5 days using the cocktail of growth factors suspended in the media, Stem Pro 34 (Invitrogen) plus 0.1% Penicillin/Streptomycin (Sigma), 0.1% glutamine (Sigma) and 1.3 ml nutrients. Growth factors (Miltenyi Biotec) are: Flt3-L 50 µg/ml, IL3 10 µg/ml, h-SCF 50 µg/ml, TPO 20 µg/ml, and IL6 10 µg/ml. After 5 days the cocktail changed to h-SCF (1 ng/ml), IL6 (7.5 ng/ml), IL9 (13.5 ng/ml) and TPO (30 ng/ml) so that CD34⁺ cells can differentiate to megakaryocytes. Following 12 days of differentiation, cells were checked for ploidy and expression of CD41 using flow cytometry and mRNA expression using RT-PCR as described in (Pinault 2013).

2.14.6. Evaluation of ploidy in megakaryocytes

Used for checking maturation stage of megakaryocytes. Cells were incubated for 45 minutes at 37 °C with 0.1% Triton X-100 (Sigma), RNase (Sigma) and propidium Iodide (Sigma) to stain for DNA. Ploidy was determined using the BD FACS Calibur flow cytometer (Protocol from A. Mazharian).

2.15. Transmission electron microscopy

PRP was prepared by centrifugation at 200g for 10 minutes and fixed with equal volume of 0.1% glutaraldehyde (GA) in 0.1 M sodium cacodylate buffer (pH 7.4) at room temperature for 15 minutes. Platelets were post-fixed with 3% GA in 0.1 M sodium cacodylate buffer at 4 °C for 30 minutes. The cells were further fixed with 1% osmium tetroxide in 0.1M sodium cacodylate buffer for 1 hour at 4 °C. Cells were then rinsed with distilled water, dehydrated in a graded series of alcohol and embedded in Epon resin. Ultrathin sections (70-90 nm) were stained with uranyl acetate and lead acetate. All imaging was done in a Tecnai G2 Spirit microscope (FEI, Eindhoven, Netherlands) using a Morada EM CCD digital camera (Olympus SIS, Münster, Germany) (Bem, Smith et al. 2015).

2.16. Statistical analysis of α -granules in electron micrographs

Results were shown as mean \pm SEM. Statistical analysis was performed with Excel software (Microsoft) using a two-way ANOVA test. $P < 0.05$ was considered statistically significant.

3. Initial characterisation of the *ANKRD18A* variant

3.1. Introduction

Prior to the commencement of this PhD, a whole exome-sequencing approach was used to identify the genetic aetiology for the severe thrombocytopenia seen in two cousins from a consanguineous family (Figure 1.14). Whole exome sequencing (WES) identified a homozygous in-frame deletion mutation (c.2401_2404delGAA) in the gene, ankyrin repeat domain 18A gene (*ANKRD18A*), of both children. To confirm that this mutation in *ANKRD18A* is causing the thrombocytopenia seen in the two patients, experiments needed to be conducted using patient platelets.

3.2. Quantitation of *ANKRD18A* mRNA expression in patient platelets

Unfortunately, because both study patients are children who received frequent platelet transfusions, it was difficult to obtain regular blood samples to study the patients' platelet defects. Prior to my joining the laboratory, some blood was obtained from both patients and the following studies were performed to characterize their platelets. Both patients had severe thrombocytopenia: their platelet counts in the platelet rich plasma (PRP) was significantly below the reference range seen in healthy volunteers (patient II:5 $1.5 \times 10^7/\text{ml}$ and patient II:3 $2.5 \times 10^7/\text{ml}$, reference range $2.1\text{--}7.1 \times 10^8/\text{ml}$ [mean \pm 2 s.d.]). Mean platelet

volume (MPV) in patient II:5 was 15.0 fl and in patient II:3, it was 10.4 fl (reference range 7.68 – 10.0 fl [mean \pm 2 s.d.]). Of note, full blood counts from both patients' parents were normal.

Prior to functional studies on the patients' platelets, one first needed to measure the abundance of *ANKRD18A* mRNA using real time quantitative PCR (qRT-PCR). This experiment compared the *ANKRD18A* mRNA levels of the patients' leukocytes with those of healthy controls. Quantitation was not performed on patient platelet mRNA because of the assumption that being anuclear, platelets would have minimal mRNA, which would further compound the results from the very low platelet counts. Instead, quantitation of *ANKRD18A* mRNA was performed using patient leukocytes from the buffy coat.

Using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA), relative gene expression quantification was performed according to the comparative C_T method using $\beta 2$ microglobulin (*B2M*) as an endogenous control. qRT-PCR product accumulation was measured by SYBR green, a DNA intercalating dye. When measuring RNA expression of target genes, to control for errors between samples, there needs to be coincident measurement of RNA expression of a reference gene (in this case, *B2M*) so that there is normalization of the target gene expression data. In this experiment, the C_T (threshold cycle) of target genes was compared to the C_T of reference genes

(relative qRT-PCR method). The threshold cycle is the intersection of the reaction curve to the threshold line, which is a level above the background fluorescence. The C_T indicates the number of cycles it takes to detect a signal above background fluorescence in the samples (www.thermofisher.com/uk/en/home/life-science/pcr/real-time-pcr/qpcr-education/pcr-understanding-ct-application-note.html). To analyse the qRT-PCR data, the double ΔC_T analysis method was used (Livak and Schmittgen 2001). This method involved averaging C_T values for the target gene and reference gene and the controls. The average C_T value for the reference gene was subtracted from that of the target gene's C_T value. The same was done for the C_T values of the controls of the target gene and reference gene. Afterwards, the difference of the above two C_T values provided a $\Delta\Delta C_T$ value, which then converted to linear form by the calculation of $2^{-\Delta\Delta C_T}$.

Compared with control leukocytes, *ANKRD18A* mRNA levels were significantly reduced by more than 80% in patient leukocytes ($p < 0.02$) (Figure 3.1). The reduction in the patients' *ANKRD18A* mRNA was presumably due to impaired stability of the mutated transcript (N.Morgan, unpublished data).

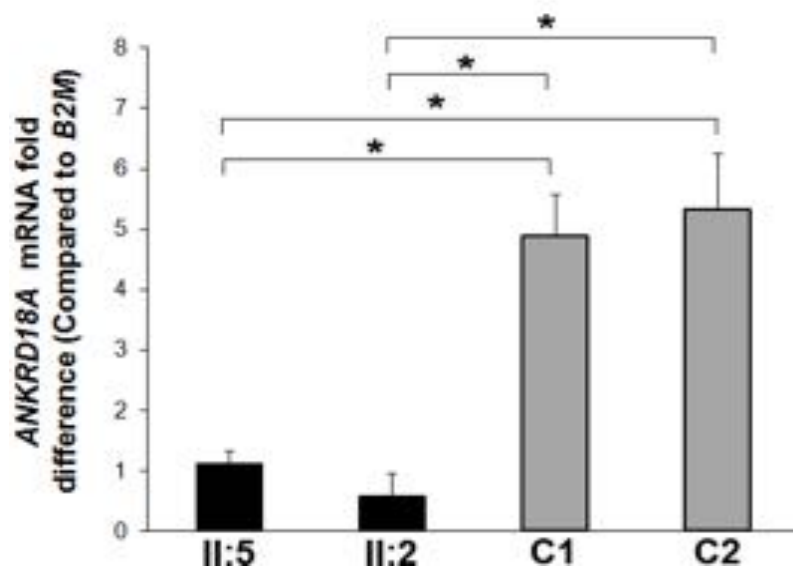
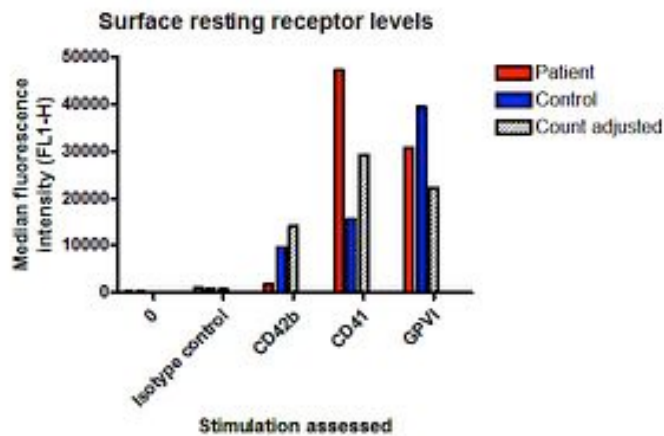


Figure 3.1 - A homozygous *ANKRD18A* mutation results in decreased expression of *ANKRD18A* mRNA in 2 patients with a severe thrombocytopenia. *ANKRD18A* mRNA production, as assessed by real-time quantitative PCR (qRT-PCR) on total RNA isolated from PBMCs from the 2 patients (II:5 and II:3) and 2 healthy controls (C1 and C2). *ANKRD18A* mRNA levels were expressed as relative expression compared to β 2 Microglobulin mRNA levels. Values are mean \pm SEM for 3 experiments for all samples. * $P < 0.05$ (N. Morgan, unpublished data).

3.3. Functional Studies using patient platelets

Prior to functional studies, the patients' resting levels of platelet surface glycoproteins: CD42b (GPIb α , a component of the GPIb-IX-V complex which binds von Willebrand factor), CD41 (integrin α IIb, which binds fibrinogen), and GPVI (binds collagen) were measured by flow cytometry (Figure 3.2).

A.



B.

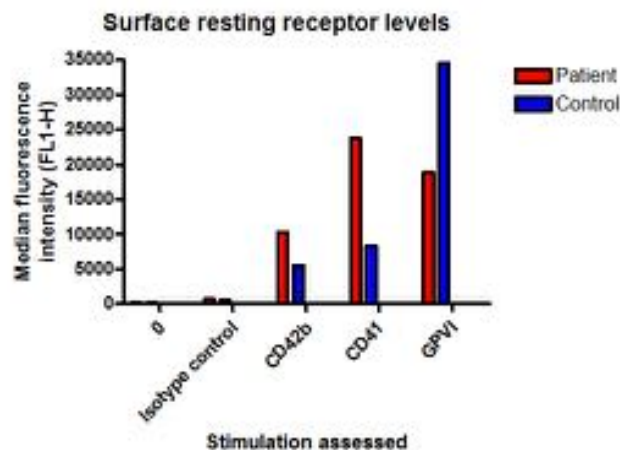


Figure 3.2 – Patients' platelets [(A) II:5 and (B) II:3] have similar surface receptor expression levels compared to control when surface area of patient platelets were normalised to control. However, CD42b levels for patient II:5 is reduced and GPVI levels are slightly reduced for patient II:3, both due to unclear reasons (N. Morgan, unpublished data).

For Patient II:5, levels of CD42b were reduced when compared to healthy control. Levels of CD41 and GPVI were similar in the control and patient (The patient's MPV was larger and the surface area of the patient's platelets were approximately 30% larger than the control). The significance of the CD42b result is uncertain as previous investigations were not compatible with a diagnosis of Bernard-Soulier disease. For Patient II:3, levels of GPVI were slightly reduced and levels of CD42b and CD41 were comparable in the control and patient (The patient's MPV is larger and the surface area of the patient's platelets is approximately 15% larger than the control). The significance of the GPVI result is uncertain (N. Morgan, unpublished data).

Platelet function was then assessed by flow cytometry using an assay that was validated using dilutions of PRP from healthy volunteers. The control PRP was diluted 1 in 10 with phosphate-buffered saline (PBS) to enable assessment of fibrinogen binding. PRP was stimulated with ADP, collagen-related peptide (CRP) and PAR-1 peptide (PAR-1 specific thrombin receptor activating peptide with the amino acid sequence SFLLRN). Both the expression of P-selectin (FITC conjugated mouse anti-human CD62 antibody, BD Pharmingen), which is a marker of dense and α -granules release upon platelet activation and binding of fluorescent fibrinogen (DAKOCytomation Polyclonal Rabbit Anti-Human Fibrinogen/FITC) to the activated form of the platelet surface α IIb β 3 (GPIIb/IIIa) complex were assessed by flow cytometry on an Accuri C6 flow cytometer. Both

patients' PRP showed markedly reduced response to ADP, CRP and PAR-1 peptide when assessing for expression of FITC-conjugated anti-CD62P. Strikingly, patient II:5 showed an abolition of the response to ADP, CRP and PAR-1 peptide (Figure 3.3A). The response of patient II:3 to the same agonists was markedly reduced (Figure 3.3A). In both patients, the response assessed using binding of fluorescent fibrinogen was reduced when compared with normal ranges derived from healthy volunteers PRP (Figure 3.3B). It cannot be ascertained however, whether these abnormal results were a reflection of decreased platelet numbers or because ANKRD18A plays a role in platelet activation (N. Morgan, unpublished data).

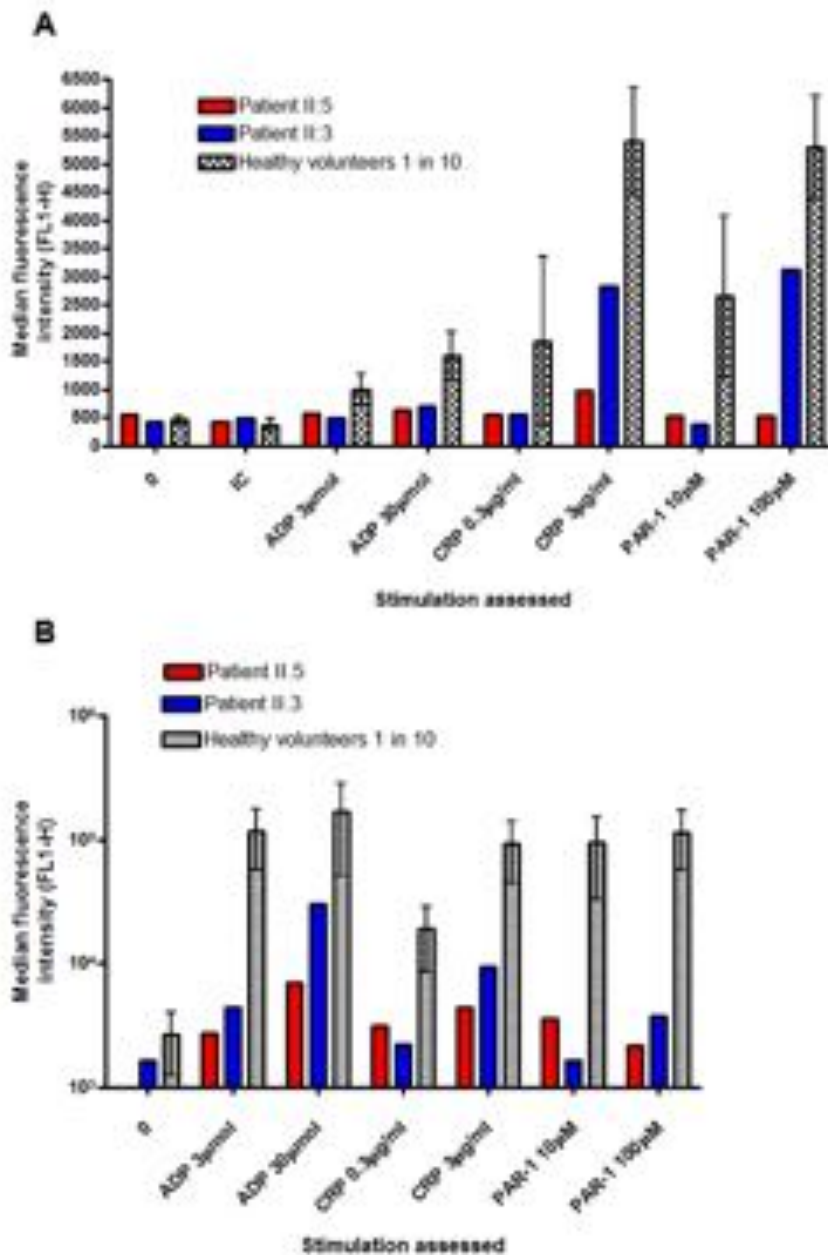


Figure 3.3 – Patients’ platelets have decreased responses to α -CD62P (A) and decreased fibrinogen binding to α IIb β 3 (B) after stimulation with low and high dose agonists: ADP, CRP and PAR-1 peptide. Healthy volunteers 1 in 10 = PRP from control individuals diluted 1 in 10 with PBS. Data for healthy volunteers shown as mean \pm 1 s.d. (n=9). Isotype control = IgGk1. (N. Morgan, unpublished data).

3.4. Electron microscopy of ANKRD18A patient platelets

To study the patients' platelet ultrastructure, transmission electron microscopy (TEM) images were obtained from patient II:3 (Figure 3.4). TEM images of the platelets from patient II:3 revealed that they were larger but fewer in number than normal subjects (macrothrombocytopenic). Also, it appeared that these large platelets may have more α -granules. However, upon further analysis whereby the surface area of the larger platelets are taken into account (number of α granules in platelet/surface area of platelet), one sees that the number of α granules is no greater than that of platelets from a healthy subject using the two-way ANOVA statistical test (Control: 1.01 ± 0.42 vs. Patient: 0.98 ± 0.32). A total of 24 platelets were counted from the patient slide and 51 platelets from the control slide. Surface area of platelets was determined by using the ImageJ analysis tool.

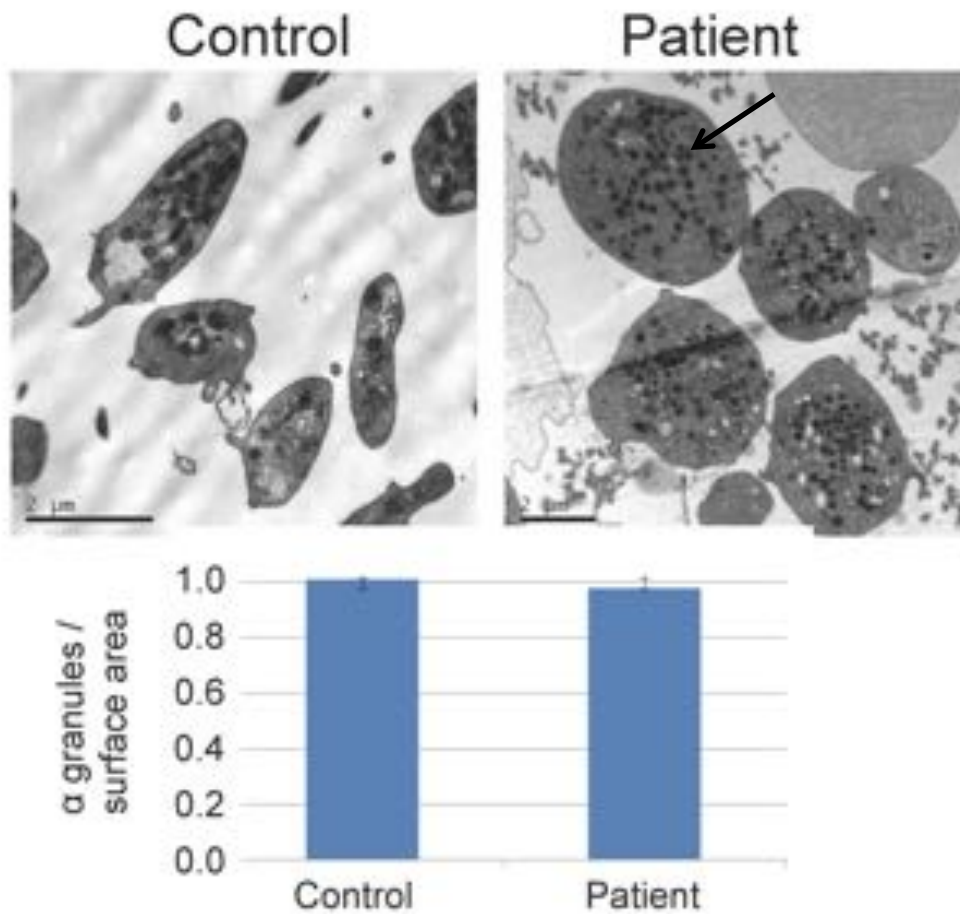


Figure 3.4 – Transmission electron microscopy image demonstrating patient platelets to be enlarged but number of α granules in patient platelets not significantly different from number of α granules in healthy control platelets. Patient with macrothrombocytopenia but the number of α granules (arrow) remain comparable to control α granule numbers when surface area of platelets were taken into account (N=1). TEM images courtesy of Denai Bem.

3.5. Utilising full-length wild-type and mutant ANKRD18A plasmids for transfections and protein expression

Because of limited patient blood that could be obtained for experimental use, it was necessary to construct cDNA plasmids with full length *ANKRD18A* wild-type and mutant ($\Delta 801$) sequences which could then be transiently transfected into cells for further characterisation of the protein, ANKRD18A. Human full-length wild-type and mutated *ANKRD18A* plasmids were constructed by cloning the respective cDNAs into pEGFP and pCMV-Myc vectors (cloning performed by I. Sanchez-Guiu). The ANKRD18A mutated construct (p.Glu801del) was created using the Quikchange Site directed mutagenesis kit.

3.5.1. Determining transfection efficiencies of full-length GFP-tagged ANKRD18A constructs

Following the construction of the tagged full-length ANKRD18A plasmids, attempts were made to transfect them into HEK293T to determine the protein expression of WT and mutant ANKRD18A by blotting with α -GFP or α -myc in Westerns. Unfortunately, no bands appeared at the correct molecular weight for any tagged recombinant proteins (~145 kD for GFP-tagged constructs). The only protein which was recognised by the α -GFP antibody was GFP protein itself (Figure 3.5B). It was thus decided to first check, using flow cytometry, what were the transfection efficiencies of the full-length plasmids. HEK293T cells, transiently transfected with DNA plasmids encoding full-length wild-type and mutant ($\Delta 799$) ANKRD18A fused to a GFP-tag, were analysed for GFP fluorescence with an excitation light source at 488 nm. Recombinant full-length

wild-type and $\Delta 801$ ANKRD18A (GFP-ANKRD18A) expressed only poorly (around 15-20% compared to vector control) (Figure 3.5A). Although it appeared that there was some transfection of these plasmids into HEK293T cells, it still proved difficult to demonstrate protein expression of GFP-tagged wild-type and $\Delta 801$ ANKRD18A when corresponding plasmids were transfected into HEK293T cells and immunoblotted with a GFP antibody (Figure 3.5B). Attempts were made to see if the protein of these full-length constructs were in the insoluble fraction rather than the whole lysates. This was done by immunoblotting Westerns loaded with the insoluble fraction. Done by myself and repeated at OPPF, both attempts proved fruitless. No expression of the proteins was seen.

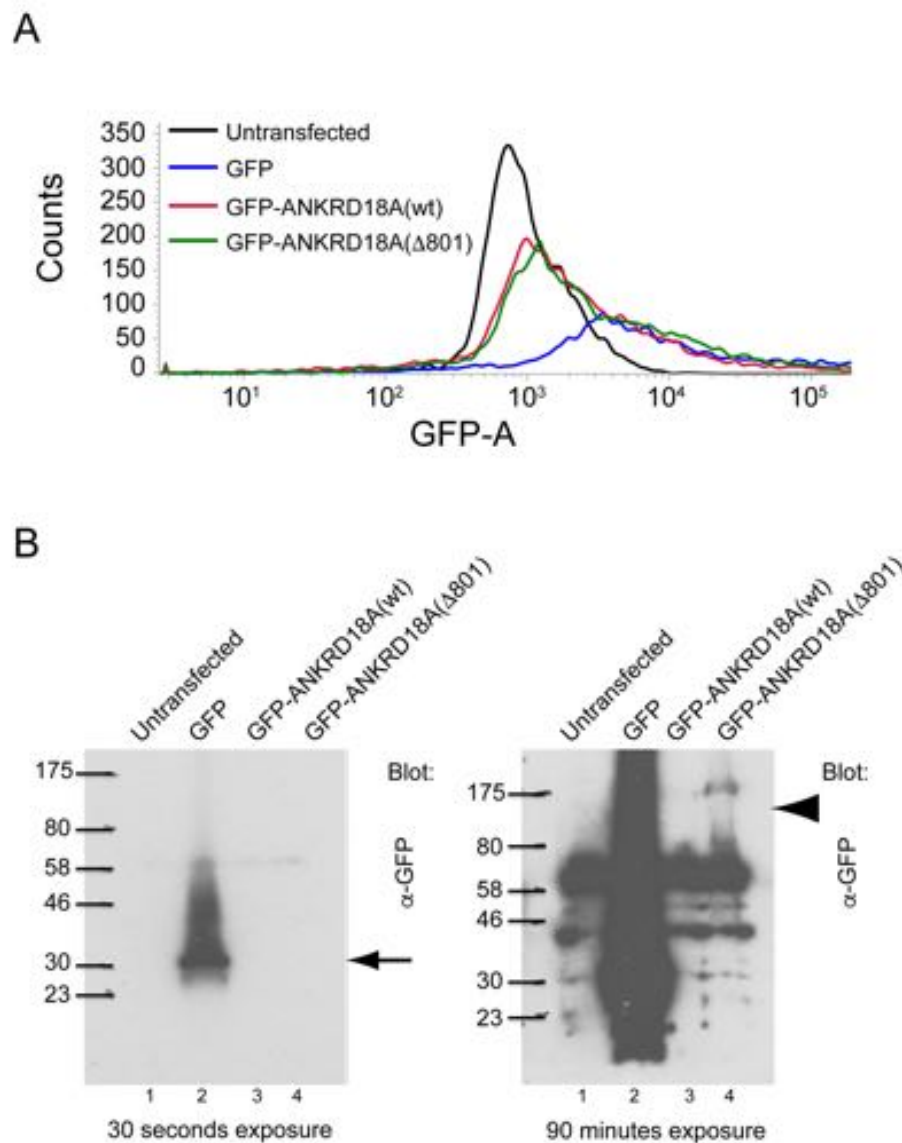


Figure 3.5 - Lower transfection efficiencies using GFP-full length wild-type and mutant ANKRD18A plasmids compared to GFP-empty vector. (A) Flow cytometry analysis of HEK293T cells untransfected and transfected with plasmids encoding GFP-ANKRD18A (WT), GFP-ANKRD18A ($\Delta 801$), GFP vector control. (B) Immunoblot of HEK293T WCLs treated as in panel A and immunoblotted with α -GFP. Untransfected (lane 1), GFP vector (lane 2) GFP-ANKRD18A (WT) (lane 3), GFP-ANKRD18A ($\Delta 801$) (lane 4). Arrow in left panel identifies the GFP (MW 27 kDa) protein and the arrowhead in the right panel demonstrates where the GFP-tagged full length ANKRD18A protein should be located, 143 kDa. (N=20)

Given the findings from the above flow cytometry experiment, which demonstrated that the full-length constructs of wild-type and mutant ANKRD18A could be transfected into cells, albeit at a very small amount, it was decided to proceed with immunofluorescence confocal microscopy. The rationale for this was because, with immunofluorescence confocal microscopy, fewer number of cells need to be transfected with the desired plasmid for one to visualise protein expression whereas with Westerns, many more cells need to be transfected in order for protein expression to be visualised by immunoblotting. The preliminary confocal experiment using the full-length GFP tagged plasmids suggested that both wild-type and mutated ANKRD18A were localised to the cytoplasm (Figures 3.6).

The question we wished to answer by using confocal microscopy was whether ANKRD18A played a role in the development of platelets from megakaryocytes. The transition from megakaryocyte to platelets is characterised by the presence of subcellular structures such as the invaginated membrane system (IMS), proplatelets, and preplatelets. These structures have been found to co-localise with components of the cell cytoskeleton, including filamentous F-actin, β 1-tubulin, microtubules, and spectrin (Machlus and Italiano 2013). Possibly, co-localisation of ANKRD18A with these and related sub-cellular membrane structures could provide a strong hint as to a mechanistic role for ANKRD18A in the development of platelets from megakaryocytes. To begin investigating the

possible association of ANKRD18A with cytoskeletal elements, it was necessary to first determine the cellular localisation of ANKRD18A. This was first accomplished by transfecting full length wild-type and mutant forms of GFP-tagged *ANKRD18A* cDNA into HEK293T cells followed by imaging of the cells by confocal microscopy (Figure 3.6).

The decision to first attempt the confocal experiment with HEK293T cells stemmed from the fact that all previous experiments had been performed on HEK293T cells. This is because they are easily transfectable and easier to manage. We wanted to perform preliminary over-expression experiments using HEK293T cells but were aware that the question which we wished to answer regarding ANKRD18A's possible role in megakaryocyte development, would require us to best answer those questions in the future using cells which are megakaryocyte-like, such as the megakaryoblastic cell lines Dami cells or MEG-01 cells, and ultimately with megakaryocytes.

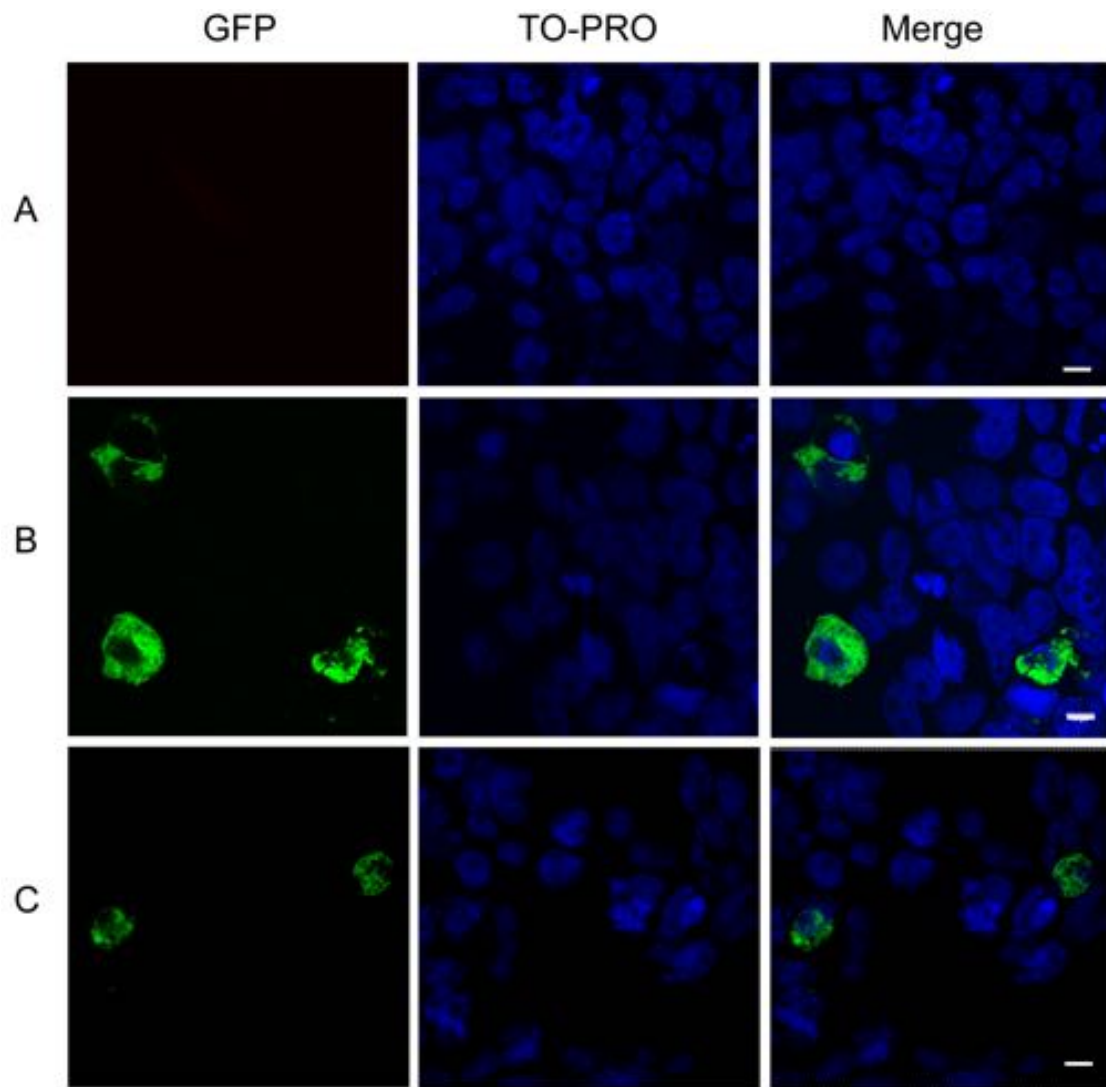


Figure 3.6 – Confocal microscopy images of HEK293T cells demonstrating cytoplasmic localisation of WT and MUT ANKRD18A following transient transfection with plasmids encoding full length (B) wild-type GFP-tagged ANKRD18A and (C) mutant GFP-tagged ANKRD18A ($\Delta 801$). Row A demonstrate images of untransfected HEK293T cells. Presence of cells is verified using the nuclear stain TO-PRO3. Scale bar = 10 μ m (N=2).

To further strengthen the observation of a potential cytoplasmic localisation of ANKRD18A, the experiments were repeated, transfecting HeLa cells with the myc-tagged full-length plasmids. After 48 hours of incubation, the HeLa cells were imaged using immunofluorescence confocal microscopy. The primary antibody used was a monoclonal anti-myc antibody, nuclear staining was with TO-PRO-3, and secondary antibody was with α -mouse Alexa Fluor 488 (Figure 3.7). Again, results of this experiment suggested that WT and mutant ANKRD18A localised to the cytoplasm.

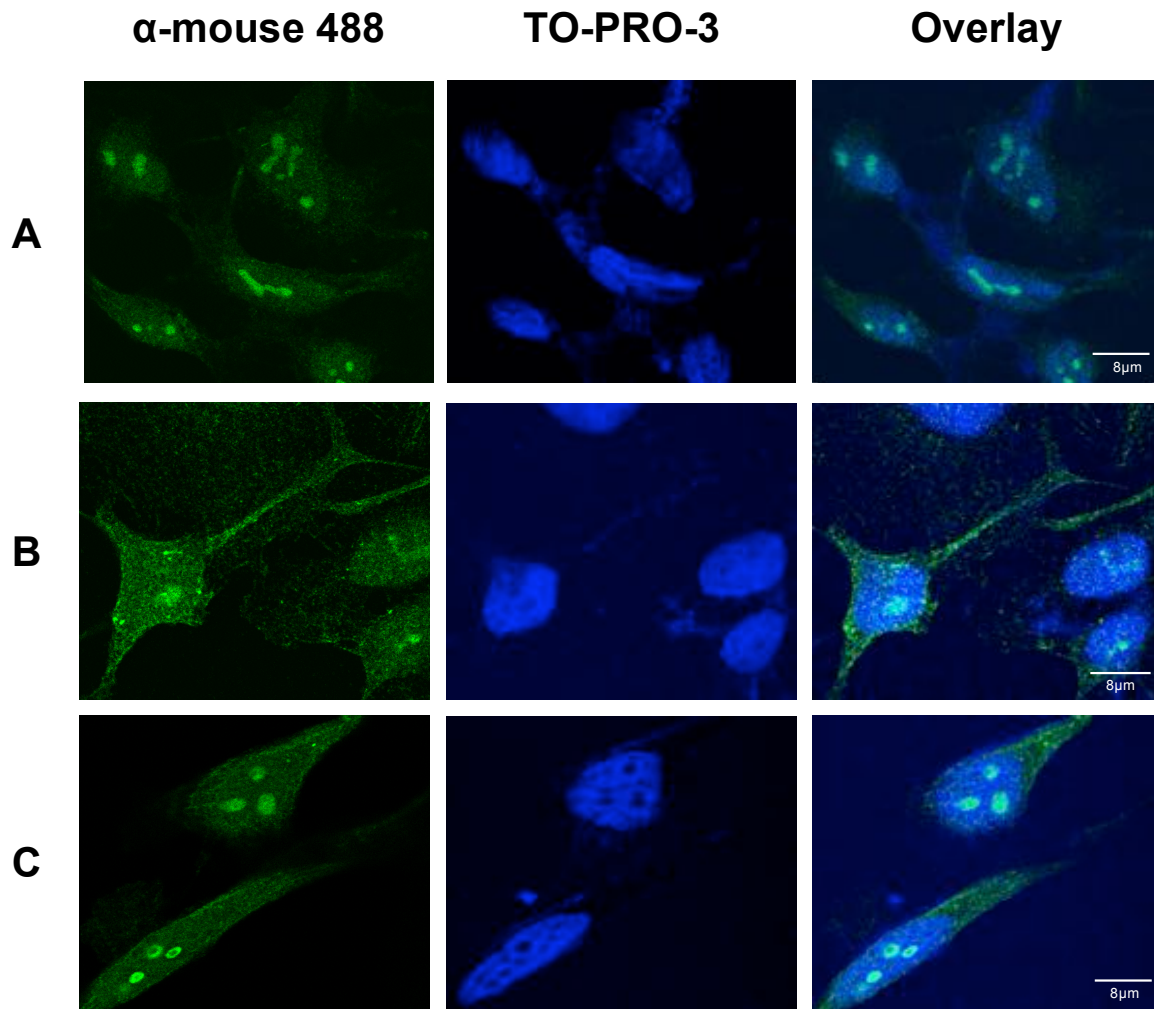


Figure 3.7 – Immunofluorescence confocal microscopy images of HeLA cells transiently transfected with plasmids encoding (A) myc-tag empty vector (B) myc-tagged WT ANKRD18A and (C) myc-tagged mutant ANKRD18A (Δ801) demonstrating cytoplasmic localisation of WT and MUT ANKRD18A. Primary antibody used was a monoclonal anti-myc antibody and secondary antibody was α-mouse Alexa Fluor 488. Nuclear staining was with TO-PRO-3 iodide. Scale bar = 8 μm (N=2)

3.6. Summary

In summary, using WES, our group has identified a mutation in a gene, which may be the genetic basis for the severe thrombocytopenia seen in a family. Preliminary TEM experiments using patient platelets indicated that the mutation results in low platelet counts and the platelets are enlarged (macrothrombocytopenia). Further analysis using patient leukocytes indicated that *ANKRD18A* mRNA levels were lower in the patients when compared to levels seen in healthy donors. Once platelet size was taken into account, platelet levels of resting surface receptor CD41 was normal for both patients whereas CD42b levels were decreased in patient II:5 and GPVI levels were slightly decreased in patient II:3. Reasons for the lower levels remain unclear. Work-up for Bernard-Soulier Syndrome was negative in patient II:5.

Assessment of patient platelet function using flow cytometry to measure expression of FITC-conjugated α -CD62P and binding of fluorescent fibrinogen demonstrated decreased responses to agonists ADP, CRP and PAR-1 peptide. One cannot be certain however, if this result was because of low platelet counts or of functionally defective platelets. Low platelet counts could have resulted in lower platelet activation because agonists cause platelet degranulation resulting in endogenous secretion of ADP and serotonin, which would lead to further platelet stimulation. If there were lower numbers of platelets, less endogenous agonist would have been produced leading to less platelet stimulation.

However, what argues against this theory were the two studies with patients suffering from ITP and WAS/XLT, where the patients were treated with elthrombopag (thrombopoietin-receptor agonist) to see if increasing patients' platelet numbers would increase platelet activity. Elthrombopag did increase platelet numbers, but did not improve platelet activity in both sets of patients (Psaila, Bussel et al. 2012, Gerrits, Leven et al. 2015).

Due to scarcity of patient platelets to work with, as well as a lack of a reliable antibody to *ANKRD18A*, cDNA plasmids of the full-length human sequences of wild-type and mutant *ANKRD18A* tagged with GFP and myc were constructed. Using cells transfected with the GFP tagged *ANKRD18A* plasmids, flow cytometry experiments demonstrated only 15-20% expression of the *ANKRD18A* wild-type and mutant proteins as compared to the GFP protein alone. Immunoblotting experiments using cells transfected with the same constructs demonstrated no expression of the *ANKRD18A* proteins whereas the control, GFP protein, was expressed. These plasmids were also used to transiently transfect HEK293T cells and HeLa cells to determine the cellular localisation of *ANKRD18A* utilising confocal immunofluorescence. Results from the transfection of both cell lines suggested that *ANKRD18A* localised to the cytoplasm.

Further imaging studies could have been performed using the peptide antibodies which recognise *ANKRD18A*. However, there was a concern that these peptide

antibodies for ANKRD18A might also recognise ANKRD18B as well, given the 93% amino acid identity. However, since it appears that *ANKRD18B* mRNA is not expressed in lymphocyte cell lines (Figure 4.6), post-transcriptional gene silencing (RNAi) could have been performed on these cells to delete *ANKRD18A*. Confocal immunofluorescence could then have been utilised using the peptide antibodies to recognise and localise ANKRD18A in cells manipulated and unmanipulated with RNAi. These images could then further substantiate the cellular localisation of ANKRD18A as well as identify whether the peptide antibody is specific to ANKRD18A.

4. Characterisation of ANKRD18A peptide-derived antibodies

4.1. Introduction

The overall objective of this study was to characterise ANKRD18A. To achieve this objective it was essential to have reagents, which could be used to probe for the presence of ANKRD18A in the proteome. At the initiation of this project, ANKRD18A polyclonal antibodies had been purchased, but when used in immunoblotting experiments, gave conflicting results. It was difficult to discern if the polyclonal peptide antibodies made in-house (α -21233 and α -21234) and a commercial (Abgent) polyclonal antibody were specific to ANKRD18A. When used to probe HEK293T and Dami whole cell lysates, the peptide antibodies identified a predominant band at an incorrect molecular weight; and, the commercial polyclonal antibody purchased from Abgent, when used to probe platelet lysates produced no bands at all. Additionally, there was confusion because of the existence of another protein, ANKRD18B, which has an amino acid sequence, which is 93% identical to ANKRD18A. Because of the need for an antibody, which is specific to ANKRD18A, it was decided to generate an antibody that recognises an epitope on the tertiary structure of ANKRD18A.

4.2. Comparison of ANKRD18B and ANKRD18A

ANKRD18B belongs to the same *POTE* family of genes as *ANKRD18A*. *ANKRD18B* has a 93% amino acid sequence identity to *ANKRD18A* (Figure S3) and is found in close proximity to *ANKRD18A* on human chromosome 9. The two genes most likely arose from ancestral duplication (N. Morgan). Because *ANKRD18B*'s amino acid sequence is so similar to *ANKRD18A*, it was essential that antibodies to *ANKRD18A* could discriminate between these two closely related proteins. The uniprot database indicates that *ANKRD18B* has two isoforms: the long isoform has 1401 amino acids with a calculated molecular weight of ~160 kDa (<http://web.expasy.org/protparam/>) and a shorter isoform, which has 1011 amino acids with a calculated molecular weight of ~116 kDa.

The peptide sequence of *ANKRD18A* that was used to raise α -21233 and α -21234 differed from the corresponding sequence in *ANKRD18B* by only one amino acid. Thus, it is conceivable that α -21233 and α -21234 recognise *ANKRD18A* and both isoforms of *ANKRD18B*. Because of this uncertainty, it was necessary to characterise *ANKRD18B* to ascertain how it differs from *ANKRD18A*.

Within the Ensembl genome browser is the Expression Atlas, which provides from various projects and consortiums, microarray and RNA-sequencing data on gene expression patterns in a particular anatomical part of an organism or a

particular cell. Of these projects, one, which is sponsored by the National Institutes of Health in Bethesda, Maryland, is the Genotype-Tissue Expression Project (GTEx), (www.genome.gov). Based on data from GTEx, *ANKRD18B* mRNA is expressed in fewer human tissues than *ANKRD18A* (at levels above 0.5 FPKM and Section 1.6.1). *ANKRD18B* mRNA is only expressed in testis (2 FPKM), cerebellum (3 FPKM), cerebellar hemisphere (2 FPKM), and cerebral cortex (0.5 FPKM). Also within the Expression Atlas is data from the BLUEPRINT Epigenome Project (www.blueprint-epigenome.eu), which examined cells from differentiated haematopoietic lineages. *ANKRD18B* mRNA expression in healthy haematopoietic cells, unlike *ANKRD18A* mRNA expression, did not appear to be expressed in leukocytes (if it is expressed, it is expressed below the 0.5 FPKM cut-off); rather, it is expressed only in CD34⁻CD41⁺CD42⁺ MK cells (1 FPKM) as determined by strand-specific RNA sequencing of rRNA-depleted total RNA from common types of cultured or uncultured primary cells of differentiated haematopoietic lineages from healthy individuals in the BLUEPRINT Epigenome Project. In addition, sequencing of poly A mRNA from cultured or uncultured primary cells of different haematopoietic lineages from healthy individuals in the BLUEPRINT Epigenome Project also demonstrated expression of *ANKRD18B* mRNA in CD34⁻CD41⁺CD42⁺ MK cells (0.6 FPKM). However, according to the PlateletWeb, a systems biology analysis of signalling networks in human platelets, (<http://plateletweb.bioapps.biozentrum.uni-wuerzburg.de>), no *ANKRD18B* is expressed in platelets.

Because of the conflicting information derived from the genome and proteome databases, it was necessary to ascertain myself whether *ANKRD18B* mRNA is present in platelets of healthy donors and in cultured cells. A set of primers were designed, ANKRD18B_CF and ANKRD18B_CR (sequences in Table 2.6), which were used in RT-PCR experiments to establish whether *ANKRD18B* mRNA was expressed in healthy platelets and cultured cells. Based on my RT-PCR experiment, *ANKRD18B* mRNA did not appear to be present in healthy donor platelets (Figure 4.1). The expected size of *ANKRD18B* is 371 bp. The PCR appeared to have worked because using β -actin primers (the positive control) resulted in a band at the correct size of 223 bp.

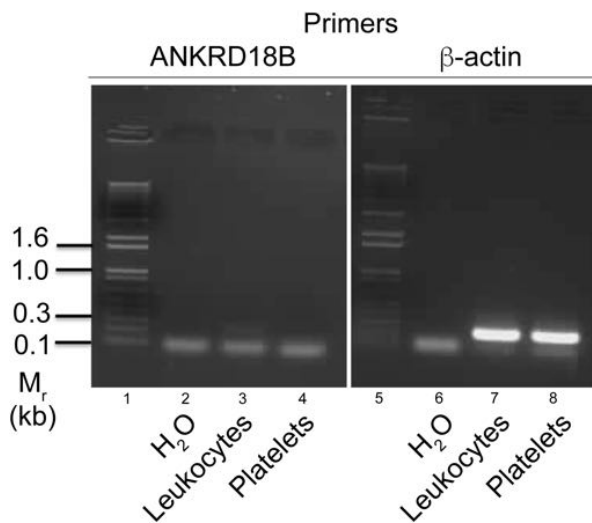


Figure 4.1 - *ANKRD18B* mRNA is not expressed in platelets or leukocytes as determined by RT-PCR. Expected size of *ANKRD18B* is 371 base pairs (bp). Expected size of β -actin is 223 bp. Bands seen in left gel are primer dimers. Lower bands seen in both gels are primer dimers. (N=1)

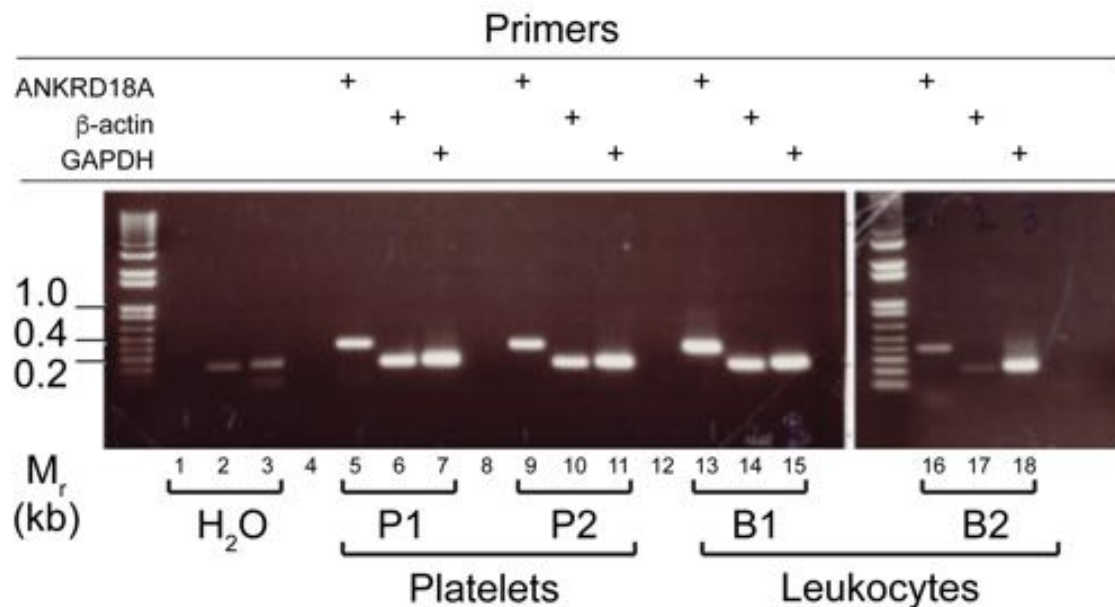


Figure 4.2 – *ANKRD18A* mRNA is expressed in platelets and leukocytes as determined by RT-PCR. Lanes 1, 5, 9, 13, 16 use *ANKRD18A* primers. Lanes 2, 6, 10, 14, 17 use β -actin primers and Lanes 3, 7, 11, 15, 18 use *GAPDH* primers. P = platelets. B = buffy coat. Expected size of *ANKRD18A* is 394 bp. Expected size of β -actin is 223 bp. Expected size of *GAPDH* is 225 bp. (N = 3)

The above results contrasts with my RT-PCR experiment for *ANKRD18A*, which suggested that there is *ANKRD18A* mRNA expression in platelets; however, one cannot ascertain whether this result is due to leukocyte contamination or actually there is *ANKRD18A* mRNA in platelets. This is especially relevant since according to the BLUEPRINT Epigenome Project, *ANKRD18A* mRNA is predominantly expressed in myeloid and lymphoid cells, with negligible expression in megakaryocytes (www.blueprint-epigenome.eu). To determine whether there is leukocyte contamination in this experiment, one would need to

check by flow cytometry for the presence of a cell population which is positive for CD45 (leukocyte common antigen) following the isolation of platelets (platelet rich plasma, PRP) from whole blood. To remove leukocyte contamination one would need to perform CD45⁺ bead selection after the preparation of PRP (Denis, Tolley et al. 2005).

When probing for the presence of *ANKRD18A* (394 bp) and *ANKRD18B* (371 bp) mRNA expression in a variety of cell lines, it appeared that *ANKRD18A* mRNA was present in all the cell lines tested. These cell lines included lymphocyte cell lines (1654, Lgn, RY, and HK), HEK293T cell line derivatives (D+ and YTW) and HeLA, Dami and HEK293T cells. On the other hand, *ANKRD18B* mRNA expression was only present in HeLA, Dami cells and HEK293T cells plus its derivatives (Figure 4.3).

The above RT-PCR results which demonstrated that *ANKRD18B* mRNA is not expressed in any of the lymphoid cell lines is consistent with the BLUEPRINT databases which suggested that *ANKRD18B* mRNA is not expressed in myeloid and lymphoid cells.

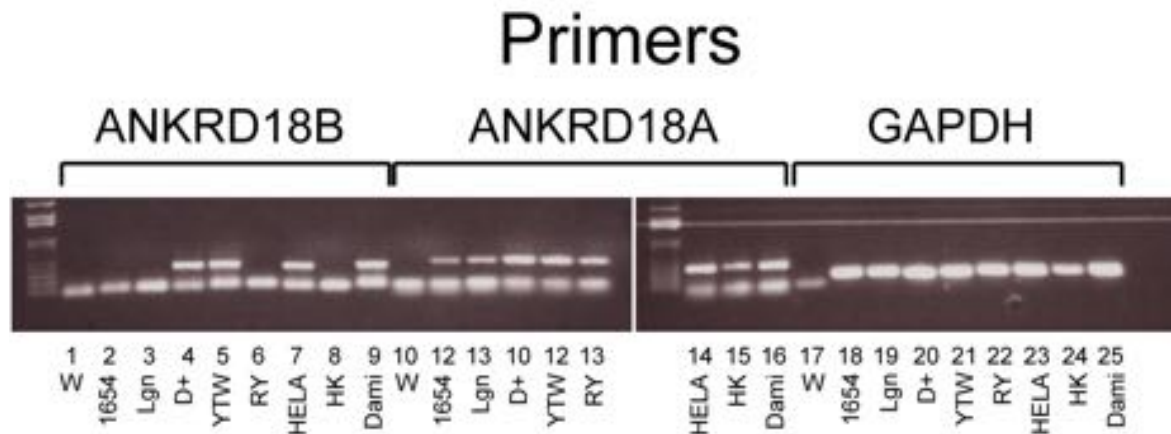


Figure 4.3 – *ANKRD18A* mRNA is expressed in lymphoid cells lines, Damis, HeLAs and HEK293T cells and its derivatives. *ANKRD18B* mRNA is expressed in Damis, HeLAs and HEK293T cells and its derivatives. Lanes 1=H₂O, lane 2= 1654, a B-cell line, lane 3= Lgn, a lymphocyte cell line, lane 4=D+, a HEK293 cell line derivative, lane 5=YTW, a HEK293 derivative, 6= RY, a lymphocyte cell line, 7= HeLa, 8= HK, a lymphocyte cell line, 9=Dami. Upper band is *ANKRD18B* or *ANKRD18A* mRNA. Expected size of *ANKRD18B* is 371 bp. Expected size of *ANKRD18A* is 394 bp. Expected size of *GAPDH* is 225 bp. Lower bands are primer dimers. (N=1)

4.3. Initial characterisation of existing ANKRD18A antibodies

Prior to the initiation of my PhD studies, two ANKRD18A polyclonal antibodies, which were directed against two non-overlapping peptides, were available in the laboratory. One was a commercial antibody purchased from Abgent, a rabbit polyclonal antibody which recognised the peptide sequence 260-289; the other were two bespoke antibodies which were two separate bleeds from a rabbit (antibodies α -21233 and α -21234) that had been immunised with the ANKRD18A-derived peptide encompassing residues 53-65 (Figure 4.4). The peptide antibodies had been purified on an affinity column, which had the

ANKRD18A peptide bound to its matrix so that monospecific IgG to ANKRD18A was obtained. These antibodies were used to immunoblot for the presence of ANKRD18A in whole cell lysates (WCLs) of HEK293T and Dami (a megakaryocytic cell line) cells and from platelets of different healthy donors. In these Western experiments, Dami WCLs were divided into adherent Damis and Damis in suspension because when cultured, Damis are initially adherent but go into suspension when more mature (Greenberg, Rosenthal et al. 1988). The question arose whether expression of ANKRD18A would differ between adherent Damis vs. Damis in suspension. However, the primary question, which we needed to answer, was whether α -21233 and α -21234 would recognise ANKRD18A as well as both isoforms of ANKRD18B since the peptide sequence of ANKRD18A that was used to raise α -21233 and α -21234 differed from the corresponding sequence in ANKRD18B by only one amino acid. Another concern was that the immunoblotting results obtained when using the commercial Abgent antibodies were not consistent with the results obtained from using α -21233 and α -21234. (Results from immunoblots using α -21233 and α -21234 did not differ so only the immunoblots using α -21234 will be included in this thesis.) When the Abgent antibody was used to probe for ANKRD18A in the WCLs from HEK293Ts and Damis, a band appeared at the correct molecular weight, ~ 116 kDa (Figure 4.5A). As expected, the Abgent antibody did not detect clone G2, the purified recombinant protein E3C Halo-ankyrin repeat domain (amino acids 1-260) because this peptide antibody recognises the

antigen sequence from residues 260 to 289 (lane 4, Figure 4.5A). When using the bespoke peptide antibodies α -21233 and α -21234 to probe for ANKRD18A in WCLs of HEK293T and Dami cells, a clear band was apparent in the region between the molecular weight markers of 135 and 190 kDa (Figure 4.5B), which is higher than the expected molecular weight for ANKRD18A. Because the amino acid sequences of the peptides used to raise α -21233 and α -21234 are virtually identical to the amino acid sequence of the two isoforms of ANKRD18B at the amino terminus, it is quite possible that α -21233 and α -21234 were recognising the long isoform of ANKRD18B (ANKRD18B_1401) in Figure 4.5B since the calculated molecular weight of this protein is ~160 kDa (<http://web.expasy.org/protparam/>). It did not appear that α -21233 and α -21234 were recognising the short isoform of ANKRD18B (ANKRD18B_1101, MW 118 kDa) as well.

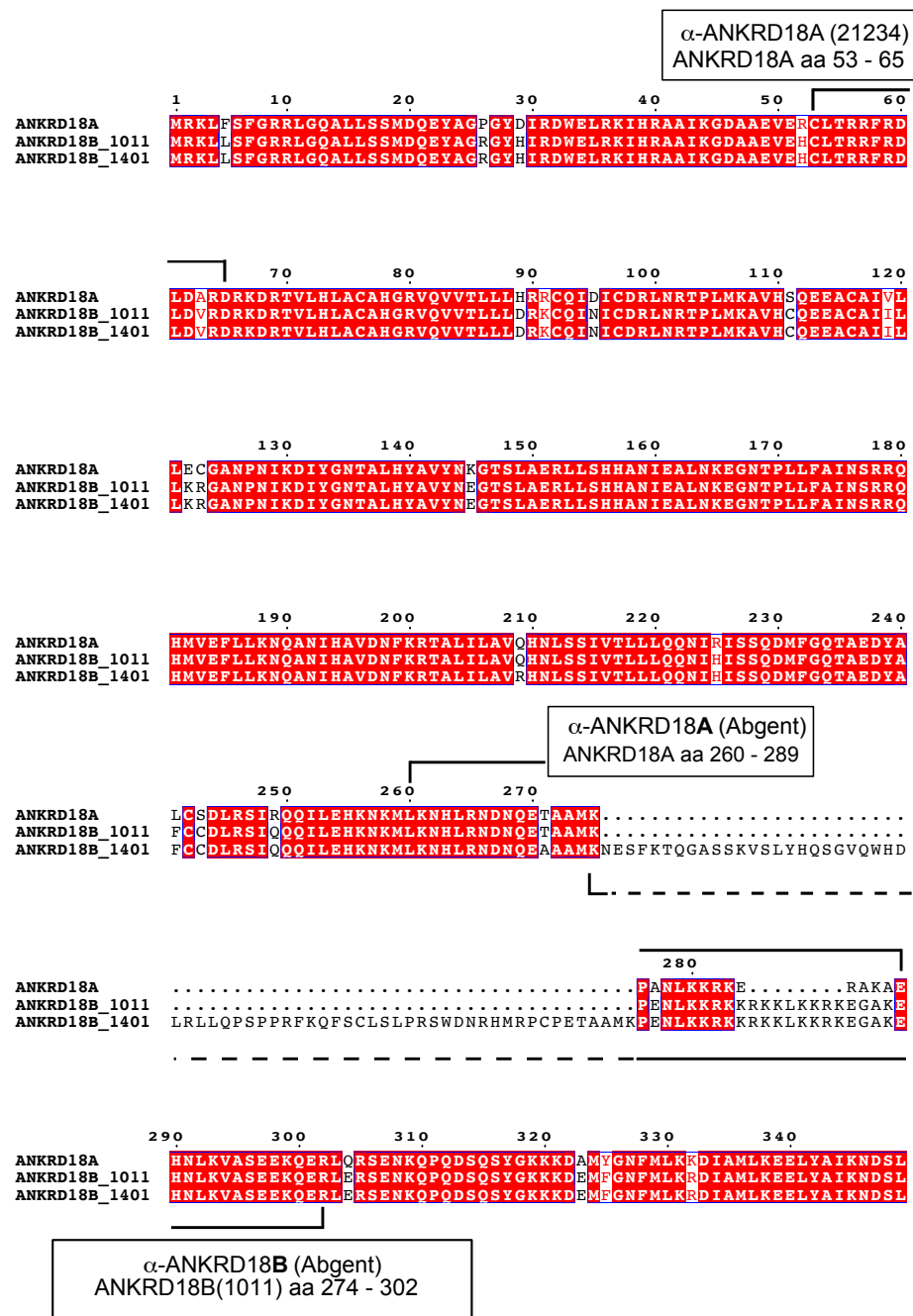


Figure 4.4 – Sequence alignment of ANKRD18A and ANKRD18B for residues 1-329 (of ANKRD18A amino acid sequence). ANKRD18B exists in two isoforms (1011 and 1401 amino acids, respectively). The peptide sequence used as antigens to raise polyclonal antibodies are highlighted by horizontal bars.

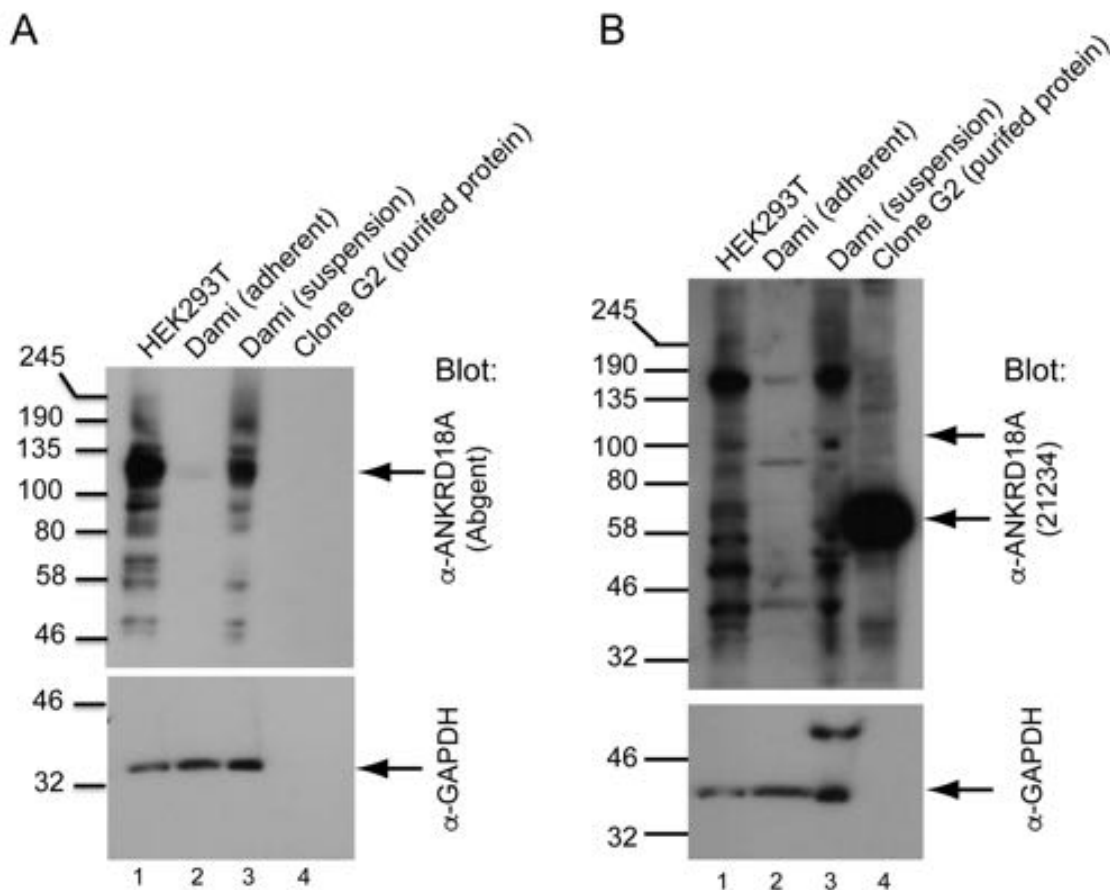


Figure 4.5 – Western blots demonstrating Abgent antibody detecting (A) ANKRD18A expression in HEK293T and Dami cells at correct molecular weight of 116 kDa whereas (B) α -21234 detects a dominant band at the incorrect molecular weight. WCLs were analysed by SDS-PAGE and immunoblotted with two different α -ANKRD18A, (A) Abgent antibody and (B) α -21234. Lane 1- HEK293T cells, lane 2-adherent Dami cells, lane 3- Dami cells in suspension, lane 4- purified recombinant protein, clone G2 ($M_r = 65$ kDa). Bottom panel: loading control, blotted with α -GAPDH ($M_r = 37$ kDa). The expected MW of ANKRD18A is 116kDa. Arrows indicate the expected positions of antigens (N=5).

When the Abgent ANKRD18A antibody was used to probe the WCLs from platelets of healthy donors, no ANKRD18A protein was detected (Figure 4.6A). However, when α -21233 and α -21234 were used to probe the WCLs of healthy donor platelets, a band appeared at the correct molecular weight of ~116 kDa (Figure 4.6B). The protein that α -21233 and α -21234 recognised is probably ANKRD18A since *ANKRD18B* mRNA is not present in platelets (Figure 4.1), according to my experimental data. And, as expected, α -21234 recognised clone G2 (lane 4, Figures 4.6B).

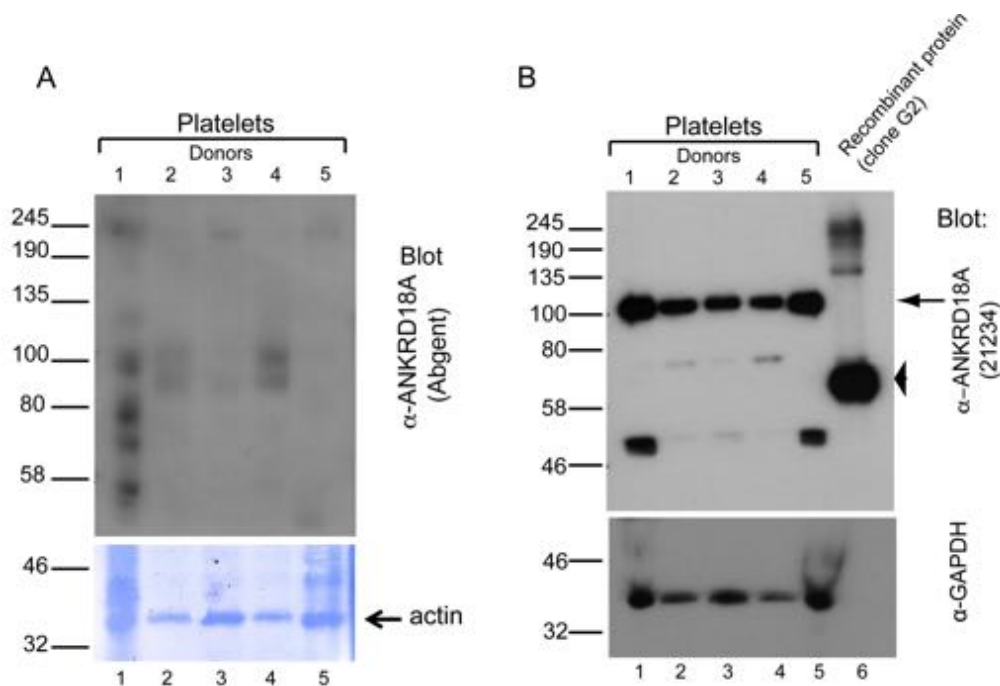


Figure 4.6 – Western blots demonstrating Abgent antibody detecting (A) no ANKRD18A expression in platelets whereas (B) α -21234 detects ANKRD18A at correct molecular weight. Platelet WCLs analysed by SDS-PAGE immunoblotted with (A) Abgent antibody and (B) α -21234. Lanes 1-5 different donor platelets, lane 6-purified recombinant protein, clone G2. Bottom panels, loading controls. Left bottom panel, Coomassie stained blot of above blot. Right bottom panel, immunoblot for GAPDH. Arrow in top right panel indicates ANKRD18A. Arrowhead in top right panel indicates clone G2. (N=5).

To confirm that in the above Western blots the protein that α -21234 recognised is ANKRD18A, one could silence ANKRD18A expression by RNAi whereby siRNA is transfected into HEK293T cells or Dami cells and WCLs from these cells are loaded onto SDS-PAGE and immunoblotted with α -21234 to see if the dominant band disappeared. The only concern is that in HEK293T and Dami cells, α -21234 recognised a protein whose molecular weight is between the molecular weight markers of 135-190 kDa. The long isoform of ANKRD18A is only ~117 kDa, so it remains unclear what protein is being recognised by α -21234 in these cells lysates. However, should this 160 kDa disappear after RNAi, one could assume that it is ANKRD18A. The next question would then be why and how, is ANKRD18A 160 kDa in these cells.

Although α -21234 recognised a protein with the correct molecular weight in platelets, because it is unlikely that platelets can be transfected, it would be difficult to perform RNAi experiments on platelets. Since platelets cannot be transfected, one would need to use megakaryoblastic cell lines such as MEG-01 or SET-02 cells to see if α -21234, when used to immunoblot for their WCLs, will identify a protein at the correct molecular weight of 116 kDa. If yes, one could then also perform RNAi to silence ANKRD18A expression and then immunoblot the WCLs of these cells using α -21234 again, thus proving α -21234's specificity for ANKRD18A.

4.4. *ANKRD18A* and *ANKRD18B* in megakaryocyte development

Since *ANKRD18A* mRNA could be detected in platelets from healthy donors, the question then arose regarding when *ANKRD18A* mRNA would be present during megakaryocyte development. Utilising quantitative real-time PCR (qRT-PCR) with *ANKRD18A* primers BR and CF (Table 2.6), it was demonstrated that *ANKRD18A* mRNA expression was present at day 0 of development and increased with progressive development of megakaryocytes (Figure 4.7). In contrast, no *ANKRD18B* mRNA was detected in megakaryocytes during megakaryocytic development, as probed by qRT-PCR (Figure 4.8).

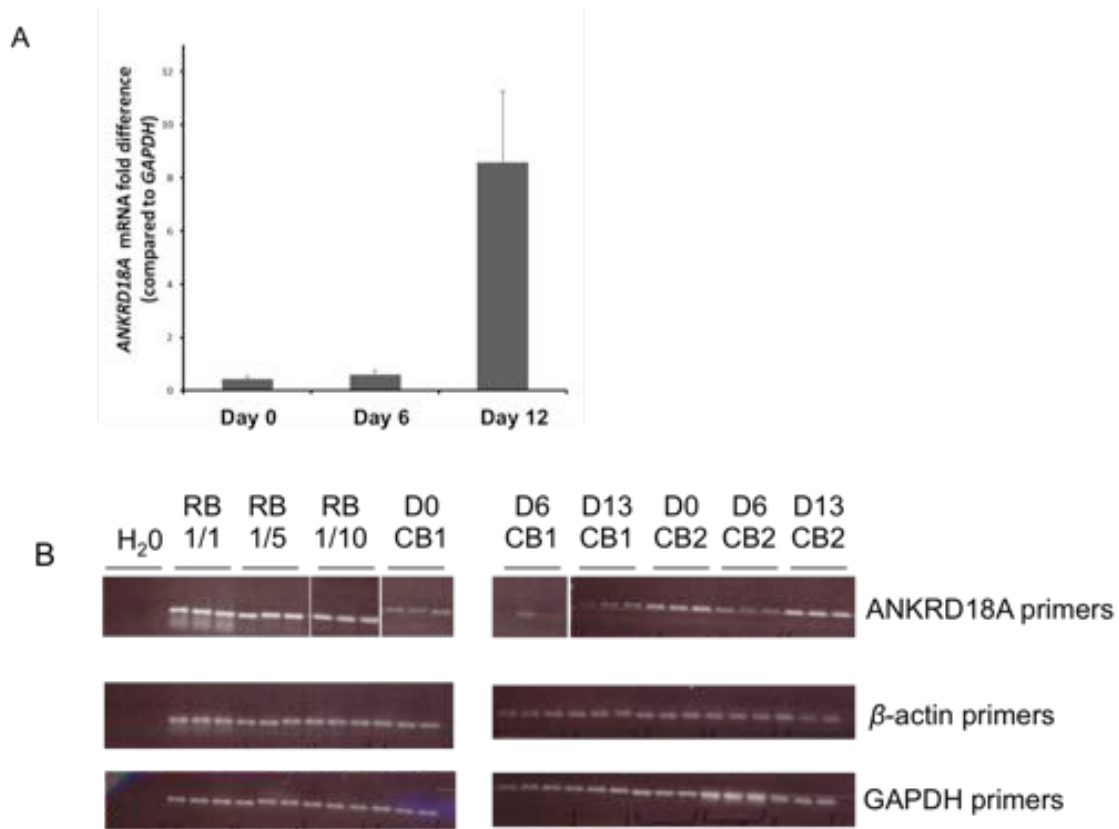


Figure 4.7 – ANKRD18A mRNA expression increases with megakaryocytic development as determined by qRT-PCR. (A) Determining the fold difference in ANKRD18A mRNA expression compared to the *GAPDH* reference during 3 stages of megakaryocyte differentiation. Samples are run in triplicate from 2 different cord blood (CB) preparations and retinoblastoma (RB) cell line as a control. (B) Agarose gel confirming expression of ANKRD18A mRNA from the samples that were run on the qRT-PCR plate. 1/5=1:5 dilution (N=3).

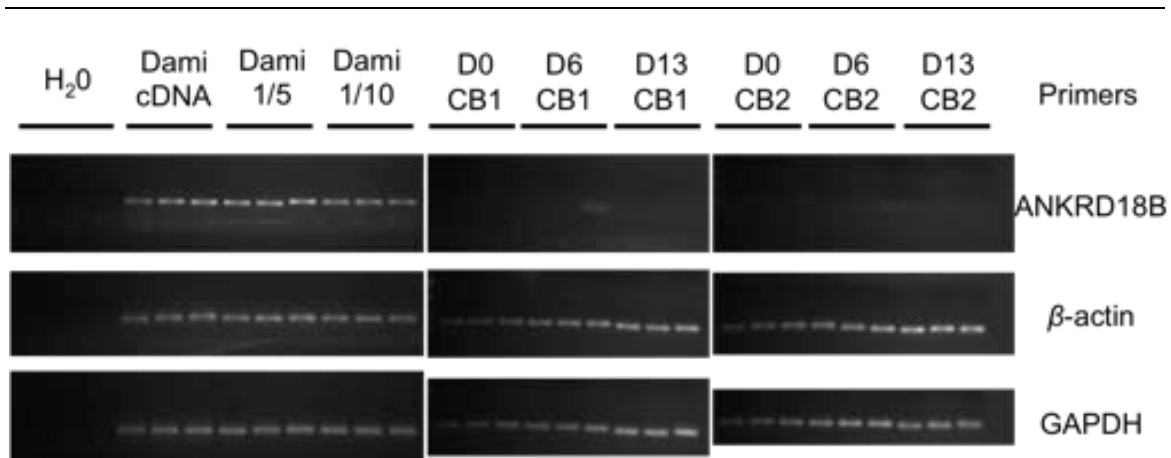


Figure 4.8 – No *ANKRD18B* mRNA expression in developing megakaryocytes. Dami cDNA was used as positive control. Dami 1/5 is cDNA at a dilution of 1:5, CB1 = cord blood sample 1 (N = 1).

4.5. Probing for ANKRD18A and ANKRD18B protein expression

Using a polyclonal peptide antibody, which was raised against a peptide of the short isoform of ANKRD18B (ANKRD18B_1101) (purchased from Abgent), expression of ANKRD18B in HEK293T, Dami cells and in platelets was determined by Western blotting. The sequence-derived molecular weight of ANKRD18B_1101 is about 118 kDa. Similar to what was seen using the Abgent α -ANKRD18A (Figures 4.5A), one saw expression of ANKRD18B in the WCLs of HEK293T and Dami cells (Figure 4.9A) at the molecular weight of 118 kDa. It is unclear why the Abgent α -ANKRD18B recognised only the short isoform of ANKRD18B and not the long isoform as well since the amino acid sequences of both isoforms are not very different.

Similar to the immunoblot probing for ANKRD18A expression in platelets using the Abgent α -ANKRD18A (Figure 4.6A), when using the Abgent α -ANKRD18B, there is no expression of ANKRD18B in platelets (Figures 4.9B). Therefore, the Western blots indicate clear expression of ANKRD18B in HEK293T and Dami cells, but no expression in platelets (Figures 4.9A and 4.9B). This correlates with the RT-PCR and qRT-PCR data above, which indicated no *ANKRD18B* mRNA expression in platelets (Figure 4.1) and megakaryocytes (Figure 4.8); but there is *ANKRD18B* mRNA expression in HEK293T and Dami cells (Figure 4.3).

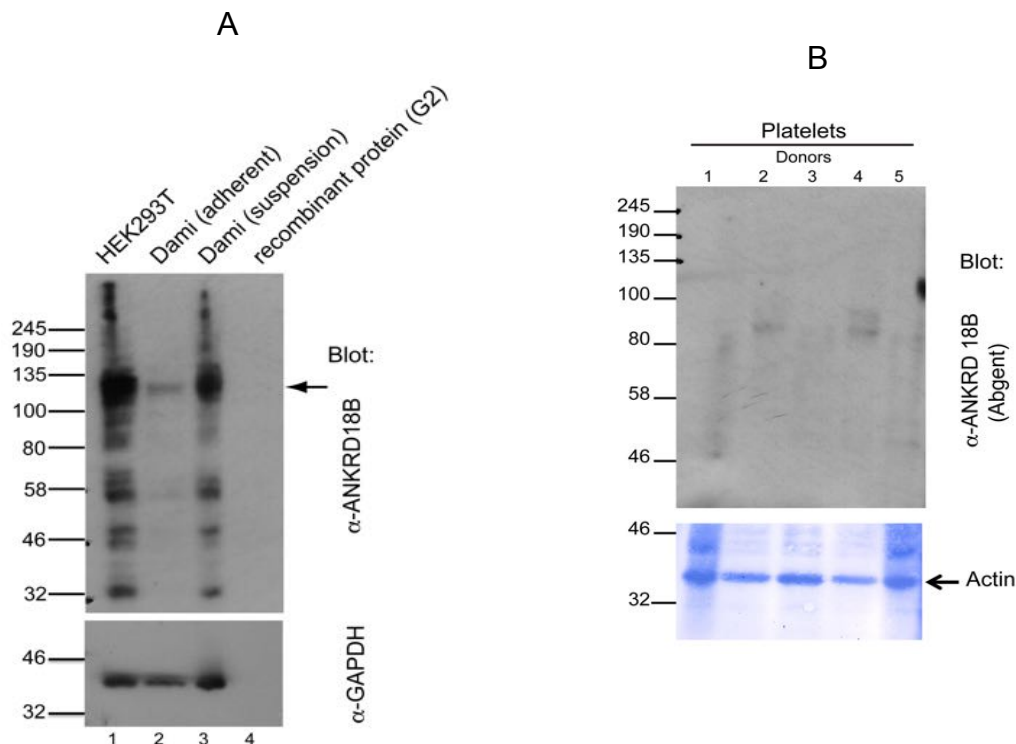


Figure 4.9 – Western blots demonstrating (A) ANKRD18B expression in HEK293T and Dami cells but (B) no ANKRD18B expression in platelets. HEK293T, Dami and platelet WCLs are immunoblotted with α -ANKRD18B (top panel) and for loading control, α -GAPDH (left bottom panel). Right bottom panel is Coomassie stain for actin, which is the loading control. (N=5)

4.6. Summary

In summary, the following observations have been made:

1. ANKRD18A and ANKRD18B are 93% identical at the amino acid sequence level.
2. Inconsistent results were obtained when immunoblotting with the Abgent α -ANKRD18A and α -21234 antibodies. When HEK293T and Dami WCLs were immunoblotted with the Abgent α -ANKRD18A (Figure 4.5A), the most prominent band that appeared was between the molecular weight markers of 100 to 135 kDa (~116 kDa). When using α -21234 (Figure 4.5B), the most prominent band was between the molecular weight markers of 135 and 190 kDa (~160 kDa). When immunoblotting platelet WCLs with the Abgent α -ANKRD18A (Figure 4.6A), no bands appeared. When α -21234 was used to immunoblot platelet lysates (Figure 4.6B), bands appeared at the correct molecular weight of 116 kDa.
3. *ANKRD18A* mRNA expression was present in platelets and all cell lines tested (Figure 4.2, Figure 4.3) whereas for *ANKRD18B* (Figure 4.1, Figure 4.3), no mRNA expression was detected in platelets and only in selected cell lines (HeLA, Dami, and select HEK293T derivatives).
4. The amount of *ANKRD18A* mRNA expression increased as megakaryocytes matured (Figure 4.7). No *ANKRD18B* mRNA expression was detected in megakaryocytes (Figure 4.8).

5. When immunoblotting HEK293T and Dami WCLs using the Abgent α -ANKRD18B (Figure 4.9A), the most prominent band was located between the molecular weight markers of 100 to 135 kDa (116 kDa). When immunoblotting platelet WCLs (Figure 4.9B), no bands appeared. This is consistent with the RT-PCR data, which suggested that there is no *ANKRD18B* mRNA in platelets.

Thus, because the amino acid sequences of ANKRD18A and ANKRD18B are so similar at the amino terminus (Figures 4.4 and S3), this has resulted in difficulty creating peptide-generated antibodies that are specific to each protein. The amino acid sequences of the peptides used to raise α -21233 and α -21234 are nearly identical to the amino acid sequence of the two isoforms of ANKRD18B at the N terminus - the ANKRD18A peptide sequence differs from ANKRD18B amino acid sequence by one amino acid, alanine to valine (Figure 4.4). It is quite possible that α -21233 and α -21234 were cross-reacting with ANKRD18B (ANKRD18B_1401). This would then explain the enhanced band in the Western blot of the Dami / HEK293T WCLs where there was a dominant band between the 135 and 190 kDa markers (Figure 4.5B). The long isoform of ANKRD18B has a sequence-derived Mr of ~160 kDa. It did not appear that α -21233 and α -21234 recognised the short isoform of ANKRD18B (ANKRD18B_1101, MW 118 kDa) as well.

Another example of the difficulty in choosing peptide sequences to generate antibodies that would discriminate between ANKRD18A and ANKRD18B concerned the Abgent purchased ANKRD18A and ANKRD18B_1101 polyclonal antibodies. Both of the Abgent ANKRD18A and ANKRD18B_1101 peptide sequences which were used to generate the respective polyclonal antibody are 30 amino acids long and they overlap, having a common stretch of 8 identical amino acids over 9 consecutive residues. It is difficult to know whether each antibody recognised the epitope that is unique to each protein or to an epitope that has an amino acid sequence shared by both proteins. Therefore, cross-reactivity could not be completely ruled out for these antibodies as well. This explanation is plausible given that the immunoblots of HEK293T and Dami WCLs (Figures 4.5A vs. 4.9A) and platelet WCLs (Figures 4.6A vs. 4.9B) using Abgent α -ANKRD18A and α -ANKRD18B looked almost identical.

However, it is still possible that α -21234, the bespoke peptide antibody, is specific for ANKRD18A. To prove this, RNAi experiments would need to be performed to silence ANKRD18A expression and then the antibody can be used to immunoblot WCLs of cells transfected with siRNA to determine if the expected band at the correct molecular weight has disappeared. When HEK293T and Dami cells are to be used, it would be interesting to see if the prominent band at ~160 kDa would disappear after RNAi. If it does disappear, the question would be why ANKRD18A has a molecular weight of ~160 kDa in HEK293T cells and

Damis. Since platelets cannot be transfected, RNAi experiments would need to be performed using cells most closely resembling platelets such as MEG-01 or SET-2 cells. Of course, the best cells to use would be megakaryocytes.

Hence, because of the necessity to ensure specificity and reduce the chances of cross-reactivity between ANKRD18A and ANKRD18B, what was required was an antibody, which would recognise the tertiary structure of ANKRD18A rather than a peptide antibody derived from ANKRD18A. However, should this polyclonal antibody which recognises the tertiary structure of ANKRD18A fail to be specific as well, the only way to produce a specific antibody is to generate a monoclonal antibody.

5. Generation of ANKRD18A recombinant protein and antibody production

5.1. Introduction

The mammalian adaptive immune system is able to respond to almost any type of molecular antigen. When raising antibodies against a protein target, one has the choice of using either a peptide or a folded protein as an immunogenic stimulant. Peptide antigens are typically generated through chemical synthesis, making them readily available. However, the corresponding peptide sequence may not be solvent-accessible within the context of the folded protein, limiting application of a corresponding antibody to immune recognition of the target if not in the denatured state. In contrast, using a folded protein as antigen will raise antibodies directed against solvent-accessible epitopes. For a folded protein to be used as an antigen, it needs to be made through overexpression of the corresponding gene in a suitable expression host, and purified from cell extracts without denaturation. When a full-length protein antigen is not amenable to recombinant expression, one can consider expressing one or several of its structural domains. Thus, a 3D epitope (rather than a linear peptide sequence) reduces the risk of cross-reactivity.

Given the inconsistent results obtained when immunoblotting with the peptide-derived antibodies described in Chapter 4, the need for an antibody with better

selectivity was evident. Raising such an antibody against a folded protein antigen rather than a short peptide would be expected to aid selectivity for ANKRD18A. However, such an antibody is not completely fail-proof. To definitely attain an antibody which would be specific to ANKRD18A would require generating a monoclonal antibody.

5.2. Generation of ANKRD18A recombinant protein

The amount of antigen required for antibody generation is in the order of 1 mg pure protein at a concentration of 1 mg/ml or greater. Initially, it was intended to generate an antibody raised to the tertiary structure of full-length ANKRD18A protein. However, because of the difficulty in transfecting cells with enough plasmids encoding full-length ANKRD18A so that expression could be visualised on a Western blot, it was finally decided to generate recombinant protein of ANKRD18A fragments, whereby the fragment encompassed the functional domains of ANKRD18A.

5.2.1. Establishment and selection of ANKRD18A constructs to provide suitable expression levels

Against the backdrop of poor expression of full-length ANKRD18A in mammalian cells, it seemed more promising to attempt expression of the structural domains of ANKRD18A (Table 5.1 and Figure 1.16). These structural domains could then be used for immunisation and, by adding suitable affinity tags in expression plasmids, efficient purification of the recombinant protein fragments from cell extracts would be facilitated. In order to rationalise the choice of domain

boundaries, the sequence of ANKRD18A was submitted to fold prediction servers (Table 5.1).

Species Fragment	Homo sapiens AAI52435
Full length	1 – 992
Ankyrin homology (Phyre2)	35 – 235
Ankyrin homology (HHpred)	1 – 260
Tropomyosin homology	450 – 790
Colicin 1A homology	790 – 992

Table 5.1 – Functional domains of ANKRD18A. Phyre2 = protein fold prediction server Phyre2 (www.sbg.bio.ic.ac.uk/phyre2); HHpred = homology detection and structure prediction by HMM-HMM comparison (toolkit.tuebingen.mpg.de/hhpred)

There are two isoforms of ANKRD18A. The short isoform has been annotated to comprise nine ankyrin repeat motifs whereas the long isoform is comprised of 11 ankyrin repeat domains at the amino terminus and both have a sequence of coiled–coil regions in the carboxy terminal region after the ankyrin repeat domain (Figure 1.16). Given that the primary amino acid sequence of a protein determines its three dimensional structure, “threading” the sequence of ANKRD18A onto known protein structures is a powerful method to predict how a given sequence will fold (provided the database contains a structural homologue). When there is a minimum of steric clashes this indicates which particular structural template is compatible with the sequence of the target,

ANKRD18A. The fold prediction (using the Phyre2 and HHpred servers) confirmed with high confidence the amino terminal ankyrin repeat domain and detected two further domains downstream (Table 5.1): domains with homology to tropomyosin and colicin IA. This information was used to design expression plasmids encompassing the predicted structural domains of ANKRD18A. Because it was difficult to predict which functional domain will express at sufficient levels, it was necessary to make multiple DNA plasmid constructs, pairing various affinity tags with various domains. Pairing the various domains with a variety of fusion partners were needed to determine which tag when fused with a particular protein, would assist the protein to fold properly as well as be soluble in certain buffers. To expedite the cloning of these multiple constructs, the facilities of OPPF (Oxford Protein Production Facility at Harwell) which has established high-throughput procedures to allow parallel cloning of a given gene sequence into a collection of plasmid DNA vectors, were used. A total of 32 constructs with various tags fused to the four different functional domains of ANKRD18A are detailed in Table 5.2. All constructs also contained the hexahistidine affinity tag.

5.2.2. Testing expression of ANKRD18A constructs

These 32 plasmids were first tested for expression in *E. coli* under auto-inducing conditions or induced by IPTG (Figures 5.1, 5.2, 5.3). *E. coli* competent cell strains B834 (a methionine auxotroph that allows high specific activity labelling of target proteins with ³⁵S-methionine or selenomethionine for crystallographic

studies, <https://us.vwr.com>) and Rosetta (designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*, www.emdmillipore.com) were transformed with the various plasmids and then protein was purified from the cell lysates over a Ni⁺²-NTA matrix. Eluates from the Ni⁺²-NTA chromatography were analysed on a Coomassie stained SDS-PAGE gel. None of the plasmids expressed the correct protein, as determined by molecular weight: no bands appeared at the expected molecular weight except for GFP. GFP served as the positive control. When the different plasmids were transfected into the B834 strain of *E. coli*, the plasmids which expressed protein were clones F2, G2, D3 and H3. However, as shown in the Coomassie gels (Figures 5.1 and 5.2), except for clone F2, only the tag was expressed, as evidenced by the molecular weights (G2, 35.3 kDa; D3, 35.3 kDa; and H3, 13.2 kDa). For clone F2, it appears that both the full construct as well as the tag alone, were expressed (Full construct, 73 kDa; tag, 43 kDa). When the different plasmids were expressed in the Rosetta strain of *E. coli*, again, of the clones which expressed, clones B2, F2, G2 and D3, what appears to be expressed were just the tags without the functional domains (Figures 5.2 and 5.3). The most likely reason why this occurred is that the DNA mixture, which was used for the transfection, contained both vector alone (unligated DNA) and ligated DNA. Since transfection into *E. coli* is easier with smaller constructs, what was seen here were transfection and expression of just the plasmids of vector alone, which has the DNA sequences of the tags.

Clone	OPTIC	Gene name	aa_N	aa_C	OPPF No.	Vector	Tag	MW of Tag (kDa)	Domains	Domain MW
A1	13915	ANKRD-FL	2	992	14229	pOPINE	(His) ₆	Nil	Full length	116 kD
B1	13915	ANKRD-FL	2	992	14230	pOPINF	3CProt+His ₆	2.2		
C1	13915	ANKRD-FL	2	992	14231	S3C	Sumo-3C	13.2		
D1	13915	ANKRD-FL	2	992	14232	pOPINM	MBP	43		
E1	13915	ANKRD-FL	2	992	14233	E-3C-HALO	E3CHalo	35.3		
F1	13915	ANKRD-35	35	235	14234	pOPINE	(His) ₆	Nil	ANK repeat (Phyre2)	23 kD
G1	13915	ANKRD-35	35	235	14235	pOPINF	3CProt+His ₆	2.2		
H1	13915	ANKRD-35	35	235	14236	S3C	Sumo-3C	13.2		
A2	13915	ANKRD-35	35	235	14237	pOPINM	MBP	43		
B2	13915	ANKRD-35	35	235	14238	E-3C-HALO	E3CHalo	35.3		
C2	13915	ANKRD-1	2	260	14239	pOPINE	(His) ₆	Nil	ANK repeat (HHpred)	30 kD
D2	13915	ANKRD-1	2	260	14240	pOPINF	3CProt+His ₆	2.2		
E2	13915	ANKRD-1	2	260	14241	S3C	Sumo-3C	13.2		
F2	13915	ANKRD-1	2	260	14242	pOPINM	MBP	43		
G2	13915	ANKRD-1	2	260	14243	E-3C-HALO	E3CHalo	35.3		
H2	13915	ANKRD-450	450	790	14249	pOPINJ	GST	30	Tropomyosin	40 kD
A3	13915	ANKRD-450	450	790	14250	pOPINFS	Strep II	1.1		
B3	13915	ANKRD-450	450	790	14246	S3C	Sumo-3C	13.2		
C3	13915	ANKRD-450	450	790	14247	pOPINM	MBP	43		
D3	13915	ANKRD-450	450	790	14248	E-3C-HALO	E3CHalo	35.3		
E3	13915	ANKRD-450	450	790	14244	pOPINE	(His) ₆	Nil		
F3	13915	ANKRD-790	790	992	14251	pOPINJ	GST	30	Colicin 1A	24 kD
G3	13915	ANKRD-790	790	992	14252	pOPINFS	Strep II	1.1		
H3	13915	ANKRD-790	790	992	14253	pOPINS3C	Sumo-3C	13.2		
A4	13915	ANKRD-790	790	992	14254	pOPINM	MBP	43		
B4	13915	ANKRD-790	790	992	14255	pOPINE	(His) ₆	Nil		
C4	13915	ANKRD-790	790	992	14256	E-3C-HALO	E3CHalo	35.3		
D4	13915	ANKRD-450a	450	992	14257	pOPINJ	GST	30	Tropomyosin + Colicin 1A	64 kD
E4	13915	ANKRD-450a	450	992	14258	pOPINFS	Strep II	1.1		
F4	13915	ANKRD-450a	450	992	14259	pOPINM	MBP	43		
G4	13915	ANKRD-450a	450	992	14260	pOPINS3C	Sumo-3C	13.2		
H4	13915	ANKRD-450a	450	992	14261	E-3C-HALO	E3CHalo	35.3		

Table 5.2 - DNA plasmid constructs generated at OPPF. Abbreviations used: 3C = cleavage site for human rhinovirus 3C protease; SUMO = small ubiquitin-like modifier; MBP = maltose binding protein; GST = glutathione-S-transferase; HALO = Promega HaloTag.

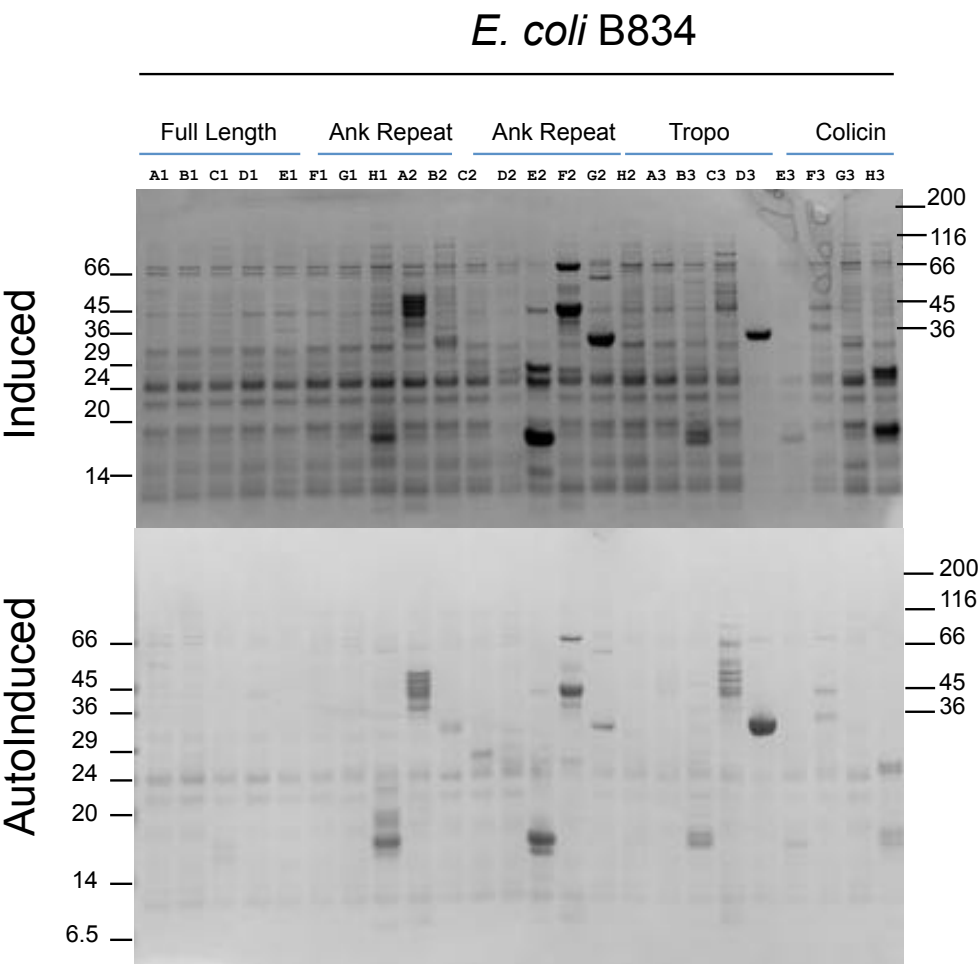


Figure 5.1 - Coomassie-stained SDS-PAGE gel of lysed *E. coli* (B834) cell cultures expressing plasmids encoding ANKRD18A or fragments thereof. Molecular weight markers are in units of kDa. For clone designations, refer to Table 5.2. (N=1)

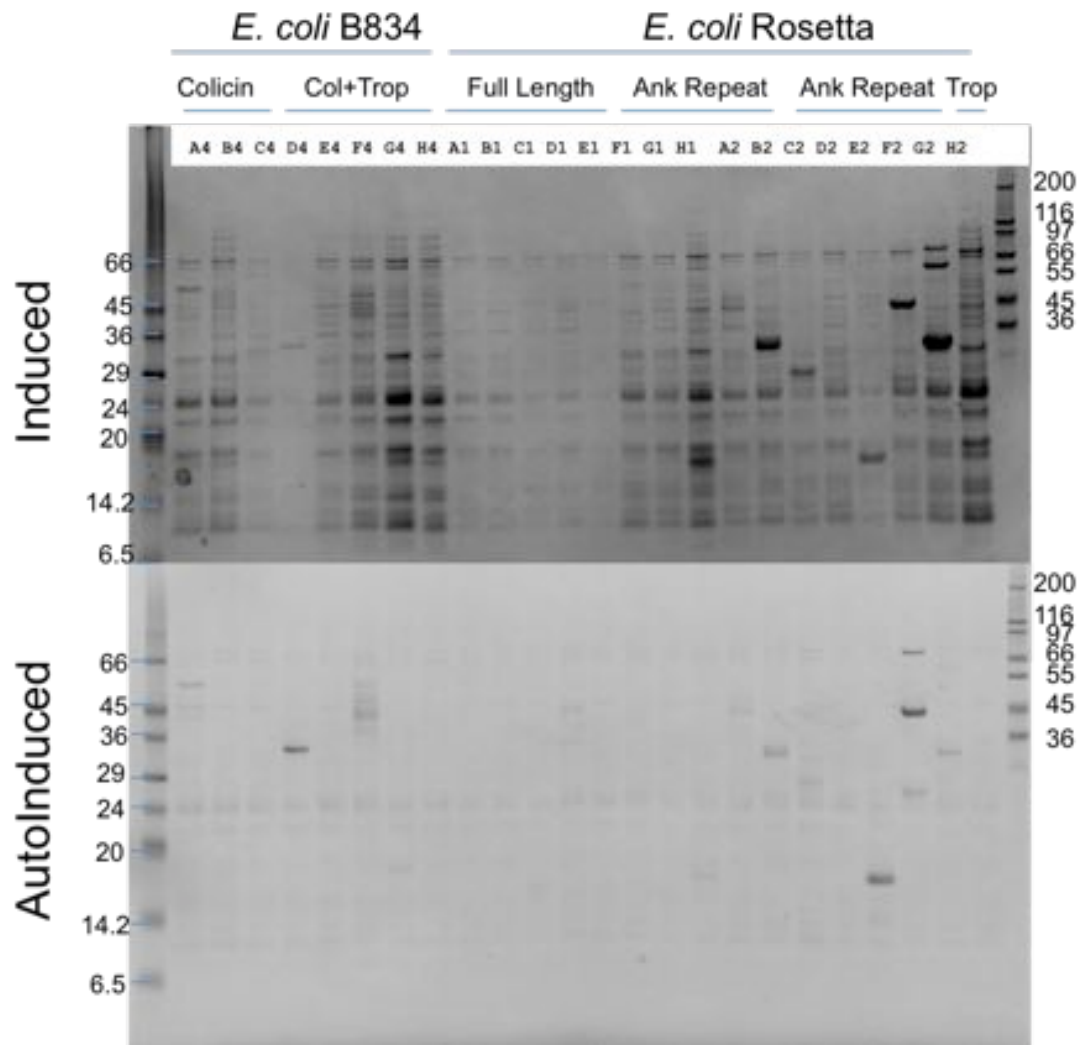


Figure 5.2 - Coomassie-stained SDS-PAGE gel of lysed *E. coli* (B834 and Rosetta) cell cultures expressing plasmids encoding ANKRD18A or fragments thereof. Molecular weight markers are in units of kDa. For clone designations, refer to Table 5.2. (N=1)

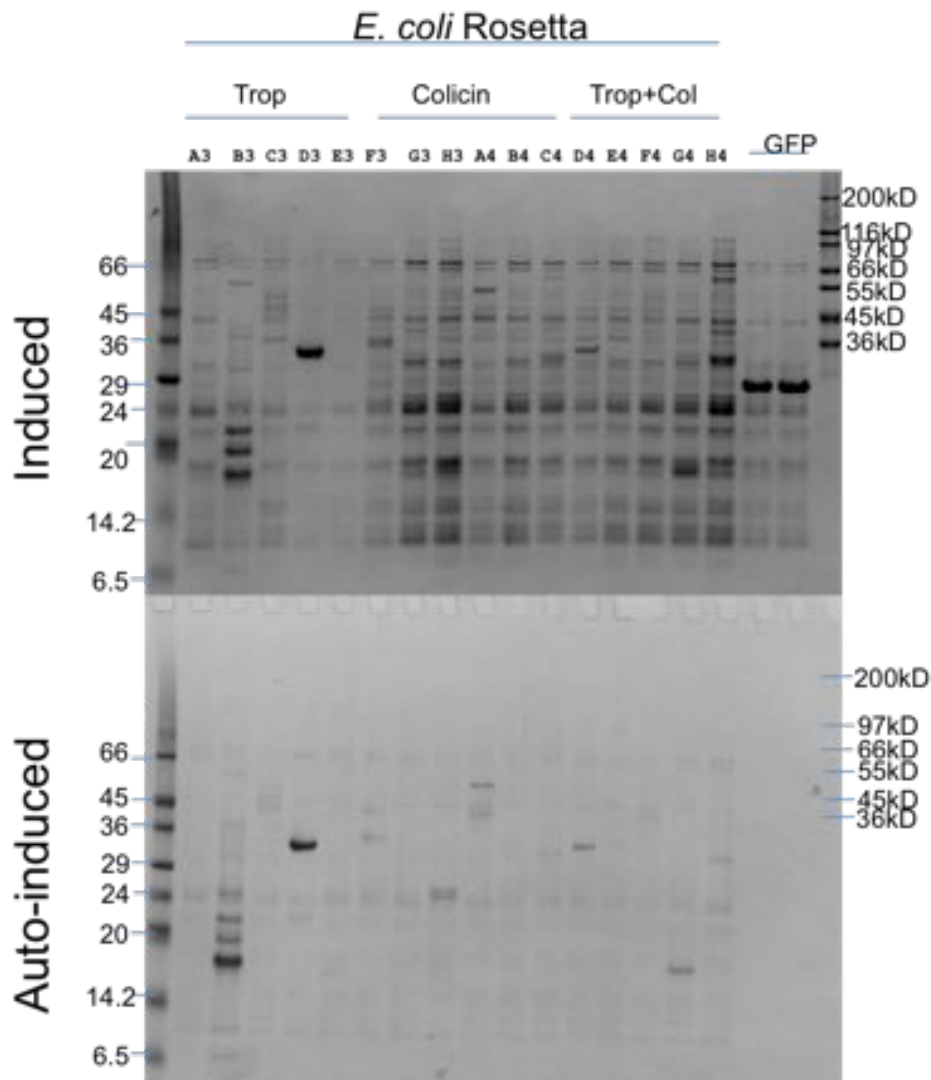


Figure 5.3 - Coomassie-stained SDS-PAGE gel of lysed *E. coli* (Rosetta) cell cultures expressing plasmids encoding ANKRD18A (or fragments thereof). Molecular weight markers are in units of 10^3 Dalton. For clone designations, refer to Table 5.2. (N=1)

Because the 32 plasmids overall did not express in bacteria, the same plasmids were transiently transfected into HEK293T cells using a proprietary transfectant, GeneJuice. WCLs from each transfection were analysed by immunoblotting SDS-PAGE gels with a monoclonal anti-His₆ antibody (Figures 5.4 and 5.5). Clones that expressed were H1 (Ankyrin repeat, 23 kDa + Sumo3C, 13.2 kDa; total, 36.2 kDa), B2 (Ankyrin repeat, 23 kDa + E3C Halo, 35.3 kDa; total, 58.3 kDa), E2 (Ankyrin repeat, 30 kDa + Sumo3C, 13.2 kDa; total, 43.2 kDa), F2 (Ankyrin repeat, 30 kDa + MBP, 43 kDa; total, 73 kDa), and G2 (Ankyrin repeat, 30 kDa + E3CHalo, 35.3 kDa; total, 65.3 kDa). Clone E2 appeared to be the highest expressing clone but may have suffered from proteolytic degradation: in addition to the expected band at ~ 43 kDa, there were two lower molecular weight bands of which the Mr corresponds approximately to the Sumo tag and the ankyrin repeat domain, respectively. Clone F2 (MBP tag + ankyrin repeat domain) also revealed two bands, the heavier one corresponding to the Mr of the fusion protein, and the lighter matching the Mr of MBP (Figure 5.4). After Sanger sequencing clone E2, it was determined that it was missing a large portion of its 5' end so the fact that its expression in HEK293T cells leads to multiple bands is irrelevant. In regards to clone F2, the lower molecular weight band corresponds to MBP, which is what one would expect if the empty vector plasmid were transfected into the HEK293T cells. The DNA mixture used to transfect HEK293T probably contained a mixture of ligated and unligated DNA. The empty vector does contain the His tag sequence and that is why the Western blot

when blotted with anti-His₆ displays bands corresponding to molecular weights of the fusion protein (73 kDa) and MBP alone (43 kDa).

In Figure 5.5, the clones that expressed encompassed B3 (Tropomyosin, 40 kDa + Sumo3C, 13.2 kDa; total, 43.2 kDa), C3 (Tropomyosin, 40 kDa + MBP, 43 kDa; total, 83 kDa), E3 (Tropomyosin, 40 kDa + His tag, 2.2 kDa; total, 42.2 kDa), and A4 (Colicin1A, 24 kDa + MBP, 43 kDa; total, 67 kDa). The positive control was His-tagged GFP.

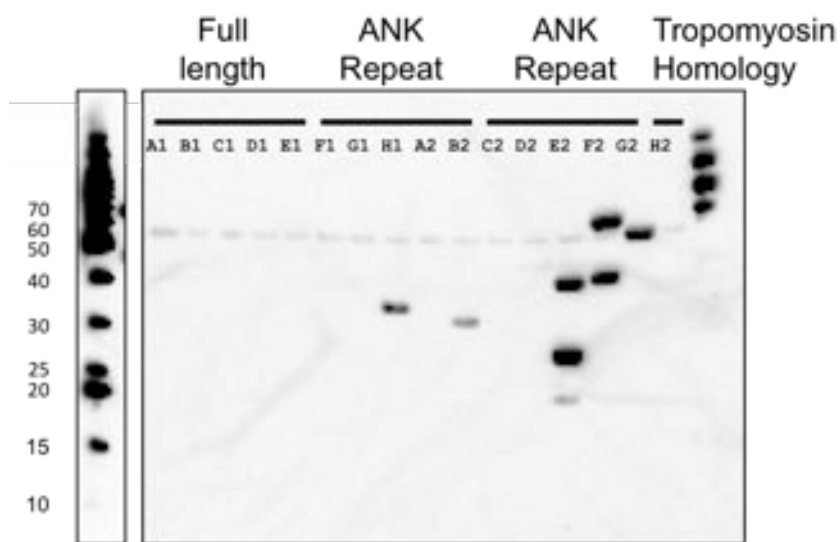


Figure 5.4 – Western blot analysis of whole cell lysates from HEK239T cells transiently transfected with expression plasmids encoding ANKRD18A (and fragments thereof) and immunoblotted with anti-His₆. Refer to Table 5.2 for clone designations. Molecular weight markers are given in units of kDa. (N=1)

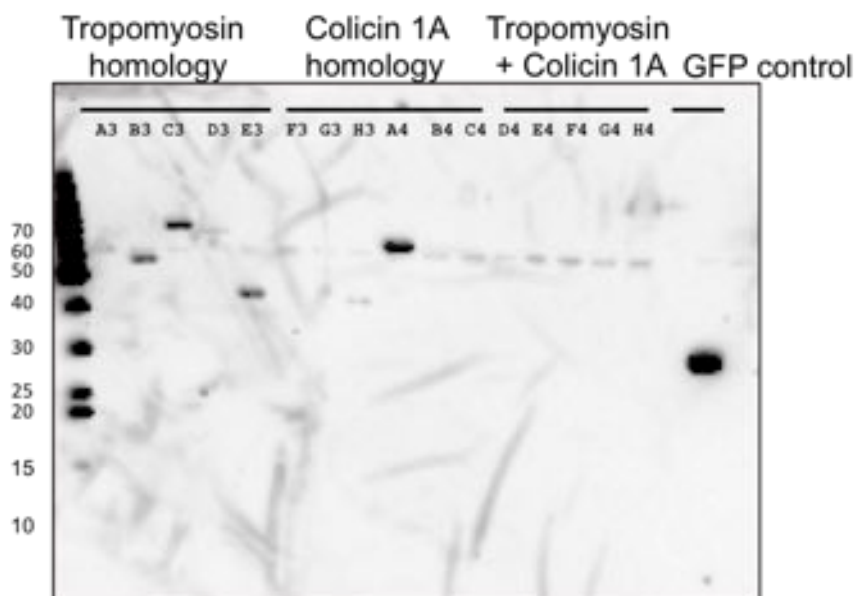


Figure 5.5 - Anti-His₆ blot of whole cell lysates from HEK239T cells transiently transfected with expression plasmids encoding ANKRD18A (and fragments thereof). Refer to Table 5.2 for clone designations. Molecular weight markers are given in units of kDa. (N=1)

These nine plasmids, which exhibited promising expression levels, were sequenced by Sanger sequencing. Results from sequencing revealed that clone E2 was missing a large portion of its DNA sequence at the 5' end and therefore this clone was discarded.

5.2.3. Testing of transfection agents

The preliminary expression experiments at OPPF used GeneJuice, a proprietary transfection agent while subsequent experiments performed in Birmingham used PEI. Given that mg amounts of purified protein were needed for immunisation relatively inexpensive PEI was preferred as the transfecting agent. To confirm

that plasmid constructs tested at OPPF (Figures 5.4 and 5.5) also expressed when using PEI as a transfectant, HEK293T cells were transiently transfected with 2 μ g of each of the eight plasmids which had expressed protein at OPPF (Clones as designated in Figure 5.6 and Table 5.3). HEK293T cell lysates were analysed on a SDS-PAGE gel and immunoblotted with a monoclonal His₆ antibody. All eight clones exhibited protein expression, with clones G2 and F2 having the highest amount of expression (Figure 5.6). It should be noted that clone G2 contains a second band, which has a molecular weight of its tag, E3C Halo (~35.3 kDa).

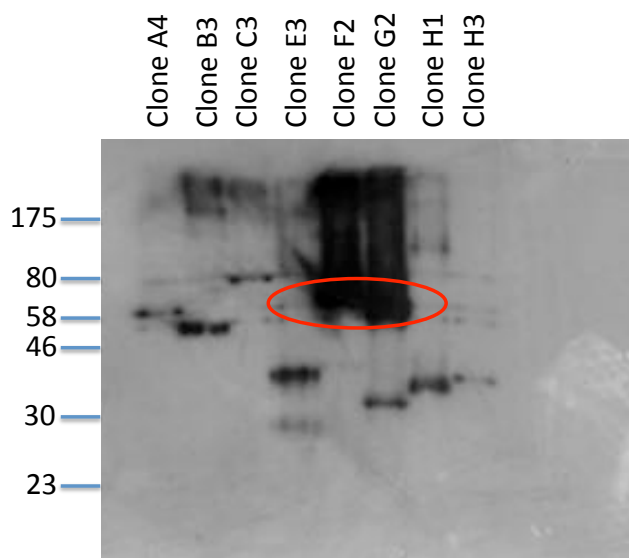


Figure 5.6 – Anti-His₆ immunoblot of HEK293T cells transiently transfected with plasmids encoding ANKRD18A fragments and its respective tags using PEI. Refer to Table 5.2 for clone designations and sizes. The highest expression clones (F2 and G2) are encircled in red. (N=2)

Clone	Functional Domain	Molecular Weight (kDa)	Residue range	Purification tag	Total molecular weight (kDa)
H1	Ankyrin Repeat(Phyre2)	23	35-235	His ₆ -Sumo3C	36.2
B3	Tropomyosin	40	450-790	His ₆ -Sumo3C	53.2
C3	Tropomyosin	40	450-790	His ₆ -MBP	83
E3	Tropomyosin	40	450-790	His ₆	40
H3	Colicin 1A	24	790-992	His ₆ -Sumo3C	37.2
A4	Colicin 1A	24	790-992	His ₆ -MBP	67

Table 5.3 – Summary of ANKRD18A recombinant proteins with detectable expression when transiently transfected into HEK293T cells. The clones highlighted in green were chosen for purification. MBP = maltose binding protein, Sumo = small ubiquitin-like modifier, Halo = modified haloalkane dehalogenase; all constructs, except for E3, include a cleavage site for human rhinovirus 3C protease.

5.2.4. Purification of selected ANKRD18A recombinant proteins

While the constructs generated at OPPF included different fusion partners (i.e. MBP or HALO), they all coded for a hexa-histidine tag. Making use of this His-tag, recombinant proteins were extracted from cleared cell lysates by Nickel - Nitrilotriacetic (Ni⁺²-NTA) column chromatography, often an effective method to obtain purified protein.

Because of the need to maximise protein production, clone G2 was initially chosen for purification since it was one of the highest expressing clones. However, while purification over a Ni⁺²-NTA column was effective, the eluted

protein was subsequently not stable, and frequently precipitated during dialysis. In contrast, clone F2 proved more stable in purified form.

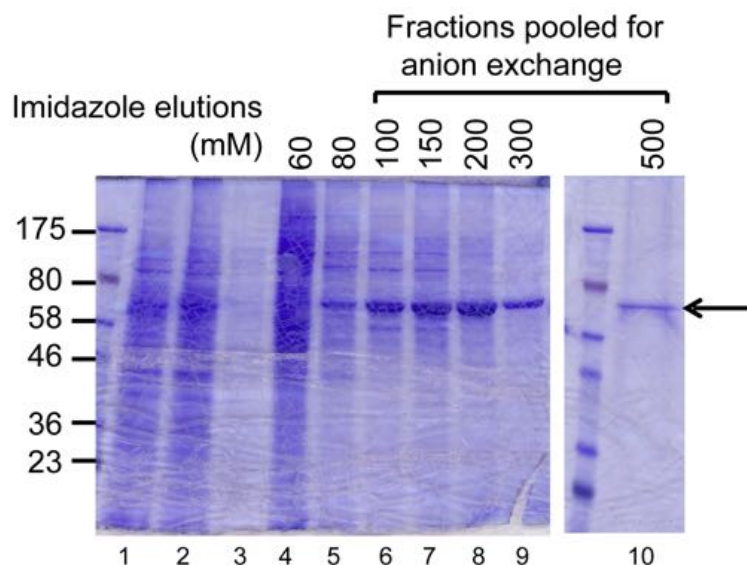


Figure 5.7 – Coomassie stained SDS-PAGE gel demonstrating purification of clone F2 using Ni^{+2} -NTA chromatography. Elution fractions using 100-500mM imidazole were collected for dialysis. Arrow points to recombinant protein, MBP-ankyrin repeat protein, clone F2. Molecular markers on the left are in units of kDa. (N=10)

Harvesting cell monolayers from ten 15 cm tissue culture plates, cells which were transiently transfected with the DNA of clone F2, were lysed by applying detergent-containing buffer and the cleared lysate was passed twice over a Ni^{+2} -NTA resin, followed by elution with an imidazole step gradient (details in section 2.7.1 of Chapter 2). Fractions eluted from using imidazole concentrations of 100-500 mM imidazole (Figure 5.7) were then pooled and dialyzed thrice against a

buffer containing 20mM Tris, 150mM NaCl, pH 7.8. Following dialysis, the dialysate was loaded onto an anion exchange column to obtain single band pure proteins.

Ion exchange chromatography is based on the principle of proteins binding through ionic interaction to the column resin. The protein can be eluted in a differential fashion either through gradually increasing the ionic strength of the buffer (thereby reducing the affinity between protein and resin) or by altering the pH (thereby changing the overall charge of the protein, and thus affecting its affinity to the resin). The choice of resin (positively vs negatively charged) is determined by the isoelectric point, pI (the pH at which the overall charge of a protein is zero) of the protein to be purified, and the availability of a buffer with a pKa value sufficiently far away from the pI (at least 1 pH unit above the pI of substance to be bound to anion exchange column) so as to maximise solubility of the protein (Duong-Ly and Gabelli 2014). Therefore, because the pI of clone F2 is 6.64 and that of clone G2 is 6.34, the dialysis buffer which was chosen after elution from the nickel column was 20 mM Tris, 150 mM NaCl, pH 7.8 and 20 mM Tris, 200 mM NaCl, pH 7.8, respectively, because the pKa (25°C) of Tris is 8.07. The NaCl concentration needed to be more concentrated than usual (normally ~50mM NaCl) to prevent precipitation of the recombinant proteins in the dialysis buffers. Fortunately, this elevated concentration did not prevent the proteins from binding the resin of the anion exchange column.

Ultimately, clone F2 was further purified using anion exchange chromatography, whereby increasing concentrations of sodium chloride were used to elute the recombinant protein (Figure 5.8). Through this two step purification, single band pure fusion protein was obtained, with a total amount of ~2.2 mg at a concentration of 2.16 mg/ml (Figure 5.13).

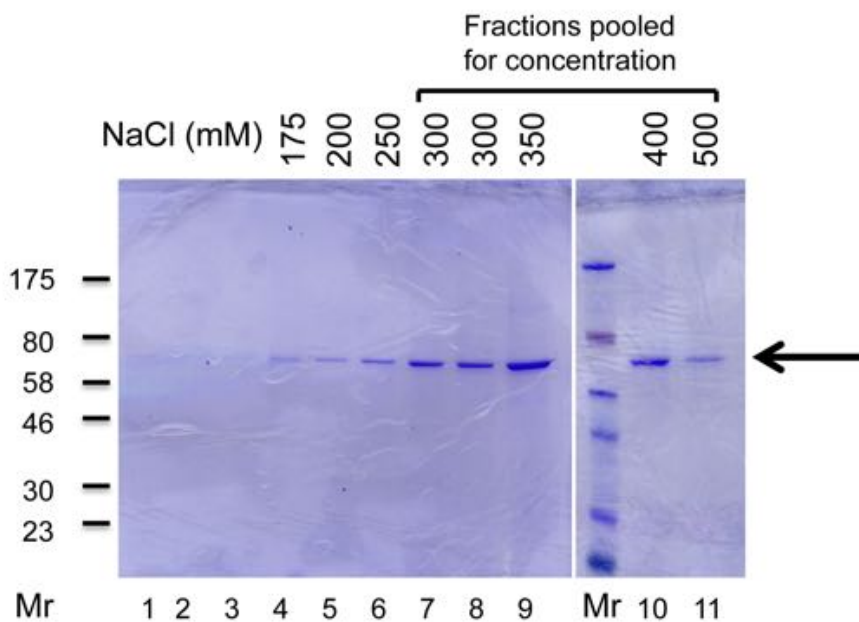


Figure 5.8 – Coomassie stained SDS-PAGE gel demonstrating purification of clone F2 using anion exchange chromatography. Fractions eluted using increasing concentrations of NaCl (250-500mM) are pooled for concentration. The arrow points to purified clone F2. (N=10)

5.2.5. Proteolytic cleavage of the MBP tag from the ANKRD18A recombinant protein (clone F2).

When immunising an animal to raise antibodies, using a fusion protein can generate antibodies specific to either fusion partner (purification tag as well as the target protein). The expression constructs had been designed such that they included a proteolytic cleavage site to cleave the fusion protein using recombinant human rhinovirus 3C (HRV3C) protease. Therefore, following concentration of the pooled purified fractions of clone F2, the fusion protein was incubated with HRV3C protease and protease activity was assessed by analysing the reaction mixtures on SDS-PAGE gels. Starting with manufacturer recommended ratios of protease to enzyme, next to no proteolytic digest was detected. Increasing the enzyme concentrations to 4 units for every 10 µg led to discernible cleavage of the fusion protein (Figure 5.9). Also, overnight incubation at 4 °C proved superior to room temperature for 1 – 6 hours.

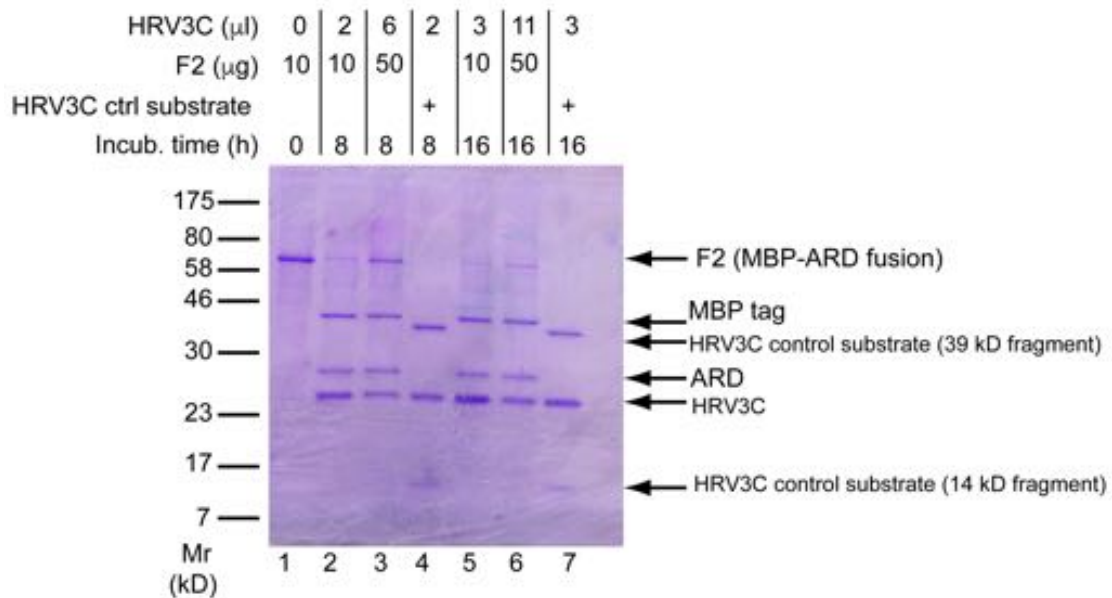


Figure 5.9 – Removing MBP tag from recombinant protein MBP-ANKRD18A (clone F2) with HRV3C protease. Lane 1- Uncut F2, lane 2-10 μg F2, 2 μl HRV3C, 8 hours incubation at 4°C, lane 3- 50 μg F2, 6 μl HRV3C, 8 hours at 4°C, lane 4- Control protein, 2 μl HRV3C, 8 hours at 4°C, lane 5- 10 μg F2, 3 μl HRV3C, overnight incubation at 4°C, lane 6- 50 μg F2, 11 μl HRV3C, overnight incubation at 4°C, lane 7- Control, 3 μl HRV3C, overnight incubation at 4°C. Each lane contains 10 μl of sample plus 30 μl of Laemmli sample buffer. (N=5)

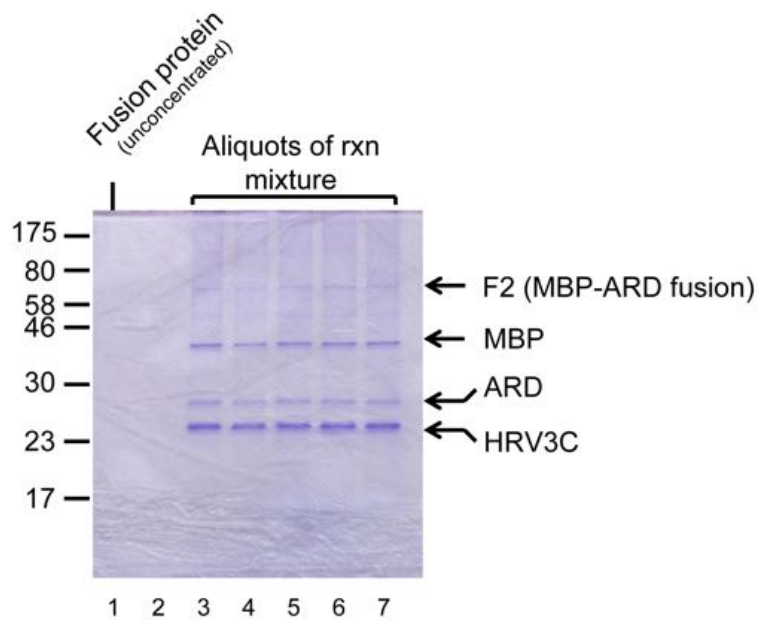


Figure 5.10 – Cleaving MBP tag from clone F2 (recombinant protein, MBP-ankyrin repeat domain). Each reaction mixture contains 200 μ g of F2 with 40 μ l (80 units) of HRV3C protease. Lane 1- unconcentrated F2, lane 2- filtrate from concentration of F2, lanes 3-7 cleaved F2. Each lane contains 10 μ l of sample and 30 μ l of Laemmli sample buffer. (N=1)

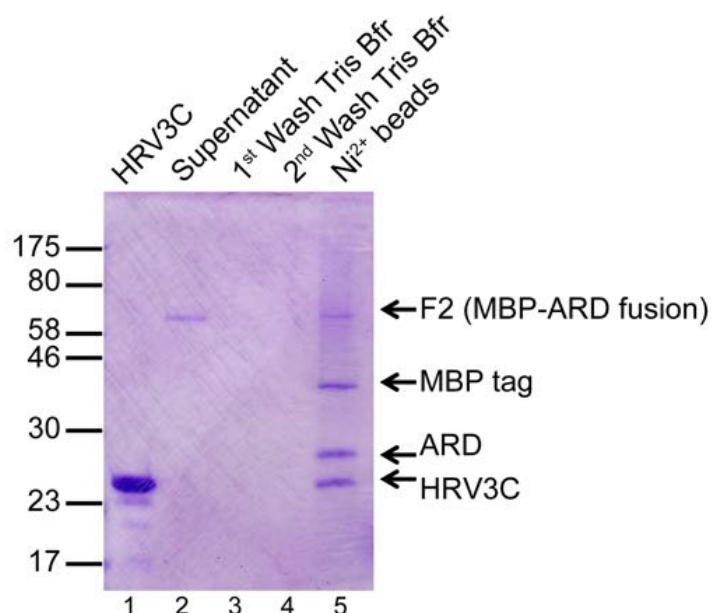


Figure 5.11 – Binding MBP tag to Ni^{+2} -NTA beads by IMAC. Clone F2, 50 μg digested with 11 μl of HRV3C protease added to 200 μl Ni^{+2} -NTA beads in suspension, rotated overnight at 4°C . Lane 1 - 1 μl (2 units) of protease alone, lane 2- supernatant after overnight incubation with beads, lane 3- 1st wash of beads with 20mM Tris/150mM NaCl, pH7.8, lane 4- 2nd wash of beads, lane 5- beads alone. Each lane contains 10 μl of sample with 30 μl of 5X Laemmli sample buffer. POI=protein of interest (ankyrin repeat domain). (N=2)

It was also noted that large concentrations of the fusion protein could not be cleaved by just increasing the amount of HRV3C protease. Rather, it appeared that the concentration of the fusion protein needed to be $\sim 200 \mu\text{g}$ and suspending in a volume greater than 1 ml in a microfuge tube for the HRV3C protease enzymatic reaction to occur (Figure 5.10). The most likely reason for this is because when too concentrated the cleavage site is obstructed.

Therefore, for large amounts of clone F2 to have its tag cleaved with HRV3C, the reaction mixture needs to be separated into many separate microfuge tubes.

In order to separate the MBP tag from the ankyrin repeat domain following incubation with HRV3C protease, Ni^{+2} -NTA beads were used, with the reasoning that the MBP domain, which includes the hexa-histidine tag, would be retained on the resin, while the ankyrin repeat domain would be released into the supernatant. This experiment was initially attempted using immobilised metal ion affinity chromatography (IMAC), whereby the POI + MBP mixture was incubated with Ni^{+2} -NTA sepharose beads overnight at 4 °C, then the beads are pelleted and the supernatant evaluated by SDS-PAGE, staining with Coomassie dye (Figure 5.11). As mentioned above, the tag was expected to remain bound to the Ni^{+2} resin, as it carried the His₆-peptide, while the purified protein of interest would be suspended in the supernatant. Contrary to expectation, the pelleted nickel beads, when eluted with SDS sample buffer, contained both the tag and the cleaved protein, while no cleaved protein was found in the supernatant. So, while the proteolysis had cleaved the MBP tag from the ankyrin repeat domain, the failure to partition the two proteins into pellet and supernatant respectively, indicated that they were still bound to each other. The failure to separate the two fusion partners after proteolytic cleavage of the peptide linker suggested that non-covalent interactions existed between these domains. By varying the buffer composition, it was tested whether this non-covalent interaction could be disrupted. This interaction was attempted to be interrupted by varying the buffer

composition, including altering pH, increasing the salt concentration to 1 M NaCl, adding 300 mM urea, adding the reducing agent dithiothreitol (DTT), or eluting the cleaved protein with increasing concentrations of glycerol. None of these strategies succeeded in separating the affinity tag from the protein of interest.

Finally, separation of the MBP tag from the ankyrin repeat domain was achieved by loading the cleaved mixture of proteins onto a Ni^{+2} -NTA column. The MBP tag did bind to the Ni^{+2} on the column matrix, and the ankyrin repeat domain

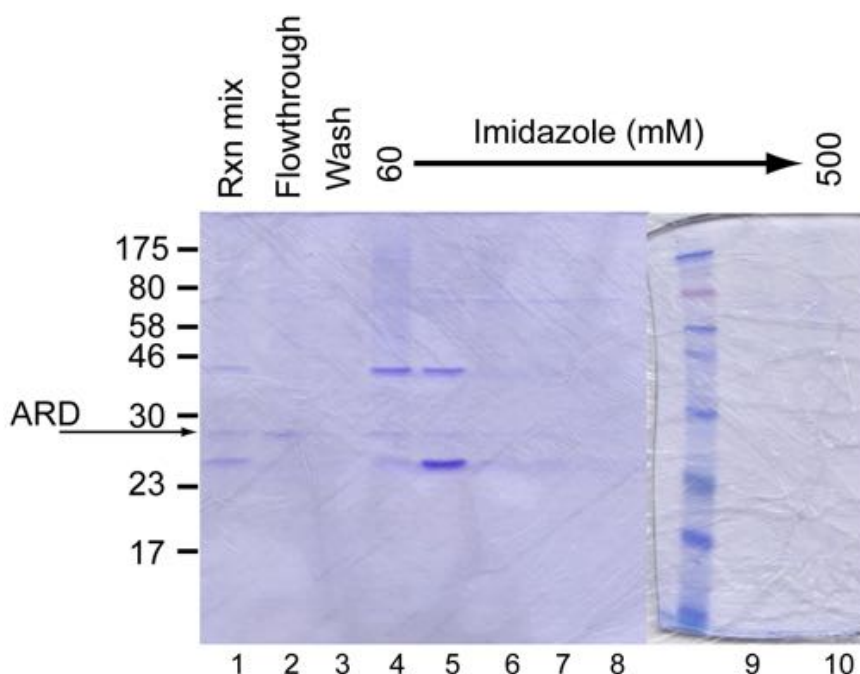


Figure 5.12 - Separation of MBP tag from ankyrin repeat domain using Ni^{+2} -NTA column. Lane 1- mixture of F2 cleaved with HRV3C protease, lane 2- FT, lane 3- wash, lane 4-8- increasing imidazole concentration from 60mM to 500mM. Arrow points to band representing ankyrin repeat domain. Each lane contains 40 μl of sample and 10 μl of Laemmli sample buffer. (N=1)

was able to be eluted with the flow-through (Figure 5.12). The ankyrin repeat domain was able to be separated from the MBP tag. However, the only caveat was that only a small amount of protein was eluted. Through these experiments, it became clear that generating a sufficient quantity of the purified and cleaved ankyrin repeat domain was probably unrealistic: attempts to proteolytically cleave the F2 fusion protein was possible although difficult and isolating the ankyrin repeat domain from the reaction mix yielded only a small amount of the target protein. Because the two step purification of the intact fusion protein was relatively successful, immunization with the fusion protein was given consideration.

Adopting this strategy, clone F2 was purified as a fusion protein. Omitting proteolytic cleavage of the MBP tag, a single band pure protein was forwarded to Biogenes for the immunisation of two rabbits to obtain ANKRD18A-specific polyclonal antibody (Figure 5.13). In order to facilitate affinity purification of the resulting antibodies, 5 mg of clone G2 was purified (at a concentration greater than 0.5 mg/ml). Clone G2 consists of the ankyrin repeat domain fused to a Halo tag (a modified haloalkane dehalogenase designed to covalently bind to synthetic ligands from Promega). The purified protein of clone G2 was used for immobilization on an immuno-affinity column. The antibody that had been raised by immunizing the rabbits with the purified protein of clone F2 was purified using

the clone G2-coated affinity column, which selected for antibody that recognized the ankyrin repeat domain of ANKRD18A.

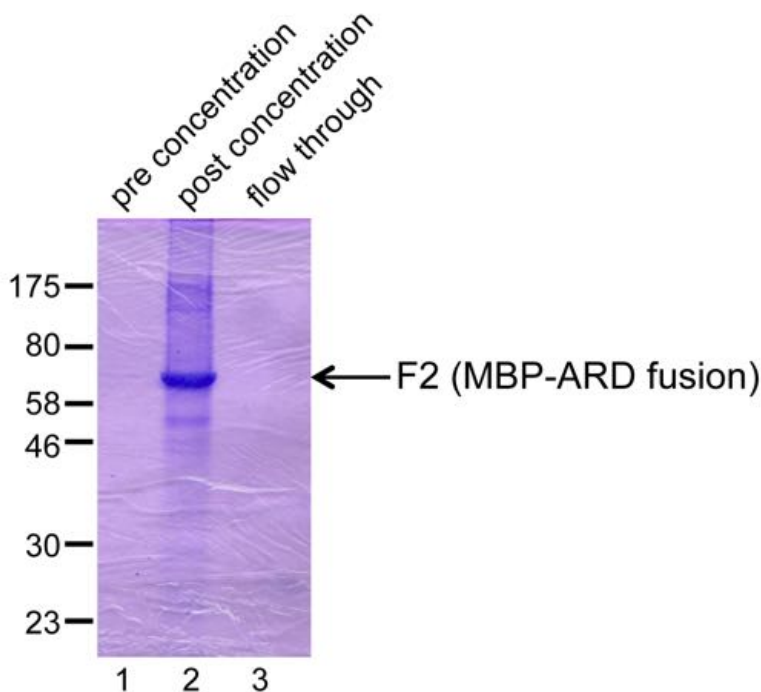


Figure 5.11 – Coomassie stained SDS-PAGE demonstrating purified protein of clone F2 sent to Biogenes for generation of antibody. Lane 1- pre-concentrated protein, lane 2-concentrated protein, lane 3- flow through from concentration. Each lane contains 5 μ l of sample and 30 μ l of Laemmli sample buffer. Total concentration of protein after concentration is 2.16 mg/ml in total volume of 1 ml. (N=1)

5.2.6. Production and purification of the ANKRD18A recombinant protein (clone G2)

The purification steps used to isolate the protein of clone G2 was similar to those of clone F2 in that the first purification step involved Ni^{+2} -NTA chromatography followed by anion exchange chromatography then concentration of protein using Amicon centrifugal filters. For Ni^{+2} -NTA chromatography, the protein of clone G2

eluted at a higher concentration of imidazole, 300 mM (Figure 5.14). Fractions obtained from imidazole elutions of 300 to 500 mM were pooled and dialysed in 20 mM Tris, 200 mM NaCl, pH 7.5. A series of trials at lower NaCl concentrations (i.e. less than 200mM) led to significant and rapid precipitations of protein during dialysis. Raising NaCl to 200 mM proved critical in preventing precipitation during dialysis, but the question arose whether this concentration of NaCl would limit clone G2's affinity to the Q-sepharose matrix for ion exchange chromatography. Fortunately, 200 mM NaCl was still compatible with G2 binding to the anion exchange column resin and elution occurred over a NaCl gradient from 350 to 500 mM NaCl (Figure 5.15). The pooled eluants were then dialysed against 20 mM sodium phosphate, 500 mM NaCl, pH 7.5.

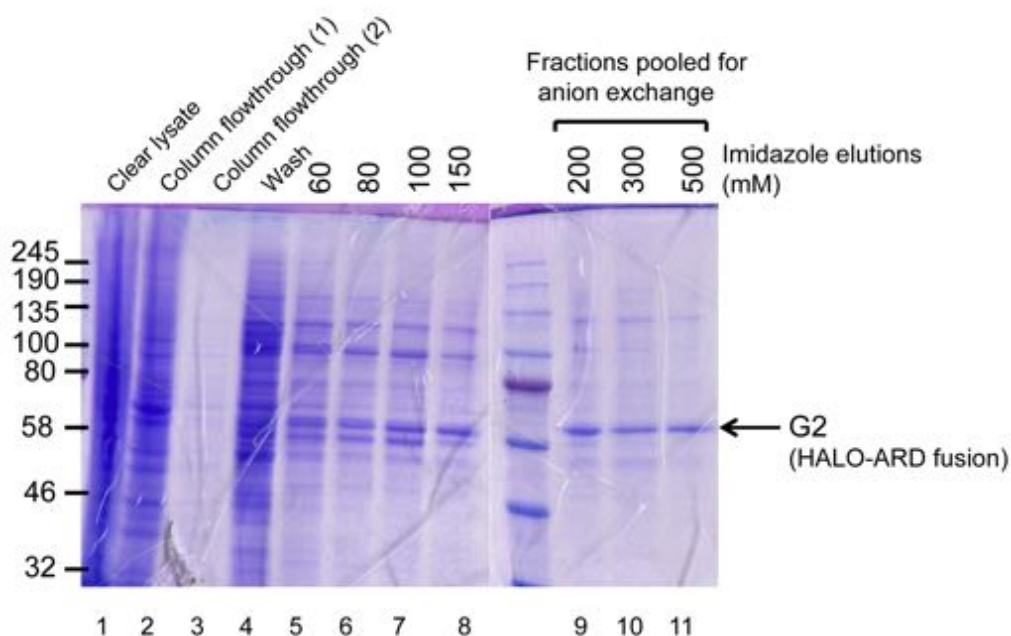


Figure 5.12 – Coomassie stained SDS-PAGE demonstrating purification of precleared lysates from thirty 15cm plates of HEK293T cells transiently transfected with cDNA from clone G2 (E3C Halo-ankyrin repeat domain). Elutions using 300-500mM imidazole are collected for dialysis and further purification using the anion exchange column. Each lane contains 40 μ l of sample and 10 μ l of Laemmli sample buffer (N=20).

Following dialysis, the pooled eluants were concentrated using Amicon Ultra-15 centrifugal filters. However, the solubility of fusion protein G2 was low and the protein was prone to precipitation if concentrated beyond 0.5 mg/ml. Adding glycerol (8% v/v) to the dialysis solution did not markedly improve solubility. In addition, the Amicon membrane was passivated (through incubation with 5% v/v solution of polyethyleneglycol 6000) to minimize stickiness of the protein to the ultrafiltration membrane, but this also proved to very little effect. Nevertheless,

after 20 preparations of clone G2, 5.05 mg of purified clone G2 (concentration 1.02 mg/ml) was finally attained and sent to Biogenes for immobilization on an immuno-affinity column.

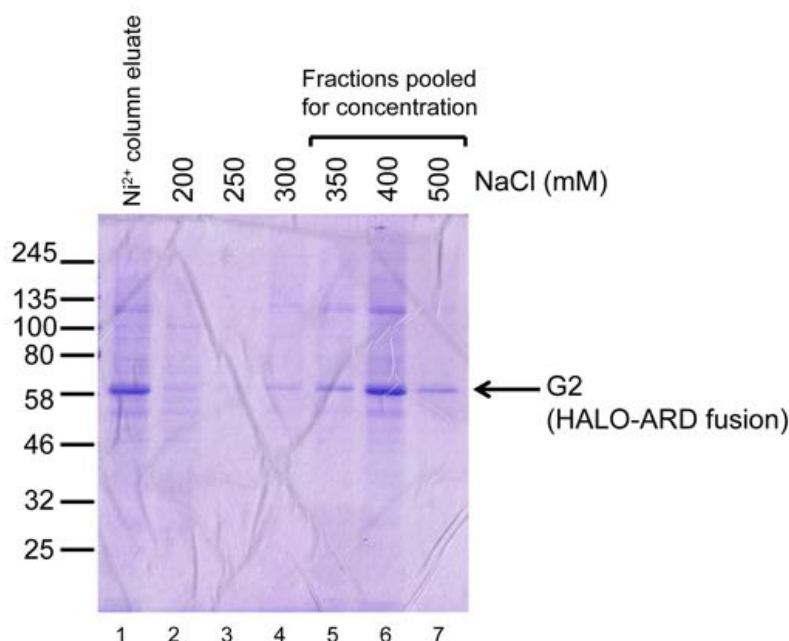


Figure 5.13 – Coomassie stained SDS-PAGE demonstrating purification of pooled Ni²⁺-NTA eluants of clone G2 on an anion exchange column. Elutions from lanes 5-7 using 350 to 500mM NaCl are pooled for dialysis and then concentration. Each lane contains 40 μ l of sample and 10 μ l of Laemmli sample buffer. Arrow points to purified elutions (N=20).

5.3. Characterisation of ANKRD18A polyclonal antibody 25277

Antibody 25277 is an affinity purified polyclonal antibody raised in rabbits to recognise the folded structure of the ankyrin repeat domain of ANKRD18A. Two polyclonal antibodies were received from Biogenes, 25277 and 25278, each raised in different rabbits. However, α -25277, which had a concentration of 0.24

mg/ml, was chosen as the one to use after data from flow cytometry and immunofluorescence experiments and immunoblotting for Westerns, suggested that α -25277 is more specific.

5.3.1. Assessing specificity of α -25277

In a flow cytometry (Accuri B6 cell sorter) experiment, following incubation with either α -25277 (1:100) or a non-specific IgG, unpermeabilised and permeabilised HEK293T cells were analysed. The cells sorting histogram (Figure 5.16) demonstrated that permeabilised HEK293T cells (Figure 5.16B), when treated with α -25277, caused an increase in fluorescence intensities (Table 5.4). This increase is reversed in a dose dependent fashion when the purified ANKRD18A fragment, clone G2, was added (Table 5.4). In contrast, unpermeabilised cells demonstrated only a modest increase in fluorescence intensities (Figure 5.16A and Table 5.4). These data suggested that ANKRD18A is located intracellularly because α -25277 only caused a large increase in fluorescence signal in the permeabilised cell. Also, the fact that the purified recombinant ankyrin repeat domain protein was able to reverse the response, supports specificity of α -25277 for ANKRD18A. Another control which should have been conducted to further support the claim that α -25277 is specific is to use a mock protein in place of clone G2 to see if there is any decrease in fluorescent intensities following the co-application of α -25277 and mock protein to the HEK293T cells.

In order to further ascertain specificity of α -25277 for its target antigen, immunofluorescence confocal microscopy was performed. These experiments

entailed a two-step process. In both experiments, Dami cells were transiently transfected with plasmid DNA encoding clone G2. In the first experiment, transfected Dami cells were stained with the primary antibody, α -His (1:750) (which recognises the hexa-histidine tag of the recombinant protein), the secondary antibody, α -mouse Alexa Fluor 488 (1:300) and a nuclear stain, TO-PRO-3. As a control, transfected Dami cells were stained with just secondary antibody. Results of this experiment (Figure 5.17) suggested that the recombinant protein, clone G2, is localised to the nucleus since the α -His stain co-localises with the TO-PRO-3 stain.

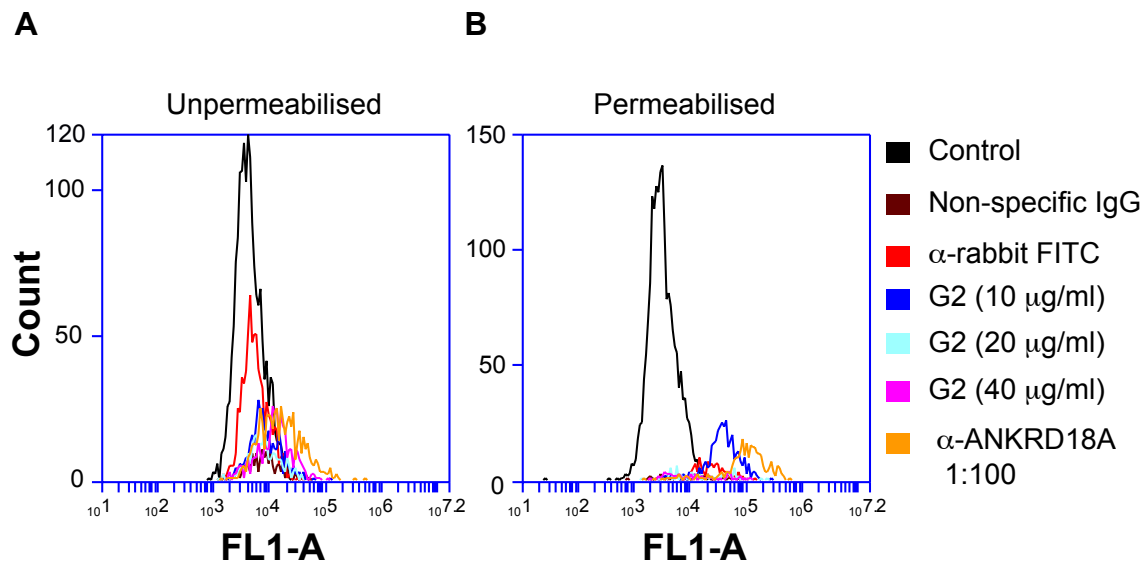


Figure 5.14 - Flow cytometry experiment probing specificity of the polyclonal α -ANKRD18A, α -25277, using HEK293T cells. A) and B) Concentrations of antibody stocks were as follows: Non-specific IgG (1 mg/ml), α -rabbit IgG conjugated to FITC (2 mg/ml), α -25277 (0.24 mg/ml). Clone G2 = purified ANKRD18A (1-260) (N=2).

	Unpermeabilised	Permeabilised
	Mean FL1-A	Mean FL1-A
Control	5015.5	3743.2
Non-specific IgG	9830.0	15218.4
α -rabbit IgG conjugated to FITC	9696.3	28937.1
α -ANKRD18A (1:100) + G2 (10 μ g/ml)	10305.2	46729.1
α -ANKRD18A (1:100) + G2 (20 μ g/ml)	9825.4	33112.0
α -ANKRD18A (1:100) + G2 (40 μ g/ml)	15830.5	23332.3
α -ANKRD18A (1:100)	24508.3	134113.3

Table 5.4 - Flow cytometry experiment probing specificity of the polyclonal α -ANKRD18A, α -25277. Values indicated are the mean FL1-A fluorescence intensities for the conditions shown in the panels in Figure 5.18 above (N=2).

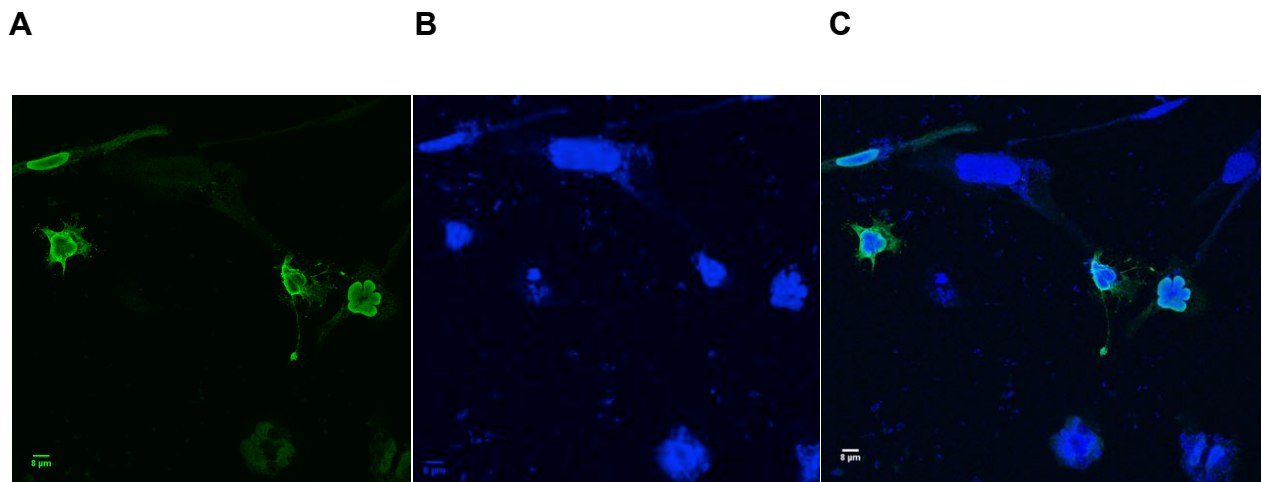


Figure 5.15 – Immunofluorescence confocal microscopy images of Dami cells stained with α -His and TO-PRO-3 demonstrating intranuclear localization of fusion protein (clone G2). Dami cells are transiently transfected with cDNA from clone G2 (E3C-Halo ankyrin repeat domain). A) primary antibody: α -His monoclonal (1:750), secondary antibody: α -mouse Alexa Fluor 488 (1:300) B) TO-PRO-3 (1:2000) C) Overlay of panels A and B.

For the subsequent experiment, a different set of transfected Dami cells were stained with both α -His (1:750) and α -25277 (1:100) as primary antibodies and the corresponding secondary antibodies, α -mouse Alexa Fluor 488 and α -rabbit Alexa Fluor 647 (both at 1:300), respectively. In this experiment, Dami cells were not stained with TO-PRO-3 as well because TO-PRO-3's far-red fluorescence has peak absorbance at 642 nm and emission at 661 nm, which overlaps with absorbance and emission peaks of Alexa Fluor 647. Again, the controls for this experiment were Dami cells stained only with the secondary antibodies: α -mouse Alex Fluor 488 and α -rabbit Alexa Fluor 647. In this

experiment, the transfected recombinant protein, clone G2, localised to the nucleus, as evidenced by the α -His stain (Figure 5.18A) and the results of the previous experiment. α -25277 also stained for protein that localised to the same area as that which was stained with α -His, i.e. the nucleus (Figure 5.18B). Since α -His is specific for the recombinant protein, co-localisation of the signals associated with α -25277 and α -His, indicated that they bind to the same protein. Results from the second experiment suggested that both the recombinant protein and ANKRD18A co-localised to the nucleus, suggesting that ANKRD18A is a nuclear protein. These experiments also further underscored the specificity of α -25277 for ANKRD18A because its staining results concurred with those of α -His, which binded to the transfected recombinant protein (Figure 5.18 and Figure S2).

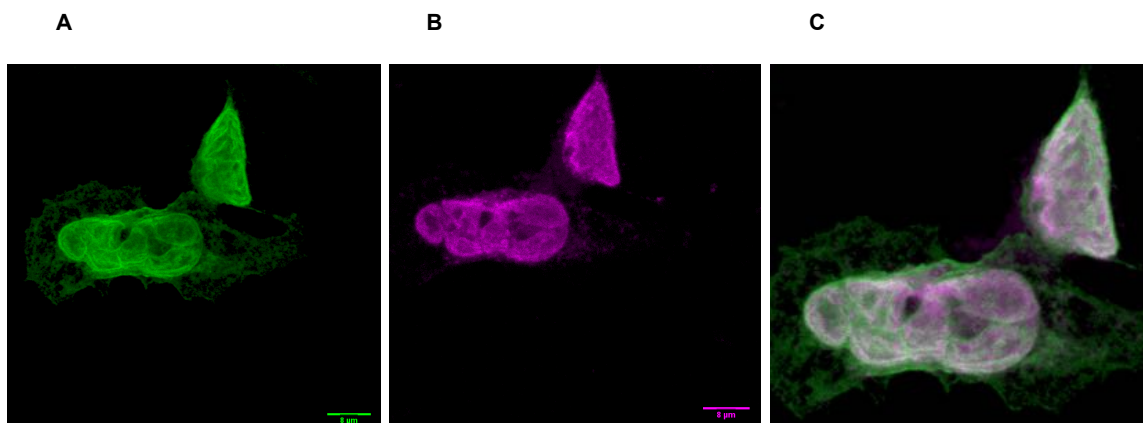


Figure 5.16 – Immunofluorescence confocal microscopy images of Dami cells stained with α -His and α -25277 demonstrating intranuclear localisation of ANKRD18A. Dami cells are transiently transfected with cDNA of plasmid for Clone G2. A) primary: α -His6, secondary: α -mouse Alexa Fluor 488, B) primary: α -25277, secondary: α -rabbit Alexa Fluor 647, C) overlay of A and B (N=2).

5.3.2. Immunoblotting to assess ANKRD18A antibody 25277

Finally, specificity of α -25277 was probed in immunoblot experiments, including dot blots, Westerns, and immunoprecipitation (IP) experiments. Dot blots assisted in determining the appropriate dilution of antibody stock to be used in immunoblotting and IP experiments. HEK293T cells lysed with RIPA buffer were used in the dot blot experiments. From the dot blots, it was ascertained that α -25277 could be used at a dilution down to 1:1500 (Figure 5.19). When α -25277 was used to immunoblot WCLs of Dami, HEK293T cells and platelets in SDS-PAGE gels, no dominant band appeared at the expected molecular weight of 116 kDa (Figure 5.20 and 5.21). Dami and HEK293T cells were lysed with RIPA buffer, which can lyse the nuclear membrane to release intranuclear proteins. The platelets were lysed with a buffer containing the detergent NP-40, which cannot lyse nuclear membranes. However, this is a mute point since platelets do not have nuclei. As is typical for a polyclonal antibody, which is used to immunoblot Westerns, multiple bands were observed, but in this case, when immunoblotting the WCLs of HEK293T and Dami cells, there were multiple bands but no predominant band to suggest that the antibody was recognizing a particular protein. Since these cells were lysed with RIPA buffer, it is quite possible that genomic material had been released upon lysis of the nucleus and when run in a Western, resulted in the “smeary” appearance seen in each lane of the gel. Nevertheless, in order to confirm that α -25277 recognised its antigen, a strong, if not dominant band was expected between the 100 and 135 kDa

markers, for the Dami/HEK293T blot (Figure 5.20). For the platelet blot (Figure 5.21), one saw a dominant band at the correct molecular weight of ~116 kD in the Donor 2 lane. This however was not seen in the other WCLs of the other donors. It appeared that the WCLs of donors 1 and 3 were very concentrated resulting in difficulty identifying individual bands in the two lanes. Also, for donors 4 and 5, the WCLs appear not concentrated enough, thus there may not be enough ANKRD18A protein in the lane to be identified with α -25277.

Thus, both of these Westerns were technically flawed and would require further modification of experimental conditions for one to obtain better blots for analysis and determination of whether α -25277 could be used for immunoblotting experiments.

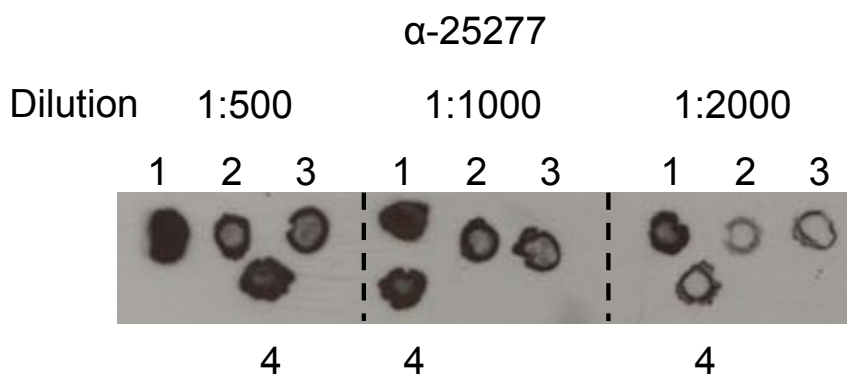


Figure 5.17 – Dot blot demonstrating the antibody dilution of α -25277 to be used for experiments. 1 = WCLs from HEK293T cells, 2 = WCLs from Dami cells differentiated for 24 hours, 3 = WCLs of Dami cells differentiated at 48 hours, 4 = WCLs of Dami cells differentiated at 72 hours. N=2

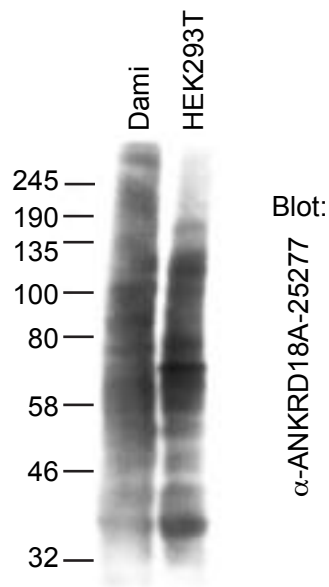


Figure 5.18 – Immunoblot of Dami and HEK293T WCLs using α-25277. Blotting antibody dilution is 1:1500 (N=3).

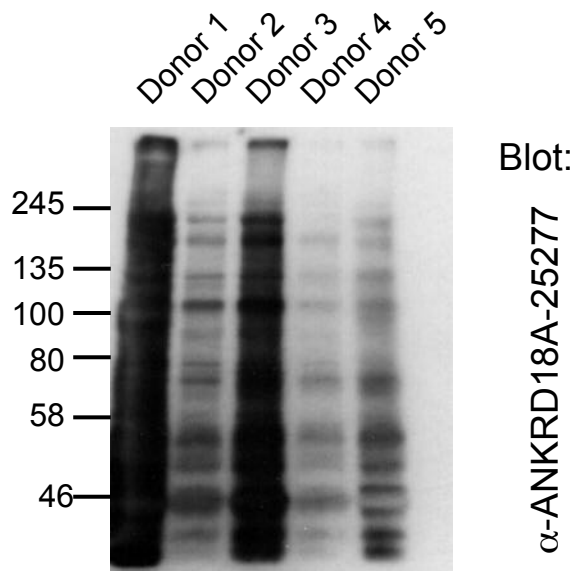


Figure 5.19 – Immunoblot of healthy donor platelet WCLs using α-25277. Blotting antibody dilution is 1:1500 (N= 3).

Because no ANKRD18A protein was detected in the Westerns immunoblotted with α -25277, the question arose whether enough protein had been present in the WCLs of HEK293T and Dami cells and platelets that had been tested (Figures 5.20 and 5.21). Immunoprecipitation experiments were thus performed with WCLs from HEK293T and Dami cells and platelets using α -25277 and non-specific IgG as a control. The IPs were loaded onto SDS-PAGE gels and immunoblotted using the same antibodies, α -25277 (Figure 5.22B) and non-specific IgG as control (Figure 5.22A), respectively. Again, Dami and HEK293T cells were lysed with RIPA buffer and platelets were lysed with buffer containing NP-40. The immunoblot blotted with α -25277 failed to show a clear-cut band at the expected molecular weight of ~116 kDa. Of note, the positive control, purified protein of clone G2, did not appear when immunoblotting with α -25277, so this experiment is mute and would need to be repeated (Figure 5.22B). In the analogous experiment for platelet WCLs, immunoprecipitation again used non-specific IgG and α -25277 followed by immunoblotting with non-specific IgG (Figure 5.23A) and α -25277 (Figure 5.23B). There was a questionable band at ~116 kDa, but its intensity was comparatively weak. Nevertheless, the lane containing the positive control (Figure 5.23B, lane 1), the purified protein of clone G2, had a band at the correct molecular weight of 65 kDa.

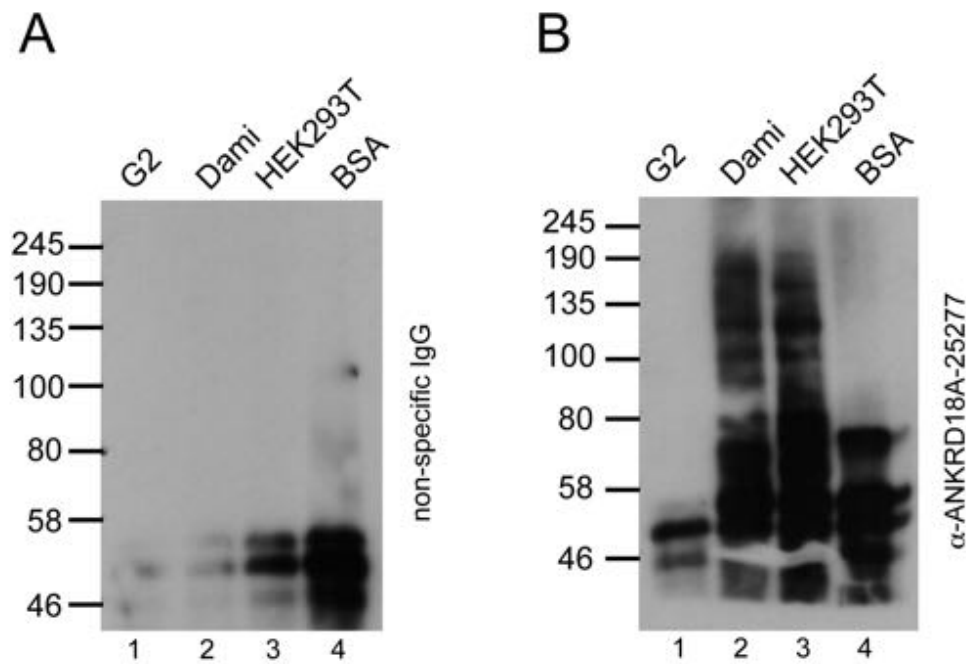


Figure 5.20 – Immunoblot of ANKRD18A immunoprecipitated from whole cell lysates of HEK293T and Dami cells. IP antibodies are (A) IgG and (B) α -ANKRD18A, α -25277. Blotting antibodies used are monoclonal (A) IgG and (B) α -ANKRD18A, α -25277 at 1:1500 dilution. Lane 1 - Clone G2, lane 2 - Dami cells, lane 3 - HEK293T cells, lane 4 - BSA (N=2).

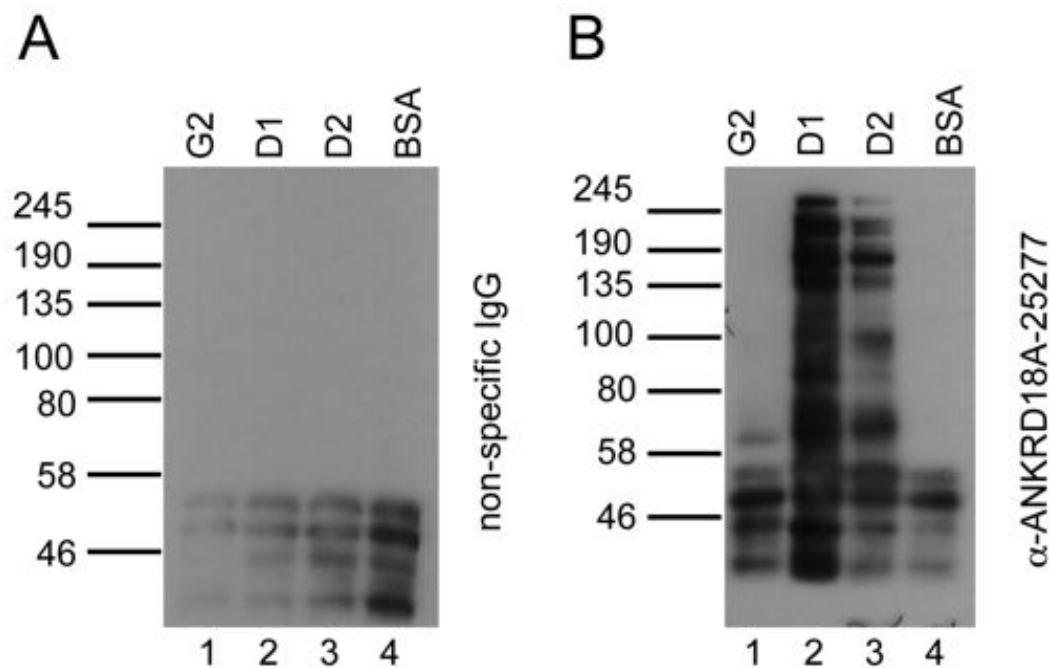


Figure 5.21 – Immunoblot of ANKRD18A immunoprecipitated from whole cell lysates of healthy donor platelets (D1 and D2). IP antibodies are (A) IgG and (B) α-ANKRD18A, α-25277. Blotting antibodies used are (A) monoclonal IgG and (B) α-ANKRD18A, α-25277 at 1:1500 dilution. Lane 1-Clone G2, lane 2-D1, lane 3- D2, lane 4-BSA (N=1).

5.4. Summary

Given these data, it is still indeterminate whether polyclonal antibody α-ANKRD18A-25277, is suitable for the detection of the ANKRD18A antigen. Although in flow cytometry and immunofluorescence confocal microscopy experiments, the specificity of α-25277 for ANKRD18A was suggestive, the fact that α-25277 did not immunoblot for ANKRD18A when using HEK293T, Dami

and platelet whole cell lysates in Westerns and immunoprecipitation experiments makes it questionable whether α -25277 is specific for ANKRD18A.

In the flow cytometry experiment, when the concentration of the immunising antigen was titrated upwards, it reversed the shift of the cell population in cell sorting experiment (Figure 5.16 and Table 5.4). And, when used in confocal experiments to immunolocalise ANKRD18A in Dami cells (Figures 5.17 and 5.18), α -25277 identified ANKRD18A by co-localising with ANKRD18A which had been transfected into Dami cells as determined by immunolocalisation using α -His. In such experiments, ANKRD18A was present in its native state and not its denatured state. This suggests that α -25277 would work equally well in immunohistochemistry and electron microscopy experiments.

In contrast, the results from the immunoblotting and immunoprecipitation experiments of whole cell lysates indicated that specific detection of ANKRD18A by α -25277 were not compelling. In these experiments however, ANKRD18A is denatured (through treatment with SDS and boiling at 95 °C), which may have disrupted the epitope that α -25277 recognised, and thus prevented the expected signal to appear in the corresponding blots. Also, it should be noted that due to time constraints, the immunoblots and IP experiments using α -25277 adhered to standard protocols used in the laboratory. More time was needed to fine-tune the experimental conditions: cell concentrations in cell lysates might have

needed to increase or decrease; there may have been genomic DNA contamination in the WCLs for the Western blots; the choice of lysis buffer might have needed to be altered i.e. the detergent in the cell lysis buffer changed or decreased in concentration or the salt concentration of the lysis buffer may have needed to be increased.

In order that specificity of α -25277 can be confirmed, ANKRD18A expression needs to be silenced by RNAi. The above flow cytometry and confocal microscopy experiments need to be repeated using HEK293T cells and Dami cells transfected with siRNA to silence ANKRD18A expression. If α -25277 is specific to ANKRD18A, the shift seen in the above flow cytometry experiment will be abrogated when these “knockdown” cells are used. Likewise, when the confocal immunofluorescence microscopy experiments are repeated with the “knockdown” cells, no protein should be localized with α -25277. This can be confirmed by transfecting in the cDNA of clone G2 into the “knockdown” Dami cells to see if after transfection of this plasmid, α -25277 can localize the recombinant protein. Overall though, given all of these conflicting data, the most realistic chance of obtaining an antibody, which is specific to ANKRD18A, is to generate a monoclonal antibody.

Although it may seem that the above efforts to purify the recombinant protein, comprised of a functional domain of ANKRD18A fused to a tag, were

superfluous, especially since α -25277 was shown to be less than effective, there are uses for the purified recombinant proteins. Because these recombinant proteins contain an ankyrin repeat domain, they can be used to “fish-out” potential binding partners for ANKRD18A. This can be accomplished by performing protein-protein interaction assays such as pull-downs or co-immunoprecipitation experiments with platelet lysates and then followed up with identification of potential binding partners using mass spectrometry. Identifying potential ANKRD18A binding partners will be helpful in determining possible functions for ANKRD18A.

6. Discussion

6.1. Conclusions from work of this thesis

6.1.1. Evolutionary context of ANKRD18A and its link to platelet disorders

The overall aim of this project was to characterise the novel protein, ANKRD18A. A mutant form of ANKRD18A is suspected to be linked to the severe thrombocytopenia identified in two children from a consanguineous family (Figure 1.14); however, the molecular mechanism causing this inherited platelet disorder remains to be clarified. By determining the function of ANKRD18A, this project sought to establish whether there is a causative link between the two patients' thrombocytopenia and the deletion mutation identified in the *ANKRD18A* gene.

A homologue of ANKRD18A, ANKRD26, has been implicated in an autosomal dominant form of thrombocytopenia, THC2 or ANKRD26-RT (Noris, Perrotta et al. 2011). ANKRD26 has 1050 amino acid residues and its amino acid sequence is 44% identical to ANKRD18A. Patients with THC2 have various single nucleotide substitutions in the 5'-UTR of *ANKRD26* (Bluteau, Glembotsky et al. 2012). These mutations contrast with the genetic alteration found in *ANKRD18A*, which is an in-frame deletion of an entire codon in the centre of exon 13 of the gene. In addition to ANKRD26, there exists another homologous protein,

ANKRD18B, which has a 93% amino acid identity with ANKRD18A. Its function remains unknown.

The domain structure of ANKRD18A gives few clues as to its physiological role in the eukaryotic cell. ANKRD18A has three distinct regions of homology to proteins of known structure (Figure 1.16) as determined by fold prediction using threading (Phyre2, HHpred). The amino terminus consists of ankyrin-like repeats, followed by a coiled coil region that appears to resemble the structures of tropomyosin and at the carboxy terminus, the colicin IA homology domain. Ankyrin repeat domains have a variety of functions, but are primarily known as mediators of protein-protein interactions, particularly in eukaryotes (Li, Mahajan et al. 2006). Tropomyosin is an integral component of actin filaments and is involved in cellular pathways which regulate the cellular cytoskeleton (von der Ecken, Muller et al. 2015). The fact that ANKRD18A has this tropomyosin homology domain further suggests that it may play a role in the cytoskeleton. Colicin IA is a transmembrane protein, which forms voltage-gated ion conducting channels across the plasma membrane of target bacteria. The C-terminal region of ANKRD18A shows remote homology to the bacterial channel-forming protein colicin IA. The homology was detected using the fold prediction server Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>) by threading. The sequence identity over the homologous region is very low (9% identity), but the prediction has a high confidence level (98%). The homology extends over the coiled-coiled region of

colicin IA, including the receptor binding region of colicin IA, but excluding the channel-forming domain of this bacterial protein. In the fold prediction, the ANKRD18A Δ 801 deletion mutation maps to a small helix in the receptor-binding region of colicin IA. However, given the very low level of sequence identity between the two proteins, a prediction of what effect the analogous deletion would have in colicin IA is not possible (Gouaux 1997, Wiener, Freymann et al. 1997, Stroud, Reiling et al. 1998).

6.1.2. Generation of ANKRD18A recombinant protein and subsequent antibodies

In order to characterise the function of ANKRD18A, a molecular probe specific for this protein is a key requirement. Prior to my joining the laboratory, the laboratory had acquired two polyclonal antibodies directed against peptides derived from amino residues 53-66 and 260-289 of the ANKRD18A sequence (Figure 4.4). Testing these set of polyclonal antibodies raised doubts as to their specificity for ANKRD18A (Chapter 4). Secondly, the closely related sequence of ANKRD18B, a protein of similarly unknown function, further confounded the results. For example, when probing WCLs of HEK293T cells with the α -ANKRD18A antibody 21234, the dominant band was between 135 and 190 kDa, suggesting that the long isoform of ANKRD18B (Mr 165 kDa), rather than ANKRD18A (Mr 116kDa) had been detected (Figure 4.5B). Thus, to overcome ambiguity and potential cross-reactivity resulting from antibodies raised against peptides, a strategy was devised to generate an antibody raised against the

folded, full-length ANKRD18A. A polyclonal antibody was decided upon with the intention that should results prove positive and interesting, a monoclonal antibody would be made later in the future.

Prior to my joining the laboratory, GFP and myc-tagged constructs had been made which encode the wild-type and a mutant form of full-length ANKRD18A. These were intended for use in over-expression experiments to determine the biological functions of ANKRD18A. However, subsequent flow cytometry experiments demonstrated poor transfection efficiencies with these plasmid cDNAs (Figure 3.5A). Likewise, α -GFP immunoblotting of whole cell lysates from HEK293T cells transiently transfected with full-length GFP-ANKRD18A in Westerns revealed no expression (Figure 3.5B). Attempts were also made to immunoblot the insoluble fractions of HEK293T cells transfected with full-length GFP-ANKRD18A and full-length constructs made at OPPF. These attempts failed to demonstrate expression by blotting with α -GFP or α -His₆.

As a first priority, my work focussed on generating one or more antibodies directed against the folded form of ANKRD18A. Given the difficulty of expressing full-length ANKRD18A, an alternative strategy was devised whereby a suitable antigen was generated by expressing and purifying the ankyrin repeat domain of ANKRD18A (Table 5.1). Approximate boundaries of the (putative) structural domains of ANKRD18A were determined through fold prediction. To reduce the

risk of erroneous domain predictions, two alternative fold prediction servers were used, which consistently predicted the ankyrin repeat domain at the amino terminus of ANKRD18A. This bioinformatics-led approach helped generate a rational basis for designing the plasmid expression constructs (Table 5.2).

The parallel processing capability of OFFP (the Oxford Protein Production Facility) enables one to rapidly identify combinations of fragments and affinity tags that will lead to expression of soluble protein. Since protein expression can be highly dependent on the expression host, a pool of 32 expression constructs was tested for expression in both *E. coli* cells (Figures 5.1, 5.2, 5.3) and the mammalian cell line, HEK293T (Figures 5.4, 5.5). In these pilot tests at OPPF, expression in HEK293T cells was generally more successful than in *E. coli*. The most likely reason why expression was more effective in HEK293T cells rather than in *E. coli* is because prokaryotic and eukaryotic gene sequences differ in their usage of the genetic code, and corresponding expression hosts differ in the prevalence of synonymous tRNAs. Thus, it is often found that mammalian gene sequences express poorly in prokaryotic hosts and vice versa. Several constructs encompassing the amino terminal ankyrin repeat domain, which is consistently identified as a conserved domain by bioinformatical analysis, demonstrated good protein expression, suggesting that this fragment indeed forms an integral structural domain. Nonetheless, even the more speculative fragments, the central tropomyosin-like domain and the carboxy terminal colicin

1A-like domain resulted in soluble protein (Figure 5.6). Consistent with my previous transfection experiments, none of the full-length constructs expressed. While it cannot be ruled out that the transfection protocol is to blame, it could also be that overexpression fails because folding of full-length ANKRD18A is not sufficiently supported by activity of cognate chaperones. In the absence of chaperone-assisted folding, the ANKRD18A polypeptide may be subject to proteolytic degradation (Goldberg 2003).

Milligram quantities of pure protein are required for generating antibodies. In order to produce large amounts of protein in mammalian cells, relatively inexpensive PEI was preferred as a transfectant. Of the clones which expressed (Table 5.3), the highest expressing clones were clone G2 (E3C Halo – ankyrin repeat domain) and clone F2 (MBP – ankyrin repeat domain). While the plasmid constructs G2 and F2, demonstrated good expression, the recombinant protein proved challenging in terms of controlling solubility in aqueous buffer. Over many expression and purification experiments, in which the protocol was varied in terms of buffer composition, it became clear that proteolytically cleaving the fusion protein to obtain the isolated ankyrin repeat domain was largely futile. Thus, it was necessary to immunise with the fusion protein clone F2 and to ‘filter’ the resulting polyclonal antibodies through affinity purification, whereby clone G2 was immobilised onto the affinity column.

The protein of clone G2 was initially chosen to be the antigen for antibody generation; however, because of precipitation difficulties encountered during the purification process, clone F2 was ultimately chosen. Purification of clone F2 was relatively clear-cut once the purification conditions were determined. One of the difficulties encountered for clone F2 was in proteolytically cleaving MBP, the affinity tag (Figure 5.9). Not only does this fusion protein require large amounts of HRV3C protease to remove the MBP tag, but also, each cleavage reaction mixture cannot contain too concentrated a protein or else the protein cannot be completely cleaved (Figure 5.10). This suggested that possibly the protein of clone F2 was folded in such a way which obstructed the cleavage site. Another difficulty encountered after proteolysis was in separating the MBP tag from the ankyrin repeat domain in the reaction mixture using the Ni^{+2} -NTA matrix. Repeated separation attempts using Ni^{+2} -NTA bead slurry failed whereas using a Ni^{+2} -NTA pre-packed column demonstrated some level of separation, albeit small (Figures 5.11 and 5.12). The most likely explanation for why separation occurred using the pre-packed column but not the slurry is that binding kinetics of pre-packed Ni^{+2} -NTA columns are different from that of Ni^{+2} -NTA beads.

The protein of clone G2 proved difficult to keep in solution following purification over the Ni^{+2} -NTA resin. During dialysis that preceded the ion exchange chromatography step or during subsequent concentration through ultrafiltration, the protein was prone to precipitate to a considerable degree. To enable affinity

purification of the antibody raised in rabbits, Biogenes required clone G2 to be at a concentration between 0.5 to 1.0 mg/ml. Even at this moderate concentration, precipitation occurred in an unpredictable fashion. In order to help increase stability of the protein, glycerol was added to the buffer. Glycerol is frequently used as a co-solute to promote protein stability and to reduce protein aggregation as it can interact with both hydrophobic and hydrophilic surface patches of the protein (Bondos and Bicknell 2003).

6.1.3. Characterisation of α -25277

Ultimately, affinity purified polyclonal antibodies were produced by immunisation of two rabbits. Of the corresponding antibodies, α -25277 and α -25278, the former proved more sensitive and specific, and was hence used in subsequent experiments. Binding of α -25277 to the ankyrin repeat domain was demonstrated through flow cytometry and immunofluorescence confocal microscopy. For the flow cytometry experiment, application of α -25277 to permeabilised vs. unpermeabilised cells transfected with clone G2 cDNA were compared (Figure 5.16 and Table 5.4). Only the former showed a distinct increase in fluorescence. This increase could be reversed in a dose-dependent fashion with addition of increasing concentrations of purified recombinant protein, clone G2; thus, demonstrating that purified clone G2 competes with binding of α -25277 to ANKRD18A or the recombinant ankyrin repeat domain expressed in the

transfected cells. The other explanation for this result could be that clone G2 was competing with α -25277's ability to bind non-specifically.

Immunofluorescence confocal microscopy provided additional evidence for binding of α -25277 to ANKRD18A (Figures 5.17 and 5.18). Binding could be demonstrated by simultaneous labelling of recombinant clone G2 expressed in Dami cells with α -His, which recognises the hexa-His affinity tag of transfected G2 and α -25277, which recognises the ankyrin repeat domain of both transfected G2 and endogenous ANKRD18A. Co-localisation of these two signals in transfected cells strongly suggested that they adhered to the same entity.

Prior to the acquisition of α -25277, initial determination of ANKRD18A sub-cellular localisation was performed utilising fluorescence confocal microscopy whereby cDNA plasmids of GFP or myc-tagged wild-type and mutant full-length ANKRD18A were transiently transfected into HEK293T and HeLA cells respectively and stained with the corresponding primary antibodies, α -GFP or α -myc (Figures 3.6, 3.7). This experiment was attempted despite the low transfection efficiencies of the full-length plasmids because confocal microscopy is a relatively sensitive method of detecting proteins – at least more so than immunoblotting. The results of these experiments suggested that full-length wild-type and mutant ANKRD18A both localise to the cytoplasm. Later, upon acquisition of α -25277, these immunofluorescence confocal microscopy

experiments were repeated, although this time, some conditions altered: Dami cells, a megakaryocytic cell line, were used, α -His and α -25277 were the primary antibodies, and the cDNA of clone G2 was transfected into Dami cells. The rationale for choosing Dami cells was that this cell line is more similar to megakaryocytes since it is derived from megakaryoblastic tumor cells, whereas HEK293T cells are derived from endothelial kidney cells. Staining of transfected Dami cells in the latter experiments, firstly with α -His and TO-PRO-3, a fluorescent dye use for nuclear counterstaining, suggested that the recombinant protein, G2, which was transfected into Dami cells, is located in the nucleus (Figure 5.17); secondly, when similarly transfected cells were stained with α -His and α -25277 together as primary antibodies (secondary being α -mouse Alexa Fluor 488 and α -rabbit Alexa Fluor 647, respectively), the same area of the cell fluoresced, suggesting that both the recombinant protein and endogenous ANKRD18A were localised to the same area (Figure 5.18) and, since the recombinant protein was known to be in the nucleus from the first experiment, one could conclude that endogenous ANKRD18A also localised to the nucleus. In the second experiment, TO-PRO-3 is not used because its peak absorption and emission spectra overlap with Alexa Fluor 647. These findings thus contrast with those from the initial experiments using full-length ANKRD18A plasmids. It is possible that the results of prior experiments are questionable since HEK293T cells were transfected with full-length ANKRD18A plasmids, which have a poor transfection efficiency (Figure 3.5A). To further substantiate this conclusion that

ANKRD18A is localised to the nucleus, one can perform another immunofluorescence confocal microscopy experiment whereby transfected Damis are stained with α -25277 as the primary antibody and TO-PRO-3 for the nuclear stain. Because TO-PRO-3 is used, a different secondary antibody would need to be used, α -rabbit Alexa-Fluor 488. Should the staining of α -25277 and TO-PRO-3 overlap, one can be certain that ANKRD18A localises to the nucleus.

6.1.4. Comparison of experimental findings with predictions by the Blueprint progenitors database

Two of my experimental observations conflict with the predictions by the Blueprint progenitors database (blueprint.haem.cam.ac.uk). Firstly, my experimental observations suggested that there is *ANKRD18A* mRNA expression in platelets (Figure 4.2). However, the Blueprint database suggests that there is negligible *ANKRD18A* mRNA expression in megakaryocytes (Figure 6.1). Secondly, my experimental observations suggested that *ANKRD18B* mRNA is absent in megakaryocytes and platelets. This again conflicts with predictions by the Blueprint progenitors database, which indicates that *ANKRD18B* mRNA is present in megakaryocytes. However, the Blueprint database conflicts with data from the PlateletWeb (plateletweb.bioapps.biozentrum.uni-wuerzburg.de), which suggests that ANKRD18A is expressed in platelets while ANKRD18B is not.

Using the Blueprint Progenitors web application one can create river plots, heat maps and bar plots from hematopoietic progenitor datasets to identify the relative

mRNA expression of genes in haematopoiesis. The Blueprint project utilised RNA sequencing to illustrate the transcriptional diversity of lineage committed human blood progenitors (Chen, Kostadima et al. 2014). When *ANKRD18A* and its homologues, *ANKRD18B* and *ANKRD26*, were entered as a query into the tools portal, the portal predicted transcript specific RNA expression of these three homologues in hematopoietic progenitors. The river plot suggested that all three mRNAs are expressed in megakaryocytes - although in relatively very small amounts. And, of the three, *ANKRD18A* mRNA is barely expressed in megakaryocytes (Figure 6.1). To interpret a river plot, one needs to realise that the amount of relative mRNA expression correlates with the thickness of the branches of the river plot. Thus, if one were to look at *ANKRD26* mRNA expression, one sees that it is highly expressed in haematopoietic stem cells and its progenitors. However, not much is expressed in the late stages of megakaryopoiesis because normally, its expression is silenced by RUNX1 and FLI1 transcription factor binding. Only when the 5'-UTR of *ANKRD26* is mutated, as seen in familial thrombocytopenia 2 (THC2), is there mRNA expression of *ANKRD26* in megakaryocytes (Bluteau, Balduini et al. 2014).

The river plot prediction contradicts my qRT-PCR data, which suggested that *ANKRD18A* mRNA expression increases with differentiation of megakaryocytes (Figure 4.7). The increase in *ANKRD18A* mRNA in developing megakaryocytes

was quite distinct; however, because the qRT-PCR method is very sensitive this result may still represent only a weak level of expression in absolute terms.

Antibody α -25277 was also used for immunoblotting and in immunoprecipitation experiments. When used to probe for denatured ANKRD18A in HEK293T, Dami and platelet WCLs in Westerns, α -25277 was not able to identify a protein with the predicted molecular weight of 116 kDa (Figures 5.20 and 5.21). The Western blot showed multiple bands but not one clear band at the correct molecular weight. When α -25277 was used to immunoprecipitate WCLs of HEK283T, Dami and platelets followed by immunoblotting with α -25277, no clear bands appeared to suggest that the antibody recognises the protein, which is denatured at the blotting stage (Figure 5.22 and 5.23). It is plausible though that α -25277 is unable to detect the unfolded protein, as it was raised against the folded ankyrin repeat domain and may require an epitope that is non-contiguous in terms of the primary sequence.

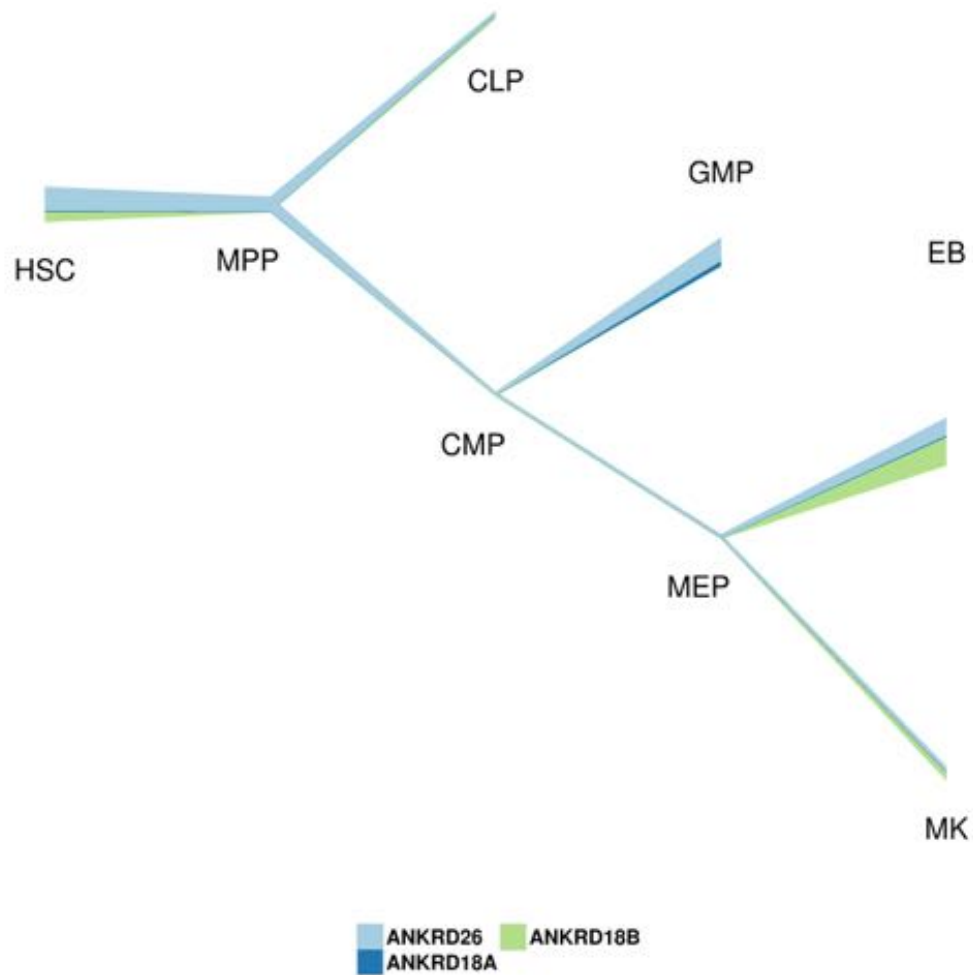


Figure 6.1 – River plot representing relative mRNA expression of *ANKRD26*, *ANKRD18B* and *ANKRD18A* across the various cell lineages of hematopoiesis. Line thickness increases with more mRNA expression. HSC = hematopoietic stem cell, MPP = multipotent progenitor, LMPP = lymphoid primed multipotent progenitor, CLP = common lymphoid progenitor, CMP = common myeloid progenitor, GMP = granulocyte-monocyte progenitor, MEP = megakaryocyte-erythrocyte progenitor, EB = erythroblast, MK = megakaryocyte (blueprint.haem.cam.ac.uk).

6.2. Future research

There are many questions, which remain to be answered regarding this project. Firstly, it is still unclear whether ANKRD18A is the gene which, when mutated, results in the thrombocytopenia exhibited by the two patients described above. Secondly, my limited characterisation of ANKRD18A has only generated more questions. For instance, it is still unclear whether ANKRD18A is localised to the cytoplasm or the nucleus given my immunofluorescence confocal microscopy results. However, given that *ANKRD18A* mRNA levels increase with differentiation of megakaryocytes, one can hypothesize that it is localised to the nucleus and may play a role in megakaryopoiesis; thus, explaining why the ANKRD18A-mutated patients develop thrombocytopenia. To confirm that *ANKRD18A* plays a role in megakaryopoiesis, megakaryocytes generated from CD34⁺ cells could be infected with short hairpin RNA (shRNA) lentiviruses targeting *ANKRD18A* to see if RNAi-mediated gene silencing will impede megakaryocyte development and thus disrupt the production of platelets. Quantitation of knockdown could then be determined using qRT-PCR with comparisons to internal controls *B-actin* and *GAPDH*. And, α -25277 or α -21234 could be used to immunoblot WCLs of megakaryocytes to determine if the protein expression of ANKRD18A has been eradicated. However, it will be necessary to clarify whether α -25277 or α -21234 are specific to ANKRD18A before either can be used for that purpose. Otherwise, since the advent of gene

editing and more specifically, the CRISPR technique, it will be possible to mimic the patient's mutations for the candidate gene variants using induced Pluripotent stem cells (iPS). Megakaryocytes can be produced by forward programming (Moreau, Evans et al. 2016) to provide a platform to study these patients' phenotype in more detail. Thirdly, it remains unclear whether the polyclonal antibody, which was generated for this project, α -25277, is specific to ANKRD18A. To determine the specificity of α -25277, an experiment which could be fairly quickly performed, is to suppress ANKRD18A expression using siRNAs. These cells could then be analysed by flow cytometry using α -25277. This experiment is essentially repeating what had been done before; however, this time, one would hope to see that the shift in the tracing seen before, using α -25277 on 'regular' HEK293T cells, would be negated by using cells in which ANKRD18A expression has been silenced. Suppression of ANKRD18A expression in Dami cells could also be performed and again, analysis on the effectiveness of gene silencing could be performed by repeating the confocal experiment. Anti-25277 would be used to localise for any protein expression in these cells. If possible, the ANKRD18A-silenced cells could also be transiently transfected with the cDNA plasmid of clone G2 to demonstrate that the 'knockdown' is reversed. In addition, α -25277 could be used for immunoblotting WCLs of cells with and without silencing by RNAi once the immunoblotting conditions for using α -25277 have been clarified. Finally, to help further characterise the functions of the novel protein ANKRD18A, one could identify its

possible binding partners whose functions may thus reflect what functions ANKRD18A may have. Using either of the purified recombinant proteins (clones F2 and G2), both of which contain the ankyrin repeat domain, which mediates protein-protein interactions, one could perform pull-down experiments or co-immunoprecipitation experiments with platelet WCLs to identify proteins which may interact with ANKRD18A. Mass spectrometry would then be utilised to identify the potential interacting partners.

6.3. Final conclusion

The purpose of this project was answer the hypothesis that the mutation identified in the novel gene, *ANKRD18A*, is the cause for the severe thrombocytopenia seen in two related children of a consanguineous family. Whole exome sequencing followed by Sanger sequencing had identified 3 possible variants that may cause this phenotype of severe thrombocytopenia: p.801_801del of ANKRD18A; p.G447R of GNE and p.A509V of FRMPD1. All three variants were present in the two affected children and were within a tightly linked region of homozygosity on chromosome 9p (Johnson, Fletcher et al. 2016). The variant in *ANKRD18A* was chosen as the likely variant to have caused the phenotype.

TEM micrographs of patient platelets showed that the patients' platelets are enlarged, but fewer in number (macrothrombocytopenic), confirming the MPV of 10.4 – 15 fL that was initially observed (Figure 3.3). Utilising qRT-PCR, it was demonstrated that the amount of *ANKRD18A* mRNA increased as the megakaryocytes differentiate (Figure 4.7) and using immunofluorescence confocal microscopy, it was demonstrated that *ANKRD18A* may localise to the cytoplasm or the nucleus (Figures 3.6, 3.7, 5.17, and 5.18). These data suggested that *ANKRD18A* may play a role in the cytoskeleton or be involved in megakaryopoiesis. Definitive proof of *ANKRD18A* involvement in either cellular functions would require silencing the gene in immature megakaryocytes to see if megakaryocytic development is altered or curtailed. Evidence to question or contradict the hypothesis that the variant of *ANKRD18A* is the cause of these patients' severe thrombocytopenia and that *ANKRD18A* has a role in the megakaryocyte /platelet cytoskeleton or is involved with megakaryopoiesis, are three arguments. Firstly, when comparing the amino acid sequences of *ANKRD18A* for *Homo sapiens* and its homologues such as *Pan troglodytes* (chimpanzee), *Gorilla gorilla*, and *Nomascus leucogenys* (gibbon), the triple glutamic residues within the amino acid sequence MLEEEVL, which is present in *H. sapiens*, is absent in gorilla; and, in *N.leucogenys*, the third glutamic acid residue, which is deleted in the two patients, is also absent. This lack of conserved residues among *ANKRD18A* homologues suggests that *ANKRD18A* and the mutation at $\Delta 801$ are not significant. Secondly, data from the RNA

sequencing of hematopoietic progenitors (blueprint.haem.cam.ac.uk), suggest that very little amount of *ANKRD18A* mRNA is present in hematopoietic progenitors. This is in contrast to the mRNA of *GNE*, which is expressed widely in hematopoietic progenitors (Figure 6.2). Finally, to cast further doubt that the variant in *ANKRD18A* is the culprit gene are the publications in 2014 by two separate groups (Izumi, Niihori et al. 2014, Zhen, Guo et al. 2014), which identified 2 sets of siblings with congenital thrombocytopenia and *GNE* myopathy. Our patients are currently of the ages ~8 and 11. The patients mentioned in the two papers developed *GNE* myopathy in their mid-adolescence in one set of siblings and in their 20's in the other set of siblings. Should we not be able to confirm by experiments that the variant in *ANKRD18A* is the causative pathological variant, we will, with the passage of time, know which is the true causative variant, once our two patients develop neuromuscular abnormalities.

Although the work for this thesis has not definitely determined that the variant in *ANKRD18A* results in severe thrombocytopenia, it has generated *ANKRD18A* purified recombinant protein directed to several functional motifs and subsequently, antibody generation. Of use are the recombinant proteins containing the ankyrin repeat domains because they can be used for protein-protein interaction experiments whereby platelet WCLs are incubated with clones F2 or G2 to “pulldown” potential binding partners. More experiments will need to be performed to confirm that the polyclonal antibody, α -25277, is specific to

ANKRD18A. Another other antibody to ANKRD18S, the peptide antibody, α -21234, may also prove to be useful once specificity has been ascertained. If, however, neither antibodies prove to be specific to ANKRD18A, the best way to obtain an antibody specific to ANKRD18A will be through generation of a monoclonal antibody.

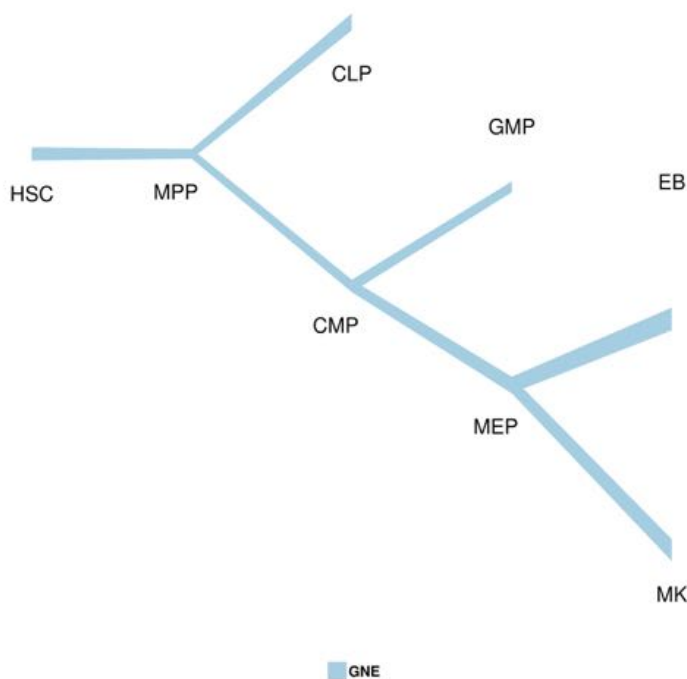


Figure 6.2 – Riverplot representing *GNE* mRNA expression across the various cell lineages of hematopoiesis (blueprint.haem.cam.ac.uk).

7. Supplementary Data

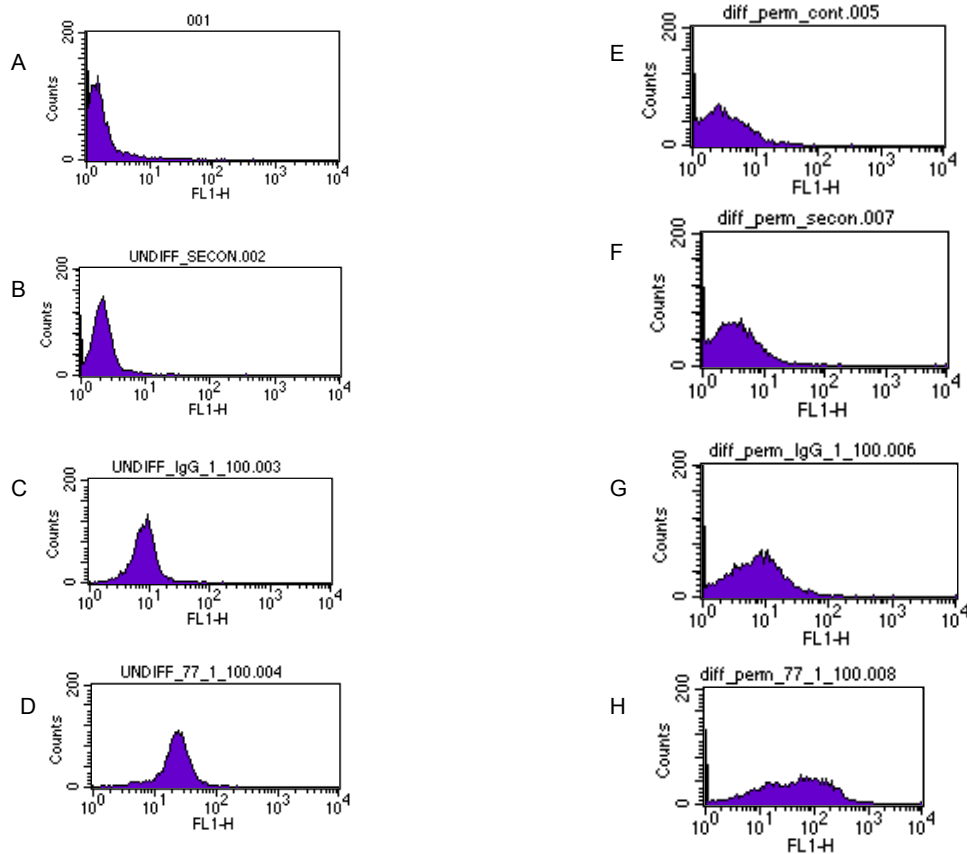


Figure S 1 – Utilising flow cytometry to probe specificity of α -ANKRD18A-25277 in Dami cells. Dami cells are undifferentiated and impermeabilised (panels A-D) or differentiated and permeabilised (panels E-H). Dami cells are differentiated using TPO (3 ng/ml) and PMA (6.2 ng/ml). Antibody stocks: IgG (1 mg/ml), α -25277 (0.24 mg/ml), and α -rabbit Alexa Fluor 488 (2 mg/ml). IgG used at a dilution of 1:100, α -25277 used at dilution of 1:100 and α -rabbit Alexa Fluor 488 used at dilution 1:500. N=1

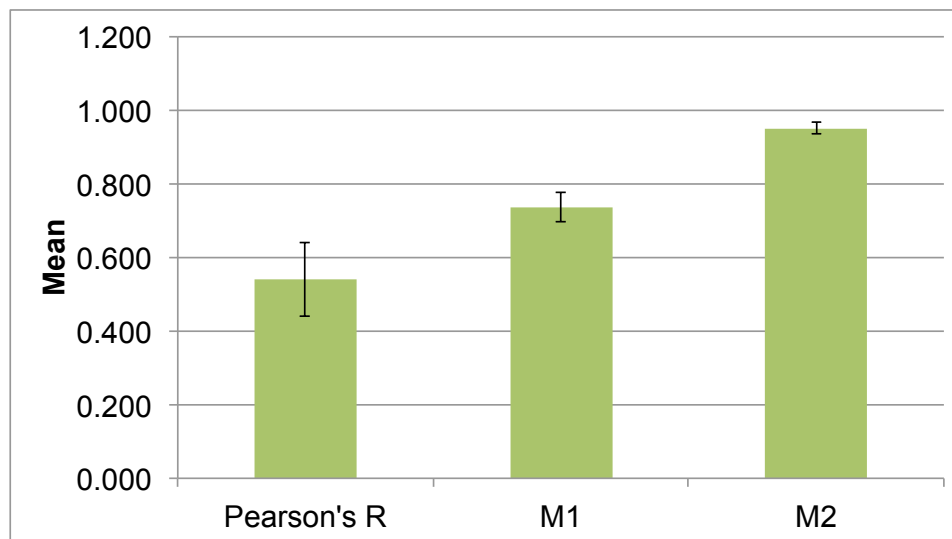


	Image 1	Image 2	Image 3	Mean	SD
Pearson's R	0.39	0.77	0.46	0.540	0.1011
M1	0.653	0.813	0.745	0.737	0.0401
M2	0.915	0.96	0.976	0.950	0.0158

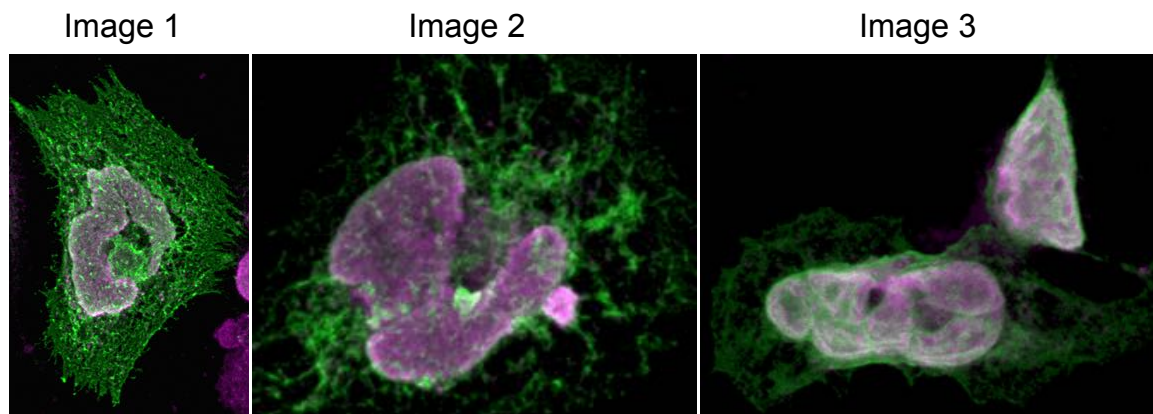
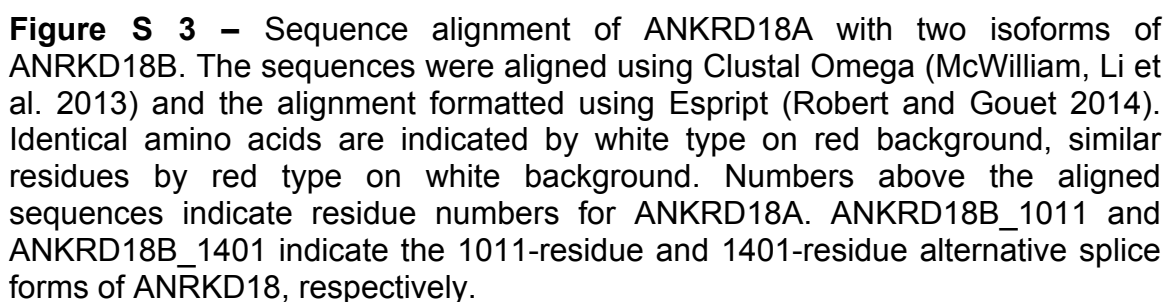


Figure S 2 – Bar graph demonstrating the mean co-localisation coefficients of confocal microscopy overlay images 1-3. M1 and M2 = Manders split coefficient. Images are differentiated Dami cells transiently transfected with cDNA of Clone G2. Cells are stained with primary antibodies: α -His and α -25277

and secondary antibodies: α -mouse Alexa Fluor 488 and α -rabbit Alexa 647.
N=3

ANKED18_1011	1	10	20	30	40	50	60	70	80	90	100	110	120
ANKED18_1011	2	10	20	30	40	50	60	70	80	90	100	110	120
ANKED18_1401	3	10	20	30	40	50	60	70	80	90	100	110	120
ANKED18_1401	4	10	20	30	40	50	60	70	80	90	100	110	120
ANKED18_1011	5	130	140	150	160	170	180	190	200	210	220	230	240
ANKED18_1011	6	130	140	150	160	170	180	190	200	210	220	230	240
ANKED18_1401	7	130	140	150	160	170	180	190	200	210	220	230	240
ANKED18_1401	8	130	140	150	160	170	180	190	200	210	220	230	240
ANKED18_1011	9	250	260	270							280		
ANKED18_1011	10	250	260	270							280		
ANKED18_1401	11	250	260	270							280		
ANKED18_1401	12	250	260	270							280		
ANKED18_1011	13	290	300	310	320	330	340	350	360	370	380	390	400
ANKED18_1011	14	290	300	310	320	330	340	350	360	370	380	390	400
ANKED18_1401	15	290	300	310	320	330	340	350	360	370	380	390	400
ANKED18_1401	16	290	300	310	320	330	340	350	360	370	380	390	400
ANKED18_1011	17	410	420	430	440	450	460	470	480	490	500	510	520
ANKED18_1011	18	410	420	430	440	450	460	470	480	490	500	510	520
ANKED18_1401	19	410	420	430	440	450	460	470	480	490	500	510	520
ANKED18_1401	20	410	420	430	440	450	460	470	480	490	500	510	520
ANKED18_1011	21	530	540	550	560	570	580	590	600	610	620	630	640
ANKED18_1011	22	530	540	550	560	570	580	590	600	610	620	630	640
ANKED18_1401	23	530	540	550	560	570	580	590	600	610	620	630	640
ANKED18_1401	24	530	540	550	560	570	580	590	600	610	620	630	640
ANKED18_1011	25	650	660	670	680	690	700	710	720	730	740	750	760
ANKED18_1011	26	650	660	670	680	690	700	710	720	730	740	750	760
ANKED18_1401	27	650	660	670	680	690	700	710	720	730	740	750	760
ANKED18_1401	28	650	660	670	680	690	700	710	720	730	740	750	760



1 10 20 30 40 50 60 70
18A_Hs MRKLFSFGRRIGQALLSSMDQEVASPGYITDWELEKTIHRAAIKGDAAEVERCLTRRFQLDARDRKDRIT
18A_Gg MRKLFSFGRRIGQALLSSMDQEVASPGYITDWELEKTIHRAAIKGDAAEVERCLTRRFQLDARDRKDRIT
18A_Pg MRKLFSFGRRIGQALLSSMDQEVASPGYITDWELEKTIHRAAIKGDAAEVERCLTRRFQLDARDRKDRIT
18A_Nl MRKLFSFGRRIGQALLSSMDQEVASPGYITDWELEKTIHRAAIKGDAAEVERCLTRRFQLDARDRKDRIT

80 90 100 110 120 130 140
18A_Hs VLEHACASGRVYVVTLLIRRCQIDCDRLRTPLMKAVHSQEEACATLLSCGANPNIDYIGNTALHY
18A_Gg VLEHACASGRVYVVTLLIRRCQIDCDRLRTPLMKAVHSQEEACATLLSCGANPNIDYIGNTALHY
18A_Pg VLEHACASGRVYVVTLLIRRCQIDCDRLRTPLMKAVHSQEEACATLLSCGANPNIDYIGNTALHY
18A_Nl VLEHACASGRVYVVTLLIRRCQIDCDRLRTPLMKAVHSQEEACATLLSCGANPNIDYIGNTALHY

150 160 170 180 190 200 210
18A_Hs AVYNSTSLAEQLSHHANEAANEENPTTFFAINSRQHMVEFLKNQANRAVDNFRALILAVOR
18A_Gg AVYNSTSLAEQLSHHANEAANEENPTTFFAINSRQHMVEFLKNQANRAVDNFRALILAVOR
18A_Pg AVYNSTSLAEQLSHHANEAANEENPTTFFAINSRQHMVEFLKNQANRAVDNFRALILAVOR
18A_Nl AVYNSTSLAEQLSHHANEAANEENPTTFFAINSRQHMVEFLKNQANRAVDNFRALILAVOR

220 230 240 250 260 270 280
18A_Hs NISSIVLLQQNTISQDQFGQTAEDYACDLRSIQQILEHKNKLKNSLRNDNQETA MKANR
18A_Gg NISSIVLLQQNTISQDQFGQTAEDYACDLRSIQQILEHKNKLKNSLRNDNQETA MKANR
18A_Pg NISSIVLLQQNTISQDQFGQTAEDYACDLRSIQQILEHKNKLKNSLRNDNQETA MKANR
18A_Nl NISSIVLLQQNTISQDQFGQTAEDYACDLRSIQQILEHKNKLKNSLRNDNQETA MKANR

290 300 310 320 330 340 350
18A_Hs KRKEGASKEENKVASDEKQERISENKKQPQDSQSGKKKDAKMGNMKQDIAMIKKEELYAIKNDLSR
18A_Gg KRKEGASKEENKVASDEKQERISENKKQPQDSQSGKKKDAKMGNMKQDIAMIKKEELYAIKNDLSR
18A_Pg KRKEGASKEENKVASDEKQERISENKKQPQDSQSGKKKDAKMGNMKQDIAMIKKEELYAIKNDLSR
18A_Nl KRKEGASKEENKVASDEKQERISENKKQPQDSQSGKKKDAKMGNMKQDIAMIKKEELYAIKNDLSR

360 370 380 390 400 410 420
18A_Hs KRKYVEIKSITEINAFKSVRLNEMITKVAYSQDLDLKAENRLNSLEKEHNKKERLEAEVE
18A_Gg KRKYVEIKSITEINAFKSVRLNEMITKVAYSQDLDLKAENRLNSLEKEHNKKERLEAEVE
18A_Pg KRKYVEIKSITEINAFKSVRLNEMITKVAYSQDLDLKAENRLNSLEKEHNKKERLEAEVE
18A_Nl KRKYVEIKSITEINAFKSVRLNEMITKVAYSQDLDLKAENRLNSLEKEHNKKERLEAEVE

430 440 450 460 470 480 490
18A_Hs SLHSRLATAANNYEIERKDLLELVLRADVSRSEKMSNISLTDNNEFLTOVKARVKFSLKAKL
18A_Gg SLHSRLATAANNYEIERKDLLELVLRADVSRSEKMSNISLTDNNEFLTOVKARVKFSLKAKL
18A_Pg SLHSRLATAANNYEIERKDLLELVLRADVSRSEKMSNISLTDNNEFLTOVKARVKFSLKAKL
18A_Nl SLHSRLATAANNYEIERKDLLELVLRADVSRSEKMSNISLTDNNEFLTOVKARVKFSLKAKL

500 510 520 530 540 550 560
18A_Hs RSTRDALREKTLAASVQLDIQAQHRIKQKMHFNQAKESQS GKONSLEERIQLINLLLRQI
18A_Gg RSTRDALREKTLAASVQLDIQAQHRIKQKMHFNQAKESQS GKONSLEERIQLINLLLRQI
18A_Pg RSTRDALREKTLAASVQLDIQAQHRIKQKMHFNQAKESQS GKONSLEERIQLINLLLRQI
18A_Nl RSTRDALREKTLAASVQLDIQAQHRIKQKMHFNQAKESQS GKONSLEERIQLINLLLRQI

570 580 590 600 610 620
18A_Hs DAKREGNKKEVINIRCLENGKEDLLERNNKELMRYNYLKEKLLQEKKAEREVIVRQOSELV
18A_Gg DAKREGNKKEVINIRCLENGKEDLLERNNKELMRYNYLKEKLLQEKKAEREVIVRQOSELV
18A_Pg DAKREGNKKEVINIRCLENGKEDLLERNNKELMRYNYLKEKLLQEKKAEREVIVRQOSELV
18A_Nl DAKREGNKKEVINIRCLENGKEDLLERNNKELMRYNYLKEKLLQEKKAEREVIVRQOSELV

630 640 650 660 670 680 690
18A_Hs DHLLKFSSESPLEGTSCHINISTTSKKKLQVQOPEKKHEEFKRLFELISLIVADQIRKKNS
18A_Gg DHLLKFSSESPLEGTSCHINISTTSKKKLQVQOPEKKHEEFKRLFELISLIVADQIRKKNS
18A_Pg DHLLKFSSESPLEGTSCHINISTTSKKKLQVQOPEKKHEEFKRLFELISLIVADQIRKKNS
18A_Nl DHLLKFSSESPLEGTSCHINISTTSKKKLQVQOPEKKHEEFKRLFELISLIVADQIRKKNS

```

      700      710      720      730      740      750      760
18A_Hs LEEEA GKKCLEMTINMLNAF NEDFSCHDLN DQKMDILPKKKLQK FNDLVAE KAVSSE VNLAK
18A_Gg LEEEA GKKCLEMTINMLNAF NEDFSCHDLN DQKMDILPKKKLQK TPLSYFYEH...FR...
18A_Pg LEEEA GKKCLEMTINMLNAF NEDFSCHDLN DQKMDILPKKKLQK FNDLVAE KAVSSE VNLAK
18A_N1 LEEEA GKKCLEMTINMLNAF NEDFSCHDLN DQKMDILPKKKLQK FDDLMAE KAVSSE VNLAK

      770      780      790      800      810      820      830
18A_Hs DNEVLHQELLSMRNVQER EKLEKDKKMLEEEVVLNLTTHMEKDMVELGKLGQYKSELDERAVQIEKLEE
18A_Gg DNEVLHQELLSMRNVQER EKLEKDKKMLEEEVVLNLTTHMEKDMVELGKLGQYKSELDERAVQIEKLEE
18A_Pg DNEVLHQELLSMRNVQER EKLEKDKKMLEEEVVLNLTTHMEKDMVELGKLGQYKSELDERAVQIEKLEE
18A_N1 DNEVLHQELLSMRNVQER EKLEKDKKMLEEEVVLNLTTHMEKDMVELGKLGQYKSELDERAVQIEKLEE

      840      850      860      870      880      890      900
18A_Hs IHLQKQAEYEKQLEQLNKDNTASLKKKELTLKDVE KFSKMTAYEEVTTLEEFKEAFAGAVKANNSMS
18A_Gg IHLQKQAEYEKQLEQLNKDNTASLKKKELTLKDVE KFSKMTAYEEVTTLEEFKEAFAGAVKANNSMS
18A_Pg IHLQKQAEYEKQLEQLNKDNTASLKKKELTLKDVE KFSKMTAYEEVTTLEEFKEAFAGAVKANNSMS
18A_N1 IQSQRAENEKQLEQLNRDNTALLKKKELTLKDVE KFSKMTAYEEVTTLEEFKEALAITLKANNSMS

      910      920      930      940      950      960      970
18A_Hs KKLMSDKKIAVISTKLFTKQRMKYFLSTLPTRPEPELP VENLNSIELNRRKYPKTAIRIPTSNPQTS
18A_Gg KKLMSDKKIAVISTKLFTKQRMKYFLSTLPTRPEPELP VENLNSIELNRRKYPKTAIRIPTSNPQTS
18A_Pg KKLMSDKKIAVISTKLFTKQRMKYFLSTLPTRPEPELP VENLNSIELNRRKYPKTAIRIPTSNPQTS
18A_N1 KKLMSDEKIAMISTMLLMEKQIQIKYFLSTLPMTDPPELP VXLNLSIGLDRKYPKTAIRIPTSNPQTS

      980      990
18A_Hs NNKNFLTEVLLC.....
18A_Gg NNKNFLTEVLLC.....
18A_Pg NNKNFLTEVLLC.....
18A_N1 NNKNFLTEVLLC.....

18A_Hs .....
18A_Gg .....
18A_Pg .....
18A_N1 SLLLYKGVFPFSRTNRPFGGVRSPGLQLEGTSTLQWEQGP IAHPRHR

```

Figure S 4 – Alignment of ANKRD18A sequences from different primates. Hs = human, Gg = gorilla, Pg = chimpanzee, N1 = gibbon. The sequences were aligned using Clustal Omega (McWilliam, Li et al. 2013) and the alignment formatted using Esript (Robert and Gouet 2014). The EEE motif mutated in the thrombocytopenic patients is at residues 799 to 801 (human sequence numbers).

References

- Acs, P., P. O. Bauer, B. Mayer, T. Bera, R. Macallister, E. Mezey and I. Pastan (2015). "A novel form of ciliopathy underlies hyperphagia and obesity in Ankrd26 knockout mice." *Brain Struct Funct* **220**(3): 1511-1528.
- Alharbi, R. A., R. Pettengell, H. S. Pandha and R. Morgan (2013). "The role of HOX genes in normal hematopoiesis and acute leukemia." *Leukemia* **27**(5): 1000-1008.
- Ali, F. Y., S. J. Davidson, L. A. Moraes, S. L. Traves, M. Paul-Clark, D. Bishop-Bailey, T. D. Warner and J. A. Mitchell (2006). "Role of nuclear receptor signaling in platelets: antithrombotic effects of PPARbeta." *FASEB J* **20**(2): 326-328.
- An, N., J. B. Blumer, M. L. Bernard and S. M. Lanier (2008). "The PDZ and band 4.1 containing protein Frmpd1 regulates the subcellular location of activator of G-protein signaling 3 and its interaction with G-proteins." *J Biol Chem* **283**(36): 24718-24728.
- Balduini, C. L. and P. Noris (2016). "Innovation in the field of thrombocytopenias: achievements since the beginning of the century and promises for the future." *Haematologica* **101**(1): 2-4.
- Balduini, C. L., A. Pecci and P. Noris (2012). "Inherited thrombocytopenias: the evolving spectrum." *Hamostaseologie* **32**(4): 259-270.
- Balduini, C. L., A. Pecci and A. Savoia (2011). "Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias." *Br J Haematol* **154**(2): 161-174.
- Balduini, C. L. and A. Savoia (2012). "Genetics of familial forms of thrombocytopenia." *Hum Genet* **131**(12): 1821-1832.
- Balduini, C. L., A. Savoia and M. Seri (2013). "Inherited thrombocytopenias frequently diagnosed in adults." *J Thromb Haemost* **11**(6): 1006-1019.
- Ballmaier, M. and M. Germeshausen (2011). "Congenital amegakaryocytic thrombocytopenia: clinical presentation, diagnosis, and treatment." *Semin Thromb Hemost* **37**(6): 673-681.
- Bamshad, M. J., S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson and J. Shendure (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." *Nat Rev Genet* **12**(11): 745-755.
- Bem, D., H. Smith, B. Banushi, J. J. Burden, I. J. White, J. Hanley, N. Jeremiah, F. Rieux-Laucat, R. Bettels, G. Ariceta, A. D. Mumford, S. G. Thomas, S. P. Watson and P. Gissen (2015). "VPS33B regulates protein sorting into and maturation of alpha-granule progenitor organelles in mouse megakaryocytes." *Blood* **126**(2): 133-143.
-

Bera, T. K., X. F. Liu, M. Yamada, O. Gavrilova, E. Mezey, L. Tessarollo, M. Anver, Y. Hahn, B. Lee and I. Pastan (2008). "A model for obesity and gigantism due to disruption of the Ankrd26 gene." *Proc Natl Acad Sci U S A* **105**(1): 270-275.

Bluteau, D., A. Balduini, N. Balayn, M. Currao, P. Nurden, C. Deswarte, G. Leverger, P. Noris, S. Perrotta, E. Solary, W. Vainchenker, N. Debili, R. Favier and H. Raslova (2014). "Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation." *J Clin Invest* **124**(2): 580-591.

Bluteau, D., A. C. Glembotsky, A. Raimbault, N. Balayn, L. Gilles, P. Rameau, P. Nurden, M. C. Alessi, N. Debili, W. Vainchenker, P. G. Heller, R. Favier and H. Raslova (2012). "Dysmegakaryopoiesis of FPD/AML pedigrees with constitutional RUNX1 mutations is linked to myosin II deregulated expression." *Blood* **120**(13): 2708-2718.

Bodi, K., A. G. Perera, P. S. Adams, D. Bintzler, K. Dewar, D. S. Grove, J. Kieleczawa, R. H. Lyons, T. A. Neubert, A. C. Noll, S. Singh, R. Steen and M. Zianni (2013). "Comparison of commercially available target enrichment methods for next-generation sequencing." *J Biomol Tech* **24**(2): 73-86.

Bompard, G. and E. Caron (2004). "Regulation of WASP/WAVE proteins: making a long story short." *J Cell Biol* **166**(7): 957-962.

Bondos, S. E. and A. Bicknell (2003). "Detection and prevention of protein aggregation before, during, and after purification." *Anal Biochem* **316**(2): 223-231.

Born, G. V. (1962). "Aggregation of blood platelets by adenosine diphosphate and its reversal." *Nature* **194**: 927-929.

Castellano, F., C. Le Clainche, D. Patin, M. F. Carlier and P. Chavrier (2001). "A WASp-VASP complex regulates actin polymerization at the plasma membrane." *EMBO J* **20**(20): 5603-5614.

Cesarman-Maus, G. and K. A. Hajjar (2005). "Molecular mechanisms of fibrinolysis." *Br J Haematol* **129**(3): 307-321.

Chen, L., M. Kostadima, J. H. Martens, G. Canu, S. P. Garcia, E. Turro, K. Downes, I. C. Macaulay, E. Bielczyk-Maczynska, S. Coe, S. Farrow, P. Poudel, F. Burden, S. B. Jansen, W. J. Astle, A. Attwood, T. Bariana, B. de Bono, A. Breschi, J. C. Chambers, B. Consortium, F. A. Choudry, L. Clarke, P. Coupland, M. van der Ent, W. N. Erber, J. H. Jansen, R. Favier, M. E. Fenech, N. Foad, K. Freson, C. van Geet, K. Gomez, R. Guigo, D. Hampshire, A. M. Kelly, H. H. Kerstens, J. S. Kooner, M. Laffan, C. Lentaigne, C. Labalette, T. Martin, S. Meacham, A. Mumford, S. Nurnberg, E. Palumbo, B. A. van der Reijden, D. Richardson, S. J. Sammut, G. Slodkiewicz, A. U. Tamuri, L. Vasquez, K. Voss, S. Watt, S. Westbury, P. Flicek, R. Loos, N. Goldman, P. Bertone, R. J. Read, S.

Richardson, A. Cvejic, N. Soranzo, W. H. Ouwehand, H. G. Stunnenberg, M. Frontini and A. Rendon (2014). "Transcriptional diversity during lineage commitment of human blood progenitors." *Science* **345**(6204): 1251033.

Chereau, D., F. Kerff, P. Graceffa, Z. Grabarek, K. Langsetmo and R. Dominguez (2005). "Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly." *Proc Natl Acad Sci U S A* **102**(46): 16644-16649.

Chilamakuri, C. S., S. Lorenz, M. A. Madoui, D. Vodak, J. Sun, E. Hovig, O. Myklebost and L. A. Meza-Zepeda (2014). "Performance comparison of four exome capture systems for deep sequencing." *BMC Genomics* **15**: 449.

Choi, M., U. I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane and R. P. Lifton (2009). "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." *Proc Natl Acad Sci U S A* **106**(45): 19096-19101.

Clark, M. J., R. Chen, H. Y. Lam, K. J. Karczewski, R. Chen, G. Euskirchen, A. J. Butte and M. Snyder (2011). "Performance comparison of exome DNA sequencing technologies." *Nat Biotechnol* **29**(10): 908-914.

Coller, B. S. (2011). "Historical perspective and future directions in platelet research." *J Thromb Haemost* **9 Suppl 1**: 374-395.

Coller, B. S. and H. R. Gralnick (1977). "Studies on the mechanism of ristocetin-induced platelet agglutination. Effects of structural modification of ristocetin and vancomycin." *J Clin Invest* **60**(2): 302-312.

Consortium, G. T. (2013). "The Genotype-Tissue Expression (GTEx) project." *Nat Genet* **45**(6): 580-585.

Cox, K., V. Price and W. H. Kahr (2011). "Inherited platelet disorders: a clinical approach to diagnosis and management." *Expert Rev Hematol* **4**(4): 455-472.

Crooks, G. M., J. Fuller, D. Petersen, P. Izadi, P. Malik, P. K. Pattengale, D. B. Kohn and J. C. Gasson (1999). "Constitutive HOXA5 expression inhibits erythropoiesis and increases myelopoiesis from human hematopoietic progenitors." *Blood* **94**(2): 519-528.

Dasouki, M. J., S. K. Rafi, A. J. Olm-Shipman, N. R. Wilson, S. Abhyankar, B. Ganter, L. M. Furness, J. Fang, R. T. Calado and I. Saadi (2013). "Exome sequencing reveals a thrombopoietin ligand mutation in a Micronesian family with autosomal recessive aplastic anemia." *Blood* **122**(20): 3440-3449.

Dawood, B. B., G. C. Lowe, M. Lordkipanidze, D. Bem, M. E. Daly, M. Makris, A. Mumford, J. T. Wilde and S. P. Watson (2012). "Evaluation of participants with suspected heritable platelet function disorders including recommendation and validation of a streamlined agonist panel." *Blood* **120**(25): 5041-5049.

Dawood, B. B., J. Wilde and S. P. Watson (2007). "Reference curves for aggregation and ATP secretion to aid diagnose of platelet-based bleeding disorders: effect of inhibition of ADP and thromboxane A(2) pathways." *Platelets* **18**(5): 329-345.

Denis, M. M., N. D. Tolley, M. Bunting, H. Schwertz, H. Jiang, S. Lindemann, C. C. Yost, F. J. Rubner, K. H. Albertine, K. J. Swoboda, C. M. Fratto, E. Tolley, L. W. Kraiss, T. M. McIntyre, G. A. Zimmerman and A. S. Weyrich (2005). "Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets." *Cell* **122**(3): 379-391.

Dore, L. C. and J. D. Crispino (2011). "Transcription factor networks in erythroid cell and megakaryocyte development." *Blood* **118**(2): 231-239.

Duong-Ly, K. C. and S. B. Gabelli (2014). "Using ion exchange chromatography to purify a recombinantly expressed protein." *Methods Enzymol* **541**: 95-103.

Eisenberg, I., N. Avidan, T. Potikha, H. Hochner, M. Chen, T. Olender, M. Barash, M. Shemesh, M. Sadeh, G. Grabov-Nardini, I. Shmilevich, A. Friedmann, G. Karpati, W. G. Bradley, L. Baumbach, D. Lancet, E. B. Asher, J. S. Beckmann, Z. Argov and S. Mitrani-Rosenbaum (2001). "The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy." *Nat Genet* **29**(1): 83-87.

Fedorov, R., E. Hartmann, D. K. Ghosh and I. Schlichting (2003). "Structural basis for the specificity of the nitric-oxide synthase inhibitors W1400 and Nomega-propyl-L-Arg for the inducible and neuronal isoforms." *J Biol Chem* **278**(46): 45818-45825.

Fei, Z., T. K. Bera, X. Liu, L. Xiang and I. Pastan (2011). "Ankrd26 gene disruption enhances adipogenesis of mouse embryonic fibroblasts." *J Biol Chem* **286**(31): 27761-27768.

Fletcher, S. J., M. Iqbal, S. Jabbari, D. Stekel and J. Z. Rappoport (2014). "Analysis of occludin trafficking, demonstrating continuous endocytosis, degradation, recycling and biosynthetic secretory trafficking." *PLoS One* **9**(11): e111176.

Forstermann, U., H. H. Schmidt, J. S. Pollock, H. Sheng, J. A. Mitchell, T. D. Warner, M. Nakane and F. Murad (1991). "Isoforms of nitric oxide synthase. Characterization and purification from different cell types." *Biochem Pharmacol* **42**(10): 1849-1857.

Geddis, A. E. (2010). "Megakaryopoiesis." *Semin Hematol* **47**(3): 212-219.

Gerrits, A. J., E. A. Leven, A. L. Frelinger, 3rd, S. L. Brigstocke, M. A. Berny-Lang, W. B. Mitchell, S. Revel-Vilk, H. Tamary, S. L. Carmichael, M. R. Barnard, A. D. Michelson and J. B. Bussel (2015). "Effects of eltrombopag on platelet

count and platelet activation in Wiskott-Aldrich syndrome/X-linked thrombocytopenia." *Blood* **126**(11): 1367-1378.

Goldberg, A. L. (2003). "Protein degradation and protection against misfolded or damaged proteins." *Nature* **426**(6968): 895-899.

Goley, E. D. and M. D. Welch (2006). "The ARP2/3 complex: an actin nucleator comes of age." *Nat Rev Mol Cell Biol* **7**(10): 713-726.

Gouaux, E. (1997). "The long and short of colicin action: the molecular basis for the biological activity of channel-forming colicins." *Structure* **5**(3): 313-317.

Greenberg, S. M., D. S. Rosenthal, T. A. Greeley, R. Tantravahi and R. I. Handin (1988). "Characterization of a new megakaryocytic cell line: the Dami cell." *Blood* **72**(6): 1968-1977.

Gresele, P., E. Falcinelli, S. Giannini, P. D'Adamo, A. D'Eustacchio, T. Corazzi, A. M. Mezzasoma, F. Di Bari, G. Guglielmini, L. Cecchetti, P. Noris, C. L. Balduini and A. Savoia (2009). "Dominant inheritance of a novel integrin beta3 mutation associated with a hereditary macrothrombocytopenia and platelet dysfunction in two Italian families." *Haematologica* **94**(5): 663-669.

Hahn, Y., T. K. Bera, I. H. Pastan and B. Lee (2006). "Duplication and extensive remodeling shaped POTE family genes encoding proteins containing ankyrin repeat and coiled coil domains." *Gene* **366**(2): 238-245.

Hajjar, K. A. (2003). Molecular basis of fibrinolysis. *Hematology of infancy and childhood*. D. G. O. Nathan, S.H.; Ginsburg, D.; Look, A.T. London, Saunders: 1497-1514.

Hanafy, K. A., J. S. Krumenacker and F. Murad (2001). "NO, nitrotyrosine, and cyclic GMP in signal transduction." *Med Sci Monit* **7**(4): 801-819.

Huang, H., M. Yu, T. E. Akie, T. B. Moran, A. J. Woo, N. Tu, Z. Waldon, Y. Y. Lin, H. Steen and A. B. Cantor (2009). "Differentiation-dependent interactions between RUNX-1 and FLI-1 during megakaryocyte development." *Mol Cell Biol* **29**(15): 4103-4115.

Iglesias, A., K. Anyane-Yeboah, J. Wynn, A. Wilson, M. Truitt Cho, E. Guzman, R. Sisson, C. Egan and W. K. Chung (2014). "The usefulness of whole-exome sequencing in routine clinical practice." *Genet Med* **16**(12): 922-931.

Italiano, J. E., Jr., P. Lecine, R. A. Shivdasani and J. H. Hartwig (1999). "Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes." *J Cell Biol* **147**(6): 1299-1312.

Izumi, R., T. Niihori, N. Suzuki, Y. Sasahara, T. Rikiishi, A. Nishiyama, S. Nishiyama, K. Endo, M. Kato, H. Warita, H. Konno, T. Takahashi, M. Tateyama, T. Nagashima, R. Funayama, K. Nakayama, S. Kure, Y. Matsubara, Y. Aoki and

M. Aoki (2014). "GNE myopathy associated with congenital thrombocytopenia: a report of two siblings." *Neuromuscul Disord* **24**(12): 1068-1072.

Jackson, C. W. and C. C. Edwards (1977). "Biphasic thrombopoietic response to severe hypobaric hypoxia." *Br J Haematol* **35**(2): 233-244.

Johnson, B., S. J. Fletcher and N. V. Morgan (2016). "Inherited thrombocytopenia: novel Insights into megakaryocyte maturation, proplatelet formation and platelet lifespan." *Platelets*: 1-7.

Kaksonen, M., Y. Sun and D. G. Drubin (2003). "A pathway for association of receptors, adaptors, and actin during endocytic internalization." *Cell* **115**(4): 475-487.

Kasahara, K., M. Kaneda, T. Miki, K. Iida, N. Sekino-Suzuki, I. Kawashima, H. Suzuki, M. Shimonaka, M. Arai, Y. Ohno-Iwashita, S. Kojima, M. Abe, T. Kobayashi, T. Okazaki, M. Souri, A. Ichinose and N. Yamamoto (2013). "Clot retraction is mediated by factor XIII-dependent fibrin- α IIb β 3-myosin axis in platelet sphingomyelin-rich membrane rafts." *Blood* **122**(19): 3340-3348.

Kashiwagi, H., S. Kunishima, K. Kiyomizu, Y. Amano, H. Shimada, M. Morishita, Y. Kanakura and Y. Tomiyama (2013). "Demonstration of novel gain-of-function mutations of α IIb β 3: association with macrothrombocytopenia and glanzmann thrombasthenia-like phenotype." *Mol Genet Genomic Med* **1**(2): 77-86.

Kaushansky, K. (2008). "Historical review: megakaryopoiesis and thrombopoiesis." *Blood* **111**(3): 981-986.

Ku, C. S., N. Naidoo and Y. Pawitan (2011). "Revisiting Mendelian disorders through exome sequencing." *Hum Genet* **129**(4): 351-370.

Li, H., J. Ruan and R. Durbin (2008). "Mapping short DNA sequencing reads and calling variants using mapping quality scores." *Genome Res* **18**(11): 1851-1858.

Li, J., A. Mahajan and M. D. Tsai (2006). "Ankyrin repeat: a unique motif mediating protein-protein interactions." *Biochemistry* **45**(51): 15168-15178.

Li, Z., M. K. Delaney, K. A. O'Brien and X. Du (2010). "Signaling during platelet adhesion and activation." *Arterioscler Thromb Vasc Biol* **30**(12): 2341-2349.

Liu, X. F., T. K. Bera, L. J. Liu and I. Pastan (2009). "A primate-specific POTE-actin fusion protein plays a role in apoptosis." *Apoptosis* **14**(10): 1237-1244.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods* **25**(4): 402-408.

Lopez, J. A., R. K. Andrews, V. Afshar-Kharghan and M. C. Berndt (1998). "Bernard-Soulier syndrome." *Blood* **91**(12): 4397-4418.

Luo, S. Z., X. Mo, V. Afshar-Kharghan, S. Srinivasan, J. A. Lopez and R. Li (2007). "Glycoprotein Ibalph forms disulfide bonds with 2 glycoprotein Ibbeta subunits in the resting platelet." *Blood* **109**(2): 603-609.

Machlus, K. R. and J. E. Italiano, Jr. (2013). "The incredible journey: From megakaryocyte development to platelet formation." *J Cell Biol* **201**(6): 785-796.

Machlus, K. R., J. N. Thon and J. E. Italiano, Jr. (2014). "Interpreting the developmental dance of the megakaryocyte: a review of the cellular and molecular processes mediating platelet formation." *Br J Haematol* **165**(2): 227-236.

Magnusson, M., A. C. Brun, H. J. Lawrence and S. Karlsson (2007). "Hoxa9/hoxb3/hoxb4 compound null mice display severe hematopoietic defects." *Exp Hematol* **35**(9): 1421-1428.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure and D. J. Turner (2010). "Target-enrichment strategies for next-generation sequencing." *Nat Methods* **7**(2): 111-118.

Marchand, J. B., D. A. Kaiser, T. D. Pollard and H. N. Higgs (2001). "Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex." *Nat Cell Biol* **3**(1): 76-82.

Mardis, E. R. (2011). "A decade's perspective on DNA sequencing technology." *Nature* **470**(7333): 198-203.

Mardis, E. R. (2013). "Next-generation sequencing platforms." *Annu Rev Anal Chem (Palo Alto Calif)* **6**: 287-303.

Massaad, M. J., N. Ramesh and R. S. Geha (2013). "Wiskott-Aldrich syndrome: a comprehensive review." *Ann N Y Acad Sci* **1285**: 26-43.

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez (2013). "Analysis Tool Web Services from the EMBL-EBI." *Nucleic Acids Res* **41**(Web Server issue): W597-600.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." *Nat Rev Genet* **11**(1): 31-46.

Millikan, P. D., S. M. Balamohan, W. H. Raskind and M. A. Kacena (2011). "Inherited thrombocytopenia due to GATA-1 mutations." *Semin Thromb Hemost* **37**(6): 682-689.

Mitchell, J. A., F. Ali, L. Bailey, L. Moreno and L. S. Harrington (2008). "Role of nitric oxide and prostacyclin as vasoactive hormones released by the endothelium." *Exp Physiol* **93**(1): 141-147.

Mitchell, J. A., R. Lucas, I. Vojnovic, K. Hasan, J. R. Pepper and T. D. Warner (2006). "Stronger inhibition by nonsteroid anti-inflammatory drugs of

cyclooxygenase-1 in endothelial cells than platelets offers an explanation for increased risk of thrombotic events." *FASEB J* **20**(14): 2468-2475.

Mitchell, J. A. and T. D. Warner (1999). "Cyclo-oxygenase-2: pharmacology, physiology, biochemistry and relevance to NSAID therapy." *Br J Pharmacol* **128**(6): 1121-1132.

Moncada, S., R. Gryglewski, S. Bunting and J. R. Vane (1976). "An enzyme isolated from arteries transforms prostaglandin endoperoxides to an unstable substance that inhibits platelet aggregation." *Nature* **263**(5579): 663-665.

Moreau, T., A. L. Evans, L. Vasquez, M. R. Tijssen, Y. Yan, M. W. Trotter, D. Howard, M. Colzani, M. Arumugam, W. H. Wu, A. Dalby, R. Lampela, G. Bouet, C. M. Hobbs, D. C. Pask, H. Payne, T. Ponomaryov, A. Brill, N. Soranzo, W. H. Ouwehand, R. A. Pedersen and C. Ghevaert (2016). "Large-scale production of megakaryocytes from human pluripotent stem cells by chemically defined forward programming." *Nat Commun* **7**: 11208.

Mullins, R. D., J. A. Heuser and T. D. Pollard (1998). "The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments." *Proc Natl Acad Sci U S A* **95**(11): 6181-6186.

Nakeff, A. and B. Maat (1974). "Separation of megakaryocytes from mouse bone marrow by velocity sedimentation." *Blood* **43**(4): 591-595.

Necchi, V., A. Balduini, P. Noris, S. Barozzi, P. Sommi, C. di Buduo, C. L. Balduini, E. Solcia and A. Pecci (2013). "Ubiquitin/proteasome-rich particulate cytoplasmic structures (PaCSs) in the platelets and megakaryocytes of ANKRD26-related thrombo-cytopenia." *Thromb Haemost* **109**(2): 263-271.

Needleman, P., S. Moncada, S. Bunting, J. R. Vane, M. Hamberg and B. Samuelsson (1976). "Identification of an enzyme in platelet microsomes which generates thromboxane A₂ from prostaglandin endoperoxides." *Nature* **261**(5561): 558-560.

Noetzli, L., R. W. Lo, A. B. Lee-Sherick, M. Callaghan, P. Noris, A. Savoia, M. Rajpurkar, K. Jones, K. Gowan, C. L. Balduini, A. Pecci, C. Gnan, D. De Rocco, M. Doubek, L. Li, L. Lu, R. Leung, C. Landolt-Marticorena, S. Hunger, P. Heller, A. Gutierrez-Hartmann, L. Xiayuan, F. G. Pluthero, J. W. Rowley, A. S. Weyrich, W. H. Kahr, C. C. Porter and J. Di Paola (2015). "Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia." *Nat Genet* **47**(5): 535-538.

Noris, P. and C. L. Balduini (2015). "Inherited thrombocytopenias in the era of personalized medicine." *Haematologica* **100**(2): 145-148.

Noris, P., R. Favier, M. C. Alessi, A. E. Geddis, S. Kunishima, P. G. Heller, P. Giordano, K. Y. Niederhoffer, J. B. Bussel, G. M. Podda, N. Vianelli, R. Kersseboom, A. Pecci, C. Gnan, C. Marconi, A. Auvrignon, W. Cohen, J. C. Yu,

A. Iguchi, A. Miller Imahiyero, F. Boehlen, D. Ghalloussi, D. De Rocco, P. Magini, E. Civaschi, G. Biino, M. Seri, A. Savoia and C. L. Balduini (2013). "ANKRD26-related thrombocytopenia and myeloid malignancies." *Blood* **122**(11): 1987-1989.

Noris, P., S. Perrotta, M. Seri, A. Pecci, C. Gnan, G. Loffredo, N. Pujol-Moix, M. Zecca, F. Scognamiglio, D. De Rocco, F. Punzo, F. Melazzini, S. Scianguetta, M. Casale, C. Marconi, T. Pippucci, G. Amendola, L. D. Notarangelo, C. Klersy, E. Civaschi, C. L. Balduini and A. Savoia (2011). "Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families." *Blood* **117**(24): 6673-6680.

Nurden, A. and P. Nurden (2011). "Advances in our understanding of the molecular basis of disorders of platelet function." *J Thromb Haemost* **9 Suppl 1**: 76-91.

Odell, T. T., Jr., C. W. Jackson and T. J. Friday (1970). "Megakaryocytopoiesis in rats with special reference to polyploidy." *Blood* **35**(6): 775-782.

Offermanns, S. (2006). "Activation of platelet function through G protein-coupled receptors." *Circ Res* **99**(12): 1293-1304.

Pai, S. Y. and L. D. Notarangelo (2010). "Hematopoietic cell transplantation for Wiskott-Aldrich syndrome: advances in biology and future directions for treatment." *Immunol Allergy Clin North Am* **30**(2): 179-194.

Palmer, R. M., D. S. Ashton and S. Moncada (1988). "Vascular endothelial cells synthesize nitric oxide from L-arginine." *Nature* **333**(6174): 664-666.

Panchal, S. C., D. A. Kaiser, E. Torres, T. D. Pollard and M. K. Rosen (2003). "A conserved amphipathic helix in WASP/Scar proteins is essential for activation of Arp2/3 complex." *Nat Struct Biol* **10**(8): 591-598.

Pecci, A., G. Biino, T. Fierro, V. Bozzi, A. Mezzasoma, P. Noris, U. Ramenghi, G. Loffredo, F. Fabris, S. Momi, U. Magrini, M. Pirastu, A. Savoia, C. Balduini, P. Gresele and M. Y. H. r. d. Italian Registry for (2012). "Alteration of liver enzymes is a feature of the MYH9-related disease syndrome." *PLoS One* **7**(4): e35986.

Penner, J., L. R. Mantey, S. Elgavish, D. Ghaderi, S. Cirak, M. Berger, S. Krause, L. Lucka, T. Voit, S. Mitrani-Rosenbaum and S. Hinderlich (2006). "Influence of UDP-GlcNAc 2-epimerase/ManNAc kinase mutant proteins on hereditary inclusion body myopathy." *Biochemistry* **45**(9): 2968-2977.

Pinault, N. R., A.; Cortin, V.; Boyer, L. (2013). Ex Vivo Differentiation of Cord Blood Stem Cells into Megakaryocytes and Platelets. *Basic cell culture protocols*. C. D. M. Helgason, C.L. London, Springer. **946**: 205-224.

Pollard, T. D. and G. G. Borisy (2003). "Cellular motility driven by assembly and disassembly of actin filaments." *Cell* **112**(4): 453-465.

Psaila, B., J. B. Bussel, M. D. Linden, B. Babula, Y. Li, M. R. Barnard, C. Tate, K. Mathur, A. L. Frelinger and A. D. Michelson (2012). "In vivo effects of eltrombopag on platelet function in immune thrombocytopenia: no evidence of platelet activation." *Blood* **119**(17): 4066-4072.

Rijken, D. C. and H. R. Lijnen (2009). "New insights into the molecular mechanisms of the fibrinolytic system." *J Thromb Haemost* **7**(1): 4-13.

Robert, X. and P. Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server." *Nucleic Acids Res* **42**(Web Server issue): W320-324.

Ruggeri, Z. M. and G. L. Mendolicchio (2007). "Adhesion mechanisms in platelet function." *Circ Res* **100**(12): 1673-1685.

Sassi, C., R. Guerreiro, R. Gibbs, J. Ding, M. K. Lupton, C. Troakes, S. Al-Sarraj, M. Niblock, J. M. Gallo, J. Adnan, R. Killick, K. S. Brown, C. Medway, J. Lord, J. Turton, J. Bras, U. K. C. Alzheimer's Research, K. Morgan, J. F. Powell, A. Singleton and J. Hardy (2014). "Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT, and PRNP) in late-onset Alzheimer's disease." *Neurobiol Aging* **35**(12): 2881 e2881-2886.

Savoia, A. (2016). "Molecular basis of inherited thrombocytopenias." *Clin Genet* **89**(2): 154-162.

Savoia, A., D. De Rocco, E. Panza, V. Bozzi, R. Scandellari, G. Loffredo, A. Mumford, P. G. Heller, P. Noris, M. R. De Groot, M. Giani, P. Freddi, F. Scognamiglio, S. Riondino, N. Pujol-Moix, F. Fabris, M. Seri, C. L. Balduini and A. Pecci (2010). "Heavy chain myosin 9-related disease (MYH9 -RD): neutrophil inclusions of myosin-9 as a pathognomonic sign of the disorder." *Thromb Haemost* **103**(4): 826-832.

Savoia, A., S. Kunishima, D. De Rocco, B. Zieger, M. L. Rand, N. Pujol-Moix, U. Caliskan, H. Tokgoz, A. Pecci, P. Noris, A. Srivastava, C. Ward, M. C. Morel-Kopp, M. C. Alessi, S. Bellucci, P. Beurrier, E. de Maistre, R. Favier, N. Hezard, M. F. Hurtaud-Roux, V. Latger-Cannard, C. Lavenu-Bombled, V. Proulle, S. Meunier, C. Negrier, A. Nurden, H. Randrianaivo, F. Fabris, H. Platokouki, N. Rosenberg, B. HadjKacem, P. G. Heller, M. Karimi, C. L. Balduini, A. Pastore and F. Lanza (2014). "Spectrum of the mutations in Bernard-Soulier syndrome." *Hum Mutat* **35**(9): 1033-1045.

Schatz, M. C., A. L. Delcher and S. L. Salzberg (2010). "Assembly of large genomes using second-generation sequencing." *Genome Res* **20**(9): 1165-1173.

Schneeberger, K. (2014). "Using next-generation sequencing to isolate mutant genes from forward genetic screens." *Nat Rev Genet* **15**(10): 662-676.

Schulze, H., M. Korpai, J. Hrov, S. W. Kim, J. Zhang, L. C. Cantley, T. Graf and R. A. Shivdasani (2006). "Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis." *Blood* **107**(10): 3868-3875.

Schwarz, U. R., U. Walter and M. Eigenthaler (2001). "Taming platelets with cyclic nucleotides." *Biochem Pharmacol* **62**(9): 1153-1161.

Sessa, W. C. (2004). "eNOS at a glance." *J Cell Sci* **117**(Pt 12): 2427-2429.

Sims, D., I. Sudbery, N. E. Ilott, A. Heger and C. P. Ponting (2014). "Sequencing depth and coverage: key considerations in genomic analyses." *Nat Rev Genet* **15**(2): 121-132.

Small, K. M. and S. S. Potter (1993). "Homeotic transformations and limb defects in Hox A11 mutant mice." *Genes Dev* **7**(12A): 2318-2328.

So, C. W., H. Karsunky, P. Wong, I. L. Weissman and M. L. Cleary (2004). "Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9." *Blood* **103**(8): 3192-3199.

Sorensen, A. L., V. Rumjantseva, S. Nayeb-Hashemi, H. Clausen, J. H. Hartwig, H. H. Wandall and K. M. Hoffmeister (2009). "Role of sialic acid for platelet life span: exposure of beta-galactose results in the rapid clearance of platelets from the circulation by asialoglycoprotein receptor-expressing liver macrophages and hepatocytes." *Blood* **114**(8): 1645-1654.

Spinler, K. R., J. W. Shin, M. P. Lambert and D. E. Discher (2015). "Myosin-II repression favors pre/proplatelets but shear activation generates platelets and fails in macrothrombocytopenia." *Blood* **125**(3): 525-533.

Stasche, R., S. Hinderlich, C. Weise, K. Effertz, L. Lucka, P. Moormann and W. Reutter (1997). "A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-acetylmannosamine kinase." *J Biol Chem* **272**(39): 24319-24324.

Stegner, D., E. J. Haining and B. Nieswandt (2014). "Targeting glycoprotein VI and the immunoreceptor tyrosine-based activation motif signaling pathway." *Arterioscler Thromb Vasc Biol* **34**(8): 1615-1620.

Stitham, J., E. J. Arehart, S. R. Gleim, K. L. Douville and J. Hwa (2007). "Human prostacyclin receptor structure and function from naturally-occurring and synthetic mutations." *Prostaglandins Other Lipid Mediat* **82**(1-4): 95-108.

Strachan, T. R., A. (2011). *Human molecular genetics*. New York, Garland Science.

Stroud, R. M., K. Reiling, M. Wiener and D. Freymann (1998). "Ion-channel-forming colicins." *Curr Opin Struct Biol* **8**(4): 525-533.

Sulonen, A. M., P. Ellonen, H. Almusa, M. Lepisto, S. Eldfors, S. Hannula, T. Miettinen, H. Tynismaa, P. Salo, C. Heckman, H. Joensuu, T. Raivio, A. Suomalainen and J. Saarela (2011). "Comparison of solution-based exome capture methods for next generation sequencing." *Genome Biol* **12**(9): R94.

Thompson, A. A. and L. T. Nguyen (2000). "Amegakaryocytic thrombocytopenia and radio-ulnar synostosis are associated with HOXA11 mutation." *Nat Genet* **26**(4): 397-398.

Thon, J. N. and J. E. Italiano (2010). "Platelet formation." *Semin Hematol* **47**(3): 220-226.

Tong, Y., W. Tempel, L. Nedyalkova, F. Mackenzie and H. W. Park (2009). "Crystal structure of the N-acetylmannosamine kinase domain of GNE." *PLoS One* **4**(10): e7165.

Undas, A. and R. A. Ariens (2011). "Fibrin clot structure and function: a role in the pathophysiology of arterial and venous thromboembolic diseases." *Arterioscler Thromb Vasc Biol* **31**(12): e88-99.

van Ommen, C. H. and M. Peters (2012). "The bleeding child. Part I: primary hemostatic disorders." *Eur J Pediatr* **171**(1): 1-10.

von der Ecken, J., M. Muller, W. Lehman, D. J. Manstein, P. A. Penczek and S. Raunser (2015). "Structure of the F-actin-tropomyosin complex." *Nature* **519**(7541): 114-117.

Wang, F., D. Sambandan, R. Halder, J. Wang, S. M. Batt, B. Weinrick, I. Ahmad, P. Yang, Y. Zhang, J. Kim, M. Hassani, S. Huszar, C. Trefzer, Z. Ma, T. Kaneko, K. E. Mdluli, S. Franzblau, A. K. Chatterjee, K. Johnsson, K. Mikusova, G. S. Besra, K. Futterer, S. H. Robbins, S. W. Barnes, J. R. Walker, W. R. Jacobs, Jr. and P. G. Schultz (2013). "Identification of a small molecule with activity against drug-resistant and persistent tuberculosis." *Proc Natl Acad Sci U S A* **110**(27): E2510-2517.

Warr, A., C. Robert, D. Hume, A. Archibald, N. Deeb and M. Watson (2015). "Exome Sequencing: Current and Future Perspectives." *G3 (Bethesda)* **5**(8): 1543-1550.

Watson, S. P., G. C. Lowe, M. Lordkipanidze, N. V. Morgan and G. consortium (2013). "Genotyping and phenotyping of platelet function disorders." *J Thromb Haemost* **11 Suppl 1**: 351-363.

Whiteheart, S. W. (2011). "Platelet granules: surprise packages." *Blood* **118**(5): 1190-1191.

Wiener, M., D. Freymann, P. Ghosh and R. M. Stroud (1997). "Crystal structure of colicin Ia." *Nature* **385**(6615): 461-464.

Wiestner, A., R. J. Schlemper, A. P. van der Maas and R. C. Skoda (1998). "An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia." *Nat Genet* **18**(1): 49-52.

Witthuhn, B. A., F. W. Quelle, O. Silvennoinen, T. Yi, B. Tang, O. Miura and J. N. Ihle (1993). "JAK2 associates with the erythropoietin receptor and is tyrosine phosphorylated and activated following stimulation with erythropoietin." *Cell* **74**(2): 227-236.

Worthey, E. A., A. N. Mayer, G. D. Syverson, D. Helbling, B. B. Bonacci, B. Decker, J. M. Serpe, T. Dasu, M. R. Tschannen, R. L. Veith, M. J. Basehore, U. Broeckel, A. Tomita-Mitchell, M. J. Arca, J. T. Casper, D. A. Margolis, D. P. Bick, M. J. Hessner, J. M. Routes, J. W. Verbsky, H. J. Jacob and D. P. Dimmock (2011). "Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease." *Genet Med* **13**(3): 255-262.

Zhang, M. Y., J. E. Churpek, S. B. Keel, T. Walsh, M. K. Lee, K. R. Loeb, S. Gulsuner, C. C. Pritchard, M. Sanchez-Bonilla, J. J. Delrow, R. S. Basom, M. Forouhar, B. Gyurkocza, B. S. Schwartz, B. Neistadt, R. Marquez, C. J. Mariani, S. A. Coats, I. Hofmann, R. C. Lindsley, D. A. Williams, J. L. Abkowitz, M. S. Horwitz, M. C. King, L. A. Godley and A. Shimamura (2015). "Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy." *Nat Genet* **47**(2): 180-185.

Zhen, C., F. Guo, X. Fang, Y. Liu and X. Wang (2014). "A family with distal myopathy with rimmed vacuoles associated with thrombocytopenia." *Neurol Sci* **35**(9): 1479-1481.

Zimmet, J. and K. Ravid (2000). "Polyploidy: occurrence in nature, mechanisms, and significance for the megakaryocyte-platelet system." *Exp Hematol* **28**(1): 3-16.

Zwaal, R. F., P. Comfurius and E. M. Bevers (2004). "Scott syndrome, a bleeding disorder caused by defective scrambling of membrane phospholipids." *Biochim Biophys Acta* **1636**(2-3): 119-128.