# Three disparate studies on the CD18 integrin : expression and biological functions of the human CD18 integrin in health and disease 

Cher, Bernard Toh Hui

2011

Cher, B. T. H. (2011). Three disparate studies on the CD18 integrin : expression and biological functions of the human CD18 integrin in health and disease. Doctoral thesis, Nanyang Technological University, Singapore.
https://hdl.handle.net/10356/50537
https://doi.org/10.32657/10356/50537


Three Disparate Studies on the CD18 Integrin: Expression and biological functions of the human CD18 integrin in health and disease

Cher Toh Hui Bernard SCHOOL OF BIOLOGICAL SCIENCES

2011

# Three Disparate Studies on the CD18 Integrin: <br> Expression and biological functions of the human CD18 integrin in health and disease 

Cher Toh Hui Bernard SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in fulfilment of the requirements for the degree of Doctor of Philosophy

# Three Disparate Studies on the CD18 Integrin: Expression and biological functions of the human CD18 integrin in health and disease 

# A Thesis for the Degree of Doctor of Philosophy (Ph.D) completed at the 

Nanyang Technological University

Cher Toh Hui Bernard

Supervised by Professor Alex Law
School of Biological Sciences
Nanyang Technological University

## Acknowledgements

My scholarship tenure at NTU was completed in 2 different laboratories. It is the work completed under the supervision of Professor Alex Law that is compiled in this thesis. In this short note, I wish to thank a number of people who have touched my life in countless ways.

I am grateful to Professor Alex Law for his guidance and inspiration as my thesis supervisor. His generous nature and sharp intellect made it possible for me to complete this work within the three years I spent in his laboratory. I would like to thank Asst. Prof. Xavier Roca for sharing his expertise in 3 ' splicing, and for excellent suggestions on manuscript preparation.

I am thankful for the friendship and expertise in microscopy that Assoc. Prof. Li Hoi Yeung and Dr. Eric Wong Chi Hang have shared with me, and to Asst. Prof. Koh Cheng Gee, Assoc. Prof. Ravi Kambadur for thesis advice. In addition, I am grateful to Assoc. Prof. Valerie Lin, Assoc. Prof. Peter Dröge, Assoc. Prof. Tan Suet Mien, Asst. Prof. Koh Cheng Gee and Dr. Kenneth Yu Chin-Wen for granting me opportunities to teach classes, thus allowing me to further develop my teaching and communication skills during my time in NTU.

I would like to thank my laboratory members, and friends in SBS for their friendship and kindness, especially Dr. Xiao Zhou for unreservedly sharing with me his experiences, both scientific and situational, and Ms. Manisha Cooray for excellent technical support and even better humor.

Last but not least, I wish to thank my family; my parents Mr. Cher Lay Kee and Ms. Yeo Chye Ping, my paternal aunt, Ms. Cher Sok Kew, my brother, Mr. Lennard Cher, my beloved fiancée, Ms. Julia Eng, and my bosom friends, Mr. Du Xiao Xian, Mr. Michael Kong Zhen Liang and Mr. Mark Chong Wei Ho, for their selfless support and encouragement throughout my time in NTU.

## List of Abbreviations

| APC | Antigen Presenting Cells |
| :--- | :--- |
| 3'ss | 3' splice site |
| 5'ss | 5' splice site |
| BPS | Branch Point Sequence |
| CD11aF | FLAG tagged CD11a |
| CD18_HIS | HIS tagged CD18 |
| CD18F | FLAG tagged CD18 |
| DNA | Deoxyribonucleic Acid |
| ECM | Extracellular Matrix |
| emPAI | Exponentially Modified Protein Abundance Index |
| FSC | Forward Scatter |
| gDNA | Genomic DNA |
| HEK293 | Human Embroynic Kidney 293 cell line |
| iC3b | inactivated Complement 3b fragment |
| ICAM | Intercellular Adhesion Molecule |
| IFN-ץ | Interferon gamma |
| IL | Inter-Leukine |
| IPCR | Inverse PCR |
| ITGB2 | Integrin Beta 2 |
| LAD | Leukocyte Adhesion Deficiency |
| LFA-1 | Leukocyte Function-Associated Antigen |
| Mac-1 | Macrophage-1 antigen |
| NK cells | Natural Killer Cells |
| NMD | Nonsense Mediate Decay |
| NTR | N-terminal Region of CD18 |
| PCR | Polymerase Chain Reaction |
| PMBC | Peripheral Monocytic Blood Cells |
| PPT | Poly Pyrimidine Tract |
| PSGL-1 | P-selectin glycoprotein ligand-1 |
| PTTG1IP | Pituitary Tumor-transforming 1 interacting Protein |
| RBC | Red Blood Cells |
| RNA | Ribonucleic acid |
| SDS-PAGE | Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis |
| SSC | Side Scatter |
| t-CD18F | Truncated FLAG tagged CD18 |
| TNFa | Tumor Necrosis Factor alpha |
| U2AF | U2 Auxillary Factor |
| WBC | White Blood Cells |
|  |  |


#### Abstract

Three disparate studies relating to the CD18 integrin are described in this thesis. Firstly, a particularly elusive CD18 mutation in an Leukocyte Adhesion Deficiency Type I (LAD-I) patient was studied. Genomic DNA and cDNA from an LAD-I patient and her family was analyzed and characterized. This mutation was a gross deletion where the CD18 gene was truncated and joined to the next gene on the chromosome, PTTG1IP.

Secondly, the effect of a relatively proximal AG dinucleotide located +11 bp upstream of a cryptic 3' splice site (3'ss) on 3' splicing was analyzed. In-silico analysis using 3'ss strength prediction tools showed that the cryptic 3 'ss at the +11 position had relatively higher scores among the potential 3 'ss in intron 2 . However, when the authentic 3 'ss was lost due to $10 \mathrm{C}>\mathrm{A}$, the +113 'ss was not used. The AG dinucleotide located +11 bp upstream of the +11 cryptic 3 'ss was analyzed and demonstrated to repress use of the +11 cryptic 3 'ss.

Thirdly, the mechanism of integrin monomer retention was studied. In general, integrins are expressed on the cell surface as obligate non-covalent heterodimers. As a starting point towards understanding integrin dimerization, this study focused on intracellular retention of the CD18 subunit expressed in HEK293 cells in the absence of a corresponding CD11a subunit. The $\beta$ I domain of CD18 was demonstrated to be critical for CD18 integrin retention. Calnexin and LMAN1 (Lectin, Mannose binding protein 1) were subsequently identified as strong candidates to mediate intracellular retention of the partially unfolded CD18 protein.


## Contents

Chapter 1 : Introduction ..... 1
1.1 Overview of the human immune response. ..... 1
1.2 Cell adhesion molecules and extravasation ..... 1
1.3 Integrins ..... 4
1.4 Inside-out activation of the integrins ..... 7
1.5 The CD18 protein ..... 8
1.5.1 ITGB2 genomic DNA sequence ..... 10
1.5.2 PSI domain ..... 11
1.5.3 Hybrid domain ..... 12
1.5.4 I-like domain ( $\beta \mathrm{I}$ domain) ..... 13
1.5.5 IEGF domains ..... 15
1.5.6 $\beta$ TD domain ..... 16
1.5.7 TM domain ..... 16
1.5.8 Cytoplasmic domain ..... 17
1.1 Intrinsic folding of the CD11a and CD18 protein ..... 18
1.2 Leukocyte Adhesion Deficiency ..... 22
1.2.1 Leukocyte Adhesion Deficiency type I. ..... 22
1.2.2 Other Leukocyte Adhesion Deficiencies ..... 24
1.3 Pre-mRNA splicing in human genes. ..... 26
1.3.1 The major and minor spliceosome ..... 26
1.3.2 U2-type GU-AG intron splicing ..... 27
1.3.3 Disruption of pre-mRNA splicing ..... 32
1.3.4 . 3'ss mutations ..... 33
1.4 Studies in this thesis ..... 34
Chapter 2 : Materials and Methods ..... 37
2.1 Materials ..... 37
2.1.1 Reagents ..... 37
2.1.2 Cells ..... 37
2.1.3 Cell culture ..... 37
2.1.4 Minigene and plasmid construction ..... 38
2.1.4.1 Integrin expression plasmids ..... 38
2.1.4.2 Minigene construction for splicing project ..... 39
2.1.4.3 HIS tagged CD18 plasmid (CD18 HIS) construction ..... 40
2.1.4.4 FLAG tagged CD18 plasmid (CD18F) construction ..... 41
2.1.4.5 Truncated CD18 (t-CD18) construction ..... 42
2.1.4.6 FLAG tagged t-CD18 (t-CD18F) construction ..... 42
2.1.4.7 FLAG tagged CD11a (CD11aF) construction ..... 43
2.1.5 Epitope Mapping of the MEM148 antibody ..... 44
2.1.6 Antibodies ..... 45
2.1.7 General Buffers and solutions ..... 47
2.2 Methods. ..... 49
2.2.1 Cell cryofreezing and recovery ..... 49
2.2.2 Plasmid DNA extraction and purification ..... 49
2.2.3 DNA quantitation ..... 50
2.2.4 Competent cells ..... 51
2.2.5 Reverse Transcription ..... 51
2.2.6 Genomic DNA and cDNA extraction for LAD-I project ..... 52
2.2.7 Polymerase chain reaction ..... 53
2.2.8 PCR amplification for 3'ss analysis ..... 53
2.2.9 Inverse PCR for LAD-I project ..... 53
2.2.10 Cycling parameters for Site Directed Mutagenesis ..... 54
2.2.11 DNA sequencing ..... 54
2.2.12 Gel electrophoresis ..... 54
2.2.13 Gel purification of PCR products ..... 54
2.2.14 Fluorescence Activated Cell Sorting (FACS) Analysis ..... 55
2.2.15 Polyfection ..... 55
2.2.15.1 Lipofection ..... 56
2.2.16 Cell Surface Biotinylation ..... 57
2.2.17 Conjugation of rabbit anti-mouse IgG to protein A sepharose beads ..... 57
2.2.18 Immunoprecipitation ..... 58
2.2.19 Reciprocal IP of calnexin myc, CD18F and t-CD18 ..... 59
2.2.20 Sodium dodecyl sulphate polyacrylamide gel electrophoresis ..... 59
2.2.21 Western Blotting ..... 60
2.2.22 Cell adhesion assay ..... 60
2.2.23 Quantitative real-time PCR ..... 61
2.2.24 Confocal Microscopy ..... 62
2.2.25 HIS tag purification ..... 63
2.2.26 FLAG tag purification ..... 64
2.2.27 Liquid-Chromatography-Mass-Spectrometry ..... 65
Chapter 3 : Characterizing a novel CD18 mutation in an LAD-I patient ..... 66
3.1 Introduction ..... 66
3.2 Summary of results for the first LAD-I mutation ..... 68
3.2.1 Patient ..... 68
3.2.2 Two novel intronic point mutations were identified from gDNA sequencing
of the ITGB2 gene in the patient ..... 68
3.2 .3 cDNA analysis ..... 69
3.2.4 cDNA and gDNA PCR products extending beyond exon 11 in the father and patient were not polymorphic ..... 70
3.3 Characterizing the second LAD-I mutation ..... 71
3.3.1 Leukocytes from the patient did not express CD18 ..... 71
3.3.2 Inverse PCR strategies ..... 76
3.3.3 Inverse PCR using EcoRI demonstrated that the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allelein the patient possessed a large deletion of at least 27 kb79
3.3.4 A 27703 bp deletion occurred after intron 11 in the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele of the patient allele, inherited from the father ..... 84
3.3.5 The hybrid CD18-PTTG1IP cDNA is expressed at the mRNA level ..... 85
3.3.6 Summary of genetic analyses. ..... 86
3.4 Discussion ..... 91
3.4.1 Retrospective discussion of the IPCR strategies ..... 93
Chapter 4 : 3' Splice Site Strength Analysis ..... 95
4.1 Introduction ..... 95
4.2 Results ..... 95
4.2.1 In silico analysis identified 11 potential 3'ss near the authentic 3'ss in intron 2 ..... 95
4.2.2 In silico analysis of 3'ss strength showed that the 3 'ss at -43 and +11 have higher scores than the authentic 3'ss ..... 96
4.2.3 In silico analyses predicted the relative preference of +11 over the -43 'ss after ablation of the +1 AG dinucleotide ..... 98
4.2.4 FACS analysis of transfectants was consistent with predicted splicing patterns ..... 99
4.2.5 RT-PCR analysis showed activation of splicing at +11 after ablation of the authentic AG ..... 100
4.2.6 Mutating the -43 AG dinucleotide ..... 101
4.2.7 The minigenes carrying both the -433 'ss and $-10 \mathrm{C}>$ A mutations showed no membrane expression. ..... 102
4.2.8 RT-PCR showed that no detectable splicing occurred upon the -43 3'ss mutation ..... 102
4.2.9 Analysis of remnant plasmid DNA ..... 103
4.2.10 Summary of results from 3'ss analysis ..... 104
4.2.11 NN analysis of all 3'ss sites in CD18 identified 274 3'ss sequences ..... 104
4.2.12 3'ss scoring algorithms predicted authentic 3'ss in CD18 with 60-73 \% accuracy ..... 106
4.3 Discussion ..... 109
4.3.1 A review of other LAD-I mutations ..... 109
4.3.2 A review of other 3'ss mutations leading to LAD-I ..... 110
4.3.3 Potential studies of 3'ss splicing mutations in LAD-I ..... 111
4.3.4 The AG dinucleotide +11 bp upstream repressed use of the +113 'ss ..... 114
Chapter 5 : Intracellular retention and folding of the integrin proteins ..... 119
5.1 Introduction ..... 119
5.1.1 Immunoprecipitation for LC-MSMS analysis of integrin interacting proteins ..... 119
5.1.1.1 HIS tagged CD18 ..... 119
5.1.1.2 FLAG tag purification of CD18F, $\mathrm{t}-\mathrm{CD} 18 \mathrm{~F}$ and CD 11 aF ..... 120
5.2 Results ..... 121
5.2.1 Western blot on JB2.7 and SKB2.7 showed that JB2.7 did not express CD11a ..... 121
5.2.2 Antibody IP followed by LCMSMS analysis identified 14 candidate integrin interacting proteins ..... 121
5.2.3 HIS tag purification from human cell line lysate was unspecific ..... 126
5.2.4 FACS analysis of CD18F demonstrated that the FLAG tag did not interfere with dimer formation ..... 129
5.2.5 FLAG tag purification was superior to HIS tag purification ..... 129
5.2.6 Silver staining showed that unspecific proteins could bind to the M2 beads ..... 131
5.2.7 LC-MSMS consistently identified 72 integrin binding proteins over 3 biological repetitions ..... 133
5.2.8 FACS analysis of t-CD18 demonstrated that the $\beta \mathrm{I}$ domain was essential for intracellular retention of CD18 ..... 140
5.2.9 Analysis of known protein functions suggested calnexin as a candidate integrin retention protein ..... 141
5.2.10 Co-IP of calnexin from CD18F and t-CD18 demonstrated that calnexin associated with other domains of CD18, including the $\beta \mathrm{I}$ domain ..... 141
5.2.11 FACS analysis of t -CD18F and CD11aF showed that the t -CD18F protein could be expressed as a monomer ..... 143
5.2.12 Combining all FLAG purification data, analysis yielded 127 consistent hits ..... 144
5.2.13 Comparing t-CD18F and CD18F LCMS data, LMAN1 did not bind the $\beta$ I domain ..... 147
5.3 Discussion ..... 152
5.3.1 Integrin IP using antibodies ..... 152
5.3.2 HIS tag purification ..... 152
5.3.3 FLAG tag purification. ..... 153
5.3.3.1 CD18F vs CD11a/CD18F ..... 154
5.3.3.2 Analysis of CD11a/CD18F vs CD11a/t-CD18F showed that CD11a associated intracellularly with t-CD18F ..... 156
5.3.3.3 Chaperone and Protein modification proteins ..... 158
5.3.3.4 Proteasome and ribosomal proteins ..... 158
5.3.3.5 Calnexin as an integrin retaining protein ..... 159
5.3.3.6 Mitochondria lucosiproteins ..... 161
5.3.4 LMAN1 may mediate retention of CD18 by binding the $\beta$ I domain ..... 162
5.4 A brief conclusion on the studies of integrin monomer retention ..... 163
Chapter 6 : Miscellaneous Experiments ..... 164
6.1 Analyzing the MEM148 epitope ..... 164
6.1.1 Introduction ..... 164
6.1.2 Results ..... 167
6.1.2.1 Increasing M/E or MEM148 concentration in both MEM148 (Serotech) and MEM148 Ascites, did not improve reporter activity ..... 167
6.1.2.2 MEM148 reporter activity was more efficient in SKW3 ..... 170
6.1.2.3 Results from epitope mapping of MEM148 were similar to previous results ..... 171
6.1.3 Discussion ..... 173
Chapter 7 : Conclusion ..... 175
References ..... 177
Appendix I: Experiments repeated for the LAD-I publication ..... 191
I. 1 Sequencing of the intron 2 cDNA product from patient cDNA ..... 191
I. 2 The C to A point mutation in intron 2 resulted in a splicing defect ..... 192
I. 3 Genomic DNA sequencing demonstrated that the patient was polymorphic at$\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$195
I. 4 cDNA and genomic DNA PCR products extending beyond exon 11 in the father and patient were not polymorphic ..... 198
Appendix II: Protein functions of all proteins consistently identified in IP-LCMS .....  203
Appendix III: Genomic DNA sequence for $I T G B 2$ variant 1 ..... 225

## Chapter 1 : Introduction

### 1.1 Overview of the human immune response

The human immune response is a robust system composed of the innate immune response and the adaptive immune response. Upon breaching the innate physical barriers of the human body, many pathogens are recognized and destroyed by macrophages of the innate immune system which reside in the submucosal tissues (Navab, et al., 1988). This is followed by the further recruitment of leukocytes such as neutrophils, eosinophils and basophils of the innate immune system to the site of infection, occurring within 4 to 96 hours after infection (Agace, et al., 1995; Murphy, et al., 2007). The innate immune system recognizes pathogens by discriminating between cell surface molecules displayed on host cells (self) and foreign antigens (Feizi, 2000).

During this time, antigen presenting cells (APC) of the adaptive immune system stimulates the adaptive immune system to mount a specific lymphocytic response against the infectious agent, culminating in antigen specific immune responses 96 hours after infection (Murphy, et al., 2007). During an immune challenge, the immune response can be further modulated by cross-talk within cells of the innate immune system and between cells of both the innate and adaptive immune response (Andoniou, et al., 2005; Dong Kim, et al., 2007; Silva, et al., 2005; Tanriver, et al., 2009).

### 1.2 Cell adhesion molecules and extravasation

Cytokines are small proteins that are released by cells in response to an activating stimulus such as a bacterial infection (Larsson, et al., 1999). In turn, chemokines are a
group of chemoattractant cytokines that serve a two-fold function. Firstly, chemokines stimulate an 'inside-out' activation of leukocyte integrins to enable firm adhesion to the endothelium at sites of infection. Secondly, the chemokine concentration gradient guides migration of the leukocytes to sites of infections (Kunkel and Butcher, 2002; Yoshie, 2000).

During local inflammation, dilation of blood vessels due to the release of local inflammatory cytokines such as Tumor Necrosis Factor alpha (TNF $\alpha$ ) results in slower blood flow (Pfeffer, 2003). This facilitates the first wave of leukocyte extravasation by the neutrophils in the innate immune response (Gompertz and Stockley, 2000), where leukocytes cross the endothelium and then continue to migrate toward the site of infection, guided along by the chemokine gradient of molecules such as CXCL8 (Bajt, et al., 2001). During an infection, neutrophils are the first leukocyte lineage to cross the blood vessel boundaries.

The engagement of leukocyte cell adhesion molecules to endothelial cell surface ligands enables leukocyte rolling, adhesion and extravasation to the site of infection. Leukocytes are initially tethered to the vascular surface by interactions between $E$ and P-selectin molecules expressed on the vascular endothelium surface and the sulfated-sialyl-Lewis ${ }^{\mathrm{x}}$ functional group found on some leukocyte glycoproteins. In particular, P-selectin binds to the P-selectin glycoprotein ligand-1 (PSGL-1), which is expressed on the surface of leukocytes (McEver, 1997).

Weak interactions between selectins and sialyl-Lewis ${ }^{x}$ allows rolling adhesion of the leukocytes along the endothelium, slowing leukocyte movement (Alon, et al., 1995;

Chen, et al., 1997). As a result of cytokine signaling through G-protein coupled, cytokine receptors on the leukocytes (Stein, et al., 2000), the leukocyte integrin, Lymphocyte Function-associated Antigen-1 (LFA-1) is activated through an insideout mechanism to engage endothelial Inter-Cellular Adhesion Molecule-1 (ICAM-1) (Issekutz, 1995; Rothlein, et al., 1986). In turn, the expression of ICAM-1 on the endothelium can be upregulated by inflammatory cytokines such as TNF $\alpha$, IL-1 $\beta$, and gamma interferon (IFN- $\gamma$ ) (van de Stolpe and van der Saag, 1996).

The interaction between activated LFA-1 and ICAM-1 results in firm adhesion of leukocytes to the endothelium, allowing the leukocyte to squeeze through endothelial cells (extravasate). While various other molecules are involved in the intermediate processes, intravascular crawling toward chemoattractants (chemokines) is dependent on LFA-1 and Macrophage-1 antigen (Mac-1) (Ley, et al., 2007; Long, 2011; Murphy, et al., 2007; Springer, 1990).


Figure 1.1 Illustration of leukocyte adhesion cascade from Ley et al (2007). Leukocyte recruitment to sites of infection occurs in response to chemokine gradients, and is dependent on the interaction between cell surface ligands and cell adhesion molecules to enable leukocyte rolling, adhesion and extravasation. Some of the key adhesion molecules and ligands, including the integrins (LFA-1, MAC-1) and ICAM-1 are depicted in this figure, alongside other molecules involved in the intermediate processes of leukocyte extravasation.

### 1.3 Integrins

Integrins are a family of type I transmembrane glycoproteins which are expressed as obligate non-covalent heterodimers of integrin $\alpha$ and $\beta$ subunits. Each integrin subunit contains an intracellular domain that interacts with the cytoskeleton, a transmembrane domain and an extracellular domain that interacts either with intercellular cell adhesion molecules or components of the extracellular matrix (ECM) (Hynes, 2002).

Integrins lack intrinsic enzymatic activity and instead mediate bidirectional "outsidein" and "inside out" biological signals by recruiting signaling proteins to their cytoplasmic tails, which adopt different conformations when bound by different cytoplasmic partners, influencing the conformation of the extracellular domain (ectodomain) (Gahmberg, et al., 2009; Hynes, 2002; Xiong, et al., 2003a). By mediating intracellular signals in response to the extracellular environment, integrins play critical roles in a wide range of biological processes, including cell to cell
interactions, cell to extracellular matrix interactions, cell migration, cell cycle control, cell survival, gene expression and rearrangement of the actin cytoskeleton (Cary, et al., 1999; Caswell and Norman, 2006; Clark and Brugge, 1995; Giancotti, 1997; Hynes, 1992; Hynes, 2002).

A total of 18 different $\alpha$ subunits and 8 different $\beta$ subunits have been identified in humans, from which 24 distinct combinations of integrins are known (Hynes, 2002). This indicates that only select permutations of integrin subunits can be co-expressed on the cell surface.

Among these, the $\beta 2$ integrins form a family of four heterodimeric proteins that are expressed only in leukocytes; $\alpha \mathrm{L} \beta 2$ (CD11a/CD18; LFA-1), $\alpha \mathrm{M} \beta 2$ (CD11b/CD18; MAC-1; CR3), $\alpha \times \beta 2$ (CD11c/CD18, p150,95, CR4) and $\alpha \mathrm{D} \beta 2$ (CD11d/CD18) (Sanchez-Madrid, et al., 1983; Wong, et al., 1996). $\beta 2$ integrins have been shown to be important for normal functioning of the innate and adaptive immune response by mediating leukocyte adhesion to substrates, and are critical for leukocyte extravasation (Hynes, 2002).

Genes encoding the four leukocyte specific $\alpha$ subunits are located on chromosome 16, while the gene encoding the CD18 subunit is located on chromosome 21 (Corbi, et al., 1988; Marlin, et al., 1986; Solomon, et al., 1988). Based on amino acid alignment of the protein sequence, sequence identity between $\alpha \mathrm{M}$ (Springer, et al., 1979), $\alpha \mathrm{X}$ (Springer, et al., 1986) and $\alpha \mathrm{D}$ (Wong, et al., 1996) ranges between 60-66 \% (Harris, et al., 2000). In contrast, the CD11a (Kurzinger, et al., 1981) amino acid sequence
shares only about $35 \%$ of homologous sequences with these proteins (Harris, et al., 2000).

While CD18 and CD11a are expressed in all leukocytes (Hyun, et al., 2009; Kurzinger, et al., 1981; Martz, 1987), the other leukocyte integrins are cell lineage specific. For example, CD11b and CD11c are expressed predominantly on monocytes and granulocytes (Fagerholm, et al., 2006; Hyun, et al., 2009; Springer, et al., 1979; Springer, et al., 1986), while CD11d is strongly expressed in tissue-specialized cells such as the splenic red pulp macrophages or macrophage foam cells inside atherosclerotic plaques (Van der Vieren, et al., 1995).

In addition to ICAM-1, another five LFA-1 ligands expressed on various cell types contribute to leukocyte extravasation and transmigration. These include ICAM2, ICAM-3, ICAM4, ICAM5, and the junctional adhesion molecule 1 (JAM-1) (Hyun, et al., 2009; Ostermann, et al., 2002). The primary LFA-1 ligand is ICAM-1 (CD54), which is expressed in leukocytes and endothelial cells together with ICAM-2 (Sigal, et al., 2000). Of the remaining, ICAM-3 is expressed in leukocytes, ICAM-4 is expressed in red blood cells, ICAM-5 is restricted to expression in neuronal cells and JAM-1 is expressed at tight junctions of epithelial cells (Gahmberg, et al., 2009).

Mac-1 is also involved in leukocyte extravasation and intra-luminal crawling. In addition, Mac-1 is involved in phagocytosis (Fagerholm, et al., 2006; Pluskota, et al., 2008; Schleiffenbaum, et al., 1989). Compared to LFA-1, Mac-1 binds to a much greater number of ligands. Till date, there are more than 30 known ligands for Mac-1, including ICAM-1, ICAM-2, VCAM, iC3b, factor X, fibrinogen and even denatured proteins (Hyun, et al., 2009).

### 1.4 Inside-out activation of the integrins

Integrins adopt a bent conformation when in the inactive, low avidity state (Tang, et al., 2005; Xie, et al., 2004). However, inside-out activation of leukocyte integrins can occur upon chemokine stimulation (Alon and Feigelson, 2009; Laudanna, et al., 2002) , or agonist binding to the leukocyte B and T-cell receptors (Arana, et al., 2008; Zell, et al., 1999).

Upon stimulation, upstream Guanine nucleotide Exchange Factors (GEFs) activate Rap1 GTPase (Katagiri, et al., 2000). In particular, the $\mathrm{Ca}^{2+}$ and diacylglycerolregulated GEF-1 (CalDAG-GEF1) is a signaling protein that has been implicated in the regulation of leukocyte adhesion and migration (Pasvolsky, et al., 2007). In turn, Rap1 GTPase activates its downstream effectors, Rap1-GTP-interacting adaptor molecule (RIAM) (Lafuente and Boussiotis, 2006) and Ras association domaincontaining protein 5 (RapL) (Katagiri, et al., 2004), which bind to the integrin cytoplasmic tail and encourage the association of talin (Knezevic, et al., 1996) with the cytoplasmic tail (Lee, et al., 2009).

Although the sequence of protein binding is currently unclear, other proteins such as paxillin, radixin, cytohesin, $\alpha$-actinin and kindlin-3 are then recruited to the cytoplasmic tail, facilitating separation of the integrin cytoplasmic tail and effecting conformational changes required for high-affinity ligand binding and coupling of integrins to the actin cytoskeleton (Abram and Lowell, 2009; Horwitz, et al., 1986).

Talin is an actin-interacting protein that binds actin though the I/LWEQ motif at their carboxy-terminal tail domain (Gingras, 2008; Smith and McCann, 2007). Another
domain at the N -terminus of talin, the FERM (four point one, ezrin, radixin and moesin) domain, binds the conserved NPXY motif in the cytoplasmic tail of the integrin $\beta$ subunit at the cell membrane (Calderwood, 2003; Garcia-Alvarez, 2003; Tanentzapf and Brown, 2006).

The FERM domain is a variant of the phosphotyrosine binding (PTB) domain that is also found in the kindlins (including kindlin-3), which use the PTB domain to bind NPXY motifs in the cytoplasmic tail of the integrins, at a position distinct from that used by talin. The binding of both talin and the kindlins to the membrane expressed integrin cytoplasmic tail is required for complete activation of the integrins (Ma, et al., 2008; Montanez, 2008; Moser, et al., 2009; Tadokoro, 2003).


Figure 1.2 Key signaling pathways implicated in integrin inside out signaling as described in the text. This diagram was taken from Abram and Lowell (2009).

### 1.5 The CD18 protein

The ITGB2 gene located on Chromosome 21 (21q22.3) codes for the CD18 protein (also known as the $\beta 2$ integrin subunit) (Marlin, et al., 1986; Solomon, et al., 1988).

The CD18 protein consists of eight domains in the ectodomain, namely, the PSI (Plexin-Semaphorin-Integrin) domain, hybrid domain, I-like domain (also referred to as the $\beta \mathrm{I}$ domain), IEGF-1 (Integrin-Epidermal Growth Factor), IEGF-2, IEGF-3, IEGF-4, and the tail domain ( $\beta$ TD). The ectodomain is followed by a 23 residue transmembrane domain (TD), and a 46 residue cytoplasmic domain (CD). The CD18 protein contains two domain insertions; the $\beta I$ domain is inserted into the hybrid domain, and the hybrid domain (containing the inserted $\beta I$ domain), is inserted into the PSI domain (Xiong, et al., 2004). A schematic diagram of the CD18 linear organization is presented in Figure 1.3. Recently, the crystal structure of CD18 in complex with CD11c was solved (Xie, et al., 2010). A cartoon model of the solved integrin structure is presented in Figure 1.4.


Figure 1.3 Linear organization of the CD18 protein, based on the amino acid sequence. The $\boldsymbol{\beta I}$ domain is inserted into the hybrid domain, which is in turn inserted into the PSI domain. The four respective IEGF domains are indicated by the numbers 1 through to 4.


Figure 1.4 A cartoon model of the extended CD11c/CD18 taken from Xie et al [2010]. This model was achieved by adjusting domain interfaces at and near the knees (the junction between the upper leg and lower leg).

Right panel: An adaptation of the cartoon model, represented in boxes to provide an overview of the general organization of CD11c/CD18. C1 and C2: CD11c calf-1 and calf-2; E1 to E4: CD18 IEGF1 to I-EGF-4, TD: $\boldsymbol{\beta}$ tail.

### 1.5.1 ITGB2 genomic DNA sequence

The annotated genomic DNA sequence of $\operatorname{ITGB2}$ is 47886 bp long, inclusive of the $5^{\prime}$ and 3' untranslated regions (NCBI reference sequence NG_007270.2). There are two variants for $I T G B 2$, although no information is available regarding the expression profile of each variant. Sequentially, the first transcriptional start codon is located at

5001 bp . This codon codes for variant 2 . The second transcriptional start codon is located at 12789 bp , coding for variant 1 . Variant 1 encodes the longer mRNA sequence, but since the translational start codon is located in exon 2, both variants code for the same ITGB2 protein. The positions of the exons, introns and the nucleic acid lengths of these features are summarized in Table 1.1.

| Exon no. | Start <br> Position | End <br> Position | Exon Length <br> (bp) | Intron no. | Intron Length <br> (bp) |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |
| Exon 1 (Variant 2) | 5001 | 5185 | 185 | Intron 1 (Variant 2) | 17869 |
| Exon 1 (Variant 1) | 12789 | 13018 | 230 | Intron 1 (Variant 1) | 10036 |
| Exon 2 | 23054 | 23114 | 61 | Intron 2 | 353 |
| Exon 3 | 23467 | 23555 | 89 | Intron 3 | 3189 |
| Exon 4 | 26744 | 26924 | 181 | Intron 4 | 3380 |
| Exon 5 | 30304 | 30474 | 171 | Intron 5 | 1632 |
| Exon 6 | 32106 | 32347 | 242 | Intron 6 | 1017 |
| Exon 7 | 33364 | 33519 | 156 | Intron 7 | 1158 |
| Exon 8 | 34677 | 34772 | 96 | Intron 8 | 4007 |
| Exon 9 | 38779 | 38868 | 90 | Intron 9 | 1427 |
| Exon 10 | 40295 | 40435 | 141 | Intron 10 | 1408 |
| Exon 11 | 41843 | 42030 | 188 | Intron 11 | 1587 |
| Exon 12 | 43617 | 43861 | 245 | Intron 12 | 483 |
| Exon 13 | 44344 | 44563 | 220 | Intron 13 | 381 |
| Exon 14 | 44944 | 45146 | 203 | Intron 14 | 1791 |
| Exon 15 | 46937 | 47103 | 167 | Intron 15 | 306 |
| Exon 16 | 47409 | 47886 | 478 |  | - |

Table 1.1 Summary of the positions of the exon boundaries, exon length, and intron length for ITGB2.

### 1.5.2 PSI domain

The PSI domain of the CD18 protein is a 63 residue long cysteine-rich region that shares significant homology with other proteins such as plexins and semaphorins (Bork, et al., 1999).

The integrin PSI domain is composed of a two-stranded antiparallel $\beta$-sheets, and is flanked by 2 short helices. Similar to other PSI domains, the integrin PSI domain contains four pair of cysteine residues which are connected in the following manner:

Cys 1 - Cys 4 , Cys 2 - Cys 8 , Cys 3 - Cys 6, Cys 5 - Cys 7. In contrast to other PSI
domains, the eighth cysteine residue is located after the hybrid domain in the primary amino acid structure (Xiong, et al., 2004).

### 1.5.3 Hybrid domain

The CD18 hybrid domain is composed of amino acid sequences flanking the $\beta \mathrm{I}$ domain. Based on structural analysis, this domain inserts in the last loop of the PSI domain, physically connecting the PSI and I-like domain when in the correctly folded conformation (Xiong, et al., 2001). The hybrid domain has been shown to be important for integrin activation, allowing for intra-domain propagation of activating signals, probably by swinging away from the $\beta$ subunit during activation of the integrin head region (Mould, et al., 2003; Tan, et al., 2001a; Tang, et al., 2005; Tng, et al., 2004; Xiao, et al., 2004).

Evidence in support of hybrid domain displacement during ligand binding includes the following observations. Using $\beta 2 / \beta 7$ chimeras where the $N$-terminal region (NTR), $\beta I$ domain and the mid-region (up to C426) of $\beta 2$ were replaced by those of $\beta 7$, Tan et al. [2001] mapped the epitopes of six anti-CD18 mAbs and demonstrated that all the mAb require both the NTR and mid-region for epitope expression, with the exception of H52. This suggested that the CD18 NTR and mid-region interacts extensively (Tan, et al., 2001a).

X-ray scattering was used to visualize the head region of integrin $\alpha_{5} \beta_{1}$ in an inactive ( $\mathrm{Ca}^{2+}$-occupied) state. This was compared to data from the head region in complex with a fibronectin fragment. Shape reconstructions of the data suggested that the conformation of the head region remains stable, but outward movement of the $\beta I$ domain occurs upon ligand binding (Mould, et al., 2003).

Furthermore, the epitope for the reporter mAb MEM148 was mapped to Pro ${ }^{374}$ on the inner face of the hybrid domain in LFA-1. MEM148 was used to report displacement of the hybrid domain with $\mathrm{Mg} /$ EGTA mediated activation combined with an activating mAb. Reporting of the MEM148 epitope coincided with a leg separated (hybrid displaced) model of integrin activation, and correlated with the high avidity state of LFA-1 required for ICAM-3 binding (Tang, et al., 2005).

Crystal structures of the platelet integrin $\alpha \mathrm{IIb} / \beta 3$ bound to fibrinogen-mimetics showed that ligand binding at the $\beta 3 \mathrm{I}$-domain alters three metal binding sites, associated loops and alpha1- and alpha7-helices and results in a 'piston-like' reorientation between the $\beta 3$ I and hybrid domains of 62 degrees (Xiao, et al., 2004).

### 1.5.4 I-like domain ( $\beta$ I domain)

Based on the primary protein sequence, the $\beta I$ domain is inserted in the hybrid domain, which is in turn inserted into the PSI domain (Xiong, et al., 2004). Although there is low sequence homology between the $\beta \mathrm{I}$ domain and the $\alpha \mathrm{I}$ domain, these 2 domains have structurally homologous $\alpha / \beta$ folds. The $\beta$ I domain consists of a central sixstranded $\beta$ sheet surrounded by eight helices, assuming a nucleotide-binding conformation that is known as the Rossmann fold.

Metal ions, including $\mathrm{Ca}^{2+}, \mathrm{Mg}^{2+}$ and $\mathrm{Mn}^{2+}$, play important roles in structural stabilization, ligand recognition and ligand binding in integrins (Plow, et al., 2000; Xiong, et al., 2003b; Yoshinaga, et al., 1994). Similar to the I-domain of the $\alpha$ subunit, the $\beta \mathrm{I}$ domain contains a putative metal-binding DXSXS sequence motif forming an $\mathrm{Mg}^{2+}$ binding MIDAS (metal ion-dependent adhesion site), located in a crevice on top of the central $\beta$ strand (Lee, et al., 1995; Luo, et al., 2007; Ponting, et
al., 2000; Springer and Wang, 2004). The $\beta$ I MIDAS is flanked on either side by the $\mathrm{Ca}^{2+}$ binding ADMIDAS (adjacent to MIDAS) and SYMBS (synergistic metalbinding sites) (Xiao, et al., 2004; Xiong, et al., 2001; Xiong, et al., 2002; Zhu, et al., 2008).

In contrast to the $\alpha \mathrm{I}$ domain, an extra acidic residue is found at the $\beta$ I MIDAS at Asp ${ }^{222}$. This residue may directly coordinate $\mathrm{Mg}^{2+}$, reducing its electrophilicity and decreasing affinity for its intrinsic ligand; the first residue of the CD11c $\alpha \mathrm{I}$ C-linker, Glu ${ }^{297}$. Mutation of this invariant Glu residue abolishes integrin activation (Huth, et al., 2000). When the Glu residue is mutated to Cys and combined with the mutation of the MIDAS-loop residue $\mathrm{Ala}^{189}$ to Cys, constitutive ligand binding was observed, due to the formation of the intersubunit disulfide bond (Yang, et al., 2004). Taken together, integrin activation may be partly regulated by $\mathrm{Mg}^{2+}$ binding to $\mathrm{Asp}^{222}$, which in turn regulates binding to the intrinsic $\mathrm{Glu}^{297}$ ligand.

In 2010, a crystal structure of $\alpha \mathrm{X} \beta 2$ showed a bent integrin structure with the $\beta$ I domain in the low affinity, closed confirmation (Figure 1.5). The $\beta$ I domain can adopt two conformations that allow it to couple with three different conformations of the alpha I-domain, allowing structural flexibility in ligand recognition (Xie, et al., 2010).

When the NTR, mid region or the $\beta \mathrm{I}$ domain was swapped between $\beta 2$ and $\beta 7$, replacement of the $\beta 2$ with the $\beta 7 \beta$ domain abolished LFA- 1 dimer formation and surface expression. Hence, the $\beta I$ domain is important for pairing specificity between CD18 and CD11a (Tan, et al., 2001a).

In addition, the $\beta I$ domain regulates ligand binding when paired with an $\alpha$ subunit possessing an $\alpha \mathrm{I}$ domain, but is responsible for ligand binding when paired with an $\alpha$
subunit lacking the $\alpha \mathrm{I}$ domain. Notably, all CD18 based heterodimers possess an $\alpha \mathrm{I}$ domain (Barczyk, et al., 2010).


Figure 1.5 Structure of the bent CD11c/CD18 structure obtained by Xie et al. [2010].

### 1.5.5 IEGF domains

There are four integrin epidermal growth factor (IEGF) domains arranged in tandem, each of which is composed of a cysteine-rich repeat, where the internal cysteines are paired as follows: Cys 1 - Cys 5, Cys 2 - Cys 4, Cys 3 - Cys 6, Cys 7 - Cys 8 (Takagi, et al., 2001; Tan, et al., 2001b). This domain is 174 residues long. While IEGF2,

IEGF3, and IEGF4 possess eight cysteine residues each, the IEGF1 domain contains only six cysteines engaged in three disulfide bonds, as the C2-C4 disulfide pair is not present (Beglova, et al., 2002; Shi, et al., 2007a; Shi, et al., 2005).

When a disulfide bond was introduced between the PSI and I-EGF2 domains (at Gly ${ }^{33}$ in the PSI domain and Gly ${ }^{486}$ in the I-EGF2 domain), the resultant LFA-1 variant was locked in a bent state and could not be detected with the monoclonal antibody KIM127 in M/E, even while retaining ICAM-1 binding capacity. Taken together with the bent and extended structures of PHE2 and PHE3 respectively, conformational changes within the I-EGF2 domain may be important for a transition between the bent (inactive) to extended (active) conformation in LFA-1 (Shi, et al., 2007a).

### 1.5.6 $\beta$ TD domain

The 78 residue CD18 tail domain ( $\beta$ TD) consists of an N -terminal $\alpha$ helix and a fourstranded $\beta$ sheet held together by five disulphide bonds (Xiong, et al., 2001). As this domain is located C terminal to the IEGF domains, and only weak hydrophobic contacts are found between the $\beta$ TD and IEGF4, this interface is flexible and probably not critical for integrin dimer formation (Butta, et al., 2003; Xiong, et al., 2001).

### 1.5.7 TM domain

The CD18 transmembrane domain is 23 residues long. Although the details and nature of TM domain movement during activation is currently unclear, this domain may play an important role in integrin activation.

Crystal structures of the platelet integrin $\alpha \mathrm{IIb} / \beta 3$ bound to fibrinogen-mimetics showed that ligand binding disrupts interaction between the legs, positioning the highaffinity head above the cell surface. This was in contrast to a previously described low-affinity bent integrin conformation (Xiao, et al., 2004). In addition, leg separation via interactions with the alpha subunit transmembrane domain has been shown to be important for LFA-1 activation and microcluster formation (Armulik, et al., 1999; Arnaout, et al., 2005; Luo, et al., 2004; Vararattanavech, et al., 2009).

### 1.5.8 Cytoplasmic domain

Compared to the 678 residue CD18 ectodomain, the cytoplasmic domain is relatively short at 46 residues. Conformational changes in the LFA-1 cytoplasmic domain were measured using fluorescence resonance energy transfer (FRET) analysis between Cterminal fused cyan fluorescent protein (CFP) in CD11a and yellow fluorescent protein (YFP) in CD18. The cytoplasmic domains were in close proximity in inactive LFA-1, but significant spatial separation was observed when inside-out signaling or outside-in signaling (ligand binding) occurred. Hence, separation of the cytoplasmic tail is important to propagate bi-directional signals in LFA-1 (Kim, et al., 2003).

Phosphorylation of the cytoplasmic tail modulates binding of other proteins such as 14-3-3, filamin and talin (Takala, et al., 2008). In addition, the cytoplasmic tail binds with a multitude of proteins including enzymes, adaptor proteins, actin-binding proteins such as the kindlins, and transcriptional co-activators (Horwitz, et al., 1986; Kieffer, et al., 1995; Liu, et al., 2000; Ma, et al., 2008; Montanez, 2008).

### 1.1 Intrinsic folding of the CD11a and CD18 protein

JY lymphoblastoid and a B lymphoblastoid cell line from an LAD-I patient (no CD18 expression is detected in LAD-I patients) pulsed with 35 S cysteine for 1 hour were immunoprecipitated using a series of antibodies directed against various conformational epitopes in the I-domain and beta propeller of CD11a. As a positive control, cells were pulsed with 35 S cysteine and chased for 16 hours with unlabelled cysteine. In the absence of CD18, the nascent CD11a $\alpha$ d domain can form the correct conformation independently. However, the beta propeller domain of CD11a could not fold correctly in the absence of CD18 (Huang and Springer, 1997).

Conversely, the same methodology was used in JY lymphoblastoid cells and JB2.7 cells (which express only CD18 but not CD11a, to be described later). An antibody panel against various CD18 conformational epitopes both flanking and within the $\beta \mathrm{I}$ domain was used to immunoprecipitate the nascent CD18 protein. The $\beta$ I domain could not fold correctly in the absence of CD11a, in contrast to the $\beta$ I domain flanking regions which could be immunoprecipitated (Huang, et al., 1997). This finding is further supported by another study which showed that removal of the $\beta I$ domain does not affect folding of the CD18 NTR/mid-region complex. The truncated sequence is expressed as a soluble protein in COS7 that can be immunoprecipitated using mAbs, (Tan, et al., 2001a).

Hence, conformational epitopes in CD11a beta propeller and the $\beta I$ domain which are found in dimeric integrins, are not detectable in CD11a and CD18 integrin monomers, indicating that the integrin proteins cannot fold completely when expressed individually. In this thesis, this observation shall be referred to as 'partially unfolded'.

After creating a series of truncated CD18 mutants at the amino acid residues 301, 414, 430, 459, 481 and 701 (Figure 1.6), Tan et al (2000) demonstrated that all of the truncated mutants were able to associate with CD11a intracellularly in COS7 cells However, only mutants which consisted of at least residue 459 and longer could be expressed on the cell surface.

Taken together, these results suggest that the CD11a beta propeller and CD18 $\beta$ I domain must fold together intracellularly for membrane expression, and that the amino acid sequence from PSI to IEGF-1 of CD18 is critical for membrane expression of integrin heterodimers, both wild type and mutant.


Figure 1.6 Positions of the truncated CD18 mutants from Tan et al (2000).

With regard to the $\alpha$-subunit, mutant LFA- 1 and Mac-1 can be expressed on the cell surface of K562 cells after removal of the respective $\alpha \mathrm{I}$ domains. Deletion of the $\alpha \mathrm{I}$ domain resulted in diminished binding to denatured BSA ( $60 \%$ vs $2 \%$ ), fibrinogen ( $60 \%$ vs $30 \%$ ), iC3b ( $65 \%$ vs $20 \%$ ) and immobilized factor X ( $60 \%$ vs $25 \%$ ). The residual binding capacity to the latter three was perhaps due to a contribution of the $\beta I$ domain to ligand binding in the absence of the CD11b $\alpha \mathrm{I}$ domain (Leitinger and Hogg, 2000).

In contrast, deletion of the CD11a $\alpha$ I domain resulted in a complete loss of binding to both the LFA-1 substrates, ICAM1 ( $40 \%$ vs $2 \%$ ) and ICAM3 (18 \% vs $2 \%$ ). Mutant

LFA-1 was constitutively active, and seemed to cross-activate the $\beta 1$ integrins $\alpha 4 \beta 1$ and $\alpha 5 \beta 1$ to bind fibronectin and $\alpha 4 \beta 1$ alone to bind VCAM-1 (Leitinger and Hogg, 2000; Yalamanchili, et al., 2000). Hence, the $\alpha$ I domain, while not critical for dimer formation, is important for integrin adhesion and activation in LFA-1 and Mac-1.

Since recycling and exocytosis for membrane expression may overlap, prior work related to integrin recycling was reviewed. It is currently known that membrane expressed integrin dimers are endocytosed frequently in a highly efficient manner and recycled to the cell membrane, rather than degraded (Bretscher, 1992; Roberts, et al., 2001). Upon endocytosis, integrins are transported in Rab4-positive early endosomes to the Rab11-positive, perinuclear recycling compartment in approximately 30 mins . Recycling from this compartment to the plasma membrane occurs via a Rab11dependent mechanism. Alternatively, a rapid recycling pathway may take place from the early endosomes via a Rab4-dependent mechanism without the involvement of Rab11 (Fabbri, et al., 2005).

Membrane expression of monomeric integrins is inconsistent across different cell lines. Loss of the CD18 subunit in LAD-I patients results in intracellular retention of all three $\alpha$-subunits in patient leukocytes (Kohl, et al., 1984; Marlin, et al., 1986).

The JB2.7 and SKB2.7 cell lines were derived from Jurkat and SKW3. Both parental cell lines are of lymphoblastic origin and express LFA-1. Using ethylmethane sulphonate mutagenesis followed by anti-CD18 immunopanning, non-adherent cells were selected over several immunopanning procedures until a predominantly LFA-1 negative population was obtained. The mutated cell lines were obtained by limiting
dilution and screened for deletion of CD11a or CD18, respectively, in Western blot. Both cell lines do not express either integrin monomer on the cell surface (Weber, et al., 1997).

Similarly, HEK 293 cells transfected with only CD18, CD11a or CD11c do not support monomer expression on the cell surface, but CD11b monomers can be expressed on the cell surface of HEK293 transfectants (unpublished results from our laboratory, refer to Figure I. 3 in Appendix I). In addition, COS7 cells transfected with only the CD18 expression plasmid support CD18 monomer expression (Tan, et al., 2000; Tan, et al., 2001b).

The mechanisms that allow membrane expression of CD11b in HEK293 and CD18 in COS7 are currently unclear, but many studies on the CD11a/CD18 leukocyte integrins are performed on HEK293 transfectants. It was reasoned that retention of CD11a and CD18 in HEK 293 and leukocytic cells are probably due to similar, if not identical, ubiquitous mechanisms.

In this thesis, the mechanism of CD11a and CD18 integrin retention in HEK293 transfectants was studied. As the individual subunits are relatively well folded with the exception of the abovementioned domains, a chaperone protein may retain partially unfolded integrin monomers until the dimer is formed. Since it is the last domain to remain unfolded, this retention protein was likely to bind the $\beta I$ domain of CD18 and the beta-propeller region of CD11a.

### 1.2 Leukocyte Adhesion Deficiency

### 1.2.1 Leukocyte Adhesion Deficiency type I

Clinical observations of delayed umbilical cord separation, defective neutrophil mobility and widespread infections were first reported by Hayward et al (1979) in five patients born to two families. An early hint into the genetic basis of the disease was provided by gel electrophoresis results indicating the loss of a 110 kDa protein in the neutrophils of a patient exhibiting these clinical symptoms. Since the mother had diminished expression of the 110 kDa protein, the authors erroneously postulated that the disease was X-linked (Crowley, et al., 1980).

Further work led to the description of Leukocyte Adhesion Deficiency type I (LAD-I) deficiency as primarily due to the absence or highly diminished expression level of the CD11/ CD18 antigens. Subsequently, LAD-I was found to be a rare autosomal recessive immunodeficiency disorder resulting from genetic mutations in the $\beta 2$ subunit of the leukocyte integrins (Anderson, et al., 1985; Arnaout, 1990; Kohl, et al., 1984; Marlin, et al., 1986).

Till today, a diverse range of molecular defects leading to LAD-I have been characterized, including the expression of functionally defective CD11/CD18 antigens (Hogg, et al., 1999; Mathew, et al., 2000), and the expression of partially active integrins (Shaw, et al., 2001). A comprehensive summary of the mutations in the CD18 gene, up to 2002, can be found in the two publications by Roos and Law (Roos and Law, 2001; Roos, et al., 2002). Since then, several mutations on LAD-I have been reported, including a 36 bp deletion in exon 12 (Hixson, et al., 2004), two nonsense
mutations in exon 3 (Fiorini, et al., 2009) and exon 15 (Hixson, et al., 2004) respectively, and a gross 4.6 Mb ring chromosome 21 deletion (Fiorini, et al., 2009).

Most recently, a study on eleven Iranian LAD-I patients identified seven novel mutations in the $I T G B 2$ sequence that led to loss of the CD18 protein, including one splicing mutation (c.500-6C>A), three missense mutations (D128Y, A239Y, and G716A), and three frameshift deletions (p.282fsX41, p.382fsX9, and p.636fsX22) (Parvaneh, et al., 2010).

LAD-I patients suffer from recurrent bacterial or fungal infections, and exhibit marked increases in leukocyte number (leukocytosis), especially during periods of infection, and demonstrate a muted inflammatory response to infections as a result of poor leukocyte adhesion to the vascular wall and defective adhesion to opsonized pathogens during phagocytosis (Hayward, et al., 1979; Malech and Nauseef, 1997).

Cell surface expression of the CD18 integrins is dependent on the formation of the integrin dimer; the LAD-I related deficiencies in immune response manifest as a result of the loss of a functional CD18 protein. The severity of LAD-I symptoms correlate with residual expression levels of the CD18 integrins on the leukocyte surface, which determine the capacity of leukocytes for cellular adhesion. Patients with less than 1\% of LFA-1 expression exhibit severely impaired leukocyte migration to sites of infection and inflammation, and are susceptible to frequent and life threatening systemic infection (Anderson, et al., 1985; Roos and Law, 2001).

A retrospective study of 36 LAD-I patients who underwent Bone Marrow Transplant (BMT) treatment demonstrated a survival rate of $75 \%$, at a median follow-up of 62 months. Mortality was greatest after haploidentical transplants (transplanting bone marrow from a healthy parent with unmatched HLA), after which 4 of 8 children did not survive. In contrast, 23 out of 28 patients who received allogenic transplants (HLA matched) survived BMT (Qasim, et al., 2009).

Hence, the LAD-I condition is usually controlled with daily prophylactic doses of trimethaprim-sulfamethoxazole and broad-spectrum antibiotics during bouts of infection, but early allogenic BMT is the recommended treatment for severe cases of LAD-I (Fischer, et al., 1998).

### 1.2.2 Other Leukocyte Adhesion Deficiencies

There are another two Leukocyte Adhesion Deficiency subtypes where patients exhibit increased susceptibility to infections as a result of defective leukocyte adhesion, but these subtypes result from mutations in different genes.

Leukocyte Adhesion Deficiency type II (LAD-II) patients are normal for weight and height at birth, and there is no delay in umbilical cord separation. Clinically, LAD-II symptoms are similar to that of LAD-I. Due to a lack of fucosylated H antigen in red blood cells, LAD-II patients have the rare Bombay (hh) blood phenotype, while growth defects, mental retardation and a distinctive facial appearance will manifest after birth. Compared to LAD-I, infections in LAD-II patients are less severe, and the frequency of infection decreases after 3 years of age, such that prophylactic antibiotics are no longer required (Etzioni, et al., 1992; Etzioni, et al., 1998; Hidalgo, et al., 2003; Luhn, et al., 2001).

LAD-II is caused by mutations in the $\mathrm{SLC35Cl}$ gene located on chromosome 11 p 11.2 . This gene encodes the GDP-fucose transporter, and loss-of-function mutations may result in defective fucose entry into the Golgi apparatus and hence, an impaired fucosylation process. Hence, LAD-II is alternatively referred to as the congenital disorder of glycosylation type IIc (CDG IIc) syndrome (Hidalgo, et al., 2003; Luhn, et al., 2001).

LAD-II patients have markedly low cellular expression of fucosylated glycoproteins relative to a healthy individual. The ligand for endothelial selectin, CD15s (SialylLewis X ), is a fucosylated glycoprotein that is expressed on the leukocyte cell surface. As a result of low expression of fucosylated CD15s, LAD-II patients exhibit defective leukocyte rolling adhesion. Leukocytes are unable to slow down, and are less efficient at adhering to the endothelial wall for extravasation to sites of infection or inflammation. However, the precise mechanism leading to severe psychological retardation and growth defect is still unknown. (Hidalgo, et al., 2003; Luhn, et al., 2001).

Patients suffering from Leukocyte Adhesion Deficiency type III (LAD-III) have recurrent infections similar to LAD-I, but have the additional complications of defective platelet activation and a tendency toward severe bleeding similar to that of Glanzmann thrombasthenia (Alon, et al., 2003; Alon and Etzioni, 2003; Etzioni, et al., 1992; Etzioni, et al., 1998).

LAD-III is a rare autosomal-recessive syndrome. Patients exhibit normal expression levels of $\beta 1, \beta 2$, and $\beta 3$ integrins, but the integrins are unable to undergo GPCR-
mediated 'inside-out' stimulation (McDowall, et al., 2003). Hence, LAD-III patients exhibit defective integrin adhesion in both leukocytes and platelets due to the loss of 'inside-out' activation of membrane integrins and the concomitant loss of high-avidity binding to integrin ligands (Moser, et al., 2009).

LAD-III results from mutations in the KINDLIN3 gene, located on chromosome 11q13 (Kuijpers, et al., 2009; Malinin, et al., 2009; Svensson, et al., 2009). This gene encodes the intracellular protein, Kindlin-3, which is expressed only in leukocytes and binds to NPXY motifs in the cytoplasmic tail of the integrins. Both talin and the kindlins are required for complete activation of the integrins (Ma, et al., 2008; Montanez, 2008; Moser, et al., 2009; Shi, et al., 2007b).

### 1.3 Pre-mRNA splicing in human genes

Human genes contain an average of eight exons. Each exon has an average length of 145 nucleotides (nt). In contrast, the average length of introns is greater than 1450 nt but some introns may be much longer (Lander, et al., 2001). Through recognition of splice site sequences, removal of intronic sequences from pre-mRNA (pre-mRNA splicing) takes place in the cell nucleus, and is a necessary post-transcriptional process for eukaryotic gene expression. Notably, multiple mRNA variants occur in an estimated $59 \%$ of human genes as a result of alternative patterns of intron removal (alternative splicing), leading to translation of diverse and even antagonistic proteins (Lander, et al., 2001; Nilsen and Graveley, 2010).

### 1.3.1 The major and minor spliceosome

Following RNA polymerase II mediated transcription, pre-mRNAs are bound by components of the basal splicing machinery at classical splice site sequences;
promoting assembly of a multi-protein spliceosome complex that catalyses the two trans-esterification reactions, removing intervening introns and joining adjacent or alternative exons (Alberts, et al., 2007).

From a mechanistic viewpoint, there are two forms of spliceosomes that function at human introns. These are of the major-U2 type and the minor-U12 type. Each spliceosome is comprised of five small nuclear ribonucleoproteins (snRNPs) and more than a hundred proteins. In turn, each snRNP is composed of a single uridinerich small nuclear RNA (snRNA) and multiple proteins (Brow, 2002).

The majority of splice sites are of the U2-type GU-AG class, making up between 98$99.24 \%$ of all human splice sites. U2 type GC-AG introns are also recognized by the major spliceosome, but only make up about $0.9 \%$ of all human splice sites (Burset, et al., 2000; Collins and Penny, 2005; Sheth, et al., 2006; Thanaraj and Clark, 2001; Zhang, 1998) . In contrast, approximately 0.05 to $0.1 \%$ of rare introns bear the AUAC sequences recognized by the minor U12 spliceosome (Burset, et al., 2000; Hall and Padgett, 1996; Sheth, et al., 2006). Other variants such as AU-AG, GU-GG introns have been identified in the human genome, but these comprise less than $0.1 \%$ of known splice sites (Burset, et al., 2000; Sheth, et al., 2006).

### 1.3.2 U2-type GU-AG intron splicing

U2-type GU-AG exons (also known as AG-dependent introns) are defined by three sequence elements; the $5^{\prime}$ ' splice site (ss), branch point sequence (bps), and the 3 'ss. (Collins and Penny, 2005). In human intron/exon boundaries, the $5^{\prime}$ consensus sequence is CAG/GU (Roca, et al., 2008), and the BPS consensus sequence is

YNYURAY (Y indicates any pyrimidine, R indicates any purine, and N indicates any nucleotide) (Nelson and Green, 1989).

Compared to 5 'ss, 3 'ss recognition sequences are more complex as a greater number of sequence elements are involved. 3'ss sequence elements are also spread over longer and more variable distances in the genomic DNA sequence (Královicová, et al., 2005; Roca, et al., 2003; Yeo and Burge, 2004). A variety of sequences may define the 3 'ss, including an AG dinucleotide at the 3 'ss and an upstream, uridine-poor polypyrimidine tract (PPT). Recognition of the 3'ss also requires cooperative interactions between components of the spliceosome and sequential recognition of the same signal by different splicing components, which occurs during spliceosome assembly and catalysis (Collins and Penny, 2005; Corrionero, et al., 2011).

Via RNA-RNA interactions between snRNA and the pre-mRNA, the U1 snRNP binds the 5 'ss, stabilized by members of the serine-arginine-rich (SR) protein family and proteins of the U1 snRNP (Cho, et al., 2011). The initial stages of spliceosome assembly also involves binding of the SF1/BBP protein to the BPS (Berglund, et al., 1998; Guth and Valcarcel, 2000) and the 35 kDa U2 Auxiliary Factor small subunit $\left(\mathrm{U}^{2} \mathrm{AF}^{35}\right)$ to the AG dinucleotide sequence (Wu, et al., 1999; Zamore, et al., 1992). Binding of $\mathrm{U} 2 \mathrm{AF}^{35}$ to the AG dinucleotide sequence facilitates recruitment of the 65 kDa U2 Auxiliary Factor large subunit $\left(\mathrm{U}^{2} \mathrm{AF}^{65}\right)$ to the upstream PPT, which interacts with SF1/BBP through its C-terminal RNA recognition motif (RRM) (Guth, et al., 2001; Ruskin, et al., 1988). Together, these proteins cooperate to form the spliceosome E complex, which is important for initial recognition of the 5 'ss and 3 'ss of introns (see Figure 1.7 for a schematic diagram of the sequence elements).

After spliceosome E complex formation, $\mathrm{SF} 1 / \mathrm{BBP}$ is displaced from the BPS by U2 snRNP in an ATP-dependent manner, resulting in formation of the spliceosome A complex (Das, et al., 2000). The interactions between U2 snRNP and the BPS are further stabilized by the heteromeric protein complexes SF3a and SF3b in the U2 snRNP (Gozani, et al., 1996) and the arginine-serine-rich domain of the $U 2 \mathrm{AF}^{65}$ protein (Valcarcel, et al., 1996).

The preassembled U4/U6/U5 tri-snRNP is then recruited, forming the catalytically inactive spliceosome B complex (Weidenhammer, et al., 1997). After major conformational rearrangements and the release of U1 and U4 (Xie, et al., 1998), the activated spliceosome $\mathrm{B}^{*}$ complex cleaves the 5 'ss and catalyzes the formation of a lariat intermediate through a $2^{\prime}-5$ ' phosphodiester bond with the highly conserved Adenosine (A) of the BPS, generating the spliceosome C complex (Brow, 2002).

Additional spliceosome rearrangements occur prior to the second catalytic step (Konarska, et al., 2006). In the second trans-esterification reaction, the oxygen atom of the $5^{\prime}$ exon (at the newly formed 3' hydroxyl group) forms a bond with the phosphorus atom which links the intron to the $3^{\prime}$ exon. This bond is displaced to the lariat intron, joining the 5' and 3 ' exons and excising the intron as a lariat (Berglund, et al., 1998; Kent, et al., 2003; Selenko, et al., 2003; Sharp and Burge, 1997; Sheth, et al., 2006; Vorechovsky, 2006). After completion of the two trans-esterification steps, the spliceosome dissociates from the mRNA, releasing the U2, U5, and U6 snRNP which are recycled in other rounds of splicing (Bell, et al., 2002). A diagram of the two trans-esterification reactions is presented in Figure 1.8, and an overview of the splicing events is presented in Figure 1.9.

Both major and minor spliceosomes contain the same U5 snRNP, but minor spliceosomes are composed of different snRNPs that have analogous functions to U1, U2, U4, and U6, which are respectively called U11, U12, U4atac, and U6atac (Friend, et al., 2008; Roca, et al., 2003; Tarn and Steitz, 1996).


Figure 1.7 A schematic diagram of the sequence elements which define the U2 type GU-AG intron and the spliceosome components that bind each element prior to splicing.


Figure 1.8 Diagrammatic representation of the two trans-esterification reactions in pre-mRNA splicing, taken from Brow (2002).


Figure 1.9 Overview of the sequential events in a U2-type GU-AG class intron taken from Alberts, et al. (2007). The events depicted in the event are described in the text.

### 1.3.3 Disruption of pre-mRNA splicing

Disruptions to pre-mRNA processing and splicing can directly cause disease, modulate disease severity, or in the case of multi-factorial diseases such as cancer, be linked with increased disease susceptibility (Faustino and Cooper, 2003; Pagenstecher, et al., 2006). It has been estimated that between $50-60 \%$ of disease-causing mutations may result from disruptions to normal splicing (Cartegni, et al., 2002; Lopez-Bigas, et al., 2005; Pagenstecher, et al., 2006). When mutations outside of splice site sequences in the Human Gene Mutation Database were analyzed, 78 \% were attributable to mutations in the exons such as single-nucleotide substitutions (56.9\%), microdeletions, or microinsertions within exons (21.7\%). While some mutations in exons may affect the protein reading frame or function, it is possible that some exonic mutations may result in disease through disruptions to splicing (Wang and Cooper, 2007).

There are two possible causes of an altered splicing code. Firstly, mutations within the classical splice site sequences may disrupt consensus sequences of the sequence elements. Secondly, cis-acting mutations may disrupt components of the splicing code, for example, by creating cryptic (de novo) splice sites, disrupting authentic splice sites, enhancing cryptic splice site strength, or by altering the secondary structure of premRNA. As a result, various splice defects in the affected gene may occur, such as complete exon skipping, cryptic splice site activation, and retention of the entire intron. Furthermore, cis-acting mutations can influence constitutive or alternative splice site usage; changing splicing patterns in an inappropriate tissue or developmental stage may possibly result in disease (Pagani and Baralle, 2004; Pagani, et al., 2005; Pagenstecher, et al., 2006).

In addition to an altered splicing code, there is another class of mutation which may disrupt normal splicing. Trans-acting mutations occur in genomic DNA positions away from the genes affected themselves. Trans-acting mutations may include disruption of normal functioning and/or expression of spliceosome components, or factors involved in splicing regulation. Possibly, trans-acting mutations may also include disruptions to regulatory proteins that control or interact with another subset of proteins. Hence, trans-acting mutations can potentially affect splicing of all premRNAs. Where a regulatory factor is affected, only the pre-mRNAs controlled by the regulator may be affected (Faustino and Cooper, 2003; Pagani and Baralle, 2004; Wang and Cooper, 2007).

### 1.3.4. 3'ss mutations

In this thesis, the U2-type GU-AG class intron 2 of $I T G B 2$ was studied. In the LAD-I patient, a $-10 \mathrm{bp} \mathrm{C}>\mathrm{A}$ mutation had resulted in ablation of the authentic 3 'ss in the intron 2 (Cher, et al., 2011).

In AG-dependent 3'ss, neighboring AG dinucleotides influence 3 'ss selection by encouraging early RNA interactions with the spliceosome, and influencing the usage of nearby 3'ss (Chua and Reed, 2001). While loss of the authentic 3'ss usually results in activation of another AG dinucleotide based cryptic 3'ss (Chua and Reed, 2001; Královicová, et al., 2005), different mutations that affect a particular AG-dependent 3'ss may result in the activation of different cryptic 3 'ss, making predictions of cryptic or de novo 3'ss activation difficult (Královicová, et al., 2005; Roca, et al., 2003; Yeo and Burge, 2004).

As AG dinucleotides are a near universal signature found at the 3 ' end of introns (Moore, 2000; Sheth, et al., 2006), all AG dinucleotides in a given DNA sequence are considered as potential 3'ss. Cryptic 3'ss are 3'ss that are only used when the authentic 3 'ss is disrupted by mutation. Cryptic 3 'ss may usually be identified insilico by scores above background. Pseudo 3'ss are potential 3'ss at which splicing is not detected experimentally. Pseudo 3'ss usually have low scores (background scores) when analysed in-silico. While in-silico analysis can be used as a 'quick and dirty' method to predict 3 'ss strength, the best way to differentiate between a cryptic or pseudo 3'ss is to determine the preferred splice site usage in a given intron, experimentally.

Activation of a cryptic 3'ss due to a local cis-acting mutation is dependent on relative 3'ss strength. Possible discriminating parameters of 3'ss strength may include the strength of neighboring splice sites, the availability of putative cryptic splice sites, the length of the resulting exon, and the density of exonic and intronic splicing enhancers or silencers (Královicová and Vorechovsky, 2007; Wimmer, et al., 2007).

### 1.4 Studies in this thesis

In this thesis, the second ITGB2 mutation in a LAD-I patient was studied and characterized. While the first mutation was previously characterized by Ms. Chan Hwee Sing in 2004, the second mutation was particularly elusive, and had remained uncharacterized in the several years that followed. To characterize the second mutation, genomic DNA and cDNA from the LAD-I patient and her family was analyzed in detail.

While it was previously shown that the $-10 \mathrm{C}>\mathrm{A}$ mutation had resulted in a splicing defect, it was unclear how a point mutation had caused this splicing defect. Furthermore, analysis of the intron 2 sequence in the LAD-I patient using 3'ss strength prediction tools showed that a potential cryptic 3'ss with scores higher than that of the authentic 3 'ss was located at +11 . However, when the authentic 3 'ss was lost due to the $-10 \mathrm{C}>$ A point mutation, the +113 ' SS was not used.

As mentioned earlier, PPTs are part of the known sequence elements that define the 3'ss (Collins and Penny, 2005; Sheth, et al., 2006; Vorechovsky, 2006; Zhang, 1998). On the other hand, upstream AG dinucleotides located within 3 nt have been shown to repress 3'ss (Corrionero, et al., 2011). Here, the effect of the upstream AG dinucleotide on use of the +113 'ss was analyzed.

Later on, the mechanism of integrin monomer retention was studied. While it is well known that completed dimers were recycled from the cell membrane in a very efficient manner (Bretscher, 1992; Fabbri, et al., 2005; Roberts, et al., 2001), dimer formation from the nascent, monomeric subunits is not clearly defined. However, it is clear that the CD11a and CD18 proteins are consistently retained across different cell lines, when expressed individually (i.e. in the absence of either subunit).

Since integrin retention probably occurs in order for heterodimer formation to take place, the retention of monomeric CD11a and CD18 was studied as a starting point towards understanding integrin dimerization. In this part of the study, proteins interacting with nascent CD11a and CD18 in HEK293 transfectants were identified.

The MEM148 antibody was useful for reporting the leg extended conformation of LFA-1. This antibody was initially obtained from V. Horejsi (Institute of Molecular Genetics, Prague, Czech Republic). After the antibody was commercialized, subsequent MEM148 stocks obtained from Serotech failed to report leg extension in LFA-1 efficiently. The experiments performed to optimize and troubleshoot MEM148 reporter activity are described in the last chapter of this thesis.

## Chapter 2: Materials and Methods

### 2.1 Materials

### 2.1.1 Reagents

Analytical grade reagents and solvents were obtained from Sigma Aldrich (St. Louis, MO, USA), Becton Dickson Biosciences (BD Biosciences, San Diego, CA, USA), Pierce (Thermo Fisher Scientific, Rockford, IL, USA), Gibco (Invitrogen, Carlsbad, CA, USA), Bio-Rad (Hercules, CA, USA), USB Corp. (Cleveland, OH, USA), Novagen (Merck, Whitehouse Station, NJ, USA), Fisher Scientific (Thermo Fisher Scientific, Rockford, IL, USA), Fermentas (Burlington, ON, CA), New England Biolabs (Ipswich, MA, USA), $1^{\text {st }}$ Base (SIN), Promega (Madison, WI, USA), Amersham (Piscataway, NJ, USA), Roche (Nutley, NJ, USA)

### 2.1.2 Cells

The following cell lines were each obtained from American Type Culture Collection (ATCC, Manassas, VA, USA); MOLT-4 (human acute T lymphoblastic leukemia), Jurkat (human T lymphoblast-like), SKW-3 (post-thymic T lymphoblastic leukemia), U937 (human leukemic monocytic lymphoma), HEK 293T cells (human embryonic kidney cell with SV40 large T antigen) and COS-7 cells (African green monkey kidney cells).

### 2.1.3 Cell culture

All cells were maintained in an incubator (Binder GmbH, Tuttlingen, Germany) at 37 ${ }^{\circ} \mathrm{C}$ in humidified air containing $5 \% \mathrm{CO}_{2}$ and $95 \%$ air. HEK 293 T and COS- 7 cells were maintained in DMEM (Hyclone) supplemented with $10 \%(\mathrm{v} / \mathrm{v})$ heat inactivated FBS, $100 \mathrm{IU} / \mathrm{mL}$ penicillin and $100 \mu \mathrm{~g} / \mathrm{mL}$ streptomycin. All other cell lines were
maintained in RPMI 1640 (Hyclone) supplemented with $10 \%$ (v/v) heat inactivated FBS, $100 \mathrm{IU} / \mathrm{mL}$ penicillin and $100 \mu \mathrm{~g} / \mathrm{mL}$ streptomycin.

### 2.1.4 Minigene and plasmid construction

### 2.1.4.1 Integrin expression plasmids

The calnexin_myc clone constructed in the pCMV6-Entry vector was obtained from Origin. The $\alpha \mathrm{L}$ (CD11a), $\alpha \mathrm{M}$ (CD11b), $\alpha \mathrm{X}$ (CD11c) and $\beta 2$ (CD18) cDNA expression plasmids and minigenes were previously constructed in our laboratory by ligating the respective cDNA sequences into the pcDNA3.0 mammalian expression vector (Invitrogen) (Figure 2.1). All pcDNA3.0 based plasmids can be maintained by ampicillin selection.

Details for the CD11a and CD18 plasmids are as follows: KpnI and SpeI was used to release the CD18 fragment from the J8.1E plasmid (Douglass et al., 1998) and ligated into the $K p n \mathrm{I}$ and $X b a \mathrm{I}$ RE sites in the multiple cloning site (MCS) of the pcDNA3.0 vector. The CD11a cDNA sequence was blunt ended and cloned into the EcoRV RE site in the MCS of pcDNA3.0.


Figure 2.1 Plasmid map of the pcDNA3.0 mammalian expression vector (Invitrogen). All of the mammalian expression constructs used in this thesis were cloned into this plasmid amplification and protein expression, including the CD11a, CD11b, CD11, and CD18 cDNA sequences.

### 2.1.4.2 Minigene construction for splicing project

Two minigenes containing the ITGB2 (CD18) intron 2 sequence inserted between exon 2 and exon 3 were previously constructed by Chan Hwee Sing. One contained the wild type CD18 intron 2 sequence, while the other contained a $-10 \mathrm{bp} \mathrm{C}>\mathrm{A}$ mutation. These minigenes were respectively annotated as WT and MUT.

Using SDM, the authentic AG dinucleotide at the 0 bp position was mutated into another purine pair; AA, GA or GG, in both the WT and MUT minigenes. The -43 bp AG dinucleotide was also mutated to GG, in the context of the WT and MUT
minigenes. Only 2 minigenes were created for this series of experiments; WT_-43GG (WG) and MUT_-43GG (MG).

After SDM, parental DNA was removed using 1 hour of DpnI digestion. This was because parental DNA was amplified in E.coli, and contained methylated plasmid DNA, which the DpnI enzyme was able to selectively digest. The reaction mix was transformed into chemically competent E.coli after removal of parental DNA. Since the constructs were based on pcDNA3.0, the desired clones were selected using ampicillin. One clone from each transformation was selected for sequencing using the T7 universal primer to check for successful mutation.

### 2.1.4.3 HIS tagged CD18 plasmid (CD18_HIS) construction

A combinatorial PCR strategy was used to piece together 2 PCR products what would flank a 'PCR introduced' 6 X HIS tag. The first (front portion) product was obtained using the primers F41905/R2540_R6HIS_stop. The reverse primer consisted of 25 bp of complementary sequence to CD18 and a 6 X HIS tag was inserted before the stop codon.

The second half of the combined product was obtained using R2540 R6HISstop (F2541_F6HIS_stop, the reverse complement of R2540_R6HIS_stop), and a reverse primer designed to bind within pBluescript (pBSR242), which was the vector backbone used to construct J8.1E.

The 2 PCR products were used as templates in the third PCR reaction, using F41905 and pBSR242 as the PCR primers. A combinatory PCR product of 1500 bp was obtained.

The combined PCR product and the J8.1E plasmid were separately digested with SacII and SpeI, and the fragments were resolved in an agarose gel. The digested pBluescript vector and the digested PCR product were excised, purified and ligated using T4 DNA ligase to create J8.1E_HIS. PCR was used to screen for a clone with the appropriate HIS tag insert using a 6 X HIS primer and the pBSR242 primer. One of the positive clones was then chosen for sequencing to confirm the HIS insert was present, and amplified in maxiprep in preparation for further manipulation.

To create a HIS tagged CD18 protein in the pcDNA3.0 plasmid, J8.1E_HIS was digested with KpnI and SpeI to release the entire CD18 fragment containing the HIS tag. The released fragment was cloned into the pcDNA3.0 plasmid, which was predigested separately with $K p n \mathrm{I}$ and $X b a \mathrm{I}$ and purified. Sticky 3' overhangs generated by SpeI and $X b a \mathrm{I}$ digestion were compatible. PCR was used to screen for a positive CD18_HIS clone.

### 2.1.4.4 FLAG tagged CD18 plasmid (CD18F) construction

The primers F41905 / R2540_RFLAGstop were used to create the front portion of the combinatory PCR product. The primers F2541_FFLAGstop / pBSR242 were used to create the back portion of the combinatory PCR product. When the two fragments were obtained, the purified fragments were used as templates in a third combinatory PCR reaction. The combinatorial PCR product obtained was digested with SacII and SpeI This fragment was cloned into the J8.1E plasmid which was also digested with the same enzymes.

10 clones obtained after ampicillin selection were screened with PCR using a FLAG sequence based forward primer (Fflagstop) and pBSR242. One clone out of ten clones was found to be positive. After maxiprep amplification of the J8.1E_FLAG plasmid, KpnI and SpeI were used to excise the entire CD18 sequence containing the FLAG tag, which was subsequently cloned into pcDNA3.0. The latter was separately digested with KpnI and $X b a \mathrm{I}$ before ligation usig T 4 DNA ligase. The primer pair Fflagstop/pcDNA3_R1555 was used to screen for a positive clone. pcDNA3_R1555 was designed to bind within the pcDNA3.0 plasmid after the MCS.

### 2.1.4.5 Truncated CD18 (t-CD18) construction

Based on a 'CD18_ILess' plasmid containing a N50S mutation, created by Prof Law previously, SDM was used to remove the N50S mutation to obtain the t-CD18 plasmid using the primers F_CD18_S50N and R_CD18_S50N. In this plasmid, the entire I-domain of CD18 was removed and the adjoining sequences were spliced together. Similar to the other integrin expression plasmids, this plasmid was cloned into the pcDNA3.0 vector.

### 2.1.4.6 FLAG tagged t-CD18 (t-CD18F) construction

The CD18F sequence was analyzed and an $X b a I$ site was identified in the $5^{\prime}$ UTR region, at position 150 bp . This portion of the 5' UTR sequence had been cloned into the CD18F cDNA expression plasmid. In addition, the SacII restriction site was previously used for cloning in J8.1E. Sequence analysis showed that the SacII RE site was located downstream of the I-domain at 1990 bp, upstream of the FLAG sequence (Figure 2.2). The CD18F and t-CD18 plasmids were digested with $X b a \mathrm{I}$ and SacII to release the DNA fragment flanking the I-domain. The released fragment from t-CD18 was cloned into the CD18F backbone to create t -CD18F, using T 4 ligase. The
ampicillin selected clones were screened with RE digestion using $X b a \mathrm{I}$ and SacII double digestion, to find a clone containing the correct sized insert.

One of the clones was selected for sequencing. A NotI restriction site sequence (GCGGCCGC) was found at 593 bp , at the junction where the I-domain was previously removed in the 'CD18_ILess N50S' plasmid. The reason these extra sequences were not detected was that the 593 bp position (at the junction between the 2 hybrid domain segments) was not sequenced when the N50S mutation was corrected previously, to cut down on experimental costs.


Figure 2.2 Schematic diagram of the XbaI and SacII RE sites in CD18F and t-CD18.

It was reasoned that the 6 bp sequence was short enough to be removed using SDM. The cDNA sequence flanking the NotI sequence by 23 bp on either side was used as DNA primers in SDM. After SDM, parental plasmids were degraded using DpnI. Following ampicillin selection, one clone was selected for maxiprep and DNA sequencing. Sequencing results showed that the SDM experiment was successful; the 8 bp NotI sequence was removed in the truncated CD18 FLAG plasmid (t-CD18F).

### 2.1.4.7 FLAG tagged CD11a (CD11aF) construction

The CD11aF sequence was created using the F_CD11a_KpnI_ext / R_CD11a_flagSTOP_XbaI_ext primer set in PCR amplification. These primers contained the $K p n \mathrm{I}$ and $X b a \mathrm{I}$ RE sites respectively, with 4 bp of random overhangs at
the 5' end to facilitate RE digestion. After PCR amplification and RE digest with the respective REs, the PCR product was ligated into the pcDNA3.0 vector, which was previously digested using $K p n \mathrm{I}$ and $X b a \mathrm{I}$. The ampicillin selected clones were screened using RE digestion with $K p n \mathrm{I}$ and $X b a \mathrm{I}$. To ensure that the sequence was correct, the CD11a_FLAG clone (CD11aF) was amplified using maxiprep and sequenced from either ends using the T7 and SP6 universal primers present in pcDNA3.0.

### 2.1.5 Epitope Mapping of the MEM148 antibody

The MEM148 epitope was previously mapped to the VTHRNPQ region on the human CD18 through sequence alignment with the mouse CD18 sequence. Through this, the mouse CD18 sequence ASSIGKS was determined to be analogous to the human VTHRNQP sequence. A series of 8 knock out plasmids based on the human CD18 sequence, and a series of 9 knock in plasmids based on the chimeric $\beta 2 \mathrm{Hu} / \mathrm{Mo}$ plasmids were cloned previously by Emilia Tng (Tng, et al., 2004).

The plasmids are described here in detail for reference. Using the human CD18 expression plasmid as template, 8 expression plasmids were created by replacing the human residues in the ASSIGKS region with the corresponding mouse residues. These were designated as the knockout plasmids. A second set of plasmids were created using the $\beta 2 \mathrm{Hu} / \mathrm{Mo}$ expression plasmid. Here, mouse VTHRNQP residues were replaced with human residues, and the result plasmids were designated as the knock-in plasmids (Table 2.1). For MEM148 epitope analysis, expression plasmids were transfected individually into COS-7 cells as described in section 2.2.15 and analyzed in FACS as described in section 2.2.14 using the antibodies KIM127 (positive control) and MEM148.

| Knock out Expression Plasmids |  |
| :--- | :--- |
| Original residues on <br> human $\beta 2$ | Replacing residues |
| VT | AS |
| HR | SI |
| NQ | GK |
| P | S |
| VTHR | ASSI |
| NQP | GKS |
| HRNQ | SIGK |
| QP | KS |
| Knock in expression plasmids |  |
| Original residues on <br> chimeric $\beta 2$ Hu/Mo | Replacing residues |
| AS | VT |
| SI | HR |
| GK | NQ |
| ASSI | VTHR |
| GKS | NQP |
| SIGK | HRNQ |
| SIGKS | HRNQP |
| ASSIGKS | VTHRNQP |
| S | P |

Table 2.1 Details of residue swapping in the knock-inand knock out plasmids

### 2.1.6 Antibodies

The antibodies used in this thesis for FACS analysis, western blotting, and immunoprecipitation, are listed as follows in Table 2.2. Of note, some of the antibodies were used to activate LFA-1; KIM185, MEM48 and KIM127. The epitopes for these antibodies are located in the IEGF domains. mAb KIM185 binds to the end of IEGF4, MEM48 binds to IEGF3, while KIM127 binds to the IEGF2 (Lu, et al., 2001).

| Antibody Name | Species | Target | Source |
| :---: | :---: | :---: | :---: |
| 1B4 | Mouse anti-human mAb | $\beta 2$ heterodimers | ATCC |
| MHM23 | Mouse anti-human mAb | $\beta 2$ heterodimers | A.J. McMichael (John Radcliffe Hospital, Oxford, UK) |
| MHM24 | Mouse anti-human mAb | $\alpha$ L I-domain | A.J. McMichael (John Radcliffe Hospital, Oxford, UK) |
| H52 | Mouse anti-human mAb | $\beta 2$ (mid region) | ATCC |
| KIM127 | Mouse anti-human mAb | $\beta 2$ I-EGF 2/3 domain | M.K. Robinson (Celltech, Slough, UK) |
| MEM148 | Mouse anti-human mAb | $\beta 2$ hybrid domain | AbD Serotech, Oxford, UK |
| MEM48 | Mouse anti-human mAb | $\beta 2$ | Fitzgerald Industries |
| KIM185 | Mouse anti-human mAb | $\beta 2$ | M.K. Robinson (Celltech, Slough, UK) |
| LPM19C | Mouse anti-human mAb | $\alpha \mathrm{M}$ I-domain | $\begin{aligned} & \text { S.K. Law (NTU, SBS, } \\ & \text { Singapore) } \end{aligned}$ |
| 4B4 | Mouse anti-human mAb | $\beta 1$ | Beckman Coulter |
| 8D4 | Mouse anti-human mAb | Linearized Talin epitope | Sigma |
| Anti- CD29 | Mouse anti-human mAb | Linearized $\beta 1$ epitope | BD transduction Laboratories |
| Anti-Actin Ab-5 | Mouse anti-human mAb | Linearized Actin epitope | BD transduction Laboratories |
| CD11a (EP1285Y) | Rabbit anti-human mAb | Linearized $\alpha$ L epitope | Abcam |
| CD18 (EP1286Y) | Rabbit anti-human mAb | Linearized $\beta 2$ epitope | Abcam |
| DM1a-FITC | Mouse anti-human mAb | microtubules | Sigma |
| phospho-histone H3 | Rabbit anti-human mAb | phosphorylated histone H3 | Sigma |
| Alexa fluor 488 | Goat anti-rabbit |  | Invitrogen |
| Alexa fluor 488 | Rabbit anti-mouse |  | Invitrogen |
| Alexa fluor 594 | Goat anti-rabbit |  | Invitrogen |
| Alexa fluor 595 | Rabbit anti-mouse |  | Invitrogen |
| MEM148-FITC | Mouse anti human | $\beta 2$ hybrid domain | Serotech, Oxford, UK |
| AE1 | Mouse anti-human mAb | Mitochondria Marker | Santa Cruz |
| Anti-cMyc | Rabbit anti cMyc | C-Myc tag | Sigma |
| Anti-HIS | Rabbit anti-HIS | HIS tag | Sigma |
| M2 antibody | Mouse anti FLAG | FLAG tag | Sigma |

Table 2.2 List of antibodies used in the work performed in this thesis

### 2.1.7 General Buffers and solutions

- LB broth: $1 \%(\mathrm{w} / \mathrm{v})$ Bacto-tryptone (BD), $0.5 \%(\mathrm{w} / \mathrm{v})$ yeast extract (BD), 1 \% (w/v) NaCl into 1 litre of distilled water.
- LB agar: LB broth supplemented with $1.5 \%(w / v)$ Bacto-agar (BD) into 1 litre of distilled water.
- Transfer buffer for western blot: 6.06 g Tris, 7.70 g Glycine and 50 ml methanol into 500 ml of distilled water
- SDS-PAGE running buffer: 6 g Tris, 28 g Glycine and 1 g SDS into 1000 ml of distilled water
- Transfer buffer for protein transfer during western blotting: 6.06 g Tris, 7.70 g Glycine, 50 ml methanol to a final volume of 500 ml with water
- $10 \%$ APS: 10 g Ammonium Persulphate in 100 ml water
- 2X SDS-PAGE loading buffer for western blot: 0.16 M Tris, 1.6 \% SDS, 8 M urea, $0.10 \% \mathrm{BPB}$, adjusted to pH 8.0
- SDS-PAGE Running buffer: 6 g Tris, 28 g Glycine, 1 g SDS to a final volume of 1000 ml with water
- Upper gel (stacking) buffer for SDS-PAGE: 1.0M Tris buffer, pH 6.8
- Lower gel (resolving) buffer for SDS-PAGE: 1.5M Tris buffer, pH 8.8
- Commassie Blue SDS-PAGE Staining Solution: 500 ml Methanol, 400 ml Ultrapure water, 100 ml Glacial Acetic Acid, 2.5 g Coomassie Blue, filtered through a Whatman No. 1 filter.
- SDS-PAGE Destaining Solution (1 litre): 785 ml ultrapure water, 165 ml Ethanol, 50 ml Glacial Acetic Acid
- TBST for western blot: 50 mM Tris. $\mathrm{HCl}, \mathrm{pH} 7.4$ and 150 mM NaCl , containing $0.05 \%(\mathrm{v} / \mathrm{v})$ Tween 20.
- Cell lysis buffer: 10 mM Tris- $\mathrm{HCl}(\mathrm{pH} 8.0), 150 \mathrm{mM} \mathrm{NaCl}, 1 \%(\mathrm{v} / \mathrm{v})$ Nonidet P40 and protease inhibitors as appropriate
- Cell freezing media: 10 \% DMSO in heat-inactivated FBS
- Sodium bicarbonate buffer: 1.36 g sodium carbonate, 7.35 g sodium bicarbonate. pH was adjusted to pH 9.2 before topping up to 1 liter of distilled water
- $\mathrm{Mg} /$ EGTA for integrin activation: Stock MgCl 2 (1M); Stock EGTA (100mM); used at a final concentration of 5 mM MgCl 2 and 1.5 mM EGTA
- Tris buffer for quenching biotinylation reaction: Stock solution (1M Tris-HCl); used at a final concentration of 10 mM Tris- $\mathrm{HCl}(\mathrm{pH} 8.0)$
- 1M EDTA in PBS (for non-enzymatic adherent cell dissociation): Stock solution ( 500 mM EDTA, pH 8.0 ), used at a final concentration of 5 mM EDTA into 1000 ml PBS.
- Immunoprecipitation IgG agarose bead wash buffer ( 200 ml ): 10 mM Tris- HCl with $0.05 \%$ SDS ( pH 8.0 )
- Immunoprecipitation elution buffer: 50 mM Glycine- HCl , adjusted to pH 1.9 with 5 M HCl
- FACS wash buffer: RPMI1640 containing $5 \%(\mathrm{v} / \mathrm{v})$ FBS and 10 mM HEPES
- HIS tag equilibration buffer: 50 mM phosphate buffer, $300 \mathrm{mM} \mathrm{NaCl}, \mathrm{pH} 7.4$
- HIS tag elution buffer: 50 mM phosphate buffer, $300 \mathrm{mM} \mathrm{NaCl}, \mathrm{pH} 7.4,500$ mM Immidazole


### 2.2 Methods

### 2.2.1 Cell cryofreezing and recovery

For cryofreezing, after complete removal of cell culture media by centrifugation, pelleted cells were resuspended in 2 ml of cryofreezing media and transferred to a cryotube. The cryotubes were placed in a Cryo freezing container (Nalgene) filled with isopropanol, and placed in a $-80^{\circ} \mathrm{C}$ freezer overnight. On the next day, frozen cells were transferred to a liquid nitrogen filled cryotank.

For cell recovery, frozen cells were thawed by swirling the cryotube in a $37^{\circ} \mathrm{C}$ water bath. Upon thawing, cells were transferred into cell culture media and all traces of DMSO were removed by centrifugation. Recovered cells were cultured for at least 3 passages before use in experiments.

### 2.2.2 Plasmid DNA extraction and purification

For transformation, $1 \mu \mathrm{~g}$ of each plasmid was transformed into $10 \mu \mathrm{l}$ of chemically competent DH5a E. coli cells (Invitrogen) via incubation on ice for 20 minutes followed by heat shock for 60 seconds at $40^{\circ} \mathrm{C}$, upon which 1 mL of sterile LB broth was added to the cells. Cells were incubated with agitation at $37{ }^{\circ} \mathrm{C}$ in a shaker incubator at 140 xg at $37^{\circ} \mathrm{C}$ and plated out onto LB agar (Conda Pronadisa, Madrid, Spain) containing $100 \mu \mathrm{~g} / \mathrm{mL}$ of ampicillin (US Biologicals). These were cultured in an IncuCell incubator (MMM, Munich, Germany) at $37^{\circ} \mathrm{C}$ overnight. A single colony from each plate was picked and transferred into 3 ml of LB broth (Conda) containing $100 \mu \mathrm{~g} / \mathrm{mL}$ ampicillin (US Biologicals, Swampscott, MA, USA). This was cultured overnight in a shaker incubator at 140 xg at $37^{\circ} \mathrm{C} .2 \mathrm{~mL}$ of the bacteria broth was
transferred into 500 mL of LB broth containing $100 \mu \mathrm{~g} / \mathrm{mL}$ of ampicillin and incubated overnight in a shaker incubator 140 xg . at $37{ }^{\circ} \mathrm{C}$.

Plasmid DNA was extracted using the Nucleobond plasmid extraction kit (Macherey Nagel, Düren, Germany). Bacteria cells were pelleted in a 500 mL high speed centrifuge bottle (Nalgene) in a JA 10 rotor (Beckman Coulter, Fullerton, CA, USA) using a Beckman Coulter High Speed Centrifuge at 6000 xg for 15 min at $4{ }^{\circ} \mathrm{C}$. The pellet was then dissolved in 12 mL of cell lysis buffer S 1 by vortexing, lysed in 12 mL of Buffer S 2 and neutralised in 12 mL of Buffer S2. The lysed samples were then subjected to centrifugation at 12000 xg for 40 min and then filtered through folded filter paper into an equilibrated binding column. The column was washed with 10 mL of wash buffer N3 and then eluted with 5 mL of buffer N5. The eluted DNA is precipitated with 11 mL of isopropanol and centrifuged with a JA25.5 rotor in a Beckman Coulter centrifuge at 10000 xg for 30 min . The pellet was washed with 5 mL of $70 \%$ ethanol, centrifuged and dried for 10 min in a vacuum dryer (Eppendorf). Pelleted DNA was dissolved in sterile distilled water.

### 2.2.3 DNA quantitation

$5 \mu \mathrm{l}$ of each DNA sample was diluted into a final volume of $100 \mu \mathrm{l}$ sterile distilled water in a quartz cuvette. The absorbance at 260 nm was obtained using a UV spectrometer (DU530, Beckton Dickson Biosciences) and the amount of DNA estimated by factoring in the dilution factor and a pathlength of 1 cm . Alternatively, $1.5 \mu \mathrm{l}$ of the DNA sample was analyzed in a nanophotometer (Thermo Scientific).

### 2.2.4 Competent cells

DH5 $\alpha$ strain E.coli was used to prepare chemically competent cells. Briefly, single colonies were obtained by streaking bacteria culture onto LB agar overnight at $37{ }^{\circ} \mathrm{C}$. A single colony was picked and amplified in 10 mL LB broth overnight at $37{ }^{\circ} \mathrm{C} .5$ mL of overnight culture was then transferred into each of two flasks containing 500 mL LB broth.

These cultures were incubated at $37{ }^{\circ} \mathrm{C}$ with aeration until a cell density of 0.5 was achieved at $\mathrm{OD}_{550}$. The cells were transferred to centrifuge bottles and collected by centrifugation at $4{ }^{\circ} \mathrm{C}$ for 10 min at 4000 x g. Collected cells were washed twice with ice cold calcium chloride $\left(\mathrm{CaCl}_{2}\right.$ dissolved in $\mathrm{ddH}_{2} \mathrm{O}$ to 0.1 M$)$, resuspended in 43 mL of the same buffer and topped up with 7 mL of sterile glycerol. The chemically competent cells were stored in aliquots of 0.2 mL in cold microcentrifuge tubes at -70 ${ }^{\circ} \mathrm{C}$.

### 2.2.5 Reverse Transcription

Cells were collected by centrifugation at $300 \mathrm{~g}, 4^{\circ} \mathrm{C}$ for 4 min in a Universal 32 R centrifuge (Hettich, Kirchlengern, Germany), washed with PBS, and resuspended in cell culture media. Resuspended cells were mixed with tryptan blue (Sigma) and viable cells were counted using a haemocytometer (Reichert, Buffalo, NY, USA) in a Nikon eclipse TS100 light microscope (Nikon, Melville, NY, USA). $5 \times 10^{5}$ cells were used for each total RNA extraction.

The ZR Mini RNA Isolation Kit (ZYMO research corporation, Orange, CA, USA) was used for cell lysis and the extraction of total RNA. Briefly, cells were collected in
1.5 mL microcentrifuge tubes and the supernatant completely removed. $200 \mu \mathrm{l}$ of RNA extraction buffer was added and the cells were incubated on ice for 20 min . Following that, $200 \mu \mathrm{l}$ of $100 \%$ ethanol (sigma) was added to the lysate before 10 min of incubation on ice. The mixture was passed through a ZYMO-SPIN column and the column was washed with $200 \mu \mathrm{l}$ of RNA wash buffer twice. Total RNA was eluted from the column with $10 \mu 1$ of RNA elution buffer.

Moloney Murine Leukemia Virus (M-MLV) Reverse Transcriptase deficient in RNase H activity (Promega) was used for the synthesis of first strand cDNA according to the manufacturers instructions. $1 \mu 1$ of Oligo $(\mathrm{dT})_{15}$ or random hexamer was added to $10 \mu 1$ of total RNA. The mixture was heated to $70{ }^{\circ} \mathrm{C}$ for 5 min and cooled on ice. $5 \mu \mathrm{l}$ of M-MLV RT 5X reaction buffer, $5 \mu \mathrm{l}$ of 10 mM dNTP mix (Promega), $1 \mu \mathrm{l}$ of M-MLV RT and $3 \mu \mathrm{l}$ of DEPC-treated water (Invitrogen) were added to the tube, which was then incubated at $40^{\circ} \mathrm{C}$ for 10 min and $55^{\circ} \mathrm{C}$ for 50 min . The mixture was inactivated by heating at $70^{\circ} \mathrm{C}$ for 15 min .

### 2.2.6 Genomic DNA and cDNA extraction for LAD-I project

10 ml of blood were drawn from the patient, her parents, and an unrelated healthy individual. Genomic DNA was extracted using the QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's instructions. Total RNA was prepared with TRIZOL® Reagent (Invitrogen) and first strand cDNA was synthesized using random primer (Roche Diagnostics) with MuLV Reverse Transcriptase and RNase Inhibitor (Applied Biosystems). The genomic DNA and cDNA were sent from Austria to Singapore for molecular analyses.

### 2.2.7 Polymerase chain reaction

PfuTurbo (Stratagene) was used for PCR with an initial denaturation of $95^{\circ} \mathrm{C}$ for 2 min , and 30 cycles of $95^{\circ} \mathrm{C}$ for $1 \mathrm{~min}, \mathrm{Tm}-5$ for 30 sec and $72^{\circ} \mathrm{C}$ for $1 \mathrm{~min} / \mathrm{kb}$. A final extension step of $72^{\circ} \mathrm{C}$ for 10 min was used. Where necessary, 10 cycles of semi nested or nested PCR was performed using $1 \mu \mathrm{l}$ of PCR product as template and the PCR cycling parameters from above

### 2.2.8 PCR amplification for 3'ss analysis

Total RNA was extracted from transfected cells (see section 2.1.4.2 for minigene details), treated with DNase I to remove genomic and plasmid DNA, and reverse transcribed to cDNA using random hexamers. $2 \mu \mathrm{l}$ of cDNA product from each sample was used as PCR template, amplified for 25 cycles using a set of primers flanking exon 2 and exon 3 of the minigenes (F228/R429). After a first round of PCR amplification, PCR products were resolved on a $1 \%$ agarose gel. The gel region corresponding to 150 bp to 300 bp for each sample was excised for gel purification, and eluted into $30 \mu \mathrm{l}$ of water. $2 \mu \mathrm{l}$ of purified PCR product was used as template in an additional 15 cycles of PCR, where required.

### 2.2.9 Inverse PCR for LAD-I project

$1 \mu \mathrm{~g}$ of genomic DNA from each subject was digested for 7 hr at $37^{\circ} \mathrm{C}$ with 25 U of REs (BamHI, and EcoRI (NEB) were used). REs were heat inactivated for 20 min at $80^{\circ} \mathrm{C}$ or $65^{\circ} \mathrm{C}$. After purification (PCR Cleanup Kit, Qiagen) the digested DNA samples were ligated (circularized) at $4{ }^{\circ} \mathrm{C}$ for 18 hr using 4000 U of T4 DNA ligase (NEB) in a final volume of $200 \mu \mathrm{l}$. The DNA was precipitated with $20 \mu \mathrm{l}$ of 3 M sodium acetate and an equal volume of isopropanol, using glycogen, at $1 \mu \mathrm{~g} / \mu \mathrm{l}$, as carrier. The DNA was washed twice in $70 \%$ ethanol and resuspended in $20 \mu \mathrm{l}$ of
nuclease free water. Circularized genomic DNA (cgDNA) was then amplified with relevant PCR primers.

### 2.2.10 Cycling parameters for Site Directed Mutagenesis

Site direct mutagenesis was performed using the Stratagene PfuTurbo enzyme in 30 cycles of PCR. Briefly, SDM primers were designed with at least 15 bp flanking both sides of the mutation site. PCR parameters were as follows: $95^{\circ} \mathrm{C}$ for $1 \mathrm{~min}, 95^{\circ} \mathrm{C}$ for $30 \mathrm{sec}, 55^{\circ} \mathrm{C}$ for 1 min and $1 \mathrm{~min} / \mathrm{kb}$ of product length at $68^{\circ} \mathrm{C}$. Primers used were as described earlier.

### 2.2.11 DNA sequencing

Where necessary, $10 \mu 1$ of purified plasmid DNA ( $100 \mathrm{ng} / \mu \mathrm{l}$ ) or purified PCR product ( $30 \mathrm{ng} / \mu \mathrm{l}$ ) were used for each sequencing reaction processed by either AIT biotech (Singapore) or 1st base (Singapore). $5 \mu \mathrm{l}$ of desalted primers ( 10 mM ) synthesized at a 50 nM scale were obtained from $1^{\text {st }}$ base and provided as sequencing primers.

### 2.2.12 Gel electrophoresis

$10 \mu \mathrm{l}$ of the PCR product or purified plasmid DNA was mixed with $2 \mu \mathrm{l}$ of 6 x loading dye (Fermentas). PCR products were analyzed using a $1 \%$ agarose gel for DNA fragments larger than 1 kb , or a $1.5 \%$ agarose gel for smaller DNA fragments, immersed in Tris Borate-EDTA (TBE) buffer and subjected to electrophoresis at 100 V for 45 min .

### 2.2.13 Gel purification of PCR products

The desired PCR product was excised from the agarose gel using a clean scapel and purified using the QIAquick gel extraction kit (Qiagen). The excised agarose fragments were weighed, 3 volumes of buffer QC was added to the sample, and each
sample was incubated for 10 min at $50{ }^{\circ} \mathrm{C}$ with agitation until all agarose fragments were completely dissolved. 1 gel volume of isopropanol was added to each sample. The solution was passed through a QIAquick column, washed with 0.75 mL of buffer PE and eluted with $30 \mu \mathrm{l}$ of TE buffer.

### 2.2.14 Fluorescence Activated Cell Sorting (FACS) Analysis

Cells were incubated with $0.05 \mu \mathrm{~g} / \mu \mathrm{l}$ of primary mAb in $50 \mu \mathrm{l}$ of wash buffer for 30 $\min$ at $37{ }^{\circ} \mathrm{C}$ with or without activating agents $\mathrm{Mg} /$ EGTA ( 5 mM MgCl 2 and 1.5 mM EGTA),. The cells were washed twice with wash buffer and incubated with secondary antibody diluted in wash buffer (FITC-conjugated sheep anti-mouse $\mathrm{F}\left(\mathrm{ab}^{\prime}\right)$ 2, 1:400 dilution, Sigma) for 30 min at $4{ }^{\circ} \mathrm{C}$. Stained cells were washed once in wash buffer and fixed in $1 \%(\mathrm{v} / \mathrm{v})$ formaldehyde in PBS. Cells were analyzed on a FACS Calibur flow cytometer (Becton Dickinson), and the data analyzed using the CellQuest pro software (Becton Dickinson). Expression index was obtained by multiplying the percentage of cells gated positive with the geometric mean fluorescence intensity.

Flow cytometric analyses of patient leukocytes and transfectants were performed by our collaborators in Austria as previously described (Uzel, et al., 2008). Anti-CD18 and anti-CD45 mAbs were purchased from Beckman Coulter. The mAb IB4 (Wright, et al., 1983) was obtained from ATCC. Data were collected using the FACSCaliber flow cytometer. The data was sent from Austria to Singapore for data analysis and processing using the CellQuest Pro software (Becton Dickinson).

### 2.2.15 Polyfection

Polyfectamine (Invitrogen) was used to transfect plasmid DNA into adherent HEK 293 T or COS-7 cells. For integrin dimer expression in HEK 293 T cells grown in a

40 mm culture dish (Techno Plastic Products (TPP), Trasadingen, CH), $1.5 \mu \mathrm{~g}$ of each expression plasmid DNA was added to a microfuge tube. Where only one plasmid was transfected, a total of $1.5 \mu \mathrm{~g}$ of plasmid was transfected. Plasmid DNA was diluted to $100 \mu \mathrm{l}$ with DMEM (Hyclone, Thermo Scientific) and $15 \mu \mathrm{l}$ of polyfect was thoroughly mixed into the solution, which was incubated for 10 min at room temperature. During this time, 1.5 ml of fresh cell culture media was used to replace spent media in the 40 mm dish. Following 10 min of incubation, 0.6 mL of cell culture media was added to the Polyfectamine-DNA mixture, gently mixed, and added to the 1.5 ml of fresh cell culture media in the culture dish. Cells were incubated for at least 18 hr before further experiments. COS-7 cells were cultured in RPMI media, hence RPMI 1640 (Hyclone) replaced DMEM in the otherwise identical polyfection procedure.

### 2.2.15.1 Lipofection

HEK293FT cells were maintained in media containing antibiotics. This media was replaced with HEK media without antibiotics 1 hour before lipofection. $15 \mu \mathrm{l}$ of the plasmid DNA obtained above was mixed with $15 \mu \mathrm{l}$ of Viralpower packaging mix (Invitrogen) and diluted with 2 mL of DMEM-Glutamax (Gibco). $36 \mu \mathrm{l}$ of lipofectamine (Invitrogen) was mixed with 2 mL of DMEM-Glutamax (Gibco) and left to stand for 5 min . The two solutions were combined and left to stand for 20 min . The combined lipofectamine-DNA complexes were added to the HEK293FT cells for 6 hours and replaced with cell line specific media, depending on the target cell type for viral transduction

### 2.2.16 Cell Surface Biotinylation

PBS-A (PBS with additional salt) was prepared by dissolving $\mathrm{MgCl}_{2}$ and $\mathrm{CaCl}_{2}$ to a final concentration of 0.5 mM and 0.15 mM in PBS ( pH 7.4 ). Biotin solution was prepared by dissolving Sulfo-NHS-SS-Biotin (Pierce) in PBS to a concentration of 2 $\mathrm{mg} / \mathrm{mL}$ and diluting this solution by 10 x in PBS-A to obtain a $0.5 \mathrm{mg} / \mathrm{ml}$ biotin solution.

Cell culture media was completely removed from the cells, which were washed in warm PBS containing $0.1 \%(\mathrm{w} / \mathrm{v})$ BSA to remove all traces of media, dead cells and debris. Cells were incubated in 5 ml of warm $0.5 \mathrm{mg} / \mathrm{ml}$ sulfo-NHS-biotin (Pierce) for 30 min in a humidified $37{ }^{\circ} \mathrm{C}$ incubator. Biotin solution was removed and cells were rinsed with quenching buffer (PBS containing 10 mM Tris- $\mathrm{HCl}(\mathrm{pH} 8.0)$ and $0.1 \%$ (w/v) BSA) Cells were resuspended in quenching buffer, collected at 300 xg in a centrifuge and rinsed in quenching buffer.

### 2.2.17 Conjugation of rabbit anti-mouse IgG to protein A sepharose beads

0.5 g of Protein A sepharose (PAS) beads were suspended in 50 mL of PBS and agitated overnight at $4{ }^{\circ} \mathrm{C}$. On the next day, the swollen beads were sedimented by centrifugation at 300 xg for 5 min and the supernatant was removed. Swelled beads were rinsed twice with PBS, occupying a bead volume of 2 mL .4 mL of PBS was added to the beads together with 1 mL of rabbit anti-mouse (RAM) antibody, and agitated for 1 hour at $4{ }^{\circ} \mathrm{C}$. Following that, swelled PAS-RaM beads were collected by centrifugation, the supernatant was removed and PAS-RAM beads were rinsed with PBS. 6 mL of PBS was added to make up an $8 \mathrm{~mL} 25 \%$ bead suspension of PAS-

RAM. Before use in immunoprecipitation experiments, the required volume of swelled beads was washed and resuspended in IP buffer.

### 2.2.18 Immunoprecipitation

Biotin-labelled cells were incubated in wash buffer (RPMI 1640 containing 5 \% FBS and 10 mM HEPES) containing the appropriate antibodies ( $2 \mu \mathrm{~g}$ in $150 \mu \mathrm{l}$ of wash buffer) in the absence or presence of $\mathrm{Mg} /$ EGTA or magnesium chloride $(0.5 \mathrm{mM})$ for 30 min at $37^{\circ} \mathrm{C}$. Cells were washed with wash buffer

Cells were spun down and lysed in $200 \mu$ of lysis buffer [10 mM Tris-HCl ( pH 8.0 ), 150 mM NaCl , and $1 \%$ (v/v) Nonidet P40] containing protease inhibitors (Roche Diagnostics, Basel, Switzerland) by vortexing and incubating for 30 min on ice. Cell nuclei were removed by centrifugation for $15 \mathrm{~min}, 12,000 \mathrm{~g}$ at $4{ }^{\circ} \mathrm{C} .200 \mu \mathrm{l}$ of biotinylated cell lysate was precleared by adding $2 \mu \mathrm{~g}$ of irrelevant mAb as appropriate, and incubated at $4{ }^{\circ} \mathrm{C}$ for 45 min with rotation
$20 \mu \mathrm{PAS}-\mathrm{RaM}$ suspension was spun down, the mixture was added to the sedimented beads, and rotated at $4{ }^{\circ} \mathrm{C}$ for 30 min . PAS-RaM was sedimented by centrifugation $\left(10000 \mathrm{xg}, 2 \mathrm{~min}, 4^{\circ} \mathrm{C}\right)$ and the cleared supernatant was transferred to a fresh tube. 2 $\mu \mathrm{g}$ of appropriate mAb was added to the precleared cell lysate and incubated at $4{ }^{\circ} \mathrm{C}$ for 45 min with rotation. Subsequently, $50 \mu \mathrm{l}$ of PAS-RaM suspension was spun down in another tube, lysis buffer removed, and the mixture was incubated with sedimented PAS-RAM at $4{ }^{\circ} \mathrm{C}$ for 1 h with rotation.
$3 \mu \mathrm{~g}$ of appropriate mAb was added to the precleared cell lysate and incubated at $4{ }^{\circ} \mathrm{C}$ for 45 min with rotation. Subsequently, $70 \mu \mathrm{l}$ of PAS-RaM suspension was spun down
in another tube, lysis buffer removed, and the mixture was incubated with PAS-RAM at $4^{\circ} \mathrm{C}$ for 1 h with rotation.

Beads were collected by centrifugation at $10000 \mathrm{xg}, 4^{\circ} \mathrm{C}$, for 2 min , supernatant was discarded, and the PAS-RaM was washed in $500 \mu \mathrm{l}$ IP-buffer thrice. Proteins were eluted from the beads using loading dye ( 0.16 M Tris, 8 M Urea, 1.6 \% (w/s) SDS, $0.08 \%(\mathrm{w} / \mathrm{v})$ bromophenol blue, pH 8.0$)$ containing 30 mM DTT, heating at $100^{\circ} \mathrm{C}$ for 5 min , centrifugation at 10000 xg to collect the beads at the bottom of the tube. Immunoprecipitated proteins were revolved in using 7.5 \% SDS-PAGE.

### 2.2.19 Reciprocal IP of calnexin_myc, CD18F and t-CD18

HEK293 cells were transfected with CD18F and calnexin_myc, or t-CD18 and calnexin_myc. 24 hours after transfection, the transfected cells were harvested for cell lysate, precleared with PAS-RAM for 30 min , and then incubated with H52-PASRAM for 1 hour. The beads were washed with lysis buffer and eluted in SDS-PAGE loading dye containing 30 mM DTT. As a control, mock transfected cells were also subject to the same procedure.

### 2.2.20 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Protein estimation was performed using the Bradford assay (Biorad). $5 \mu \mathrm{~g}$ of protein sample was mixed with an equal volume of 2 x loading dye containing 30 mM DTT, heated at $100^{\circ} \mathrm{C}$ for 10 min and resolved using SDS-PAGE in a Mini Electrophoresis Set (Biorad) immersed in SDS-PAGE electrophoresis buffer.

### 2.2.21 Western Blotting

Proteins resolved using SDS-PAGE were transferred onto polyvinylidene diflouride (PVDF) membranes (Immobilon-P, Millipore) by western blotting. The membrane was equilibrated in methanol for 20 sec , washed with distilled water, and soaked in transfer buffer ( 12 mM Tris- $\mathrm{HCl}, 95 \mathrm{mM}$ glycine, $10 \%(\mathrm{v} / \mathrm{v})$ methanol) for 5-10 min. The gel was equilibrated by soaking in transfer buffer for at least 15 min .

Protein samples were transferred using wet transfer apparatus (Bio-Rad) at 200 mA at $4{ }^{\circ} \mathrm{C}$ for 1 hour or at 10 V at $4{ }^{\circ} \mathrm{C}$ overnight. Membranes were blocked for 30 min at RT with $1 \%(\mathrm{w} / \mathrm{v})$ BSA diluted in Tris Buffered Saline with $0.1 \%$ Tween-20 (TBST), followed by incubation in primary antibody in $1 \%$ BSA ( $\mathrm{w} / \mathrm{v}$ ) diluted in TBST for 2 hours at RT or overnight at $4{ }^{\circ} \mathrm{C}$.

After primary antibody incubation, membranes were washed thrice with TBST and incubated with streptavidin - horse radish peroxidase (HRP) conjugated secondary antibody (1:5000, Amersham) for 1 hr at RT. Membranes were washed thrice with TBST, developed with with Enhanced Chemiluminescence (ECL) Plus Detection Kit (Amersham) and visualized on x-ray film (Konica Minolta Inc., Japan). The films were developed in a Kodak X-OMAT 2000 processor (Kodak, Ontario, Canada).

### 2.2.22 Cell adhesion assay

For ICAM (ICAM-1 and ICAM-3) binding assays, polysorb microtire well plates (Nunc) were coated with $100 \mu \mathrm{l} /$ well of $5 \mu \mathrm{~g} / \mathrm{ml} \mathrm{Fc}$-specific IgG (goat anti-human, Sigma) diluted in Bicarbonate ( $\mathrm{pH} 9.2,100 \mathrm{mM}$ ), and left overnight at $4^{\circ} \mathrm{C}$. On the next day, the wells were rinsed twice with $150 \mu 1$ of PBS per well. Unspecific binding
sites were blocked using $150 \mu \mathrm{l} /$ well of $0.5 \%$ BSA in PBS at $37{ }^{\circ} \mathrm{C}$ for 30 min and wells were rinsed with $150 \mu 1 /$ well of PBS. Each well was coated with $100 \mu 1$ of $\operatorname{ICAM}(1 \mu \mathrm{~g} / \mathrm{ml})$ in $0.1 \%$ BSA (PBS) at room temperature for 2 hr or $37^{\circ} \mathrm{C}$ for 30 min . The plates were washed twice with wash buffer before use.

For iC3b binding assays, iC3b was diluted to a concentration of $7.5 \mu \mathrm{~g} / \mathrm{ml}$ in bicarbonate buffer. This solution was aliquoted into each well and incubated at $4{ }^{\circ} \mathrm{C}$ overnight. On the next day, the plate was rinsed twice with $150 \mu \mathrm{PBS} /$ well blocked with 0.2 \% PVP (polyvinyl-pyrrolidone) in PBS at $37{ }^{\circ} \mathrm{C}$ for 30 min . Each plate was washed twice with wash buffer before use.

Cells were collected, resuspended in 3.5 ml of wash buffer, counted using a hemocytometer and the required amount ( $30 \times 10^{4}$ in 12 wells) was transferred to a new tube and resuspended in 2.5 ml of wash buffer. $3.5 \mu \mathrm{~g}$ of BCECF ( $2^{\prime}, 7^{\prime}$-bis-(2-carboxyethyl)-5(6)-carboxyfluorescein) was dissolved in $50 \mu \mathrm{l}$ of DMSO, and was diluted to $1 \mu \mathrm{~g} / \mathrm{ml}$ in wash buffer containing the cells, and the cells were incubated at $37^{\circ} \mathrm{C}$ for 20 min . BCECF was removed and the cell pellet was resuspended in wash buffer according to the number of conditions for the assay. $50 \mu \mathrm{l}$ of cell suspension was added into each well. $\mathrm{M} / \mathrm{E}, \mathrm{Mn}$ and $\mathrm{mAbs}(10 \mu \mathrm{~g} / \mathrm{ml})$ as were also added as required. Plates were incubated at $37{ }^{\circ} \mathrm{C}$ for 30 min . The plates were read at absorbance/emission of $480 / 530 \mathrm{~nm}$ with a gain of 75 in a FL600 fluorescence plate reader (Bio-Tek) before washing, after 1 wash and after 2 washes.

### 2.2.23 Quantitative real-time PCR

For analysis of integrin expression at the mRNA level, the applied biosystems power SYBR green Cells-to- $\mathrm{C}_{\mathrm{T}}$ kit was used. qRT-PCR analysis of gene expression in

MOLT-4, Jurkat, SKW-3 and U937 was performed. Cells were counted using a haemocytometer to obtain $1 \times 10^{5}$ cells per lysis reaction. The required amount of cells was aliquoted into a new tube, washed twice with PBS and the supernatant was completely removed. Cells were lysed by incubating for 5 min at room temperature in $50 \mu \mathrm{l}$ of lysis buffer containing DNase I to completely remove genomic DNA, and the lysis reaction was inhibited with $5 \mu \mathrm{l}$ of stop solution for 2 min at room temperature. $10 \mu \mathrm{l}$ of the cell lysate was used for random hexamer based reverse transcription with the provided $20 \times \mathrm{RT}$ mix, $2 \times \mathrm{RT}$ buffer and sterile nuclease free water, by incubating at $37{ }^{\circ} \mathrm{C}$ for 60 min . The RT enzyme was denatured with 5 min of incubation at $95{ }^{\circ} \mathrm{C} .4 \mu \mathrm{l}$ of synthesized cDNA was analysed in an applied biosystems 7500 real time PCR system for gene expression levels relative to actin using the SYBR green premix provided. The cycle parameters used were an initial denaturation at $95{ }^{\circ} \mathrm{C}$ for $10 \mathrm{~min}, 40$ cycles of $95{ }^{\circ} \mathrm{C}$ for 15 sec and $60^{\circ} \mathrm{C}$ for 1 min , and an additional dissociation curve cycle to analyse for reaction specificity.

### 2.2.24 Confocal Microscopy

For confocal microscopy, $1 \times 10^{6}$ cells (MOLT-4, Jurkat, SKW-3 and U937) were collected into $10 \mu \mathrm{l}$ of cell culture media and allowed 10 min to adhere to poly-Llysine coated (PLL) slides at $37{ }^{\circ} \mathrm{C}$ in a humidified chamber. After adhesion, cells were fixed using a $3.7 \%$ formyldehyde solution diluted in PBS, and permeabilized with $0.5 \%$ triton-X diluted in PBS containing $2 \mathrm{mg} / \mathrm{ml}$ BSA and $1 \mathrm{mM} \mathrm{NaN}_{3}$.

For double immunoflourescence staining, cells were blocked with blocking solution (PBS containing $0.05 \%$ Tween $20,2 \mathrm{mg} / \mathrm{ml} \mathrm{BSA}, 1 \mathrm{mM} \mathrm{NaN} 3$ ) for 1 hr at room temperature. The required primary antibodies ( 1 to $10 \mu \mathrm{~g} / \mathrm{ml}$ ) and control reagents
were diluted into blocking solution, and cells were incubated overnight at $4{ }^{\circ} \mathrm{C}$ with primary antibodies in a humidity chamber.

On the next day, cells were washed thrice for 10 min with blocking solution, and incubate for 1 hr with fluorophore conjugated secondary antibodies ( $2 \mathrm{ug} / \mathrm{ml}$ ) in blocking solution. Cells were washed cells thrice for 10 min with blocking solution and incubated with the second primary antibody for 1 hr at room temperature in a humidity chamber. Cells were then washed cells thrice for 10 min with blocking solution and incubated with the second fluorophore conjugated secondary antibody for 1 hr at room temperature in a humidity chamber. After 3 washes with blocking solution, DAPI antifade solution (Invitrogen) was used to mount glass coverslips onto the cells and the coverslips were sealed with nail polish.

Slides were analyzed with an Axiovert 200M fluorescence microscope (Zeiss) at 63x oil immersion, Z-sections were obtained using a confocal attachment (LSM 510 META, Zeiss) and captured images were processed using LSM ZEN (Zeiss).

### 2.2.25 HIS tag purification

For HIS tag purification of HIS-tagged CD18 protein, the plasmid was transfected into HEK 293 T cells together with the CD11a expression vector. Transfected cells were analyzed for cell surface expression of the integrins in FACS before cell lysis in lysis buffer containing protease inhibitors, and diluted to a total of 10 ml in HIS tag equilibration buffer.

For HIS tag purification, a nickel column (GE healthsciences) was initially used. The column was equilibrated in 8 column volumes (CV) of equilibration buffer using a

Fast Protein Liquid Chromatography machine (FPLC, Amersham Biosciences, UPC900) before the diluted cell lysate was injected into the injection lobe, and passed through the column. The column was then washed with 8 CV of equilibration buffer before elution buffer was pumped into the column over a gradient of $30 \mathrm{~min} .500 \mu \mathrm{l}$ fractions were collected and probed using western blot. Later on, a cobalt column (HIStalon, Clonetech) was used for the purification, using the same wash and elution parameters.

For monoQ purification, a 1 ml monoQ column (GE healthcare) was used to purify CD18-HIS. For this, cell lysate was diluted into 10 ml 20 mM Tris-HCl, pH 7.4 , and injected into a MonoQ column at $0.5 \mathrm{ml} / \mathrm{min}$. According to the manufacturer's instructions, after washing with loading buffer for 5 column volumes (CV), Tris-HCl ( 20 mM ) containing $1 \mathrm{M} \mathrm{NaCl}, \mathrm{pH} 7.4$, was injected into the column using a 60 min gradient. Fractions were collected at 1 ml volumes

### 2.2.26 FLAG tag purification

For FLAG purification, M2 conjugated agarose beads were obtained from Sigma. Transfected cells were analysed in FACS for integrin expression level, and lysed in lysis buffer containing protease inhibitiors. However, as the FLAG beads were not available in column format, purification was not performed on a FPLC.

Cell lysate was precleared twice using PAS-RAM at $4{ }^{\circ} \mathrm{C}, 30 \mathrm{~min}$ each time. After preclearing, $20 \mu \mathrm{l}$ of $50 \%$ M2 bead slurry was added to the lysate, and mixed at $4{ }^{\circ} \mathrm{C}$ for 4 hours. The beads were washed once with high salt lysis buffer, transferred to a microbiospin chromatography column (Biorad) and washed twice with lysis buffer.

The FLAG protein was eluted using a $5 \mathrm{mg} / \mathrm{ml}$ solution of FLAG peptide (Sigma) dissolved in lysis buffer, by shaking at room temperature for 20 min .

### 2.2.27 Liquid-Chromatography-Mass-Spectrometry

LCMSMS was performed by Dr. Newman Sze as part of the core facilities services in the School of Biological Sciences. For mass spec analysis, $400 \mu 1$ of FLAG peptide eluted proteins were concentrated using a 3 K filter (Amicon) by centrifugation at 14000 xg for 1 hour. The concentrated eluates were mixed with SDS-PAGE loading dye containing 30 mM DTT and boiled for 10 min before loading onto a $10 \%$ SDSPAGE. Resolved proteins were stained using commassie blue, and diced into 1 mm by 1 mm cubes before submission for LCMSMS analysis.

As the FLAG tagged plasmids were of human origin, and had been transfected into Human Embryonic Kidney cells (HEK 293), the tryptic peptide fragments detected in LCMSMS were matched against the International Protein Index (IPI) or the UniProt human protein databases for protein hit identification. Relative abundance of the identified proteins was estimated using the exponentially modified protein modification index (emPAI) scores (Ishihama, et al., 2005).

## Chapter 3 : Characterizing a novel CD18 mutation in an LAD-I patient

### 3.1 Introduction

A female patient was identified to suffer from Leukocyte Adhesion Deficiency type 1 (LAD-I) in the Elisabethinen Hospital in Linz, Austria, by our collaborators, Georg F. Klein, Jörg Jabkowski, Gabriela Schadenböck-Kranzl and Otto Zach. In 2002, our laboratory was approached by her attending clinician (Jörg Jabkowski) to characterize the associated genetic mutations.

One mutation was characterized by Ms. Chan Hwee Sing and the results were presented as her M.Sc. thesis in the School of Biological Sciences, Nanyang Technological University in 2004. She was unable to find the other mutation. In the following years, several members of our laboratory, including Zhou Shuang, Florence Lim, and Cheng Ming, were also unsuccessful. In 2009 December, I identified the second mutation using a new strategy.

This chapter will first summarize results results from Chan Hwee Sing, useful to provide an understanding of this study. To obtain results in a publishable form, I repeated and extended experiments where necessary (presented in Appendix I for reference). This is followed with work carried out by me after October 2009, leading to detection and characterization of the other mutation in the patient.

Many primers were used in this project. These primers were used for nest and seminested PCR and as sequencing primers to study genomic DNA (gDNA) and cDNA.

The primer binding positions of all primers used for ITGB2 PCR are displayed in Figure 3.1. Througout this chapter the abbreviations $\mathrm{C}, \mathrm{F}, \mathrm{M}, \mathrm{P}$ and L are used to annotate experimental results associated with the normal (healthy) Control, Father, Mother, $\underline{P}$ atient and DNA Ladder (where appropriate) respectively.


Figure 3.1 Schematic diagram of the ITGB2 sequence and the respective primer binding positions. The ITGB2 exons are displayed in green, with intronic regions in grey. The orientations of the primers are indicated by the arrows. Where multiple primer binding locations are in nearby positions, the primer names are arranged in sequential order, starting from the nearest to the farthest primer.

### 3.2 Summary of results for the first LAD-I mutation

### 3.2.1 Patient

The patient is a Caucasian female born in 1996. Her parents are not related. There was no significant delay in separation of the umbilical cord, which is generally taken as the first hint of LAD-I. She developed normally until an episode of a severe facial soft tissue and skin infection which required surgical intervention in 1999. A non-healing ulcer on her right arm was treated in 2002, while severe periodontitis and gingivitis were also observed. When further examined for leukocytosis, the child had an elevated white blood cell (WBC) count of $25000 \mathrm{WBC} / \mu \mathrm{l}$.

### 3.2.2 Two novel intronic point mutations were identified from gDNA

 sequencing of the ITGB2 gene in the patientBoth gDNA and cDNA were prepared using standard commercially available kits by our Austrian collaborators, and the material was shipped to Singapore for analysis. The 16 exons of $I T G B 2$ from the patient were amplified, including approximately 50 bp of flanking sequences using PCR. Seven variations were identified to be different from the reference $I T G B 2$ genomic sequence, and are presented in Table.3.1. Three polymorphic codons had been reported previously, for Leu ${ }^{8}$ in the leader peptide, $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$ (Roos and Law, 2001; Wright, et al., 1995). The remaining variations were found in intronic regions including two previously reported variations, -11 in intron 5 and -47 in intron 8 (Roos and Law, 2001) and two novel variations at position -10 in intron 2 and position -29 in intron 5 . In all seven cases, both the wildtype (RefSeqGene NG_007270.2) and the variant nucleotides were present.

## Exonic variants

| Exon | Amino acid $^{\mathrm{a}}$ <br> affected $^{8}$ | Codon in reference <br> database $^{\mathrm{b}}$ | Codon in patient | Mutation <br> Nomenclature $^{\mathrm{c}}$ |
| :---: | :---: | :---: | :---: | :---: |
| 2 | Leu $^{8}$ | CTG | CTG / CTT | $\mathrm{g} .23080 \mathrm{G}>$ T |
| 10 | Val $^{367}$ | GTC | GTC $/ \mathrm{GTA}$ | $\mathrm{g} .40312 \mathrm{C}>\mathrm{A}$ |
| 11 | Val $^{441}$ | GTT | GTT $/ \mathrm{GTC}$ | $\mathrm{g} .41941 \mathrm{~T}>\mathrm{C}$ |

Intronic variants

| Intron | Position | Nucleotide in reference database | Nucleotide in patient | Mutation Nomenclature ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: |
| 2 | -10 | C | C / A | g.23457C>A |
| 5 | -11 | G | G / T | g.32095G>T |
| 5 | -29 | C | C / T | g.32077C>T |
| 8 | -47 | G | G / A | g.38732G>A |

Table.3.1 Summary of variants detected in the CD18 gene of the patient.
${ }^{\text {a }}$ The position of the amino acid is given as in the coded protein in which the starting methionine is numbered " 1 ".
${ }^{\mathrm{b}}$ The reference database for the CD18 gene (ITGB2) used here is RefSeqGene NG_007270.2.
c The mutation nomenclature used was as described by den Dunnen and Antonarakis (2000).

### 3.2.3 cDNA analysis

cDNA was obtained from peripheral blood leukocytes of the patient using oligo-dT as primers. An insertion of 43 bp was found between exons 2 and 3, and the 43 bp was from the 3 ' end of intron 2. Furthermore, the sequence included the mutant nucleotide "A" instead of the wildtype "C" at the -10 position. Using a forward primer in exon 2 and a reverse primer in exon 4 (F23051/R26792), amplification of cDNA yielded two bands products of 202 and 245 bp (Figure 3.2a). Sequencing of the 245 bp fragment revealed an additional 43 bp at the exon $2 / 3$ junction, identical to the 3 ' end of the mutant intron 2 (r.[59-43_59-1ins;59-10C $>\mathrm{A}]$ ). This suggested that mutation of the C at position -10 to an $\mathrm{A}(\mathrm{g} .23457 \mathrm{C}>\mathrm{A})$ may cause or contribute to
aberrant splicing. According to the nucleic acid sequence, the splice aberration would result in a premature stop codon (p.C19_V20ins11X12) (Figure 3.2b). It should also be noted that the codon for Leu ${ }^{8}$ is CTG in the longer fragment, i.e. with the mutation, but CTT in the normal fragment.


Figure 3.2 Intron 2 mutation analysis. (a) PCR products from the patient's cDNA across spanning the exon 2 / exon 3 junction. Two fragments of approximately 245 and 202 bp were obtained ( P ); $\mathbf{M}=$ marker. (b) Sequence showing the aberrant splicing of the 245 bp fragment, by activation of a cryptic 3 'ss in intron 2 (r.[59-43_59-1ins;59-10C $>A]$ ). The g. $23457 \mathrm{C}>\mathrm{A}$ mutation at position $\mathbf{- 1 0}$ is shown in bold and underlined. Normal splicing would result in the wildtype amino acid sequence of SLGCVLSQ across the exon boundary. Aberrant splicing would result in an additional 43 nucleotides from the 3 ' end of intron 2, and the irrelevant amino acid sequence of GRPRAPTHQPA (in bold) in place of the VLSQ coded in exon 3 (p.C19_V20ins11X12). It should be noted that the mature CD18 protein starts with the $\mathbf{Q}$ in the VLSQ sequence coded in exon 3. Also highlighted (bold and underlined) is the CTG codon for Leu ${ }^{8}$, which appears to be polymorphic (CTG and CTT) from the genomic sequence of the patient (Table 1).

### 3.2.4 cDNA and gDNA PCR products extending beyond exon 11 in the father

## and patient were not polymorphic

When the cDNA and gDNA was amplified using PCR with primers extending beyond exon 11, the polymorphisms at $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$ were not detected. Several
combinations of primers were tested to verify this observation, and the sequencing results are summarized in Table 3.2.

| Template | Region amplified | Val $^{367}$ codons | Val $^{441}$ codons |
| :---: | :---: | :---: | :---: |
| cDNA | Exons 6-11 | GTC / GTA | GTT / GTC |
| cDNA | Exons 1-12 | GTC | GTT |
| cDNA | Exons 9-16 | GTC | GTT |
| gDNA | Exon-10 | GTC / GTA |  |
| gDNA | Exon-11 |  | GTT / GTC |
| gDNA | Exons 10-11 | GTC / GTA | GTT / GTC |
| gDNA | Exons 11-12 |  | GTT |

Table 3.2 Polymorphisms of the $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$ codons in the cDNA and gDNA of the patient.

The polymorphic markers clearly observed in patient gDNA were not present in PCR products extending beyond exon 11. This suggested that there was a mutation after exon 11 in one of the $I T G B 2$ allele in the patient such that an unknown DNA sequence was joined to the $I T G B 2$ sequence somewhere in intron 11. This mutation could involve a sizable deletion, insertion, inversion, or a translocation

### 3.3 Characterizing the second LAD-I mutation

### 3.3.1 Leukocytes from the patient did not express CD18

To analyze cell surface expression levels of the CD18 integrins, FACS analysis of patient PMBCs (peripheral monocytic blood cells) was performed alongside PMBCs obtained from the parents and an unrelated healthy subject (as positive control). The raw data was sent by our collaborators in Austria to our laboratory in Singapore for further analysis.

Using the CellQuest Pro software (Beckon Dickson), the data was further analyzed for expression of CD18, CD11a, CD11b and CD11c in the peripheral blood granulocyte, monocyte and lymphocyte cell populations of the subjects. FSC (cell
volume) vs SSC (cell granularity) scatter plot profiles of PMBCs in all the subjects were as expected. PBMCs from each subject were positively identified with antiCD45 (human hematopoietic antigen) and gated according to side scatter (cell granularity) into granulocytes (mainly neutrophils), monocytes and lymphocytes (composed of B cells, T cells and NK cells) (Figure 3.3).

Patient granulocytes, monocytes and lymphocytes showed less than 2-fold CD18 expression over a non-specific background control (Figure 3.4), whereas those in the normal control are over 35 -fold in the three leukocyte populations. The parents, who are carriers of the LAD-I mutations, showed a normal level of CD18 expression, consistent with earlier reports on parents of LAD-I patients (Arnaout, et al., 1984; Ross, et al., 1985; Springer, et al., 1984). The lack of CD18 expression in the patient corresponded with the diagnosis of Leukocyte Adhesion Deficiency Type I (LAD-I).

Similarly, low expression levels of CD11a, CD11b and CD11c were detected in the patient. In contrast, each parent expressed detectable levels of CD18 and each integrin alpha subunit (CD11a, CD11b and CD11c), comparable to that of a healthy control, in all 3 cell populations. The percentage expression of each integrin protein is summarized in Table 3.3.


Figure 3.3 Scatter dot plots of PBMCs obtained from a healthy control (C), the father (F), mother $(M)$ and patient ( P ). The stained cells were first analyzed for forward scatter (FSC) vs side scatter (SSC). The scatter plot profile of the PMBCs obtained from each subject was as expected (top panel). Cells were then gated according to CD45 expression and granularity (SSC). For each subject, granulocytes are indicated in red, monocytes are indicated in green, while the lymphocytes are indicated in pink.


Figure 3.4 Frequency histogram distributions of CD18 staining intensity for each cell population from a healthy control (C), the father (F), mother (M) and patient (P). The mean fluorescence intensity of each cell population was plotted against the background mean fluorescence intensity (BG MFI) obtained from an anti-mouse FITC antibody as background control. The MFI of cells obtained from the patient were only slightly above that of BG MFI.












CD11b


Figure 3.5 Frequency histogram distributions of CD11a CD11b and CD11c staining intensity for each cell population from a healthy control ( $C$ ), the father ( $F$ ), mother ( $M$ ) and patient. The mean fluorescence intensity of each cell population was plotted against the background mean fluorescence intensity (BG MFI) obtained from an anti-mouse PE antibody as background control. The MFI of cells obtained from the patient were only slightly above that of the BG MFI.

| Percentage expression relative to normal control |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Subject | Cell Type | CD18 | CD11a | CD11b | CD11c |
| Father | Granulocytes | 84.95 | 91.75 | 74.49 | 127.7 |
| Mother | Granulocytes | 90.62 | 91.94 | 82.13 | 189.15 |
| Patient | Granulocytes | 0.6 | 0.96 | 0.06 | 0.97 |
| Father | Monocytes | 153.93 | 105.82 | 335.93 | 180.88 |
| Mother | Monocytes | 120.36 | 99.33 | 178.23 | 145.89 |
| Patient | Monocytes | 0.88 | 0.96 | -0.09 | 0.79 |
| Father | Lymphocytes | 92.56 | 77.03 | 63.47 | 77.3 |
| Mother | Lymphocytes | 92.7 | 77.33 | 73.43 | 160 |
| Patient | Lymphocytes | 0.4 | 0.2 | 2.58 | 2.16 |

Table 3.3 Percentage expression of CD18, CD11a, CD11b and CD11c relative to normal control. The percentage expression relative to a healthy control was obtained according to the following method: Percentage Relative Expression $=[($ Experiment MFI - Experiment BG MFI) $/($ Control MFI - Control BG MFI)] x $100 \%$.

### 3.3.2 Inverse PCR strategies

An idea to characterize the mutation based on unique RE sites in the gDNA sequence, was conceived. Based on these RE sites, gDNA segments would be circularized for PCR amplification using primers orientated outward from defined positions before exon 12 in the gDNA sequence. As it would turn out, this method of amplifying a circular strand of DNA, was previously described by Ochman et al (1988) as inverse PCR.

An initial search was conducted. To limit the number of restriction enzymes under consideration, only enzymes with less than four RE sites within the gene segment from the start of exon 8 to the intergene region after exon 16 were considered. In order to retain exon 11 such that the polymorphic exon of Val ${ }^{441}$ could be used as allele markers and the unknown DNA after intron 11, the RE sites should be located 5 ' to exon 11 but as close to exon 11 as possible, and no site should be found before exon 12 (highlighted box in Figure X). Using these criteria, the following enzymes were identified as "suitable", including AleI, PshAI, ScaI, AclI, BspHI, BsrBI, BbeI, BspRI, FspAI, KasI, NarI, SfoI, TaqI,AloI, BmtI, and XmnI.

Two other enzymes, KpnI and BamHI. were found to be suitable, but were not on the original list because KpnI does not have a site from the region spanning exon 8 to intron 10, and BamHI has 5 sites, respectively, in the region selected. The restriction sites and the possible circularized fragments are shown in Figure 3.7 and Figure 3.8.


Figure 3.6 Diagram of the shortlisted restriction enzymes with four to one RE sites in the ITGB2 gene sequence between exon 8 to intron 16. The ideal RE would not have a cut site with exon 11 and intron 11 (represented by the yellow box. This diagram is drawn to scale. EcoRI, the only shortlisted enzyme in this list, is highlighted in red bold.


Figure 3.7 Positions of the RE sites for KpnI, BamHI and EcoRI. Only the gDNA segment between intron 7 to intron 16 was considered in this analysis.


Figure 3.8 Possible circularized fragments of EcoRI (E1), BamHI (B1) and KpnI (K1) digested gDNA. A RE site was assumed to be present inside the unknown DNA, since each of the enzymes are six cutters and therefore have a high probability of occurring regularly in gDNA sequences.

### 3.3.3 Inverse PCR using EcoRI demonstrated that the A (Val ${ }^{367}$ ) (Val ${ }^{441}$ )

allele in the patient possessed a large deletion of at least 27 kb
After several unsuccessful attempts to amplify the mutant allele using KpnI and BamHI, EcoRI was used in the next phase of IPCR experiments. With EcoRI, a 'wild type' fragment of 8280 bp would be circularized; the long fragment was possibly longer than the mutant allele and was an advantage since PCR is biased toward shorter fragments. However, the size of the paternal (mutant) allele after digestion could not be predicted. PCR primers were designed to bind in exon 11 to ensure that the primers bound to known, existing positions in the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ paternal allele.

After T4 ligase re-ligation of EcoRI digested gDNA, successful circularization was verified using the F47317-B/R42092 primer set. The 'wild type' PCR product (3064 bp) was obtained in the father, mother and patient, demonstrating that circularized gDNA corresponding to the 'wild type' fragment (CD18 intron 11 to intron 15) was present in all the samples (Figure 3.9). These PCR products were not sequenced; they were used to verify that the circularization procedure was successful in the 'wild type' allele such that the deletion allele should have formed a circularized fragment as well.


Figure 3.9 Gel electrophoresis of the 3064 bp PCR product obtained from circularized gDNA digested with EcoRI. The primers (F47317B/R42092) were designed to flank the expected religation site. The schematic diagram on the right panel depicts the orientation and location of the primers used.

For IPCR, the primer set F41978/R41928 was designed to amplify the PCR product that spanned the EcoRI re-ligation site. These primers were located in exon 11 in opposite directions (Figure 3.12).

In order to amplify the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele, an arbitrary extension time of 10 minutes was initially used. No product was obtained in all 3 samples, not even from the 'wild type' allele. This was probably due to the size of the 'wild type' fragment, which was rather long to be amplified efficiently by PCR.

Keeping all other factors constant, working both upwards and downwards from the 10 minute extension time, each possible length of the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right) \mathrm{PCR}$ product was eliminated by increasing or decreasing the extension time by 30 seconds at a time. For the longer extension times, as to be expected, no PCR product was obtained. However, when the extension time was shortened to 6 minutes, a PCR product was obtained in the patient and father, which was not detected in the mother.

A 4385 bp product was obtained in the patient (Figure 3.10). EcoRI digestion of this PCR product was performed to verify the presence of the EcoRI site in the product, and the size ( 756 bp ) of the smaller restriction fragment spanning R41928 to the EcoRI site (Figure 3.11). Indeed, the 756 bp product was observable after EcoRI digestion (Figure 3.12). Reverse sequencing of the PCR product using the primer R11 indicated that this was derived from the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele (Figure 3.13).

The 4385 bp product was also obtained in the father (Figure 3.10), and products from both subjects were cloned into the invitrogen pGEM-T vector for amplification and sequencing.

DNA sequencing of a single IPCR product clone from the father showed that the EcoRI RE site in intron 10 was ligated to intron 3 of the next gene on the chromosome, Pituitary Tumor-Transforming Gene 1 protein-Interacting Protein (PTTG1IP) (Figure 3.14); the specific size of the IPCR product was 4434 bp . Sequencing results also indicated that the 5 EcoRI RE sites prior to exon 3 of PTTG1IP were deleted, and that the deletion spanned an approximate gDNA region of 27 kb in the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}$ $\left(\mathrm{Val}^{441}\right)$ allele. The patient had inherited this allele from her father.


Figure 3.10 Gel electrophoresis of the 4385 bp PCR product obtained from circularized gDNA digested with EcoRI. The primers (F41978/R41928) were located in exon 11 in opposite orientations. The expected size was 8231 bp in a wildtype sequence, although no product was obtained from the mother, probably due to the product size.


Figure 3.11 In the 'wild type' F41978/R41928 product schematically depicted in the left panel, two expected bands of 7475 bp (F41905 to R1) and 756 bp (R41928 to R1) would be obtained after EcoRI digestion of the PCR product. However, in the approximately 4500 bp mutant product containing an unknown deletion (right panel), the 756 bp gDNA segment should remain intact while the unknown gDNA segment was expected to resolve at 4000 bp . As seen in Figure 3.12, the EcoRI digested fragments corresponded exactly to the product containing the unknown DNA.


Figure 3.12 To verify that the F41978/R41928 PCR product was specific, EcoRI digestion was performed to cleave the product across the ligation site used for gDNA circularization. In the agarose gel, a 4000 bp product was obtained together with an expected 756 bp product.


Patient Reverse Seauencing - R11


Figure 3.13 Reverse sequencing chromatograms of the $\mathrm{F} 41978 /$ R41928 product obtained from circularized gDNA. At Val ${ }^{441}$, both the father and the patient had a C. In contrast to IPCR using circularized BamHI digested gDNA as template, where a $C$ and a $T$ was found in the father and patient at Val441 respectively, this experiment showed that the paternal A (Val ${ }^{367}$ ) $\mathbf{C}\left(\mathbf{V a l}^{441}\right)$ allele in the patient was amplified.


Figure 3.14 Reverse sequencing chromatogram of the cloned 4434 bp product from the father using R41978. The EcoRI ligation site is depicted with a broken red line. Past the EcoRI site, the DNA sequence corresponds to intron 3 of PTTG1IP in the reverse orientation.

### 3.3.4 A 27703 bp deletion occurred after intron 11 in the $A\left(\right.$ Val $\left.^{367}\right) C\left(V a l^{441}\right)$

## allele of the patient allele, inherited from the father

To determine the boundaries of the deletion in gDNA, the primer set F41905/R447 was used to amplify a PCR product extending from exon 11 of CD18 to exon 3 of PTTG1IP in gDNA. A 3000 bp gDNA PCR product was obtained in the father and patient, but not the mother (Figure 3.15). Sequencing of this DNA product indicated that there was a 27703 bp deletion +1170 bp from exon 11 ( 1566 bp ) of CD18. Intron 11 of CD18 was joined directly to intron $2(4125 \mathrm{bp})$ of PTTG1IP at the position 1664 bp from exon 3 (Figure 3.16).


Figure 3.15 Gel electrophoresis of the 3000 bp product obtained from gDNA. The primers F41905 and R447 were located in exon 11 of CD18 and exon 3 of PTTG1IP.


Figure 3.16 Sequencing chromatogram of the 3000 bp PCR product in the father and patient. The primer set F41905/R447 was used to amplify the truncated gDNA region spanning exon 11 of CD18 to exon 3 of PTTG1IP. The resulting PCR product from the father and child was sequenced. The position where intron 11 of CD18 transits to intron 2 of PTTG1IP is indicated by a dotted line. In both subjects, intron 11 of CD18 was truncated after +1170 bp and joined directly to intron 2 of PTTG1IP at the position $\mathbf{- 1 6 6 4} \mathbf{b p}$ from exon 3.

### 3.3.5 The hybrid CD18-PTTG1IP cDNA is expressed at the mRNA level

Earlier results from PCR amplification of cDNA indicated the existence of a mRNA product from the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele in the patient, since the patient was polymorphic at these 2 positions when the cDNA template spanned the region before exon 11.

Using the primer sets F41905/R2695 and F41905/R2541 to amplify the cDNA region between exon 11 of CD18 and exon 6 of PTTG1IP, a specific gene product of 2497 bp and 2343 bp was obtained in the patient, respectively. DNA sequencing of the 2343 bp PCR product indicated that exon 11 from CD18 was joined to exon 3 of PTTG1IP to form a hybrid mRNA product (Figure 3.17).

On retrospect, this was not a surprising result; based on gene organization of the deletion allele, an intact CD18 promoter region and an intact PTTG1IP polyadenylation signal would have been retained.


Figure 3.17 Gel electrophoresis of the F41905/R2541 and F41905/R2695 PCR products. As expected, the PCR product was only detected in the father and patient.


Figure 3.18 The primer set F41905/R2541 was used to amplify the cDNA region spanning exon 11 of CD18 to exon 6 of PTTG1IP. The PCR product was sequenced, and the position where exon 11 of CD18 transits to exon 2 of PTTG1IP is indicated by a dotted line. This demonstrated that the CD18-PTTG1IP hybrid mRNA product was present in both father and patient.

### 3.3.6 Summary of genetic analyses

Polymorphic markers in the parental and patient alleles at intron 5, intron 8, exon 10 $\left(\mathrm{Val}^{367}\right)$ and exon 11 positions $\left(\mathrm{Val}^{441}\right)$ were identified. With regard to the intron 2 splice mutation, gDNA sequencing was performed by Chan Hwee Sing previously to identify the source of the $-10 \mathrm{bp} \mathrm{C}>\mathrm{A}$ mutation. Hence, to show that the intron 2 splice mutation was inherited from the mother, cDNA was amplified using a forward primer designed to bind within the extra 43 bp , using the primers F23434/R26793.
gDNA from the parents and patient were amplified using F23051/R26792 (exon
2/exon 4), F31802 / R32752 (intron 5/intron 6), F8/R9 (intron 7/intron 9),
F10/A8913C (flanking exon 10 and 11).


Figure 3.19 Gel electrophoresis of the respective PCR products flanking the intron 2 splice mutation, intron 5, intron 8, and the gDNA segment between exon 10-and exon11.


Figure 3.20 Sequencing chromatograms of the 950 bp F31802/R32752 (intron 5 flanking) product in the father, mother and patient. The $\mathbf{- 2 9} \mathrm{bp}$ and $\mathbf{- 1 1} \mathrm{bp}$ positions relative to exon $\mathbf{6}$ are indicated.


Figure 3.21 Sequencing chromatogram of the F8/R9 (intron 8 flanking) PCR product, sequenced using the R9 primer. Polymorphic markers at the $\mathbf{- 4 7} \mathbf{~ b p}$ position are as indicated by the red arrow.

A8913C reverse sequencing

F10 forward sequencing

A8913C reverse sequencing



30 3 3 4 4 40 20 6 CAhACGhC.


Figure 3.22 Sequencing chromatograms of the F10/A8913C gDNA PCR product (flanking exon 10 and exon 11) for the father (top panel), mother (middle panel) and patient (bottom panel). All the subjects were polymorphic at the Val ${ }^{367}$ and Val ${ }^{41}$ positions.

Hence, it was established that the allele with the mutation in intron 2 was inherited from the mother whereas the deletion allele was inherited from the father. Combining the sequencing data, the mutation of the two alleles, as well as the polymorphic markers at each of the positions, the characteristics of the paternal and maternal mutant alleles are summarized in Table 3.4.

|  | Intron- $2^{(-10)}$ | Intron- $5^{(-29)}$ | Intron- $5^{(-11)}$ | Intron-$8^{(-47)}$ | $\mathrm{Val}^{367}$ | Val ${ }^{441}$ | $\begin{gathered} 27 \mathrm{~kb} \\ \text { del } \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Father | C | C | T | A | GTA | GTC | + |
|  | C | C | G | G | GTC | GTC | - |
| Mother | A | T | G | G | GTC | GTT | - |
|  | C | C | G | G | GTC | GTC | - |
| Patient maternal allele | A | T | G | G | GTC | GTT | - |
| Patient paternal allele | C | C | T | A | GTA | GTC | + |

Table 3.4 The two CD18 mutant alleles and their polymorphic markers.

### 3.4 Discussion

Leukocyte Adhesion Deficiency Type-1 was first identified in 1980 (Crowley, et al., 1980) and to date, no less than 30 different deleterious sequence defects were identified in the $I T G B 2$ gene of the patients (Fiorini, et al., 2009; Hixson, et al., 2004; Parvaneh, et al., 2010; Roos and Law, 2001; Roos, et al., 2002). Together with the $2^{\text {nd }}$ LAD-I mutation characterized in this thesis, a total of two novel mutations in the compound heterozygote LAD-I patient were described (Cher, et al., 2011). This discussion will focus on the $2^{\text {nd }}$ LAD-I mutation, which does not contribute to understanding the functional role of the CD11/ CD18 integrins, but are interesting in other aspects.

Both the ITGB2 allele causing cryptic 3' splice-site activation in intron 2 and the allele from the deletion fusing the $I T G B 2$ and PTTG1IP genes result in frameshifted mRNAs and PTCs, so they are potential targets of NMD. Indeed, the mutant mRNAs are difficult to detect by RT-PCR from the parents, as their presence is masked by the normal alleles. In order to trace the origin of the mutant allele, mutant specific primers had to be used (Figure 3.2c). It is probably incidental that both mutant mRNAs are detected at similar levels in the patient (Figure 3.2a) which may be attributed to their degradation by NMD to similar degrees.

The discrepancies between the polymorphic codons for $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$ between the cDNA and the gDNA led us to conclude that there is gross defect in one of the ITGB2 alleles of the patient. Using a strategy which was later found to have been described as inverse PCR (Ochman, et al., 1988), the $27,703 \mathrm{bp}$ deletion was found to have resulted in a fusion gene between $I T G B 2$ and PTTG1IP. The fusion cDNA was also
identified. Since the two adjoining exons are of different phases, the resultant cDNA, after exon 11 of $I T G B 2$, would code for 27 irrelevant amino acids before a stop codon (p.Arg472fsSerX28).

The Pituitary Tumor-Transforming Gene 1 Interacting Protein (PTTG1IP) binds to and facilitates nuclear translocation of PTTG1 (Chien and Pei, 2000). In turn, PTTG1 regulates the separation of sister chromatids in dividing cells (Zou, et al., 1999), and induces the transcriptional activation of growth factors such as the fibroblast growth factor (FGF)-2 (Ishikawa, et al., 2001) and vascular endothelial growth factor (VEGF) (McCabe, et al., 2002). Mice deficient in PTTG1 exhibit a number of cell growth abnormalities (Wang, et al., 2001). Thus far, there is no report on the deficiency of PTTG1IP in animal models or humans. The patient, because of the deletion allele spanning the ITGB2 and PTTG1IP genes, has a fusion gene leading to the production of a fusion mRNA coding for an abnormal CD18 protein. There was no evidence that the fusion CD18-PTTG1IP protein was produced. Presumably, no PTTG1IP is produced from this allele. Since there is no evidence that the father, who passed this allele to the patient, suffer from any notable pathogenic disorder, one may deduce a carrier of a defective PTTG1IP allele is healthy.

The linkage of the known polymorphic sites and the mutation of the two ITGB2 alleles were established and are summarized in (Table 3.4). This linkage has probably no predictive value of the mutations, since both alleles were described for the first time. Nonetheless, it should be noted that if not for the two polymorphic codons of $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$, a more cumbersome method to characterize the deletion allele may be required. For example, gDNA may be digested with a restriction enzyme such as

EcoRI, radioactively labeled, and used for Southern blot analysis of restriction fragment size against a normal control. In our experiments, the amount of gDNA available for experiments was limited since the samples had to be couriered from Austria to Singapore. However, given enough gDNA, Southern blotting might have been used instead of IPCR.

### 3.4.1 Retrospective discussion of the IPCR strategies

The IPCR strategy that was used as above is versatile. It was not expected that the deletion spanned over 27 kbp of the gDNA; neither the length of the circularized gDNA fragment was known, nor which RE site would be used. In order to maximize the possibility of self ligation, all circularization reactions using T4 ligase were carried out in proportionately large reaction volumes. Since the DNA fragments would then be well diluted in a large reaction volume, this helped to ensure that most of the ligations were intramolecular. Hence, given that circularization of the fragments was robust as indicated by the 'wild type' fragment, one could then find the ideal PCR extension time to amplify the the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele by keeping all other factors constant.

While a conventional PCR amplification strategy may be useful in some shorter mutations involving inversion, insertion or translocation, IPCR would allow amplification and sequencing of the circularized fragments regardless of the nature of mutation, since most of the REs have multiple cut sites in gDNA.

The original strategy described by Ochman et al. (1988) was based on a known DNA sequence which was flanked by unknown DNA on either ends. The original strategy is useful for detecting unknown sequences that flank, for example, known
transposable elements. The method used in our experiments is similar to Ochman's strategy, although our objective was to identify downstream genetic sequences. The systematic elimination of PCR extension times contributed to the success of our experiments.

Now that the deletion is well characterized, based on the RE cut sites for BamHI relative to those for EcoRI (Figure 3.8), a circularized fragment corresponding to the deletion allele had been present in an earlier experiment using BamHI. However, the deletion allele was not detected in the BamHI experiment (data not shown). PCR is biased towards amplifying shorter products; in the BamHI strategy, 'the wild type' PCR product was shorter, and was probably amplified with higher efficiency. In contrast, in the EcoRI strategy, the 'wild type' product was longer; perhaps only the EcoRI strategy would have been worked well. Several other REs which possess a cut site before exon 11 would have been useful for IPCR amplification of the $\mathrm{A}\left(\mathrm{Val}^{367}\right)$ $\mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele as well.

## Chapter 4: 3' Splice Site Strength Analysis

### 4.1 Introduction

In this chapter, the cryptic 3 'ss within intron 2 were studied. The g. $23457 \mathrm{C}>\mathrm{A}$ mutant found in the LAD-I patient (as described in Chapter 3, otherwise annotated as $-10 \mathrm{C}>\mathrm{A}$ in this thesis), abolished normal splicing at the authentic 3 ' splice site ( 3 'ss) and resulted in activation of the -43 cryptic 3 'ss in intron 2 . At the protein level, 11 nonsense amino acid residues were inserted before a stop codon (Cher, et al., 2011). This observation prompted the following 2 questions.

Firstly, how did the single nucleotide mutation abolish splicing at the authentic 3 'ss? Secondly, why was the -43 cryptic 3 'ss activated upon $-10 C>A$ mutation? In silico analysis identified many other potential 3'ss closer to the authentic 3'ss (see below).

### 4.2 Results

### 4.2.1 In silico analysis identified 11 potential 3'ss near the authentic 3'ss in

## intron 2

To estimate the number of potential 3'ss near the authentic 3'ss position, a window of 50 nt upstream and downstream of the authentic 3 'ss was analyzed. 11 AG dinucleotides were found. In addition, the $-10 \mathrm{C}>\mathrm{A}$ mutation created a new AG dinucleotide at -8 bp . The positions of the authentic and potential 3'ss in the genomic DNA sequence are depicted in Figure 4.1. For each position, the AG dinucleotide is located 2 bp upstream of the potential 3 'ss, but is referred to with respect to the position of the potential 3'ss in this thesis.


Figure 4.1 Positions of the authentic and potential 3'ss (p 3'ss) near the authentic 3'ss between the $\mathbf{- 5 0}$ and $+\mathbf{5 0}$ bp positions. At the $\mathbf{- 1 0}$ position, a new AG dinucleotide was formed as a result of the $-10 C>A$ mutation. The authentic and potential 3 'ss AG dinucleotides, are highlighted in red.

### 4.2.2 In silico analysis of 3'ss strength showed that the 3'ss at $\mathbf{- 4 3}$ and $+\mathbf{1 1}$

## have higher scores than the authentic 3'ss

3'ss strength was estimated in silico using the MAXENT, Weight Matrix Model (WMM), First order Markov Model (MM) and Neural Network (NN) models (Reese, et al., 1997; Yeo and Burge, 2004). The Neural Network (NN) method and the Weight Matrix Model (WMM) are based on training sets or position-weight matrices, respectively, while the Maximum Entropy (MAXENT) and First-Order Markov (MM) models consider dependencies between positions. Each model returns a score that predicts the relative strength of 3'ss (Reese, et al., 1997; Yeo and Burge, 2004).

Most of the potential 3'ss found had very low scores. Only the 3 'ss at $-43,+1$ and +11 had scores above background. Surprisingly, the authentic 3'ss (+1) was not attributed
with the highest score in any of the algorithms. Instead the +113 'ss was attributed with the highest score by the NN and WMM algorithms, while the -43 3'ss was attributed with the highest score by MAXENT and MM (Table 4.1).

With the $-10 C>A$ mutation, scores at the authentic 3 'ss were reduced drastically.
Although scores at the -43 and +113 'ss remained unchanged, the MAXENT and MM tools now (correctly) returned higher scores for the cryptic 3'ss at -43 (Table 4.1).

This was consistent with our earlier observation where inactivation of the authentic
3'ss due to $-10 \mathrm{C}>$ A resulted in use of the -43 cryptic 3 'ss (Cher, et al., 2011).

| Intron | Sequence | Position | NN | MAXENT | MM | WMM |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| WT | tcccccttgggtgtgggcag ggc | -43 | 0.23 | $5.73{ }^{+}$ | $6.11{ }^{+}$ | 3.12 |
|  | gtgtgggcagggcggcccag agc* | -33 | No Score | -7.93 | -8.46 | -4.59 |
|  | gtgggcagggcggcccagag cag* | -31 | No Score | -14.48 | -14.99 | -11.34 |
|  | ccagagcacccactcaccag ccg* | -17 | 0.01 | -1.97 | -2.68 | 0.92 |
|  | ccactcaccagccggcctcg tcc | -8 | No Score | -12.65 | -13.49 | -8.07 |
|  | cagccggcctcgtccctcag TCC | 0 | 0.35 | 5.18 | 5.2 | 6.04 |
|  | gtcoctcagTССтСтСтСAG GAG | 11 | $0.95{ }^{+}$ | 4.78 | 4.83 | $10.85{ }^{+}$ |
|  | cctcagTCCTCTCTCAGGAG TGC | 13 | No score | -20.94 | -10.32 | -0.63 |
|  | tctctcaggagtgcacgaag ttc | 23 | No score | -9.60 | -10.11 | -3.20 |
|  | aggagtgcacgaagttcaag gtc | 29 | No score | -6.80 | -9.02 | -4.72 |
|  | tgcacgaagttcaaggtcag cag | 34 | No score | -10.29 | -8.79 | -4.87 |
|  | acgaagttcaaggtcagcag ctg | 37 | No score | -8.59 | -8.53 | -3.19 |
| MUT | ccactcaccagccggcetag tcc | -8 | No Score | -4.61 | -5.44 | -0.03 |
|  | cagccggcctagtccetcag TCC | 0 | 0.08 | -2.36 | -1.85 | 3.64 |

Table 4.1 Relative scores for the potential 3'ss in the intron 2 sequence between $\mathbf{- 5 0} \mathbf{b p}$ to $+50 \mathbf{b p}$. In the mutant intron 2 sequence, only scores for -8 and +1 were augmented by the $-10 C>A$ mutation. Intronic and exonic genomic DNA sequences are in small caps and large caps respectively. ${ }^{+}$The highest score returned for each model in the WT and MUT genomic DNA sequence.

The +113 'ss has higher scores than the -43 3'ss, yet the -43 3'ss was used instead of the +11 . Therefore, it was hypothesized that the upstream AG dinucleotide of the authentic 3 'ss repressed use of the +113 'ss, resulting in activation of the next
candidate cryptic 3 'ss at -43 . Consistent with this notion, there was no nearby AG dinucleotide upstream of the -43 cryptic 3'ss (Figure 4.2).


Figure 4.2 Schematic diagram of the hypothesis. In the presence of $\mathbf{- 1 0 C}>A$, the newly formed AG dinucleotide may have repressed use of the authentic 3'ss. Splicing occurred at the -43 3'ss, which was the next best 3 'ss. It was then hypothesized that the +113 'ss was not used as a result of repression from the AG dinucleotide of the authentic 3'ss.

In this study, this hypothesis was addressed experimentally by mutating the authentic AG dinucleotide to another purine pair (AA, GA and GG) in minigenes, where the wild type (WT) or mutant ( $-10 \mathrm{C}>\mathrm{A}$, MUT) intron 2 sequence was inserted into a cDNA expression vector between exon 2 and 3 . This manipulation would retain upstream pyrimidine content at the +113 'ss, while removing the upstream AG dinucleotide. It was expected that these mutations would result in activation of the +11 3'ss.

### 4.2.3 In silico analyses predicted the relative preference of +11 over the -43

## 3'ss after ablation of the +1 AG dinucleotide

Scores for the mutant 3'ss were calculated. At the -43 cryptic 3'ss, scores were unchanged. For the +11 3'ss, scores were approximately uniform when calculated using NN and WMM, yet there was an approximately 2 -fold increase in the MAXENT and MM scores. Nonetheless, all four algorithms predicted the relative
preference of the +11 over the -43 cryptic 3 'ss after ablation of the authentic AG dinucleotide (Table 4.2).

| Intron <br> Sequence | Sequence | Position | NN | MAXENT | MM | WMM |  |
| :--- | :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| WT | gtccctcagTCCTCTCTCAG GAG | 11 | $0.95^{+}$ | 4.78 | 4.83 | $10.85^{+}$ |  |
| WT | tcccccttgggtgtgggcag ggc | -43 | 0.23 | $5.73^{+}$ | $6.11^{+}$ | 3.12 |  |
|  |  |  |  |  |  |  |  |
| WT AA | gtccctcaaTCCTCTCTCAG GAG | 11 | 0.97 | 8.28 | 8.36 | 10.3 |  |
| WT GA | gtccctcgaTCCTCTCTCAG GAG | 11 | 0.97 | 7.86 | 8 | 10.71 |  |
| WT GG | gtccctcggTCCTCTCTCAG GAG | 11 | 0.98 | 8.4 | 8.64 | 11.25 |  |
| MUT AA | gtccctcaaTCCTCTCTCAG GAG | 11 | 0.95 | 8.28 | 8.36 | 10.3 |  |
| MUT GA | gtccctcgaTCCTCTCTCAG GAG | 11 | 0.96 | 7.86 | 8 | 10.71 |  |
| MUT GG | gtccctcggTCCTCTCTCAG GAG | 11 | 0.96 | 8.4 | 8.64 | 11.25 |  |

Table 4.2 Scores for the WT and MUT based minigenes at +113 'ss, after mutation of the authentic AG dinucleotide into another purine pair (AA, GA or GG). Scores for the wild type +11 3'ss and -43 3'ss is included here for reference.

### 4.2.4 FACS analysis of transfectants was consistent with predicted splicing

## patterns

A total of 8 pairs of CD11a/CD18 clones were used for HEK293T transfection, including each of the six mutant CD18 minigenes, or the WT and MUT minigenes as positive and background controls, respectively.

HEK293 cells were used in this experiment since individual CD18 monomers cannot be expressed on the cell surface of these cells. In contrast, only CD11a/CD18 dimers can be expressed on the cell surface. Hence in this experiment, the contrast between monomer and dimer expression on the cell surface of HEK293 cells is a useful characteristic to determine if the CD18 protein expressed by each of the minigenes is functional

Cells co-transfected with CD11a and WT expressed the integrin dimer on the cell surface. In contrast, cells co-transfected with CD11a and WT_AA, WT_GA or WT_GG did not show surface expression of the dimer. In addition, none of the MUT plasmids supported surface expression of the integrin dimers when co-transfected with CD11a, suggesting that no functional CD18 protein was produced after AG dinucleotide mutation to AA, GA and GG.


Figure 4.3 FACS analysis of HEK293 cells transfected with WT or MUT based SDM minigenes, with the CD11a expression plasmid. Cells were stained with MHM23 (dimer specific), MHM24 (CD11a specific) or $\mathbf{H 5 2}$ (CD18 specific). After SDM, none of the SDM minigenes expressed a functional integrin dimer.

### 4.2.5 RT-PCR analysis showed activation of splicing at +11 after ablation of

## the authentic $A G$

Total RNA was extracted from HEK293T cells transfected with each of the 8 minigenes, reverse transcribed with random hexamers and PCR amplified using the intron 2 flanking primers, F228 and R429. The gel region corresponding to 150 bp to 300 bp was excised for gel purification and used as template for a second round of PCR, and the identity of the PCR products was confirmed by sequencing.

The RT-PCR results show that ablation of the authentic AG dinucleotide in both WT and MUT $(-10 \mathrm{C}>\mathrm{A})$ resulted in use of the +113 'ss. This result strongly suggested that the +113 'ss is repressed by the upstream AG of the authentic 3 'ss, leading to activation of the next best candidate site at the -43 cryptic 3 'ss.


Figure 4.4 PCR amplification of splicing isoforms in the modified minigenes. Gel electrophoresis of PCR products from the first and second PCR reactions. The product sizes are as indicated, and correspond to splicing at the $-43,+1$ and +113 'ss, as depicted in the schematic diagram.

### 4.2.6 Mutating the -43 AG dinucleotide

To see if the +113 'ss would be used in the absence of the -433 'ss, the -43 AG dinucleotide was mutated to another purine pair. According to the sequence, any other combination of purines (i.e. GG or AA) would result in formation of another AG
dinucleotide that could be used for splicing. Hence, only the GG mutation was introduced into the minigenes to create the $\mathrm{WT}_{-}-43 \mathrm{AG}>\mathrm{GG}$ and $\mathrm{MUT}_{-}-43 \mathrm{AG}>\mathrm{GG}$ minigenes, abbreviated as WG and MG in this thesis.

### 4.2.7 The minigenes carrying both the -43 3'ss and -10C>A mutations

 showed no membrane expressionThe WG and MG minigenes were co-transfected into HEK 293 cells together with CD11a and analyzed for surface expression of the integrin dimers. Whereas the WG minigene expressed a functional integrin, the mutant minigene did not, consistent with lack of splicing at the authentic 3 'ss.


Figure 4.5 FACS analysis of CD11a and WG or MG transfectants.
4.2.8 RT-PCR showed that no detectable splicing occurred upon the -43 3'ss mutation

The WG and MG minigenes were transfected into HEK293 cells. Minigene splicing was analyzed by RT-PCR amplification as before. After the two rounds of PCR, a 202 bp product was detected in WG, resulting from 3' splicing at the authentic position.

No product was obtained for MG, suggesting that splicing did not occur in MG (Figure 4.6).


Figure 4.6 Gel electrophoresis of RT-PCR products obtained from WG and MG. The two top and bottom panels show the DNA fragments obtained from the first and second round of PCR, respectively.

### 4.2.9 Analysis of remnant plasmid DNA

Despite DNase I treatment, a 552 bp PCR product corresponding to the size of a PCR fragment from the minigenes were consistently obtained. To determine the approximate amount of remnant plasmid DNA, the primers F41905 and pcDNA3_R1555 were used to amplify plasmid DNA from DNase I treated cDNA. A water only control was included as negative control. Results showed that remnant plasmid DNA was present in all of the cDNA samples despite DNase I treatment.


Figure 4.7 PCR amplification of cDNA samples after DNase I treatment, using the primers F41905 and pcDN3_R1555.

### 4.2.10 Summary of results from 3'ss analysis

In summary, formation of the AG dinucleotide at -10 shifts splicing equilibrium from the authentic to the -433 'ss. When the authentic 3 'ss is ablated after removal of the AG dinucleotide, splicing then occurs at the +113 'ss. Furthermore, when introduced in the context of the $-10 \mathrm{C}>\mathrm{A}$ mutation, loss of the authentic 3 'ss results in a high proportion of splicing at the +113 'ss and a low proportion of splicing at the -43 3'ss.

Lastly, when the -43 AG dinucleotide was removed in the context of the $-10 \mathrm{C}>\mathrm{A}$ mutation, no splicing between exon 2 and exon 3 could be detected in the minigene.

The results are summarized in Table 4.3.

| Plasmid name | -10 bp dinucleotide | +1 bp dinucleotide | $-43 \mathrm{bp}$ dinucleotide | Experimentally verified 3'ss position |
| :---: | :---: | :---: | :---: | :---: |
| WT | CG | AG | AG | +1 bp |
| WT_AA | CG | AA | AG | +11 bp |
| WT_GA | CG | GA | AG | +11 bp |
| WT_GG | CG | GG | AG | +11 bp |
|  |  |  |  |  |
| MUT | AG | AG | AG | -43 bp |
| MUT_AA | AG | AA | AG | -43 bp, +11 bp* |
| MUT_GA | AG | GA | AG | $-43 \mathrm{bp},+11 \mathrm{bp} *$ |
| MUT_GG | AG | GG | AG | $-43 \mathrm{bp},+11 \mathrm{bp} *$ |
|  |  |  |  |  |
| WG | CG | AG | GG | +1 bp |
| MG | AG | AG | GG | No splicing detected |

Table 4.3 Summary of the experimentally verified 3'ss positions following SDM ablation of AG dinucleotides into another purine pair at the authentic and -43 3'ss. *Indicates the main product obtained, where 2 products were present.

### 4.2.11 NN analysis of all 3'ss sites in CD18 identified 274 3'ss sequences

To estimate the number of potential 3'ss, the entire ITGB2 genomic DNA sequence was analyzed using the NN algorithm. The cutoff score was set at 0.34 since the lowest authentic 3'ss NN score in the ITGB2 genomic DNA sequence was 0.35 (for
the authentic 3 'ss sequence of intron 2). A total of 274 3'ss sites were found, inclusive of the 15 authentic 3 'ss sequences.

The 3'ss were grouped according to their positions in the $I T G B 2$ genomic DNA sequence. For example, excluding the exon 1 genomic sequence, all 3 'ss between intron 1 and the last residue of exon 2 were grouped together, and so forth. Based on the total length of the preceding intron and the entire exon, the total number of potential 3'ss in each segment was tabulated (Table 4.4).

The number of potential 3 'ss found exceeded the number of authentic 3 'ss by more than 18 fold. In addition, while some were close to the authentic 3'ss, the majority of the potential 3'ss were located in deep intronic positions (Figure 4.8).

| Start Position | End Position | Total Length (bp) | No. of Potential 3'ss |
| :--- | :--- | :---: | :---: |
| Intron 1 (Variant 2) | Exon 1 (Variant 1) | 7604 | 55 |
| Intron 1 (Variant 1) | Exon 2 | 10096 | 76 |
| Intron 2 | Exon 3 | 441 | 3 |
| Intron 3 | Exon 4 | 3369 | 28 |
| Intron 4 | Exon 5 | 3550 | 13 |
| Intron 5 | Exon 6 | 1873 | 6 |
| Intron 6 | Exon 7 | 1172 | 6 |
| Intron 7 | Exon 8 | 1253 | 9 |
| Intron 8 | Exon 9 | 4096 | 23 |
| Intron 9 | Exon 10 | 1567 | 14 |
| Intron 10 | Exon 11 | 1595 | 12 |
| Intron 11 | Exon 12 | 1831 | 11 |
| Intron 12 | Exon 13 | 702 | 5 |
| Intron 13 | Exon 14 | 583 | 1 |
| Intron 14 | Exon 15 | 1957 | 10 |
| Intron 15 | Exon 16 | 783 | 2 |

Table 4.4 Summary of the start and end positions of each genomic DNA segment, the length of the segments and the number of potential 3'ss found in each segment.


Figure 4.8 A graphical depiction of the potential 3'ss in ITGB2 with NN scores higher than 0.34 . The exons are highlighted by the turquoise blocks, while the authentic 3'ss are highlighted by the yellow blocks. Positions of the potential 3'ss are highlighted by the red blocks.

### 4.2.12 3'ss scoring algorithms predicted authentic 3'ss in CD18 with 60-73 \%

## accuracy

The relative accuracy of the 4 algorithms in identifying the authentic 3 'ss in all 15 introns was next studied. To focus on the potential 3 'ss in the neighborhood of the authentic 3 'ss, only the sequence from -175 bp to the last basepair of each exon were considered. The NN cut off score used in this analysis was 0.23 , since this was the score of the cryptic 3 'ss at -43 .

Of interest, two authentic 3 'ss at intron 2 and intron 10 were not attributed with the highest scores by any of the four algorithms. For intron 13, 14 and 15, no potential 3'ss with predicted NN scores higher than 0.23 were found.

In total, 45 potential 3'ss were found. The four algorithms NN, MAXENT, MM and WMM, correctly attributed $9,11,11$ and 10 authentic 3 'ss with the highest scores in the 15 exons, respectively (Table 4.5). This suggested that the MAXENT and MM algorithms had slightly better predictive value than the NN and WMM algorithms.

Furthermore, the 3'ss scoring methods correctly identified the authentic 3 'ss with an accuracy ranging between $60-73 \%$, highlighting the need for improving such prediction programs.

| Exon | Sequence | Position | NN | MAXENT | MM | WMM |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intron 1 | taacccaccccotgcagcag gga | -108 | 0.36 | 1.24 | 2.16 | 4.69 |
|  | gccccagctccttgtcccag gcc | -81 | 0.88 | 5.30 | 6.04 | 8.72 |
|  | aagacccettccctccacag gac | 0 | 0.99* | 10.64* | 9.47* | 10.44* |
| Intron 2 | tccccettgggtgtgggcag ggc | -43 | 0.23 | 5.73* | 6.11* | 3.12 |
|  | cagccggcetcgtccctcag tcc | 0 | 0.35 | 5.18 | 5.20 | 6.04 |
|  | gtccetcagtcctctctcag gag | 11 | 0.97* | 4.78 | 4.93 | 10.85* |
| Intron 3 | ctccetgtctctgtccacag gga | -130 | 0.95 | 13.89* | 12.65* | 11.67* |
|  | acttctgtcccaaattccag gcg | -108 | 0.67 | 8.59 | 7.76 | 6.94 |
|  | ggccagtcagcctcctgcag gtc | -46 | 0.78 | -0.36 | 0.02 | 7.03 |
|  | ccctgactccccotcccoag aac | 0 | 0.96* | 8.85 | 8.72 | 10.83 |
|  | ccctccccagaacttcacag ggc | 10 | 0.78 | 0.92 | 0.26 | 5.77 |
| Intron 4 | gtcctgtgtctggtcctcag ctc | -150 | 0.79* | 7.53 | 8.05 | 7.43 |
|  | gtcctcagctctcctgccag gcg | -138 | 0.69 | 5.60 | 4.33 | 8.77* |
|  | tgttctgtccccaccggcag gcc | 0 | 0.46 | 9.79* | 9.22* | 8.02 |
|  | tccatgcttgatgacctcag gaa | 103 | 0.47 | 4.73 | 4.63 | 4.57 |
| Intron 5 | tttctccacctccotgctag gag | -97 | 0.97* | 8.20 | 9.08 | 10.87* |
|  | gccetcgtgcccetttgaag ctg | -59 | 0.54 | 7.30 | 6.45 | 7.01 |
|  | ccagcctgccacctccccag ccc | -20 | 0.23 | 5.03 | 4.54 | 6.72 |
|  | cccetccatgtgccetgcag gct | 0 | 0.85 | 8.87* | 9.92* | 9.76 |
|  | cagccccogtttgccttcag gca | 115 | 0.92 | 7.03 | 7.66 | 9.66 |
| Intron 6 | catcgtctcctctgctccag agg | -57 | 0.89 | 8.09* | 7.94* | 10.16 |
|  | cccccacgcettcttctcag gag | 0 | 0.97* | 7.89 | 7.88 | 11.05* |
| Intron 7 | ggcagcctctgcctctccag cct | -121 | 0.39 | 4.08 | 3.95 | 7.91 |
|  | gcctctccagccttccccag aga | -111 | 0.85 | 3.50 | 2.21 | 9.13 |
|  | gcaggacctcctctctccag gac | 0 | 0.95* | 6.77* | 6.42* | 9.56* |
| Intron 8 | ccgtcccagccattttgcag gtg | -88 | 0.86* | 4.46 | 4.75 | 8.39 |
|  | ctggtcttcttgcccaacag aaa | 0 | 0.84 | 9.06* | 8.65* | 8.23* |
| Intron 9 | gctctcctctgtgtcattag ctg | -117 | 0.57 | 5.61 | 6.87 | 7.79 |
|  | taaccgctctctgcccgcag aaa | 0 | 0.70 | 12.07* | 10.14* | 8.08* |
|  | ccgcagaaactctcctccag ggt | 14 | 0.33 | 3.62 | 3.07 | 5.17 |
|  | acctacgactccttctgcag caa | 77 | 0.93* | 7.10 | 7.13 | 7.43 |
| Intron 10 | ccccacgetgaaccccccag gac | -92 | 0.50 | 4.81 | 5.95* | 5.61 |
|  | agcccagcactgccetgcag atc | 0 | 0.61 | 3.95 | 3.71 | 5.71 |
|  | ccctgcagatcaccttccag gtg | 12 | 0.62* | 5.46* | 4.25 | 7.63* |
|  | ccgtgcaggttcttccccag tgt | 108 | 0.45 | 3.49 | 4.16 | 6.45 |
| Intron 11 | ggggccectcctgtccccag gtg | 0 | 0.91* | 10.24* | 9.11* | 9.63* |
|  | caactccatcatctgctcag ggc | 110 | 0.40 | 4.60 | 5.39 | 5.58 |
| Intron 12 | gccccatgccttttgctcag gaa | -118 | 0.40 | 6.34 | 6.15 | 7.55 |
|  | tgctcaggaacatcccccag gct | -105 | 0.44 | 1.73 | 1.17 | 2.63 |
|  | gggcgggtctcggttcccag gcc | -38 | 0.81 | 6.01 | 3.28 | 4.56 |
|  | cgtgctgccecgtcttccag gga | 0 | 0.94* | 10.77* | 9.57* | 10.02* |
|  | accagctgcctctgtgccag gag | 182 | 0.26 | 4.78 | 4.92 | 5.32 |
| Intron 13 | tcacacctggctccccgcag ctc | 0 | 0.85* | 8.44* | 8.41* | 7.95* |
| Intron 14 | gtgtgctgcttcttccetag agt | 0 | 0.90* | 9.54* | 9.87* | 9.73* |
| Intron 15 | ctgtttgcatttcccaccag gat | 0 | 0.83* | 10.38* | 9.48* | 10.01* |

Table 4.5 Summary of 3 'ss strength predictions for the 15 genomic DNA segments analyzed. * indicates the highest 3'ss score obtained using each tool in the neighborhood of the authentic 3'ss.

### 4.3 Discussion

### 4.3.1 A review of other LAD-I mutations

A recent paper by van de Vijver et al (2012) reviewed all of the known mutations relating to LAD-I. 86 different genetic mutations leading to LAD-I were identified in a total of 123 LAD-I patients (from 109 different families). Of these, splice site mutations account for 12 alleles (approximately $14 \%$ ), including the $-10 \mathrm{C}>$ A splice mutant described in this thesis (Table 4.6).

This is in contrast to studies which have estimated that between $50-60 \%$ of diseasecausing mutations may result from disruptions to normal splicing (Cartegni, et al., 2002; Lopez-Bigas, et al., 2005; Pagenstecher, et al., 2006). Nonetheless, LAD-I is a relatively rare autosomal recessive disorder, there may be other splice mutations in LAD-I carriers that have not ben identified. Among the 12 splice mutations, 5 originated from 3'ss aberrations, 4 from 5'ss aberrations and the remaining 3 were not attributable to either class as insufficient data are available (Table 4.7).

| Mutation type | Number of unique alleles found | Percentage (\%) |
| :--- | :---: | :---: |
| Deletions | 23 | 26.7 |
| Insertions | 2 | 2.3 |
| Deletion/insertions | 4 | 4.7 |
| Nonsense mutations | 10 | 11.6 |
| Splice site mutations | 12 | 14 |
| Missense mutations | 35 | 40.7 |

Table 4.6 Distribution of CD18 mutations that result in loss of a functional CD18 protein according to the mutation type. 86 unique mutations were found in 123 LAD-I patients. Adapted from van de Vijver et al [2012].

| Mutation | Splice Site Mutation | Resulting protein | References |
| :---: | :---: | :---: | :---: |
| c.59-10C>A | 3'ss : -10 bp from exon 3 | p.Cys19_Val20ins11X12 | (Cher, et al., 2011) |
| c. $328+1 \mathrm{G}>\mathrm{A}$ | 5'ss : +1 bp from exon 4 | p.Asn50AlafsX8 | (Tone, et al., 2007) |
| c. $329-6 \mathrm{C}>\mathrm{A}$ | 3'ss : -6 bp from exon 5 | p.Gln111_Phe168del | (Parvaneh, et al., 2010) |
| c.500-12T>G | 3'ss : -12 bp from exon 6 | p.Gly167ValfsX47 | (Roos and Law, 2001; Roos, et al., 2002) |
| c.742-14C>A | 3'ss : -14 bp from exon 7 | p.Pro247_Glu248insPro SerSerGIn | (Nelson, et al., 1992; Roos and Law, 2001; Roos, et al., 2002; Uzel, et al., 2008; Wright, et al., 1995) |
| c. $897+1 \mathrm{G}>\mathrm{A}$ | 5'ss : +1 bp from exon 7 | fsX26 or fsX44 | (Matsuura, et al., 1992; Parvaneh, et al., 2010; Roos and Law, 2001; Tsai, et al., 2008) |
| c. $1083+3 \mathrm{G}>\mathrm{C}$ | 5'ss : +3 bp from exon 9 | p.Lys332_Asn361del | (Kishimoto, et al., 1989; Roos and Law, 2001) |
| c. $1083+5 \mathrm{G}>\mathrm{C}$ | 5'ss : +5 bp from exon 9 | p.Lys332_Asn361del | Unpublished |
| $\begin{aligned} & \text { c.1225- } \\ & \text { ?_1412+?del } \end{aligned}$ | (mutation is currently uncharacterized) | p.Ile409ValfsX1 | Unpublished |
| c. $1658-2 \mathrm{~A}>\mathrm{G}$ | 3'ss : -2 bp from exon 13 | p.Gly553AlafsX6 | Unpublished |
| Not identified | (mutation is currently uncharacterized) | p.Gly553AlafsX6 | (Back, et al., 1992; Roos and Law, 2001) |
| c.2080+1delG | (mutation is currently uncharacterized) | p.Ser627ValfsX44 | Unpublished |

Table 4.7 Summary of all currently known LAD-I causing splice mutations. Adapted from van de Vijver et al [2012].

### 4.3.2 A review of other 3'ss mutations leading to LAD-I

Of the 53 'ss mutants above, 3 where the nucleotide change (mutation) was described are reviewed here, including c.329-6C $>\mathrm{A}, \mathrm{c} .500-12 \mathrm{~T}>\mathrm{G}$, and $\mathrm{c} .742-14 \mathrm{C}>\mathrm{A}$ (Matsuura, et al., 1992; Nelson, et al., 1992; Parvaneh, et al., 2010; Roos and Law, 2001; Roos, et al., 2002; Tsai, et al., 2008; Uzel, et al., 2008; Wright, et al., 1995). To compare the highest score for each position as a result of the mutations, scores for the predicted 3'ss strength were calculated using the NN, WMM, MM and MAXENT tools (Reese, et al., 1997; Yeo and Burge, 2004) and summarized in Table 4.8.

### 4.3.3 Potential studies of 3'ss splicing mutations in LAD-I

c. $329-6 \mathrm{C}>\mathrm{A}, \mathrm{c} .500-12 \mathrm{~T}>\mathrm{G}$ and $\mathrm{c} .742-14 \mathrm{C}>\mathrm{A}$ had each resulted in the formation of an AG dinucleotide $6 \mathrm{nt}, 13 \mathrm{nt}$ and 14 nt upstream of the authentic 3 'ss (in intron 4,5 and 6, respectively). Based on results obtained in our study here, it is likely that the authentic 3'ss was repressed by the upstream AG dinucleotide in all three cases, leading to cryptic 3 'ss activation.

3'ss scores for $\mathrm{c} .329-6 \mathrm{C}>\mathrm{A}$ suggests activation of the -4 3'ss. However, this prediction could not be corroborated as the DNA sequence was not analyzed in the patient (Parvaneh, et al., 2010).

3'ss scores for $\mathrm{c} .500-12 \mathrm{~T}>\mathrm{G}$ predicted activation of the -973 'ss, but analysis of the cDNA sequence showed activation of a cryptic 3'ss at +149 (Roos and Law, 2001; Roos, et al., 2002), which was predicted by all four tools to have scores below background or low scores.

Scores for the authentic 3'ss remained high in c. $742-14 \mathrm{C}>\mathrm{A}$, but cryptic 3' splicing was detected to occur at -12 (Nelson, et al., 1992; Roos and Law, 2001; Roos, et al., 2002; Uzel, et al., 2008; Wright, et al., 1995). The -12 3'ss was consistently attributed with low scores all four tools. Notably, the nearest upstream AG dinucleotide is located 17 nt away from the -12 'ss.

Since canonical 3'ss signals are currently well characterized, exceptions to the norm will provide additional information that can help in refinement of 3'ss prediction. The observed cryptic 3 'ss activation in $\mathrm{c} .500-12 \mathrm{~T}>\mathrm{G}$ and $\mathrm{c} .742-14 \mathrm{C}>\mathrm{A}$ are both
exceptions to the norm, and suggest that a nearby sequence element not considered in the four tools may have improved cryptic 3 'ss strength at the +149 and -12 cryptic 3'ss of intron 5 and 6 respectively. Hence, both of these cases warrant an in depth study of cryptic 3'ss activation, and may provide additional novel information related to cryptic 3'ss activation.

| Mutation position | Nucleotide Change | Observed Phenotype | Sequence | Position | NN | MAXENT | MM | WMM | References |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intron 4 (-6) | c.329-6C>A | Not determined | gtcctgtgtctggtcctcag ctc | -150 | 0.79 | 7.53 | 8.05 | 7.43 | (Parvaneh, et al., 2010) |
|  |  |  | gtcctcagctctcctgccag gcg | -138 | 0.69 | 5.60 | 4.33 | 8.77 |  |
|  |  |  | cttttgttctgtccccacag gca | -4 | 0.99* | 12.98* | 12.66* | 12.47* |  |
|  |  |  | tgttctgtccccacaggcag GCC | 1 | No score | 3.63 | 3.05 | 5.38 |  |
|  |  |  | TCCATGCTTGATGACCTCAG GAA | 103 | 0.47 | 4.73 | 4.63 | 4.57 |  |
| Wild type intron 4 | - | - | tgttctgtccccaccggcag GCC | 1 | 0.46 | 9.79 | 9.22 | 8.02 |  |
| $\begin{aligned} & \text { Intron } 5 \text { (- } \\ & \text { 12) } \end{aligned}$ | c.500-12T>G | Skipping part of exon 6 (Activation of 3'ss at +149) | tttctccacctccctgctag gag | -97 | 0.97 | 8.2* | 9.08* | 10.87* | (Roos and Law, 2001; <br> Roos, et al., 2002) |
|  |  |  | gccctcgtgcccctttgaag ctg | -59 | 0.54 | 7.30 | 6.45 | 7.01 |  |
|  |  |  | cacctccccagccectccag gtg | -11 | 0.98* | 3.76 | 3.05 | 10.50 |  |
|  |  |  | cccctccaggtgccetgcag GCT | 1 | 0.58 | 3.60 | 3.94 | 7.78 |  |
|  |  |  | CAGCCCCCGTTTGCCTTCAG GCA | 115 | 0.92 | 7.03 | 7.66 | 9.66 |  |
|  |  |  | TGACCAACAACTCCAACCAG TTT | 149 | No score | 1.89 | 1.88 | -0.87 |  |
| Wild type intron 5 | - | - | cccctccatgtgccetgcag GCT | 1 | 0.85 | 8.87 | 9.92 | 9.76 |  |
| $\begin{aligned} & \text { Intron } 6 \text { (- } \\ & \text { 14) } \end{aligned}$ | c.742-14C>A | Inclusion of 12 bp from intron 6 (p.Pro247_ Glu248ins4) | cctccgggccaccccaccag ctc | -82 | 0.00 | 3.68 | 2.76 | 5.13 | (Nelson, et al., 1992; <br> Roos and Law, 2001; <br> Roos, et al., 2002; <br> Uzel, et al., 2008; <br> Wright, et al., 1995) |
|  |  |  | catcgtctcctctgctccag agg | -57 | 0.89 | 8.09* | 7.94* | 10.16* |  |
|  |  |  | aggtaaccetgccecccaag cct | -12 | 0.00 | -0.78 | 1.11 | 1.89 |  |
|  |  |  | cccccaagccttcttctcag GAG | 1 | 0.95* | 3.96 | 4.46 | 9.15 |  |
| Wild type intron 6 | - | - | cccccacgccttcttctcag GAG | 1 | 0.97 | 7.89 | 7.88 | 11.05 |  |

Table 4.8 3'ss mutations characterized in LAD-I patients. Exon and intron sequences in the wild type sequence are shown in large and small caps respectively. The positions of the mutation are highlighted in red. The nucleotide positions are as calculated with A of the ATG start codon considered as $\mathbf{n t}=1$. * indicates the cryptic 3 'ss attributed with the highest score by each algorithm, among all the potential 3'ss identified.

### 4.3.4 The $A G$ dinucleotide +11 bp upstream repressed use of the +113 'ss

The authentic 3'ss of intron 2 has a low score and was not attributed with the highest scores by any of the 4 methods (Table 4.5). Remarkably, the $-10 \mathrm{C}>\mathrm{A}$ mutation in intron 2 further reduces the scores for the authentic 3'ss in all algorithms, consistent with the splicing disruption seen in the LAD-I patient and the minigenes. Molecular analysis of the CD18 minigene recapitulated the splicing patterns of intron 2 from endogenous transcripts, including activation of the cryptic 3 'ss at position -43 in the $-10 \mathrm{C}>\mathrm{A}$ mutant. Such experiments confirmed the molecular diagnosis of this mutation, and suggested that missplicing caused by the mutation is not leukocyte specific (Cher, et al., 2011).

In silico analyses revealed the presence of another potential 3'ss 11 nt downstream of the authentic 3 'ss in intron 2, with scores that were similar or even higher than the cryptic 3 'ss at -43. However, the +113 'ss was not used in either patient cells or in the mutant CD18 minigene. We experimentally tested the hypothesis that use of the +113 'ss was repressed by the upstream AG at the authentic 3'ss. By mutating the AG dinucleotide at the authentic 3'ss, activation of the +113 'ss was detected. Furthermore, when this mutation was combined with $-10 \mathrm{C}>\mathrm{A}$, the +113 'ss was used more efficiently than the -43 3'ss. Hence, different mutations may result in activation of different cryptic 3 'ss, further complicating an accurate prediction of 3'ss activation.

To examine whether the +113 'ss would be activated in the absence of the competing -433 'ss, the -43 AG dinucleotide was abolished in the $-10 \mathrm{C}>\mathrm{A}$ mutant. As a result, no detectable splicing was observed, demonstrating that the +113 'ss is not competent for splicing when the upstream AG is present, even in the absence of any competing 3 'ss.

Typically, AG dinucleotides are excluded from the upstream region of 3'ss, possibly because of interference in discrimination of the correct intron-exon boundary (Královicová, et al., 2005). This feature is captured by the MAXENT and MM methods. Changing the upstream AG of the +113 'ss (i.e. the authentic AG dinucleotide of the intron/exon boundary) to any other pair of purines nearly doubles the scores by these methods, without much change in the NN and WMM scores. This indicates that the MAXENT and MM methods are better predictors of cryptic 3'ss activation. Nevertheless, these algorithms give scores well above background for the +113 'ss, yet this 3 'ss is not used even in the absence of other competing 3'ss, as seen with the -43 'ss mutants. This observation reveals that the MAXENT and MM tools are still inaccurate.

Nonsense Mediated mRNA Decay (NMD) pathway partially degrades mRNAs harboring premature termination codons (PTCs), defined as in-frame stop codons located more than 50 nucleotides upstream of the final exon-exon junction (Isken and Maquat, 2007). While the mutant product would be a target of NMD in the LAD-I patient, NMD would not degrade the -43 and +113 'ss products because these minigenes only have one intron. The observed difference in mRNA adundance may be attributed to lower splicing efficiency at the -43 relative to the +113 'ss, but not NMD.

Of note, at the protein level, splicing at +11 would result in the addition of 34 irrelevant amino acids before a premature stop codon (p.C19_V20ins34X35), after the wild type amino acid sequence SLGC. This was consistent with FACS analysis showing that that no functional dimer was expressed in the SDM transfectants.

DNase I treatment of the RNA samples could not remove all traces of plasmid DNA. However, unspliced mRNA resulting from intron retention in the minigenes may also resolve at the same molecular weight in gel electrophoresis.

The mechanism of 3 'ss recognition is currently unclear. In addition to the other interactions described earlier (in Chapter 1), the DEK protein (DEK oncogene) was shown to proofread interaction of U2AF with the 3's splice site. The DEK protein is activated by phosphorylation at two serine residues at the amino-terminus. After activation, DEK interacts with $\mathrm{U} 2 \mathrm{AF}^{35}$, inducing release of the $\mathrm{U} 2 \mathrm{AF}^{65 / 35}$ complex from 3 'ss sequences that do not contain the consensus AG dinucleotide (Soares, et al., 2006). This contributes to AG discrimination by the U2AF complex in pre-mRNA splicing.

Optimally, the 3'ss AG dinucleotide is located within 19-23 nt downstream of the BPS. (Chua and Reed, 2001). While the first downstream AG dinucleotide after the BPS is usually selected for exon ligation, several exceptions have also been found. For example, distal AG dinucleotides located 6 nt downstream of a proximal AG can compete efficiently for $3^{\prime}$, splicing (Chua and Reed, 2001). In addition, the upstream -43 and downstream +113 'ss products were detected concurrently in the $-10 \mathrm{C}>\mathrm{A}$ mutated minigenes. Results from the current study indicate that 3 ' splicing is relatively more efficient at the nearer +113 'ss, consistent with a hypothesized scanning mechanism initiated from the authentic 3'ss upon repression or mutation of the authentic 3 'ss, such that distant AG dinucleotides compete for 3' splicing less efficiently (Královicová, et al., 2005).

Overall, the 3 'ss tools correctly attributed $60-73 \%$ of the authentic 3 'ss with the highest scores, highlighting their limitations. Hence, prediction inaccuracy may be due to additional
discriminating parameters not considered in the models, but which contribute to selection of the correct (authentic) 3'ss, such as the strength of neighboring splice sites, the availability of potential splice sites, the length of the resulting exon, and the density of exonic and intronic splicing enhancers or silencers (Královicová and Vorechovsky, 2007; Wimmer, et al., 2007).

Between 50-60 \% of disease-causing mutations may result from disruptions to pre-mRNA processing and splicing (Cartegni, et al., 2002; Lopez-Bigas, et al., 2005; Pagenstecher, et al., 2006), possibly resulting in disease, modulate disease severity, or in the case of multifactorial diseases such as cancer, be linked with increased disease susceptibility (Faustino and Cooper, 2003; Pagenstecher, et al., 2006). The human genome has been sequenced (Lander, et al., 2001). While mutations in exons leading to changes in protein function are obvious, a large proportion of the single nucleotide polymorphisms in the human genome may disrupt pre-mRNA splicing, resulting in disease or increased disease susceptibility (Cartegni, et al., 2002; Lopez-Bigas, et al., 2005; Pagenstecher, et al., 2006).

With respect to 3 'ss mutations, disease severity may vary with activation of an in-frame or out-of-frame cryptic 3 'ss, and the distance of an in-frame cryptic 3 'ss from the authentic 3 'ss, which correlates with amount of residues deleted or added to the protein sequence. Depending on the protein function of the mutated gene, such mutations may have a local (cis) effect on a single protein, or a global (trans) effect over the subset of other proteins controlled by the mutated gene (Faustino and Cooper, 2003; Pagani and Baralle, 2004; Wang and Cooper, 2007).

In conclusion, this study has demonstrated that an AG dinucleotide located up to 11 nt upstream can repress cryptic 3 ' splicing in the +11 3'ss. Furthermore, this repression was
strong; the -113 'ss was not used even in the absence of competing cryptic 3 'ss. Lastly, a comprehensive understanding of 3 'ss activation both in authentic and mutant sequences, combined with accurate prediction tools, can assist in diagnosis of disease severity and facilitate timely genetic counseling. Further studies that highlight current flaws in the prediction tools or exceptions to the norm, such as the one presented here, may be useful to provide additional information to refine the current 3'ss prediction tools.

## Chapter 5: Intracellular retention and folding of the integrin proteins

### 5.1 Introduction

In this chapter, a study of CD11a and CD18 retention in HEK293 transfectants was performed as a first step towards understanding integrin retention and dimerization. Candidate integrin interacting proteins with chaperone or protein modification functions were identified. In this thesis, these proteins shall be referred to as chaperone/modification proteins.

### 5.1.1 Immunoprecipitation for LC-MSMS analysis of integrin interacting proteins

Immunoprecipitation combined with LC-MSMS (IP-LCMS) was used to study the integrin interacting proteins. Presumably, integrin retention must be an important process that facilitates heterodimer formation. Hence, this technique may also be useful to identify proteins involved in integrin dimerization. Similarly, candidate proteins involved in dimerization may have chaperone or protein modification functions. In particular, retained monomers may present the best opportunity to capture transient protein interactions. In this system, integrin proteins were probably stalled en route to cell surface expression by interacting proteins responsible for integrin retention.

### 5.1.1.1 HIS tagged CD18

Antibodies specific for conformational epitopes in CD18 and CD11a were initially used to identify chaperone/modification proteins in JB2.7 and SKB2.7. Later on, to avoid the use of antibodies, a 6 X HIS tag was cloned behind the C-terminal of the CD18 protein.

The HIS tag was previously used to purify the truncated CD18 proteins, PHE1, PHE2 and PHE3, which were secreted proteins that did not contain the $\beta$ I domain sequence truncated before the IEGF4 domain (Figure 5.1) (2007a; Shi, et al., 2005).


Figure 5.1 Schematic diagram of the PHE constructs. Diagram taken from Shi et al (2007a).

### 5.1.1.2 FLAG tag purification of CD18F, $t$-CD18F and CD11aF

Later on, a FLAG tag was cloned behind the C-terminal of CD18 (CD18F) instead. In this system, CD18F was initially analyzed as this was the common partner for all four of the leukocyte integrins (CD11a/CD18, CD11b/CD18, CD11c/CD18 and CD11d/CD18). For a IP-LCMS based comparison with CD18F, a FLAG tagged 'CD18_ILess' construct (t-CD18F) was created where the $\beta \mathrm{I}$ domain was removed based on the PHE sequence (Shi, et al., 2007a). In addition, to study CD11a binding chaperone/modification proteins, a FLAG tagged CD11a plasmid was created (CD11aF) where a FLAG tag was appended behind the CD11a sequence.

### 5.2 Results

### 5.2.1 Western blot on JB2.7 and SKB2.7 showed that JB2.7 did not express CD11a

The JB2.7 (CD11a deficient) and SKB2.7 (CD18 deficient) cell lines were derived from Jurkat and SKW3, respectively (as described in Chapter 1). To confirm that CD11a was not expressed in JB2.7, both JB2.7 and SKB2.7 were analyzed in Western blot. The Western bloting results were as expected; no CD11a was detected in JB2.7 (Figure 5.2).

Several commercially available anti-CD18 antibodies were tested in Western blotting (results not shown). While two of the antibodies were able to detect purified CD18 (unpublished data), specific detection of the protein from cell lysate in Western blot was poor. Hence, CD18 expression in the cell lines was not analyzed further.


Figure 5.2 Western blot of cell lysate obtained from each cell line, probed with an anti-CD11a antibody.

### 5.2.2 Antibody IP followed by LCMSMS analysis identified 14 candidate integrin

## interacting proteins

Integrins were immunoprecipitated from cell lysates using CD11a (MHM24) or CD18 (H52) specific antibodies. The eluates were resolved in SDS-PAGE and stained with Commassie blue. Each lane was divided into 5 portions for LC-MS analysis (Figure 5.3).


Figure 5.3 Commassie blue stained SDS-PAGE. Each sample lane on this gel was separately divided into 5 portions as shown in the picture, and submitted for LC-MS analysis.

The exponentially modified protein abundance index (emPAI) score was used to estimate the relative abundance of the identified proteins. This is a commonly used method in mass spectrometry to estimate the relative amount of a specific protein in a protein mixture (Ishihama, et al., 2005).

This was a pilot experiment to identify patterns in integrin interacting proteins for CD18 and CD11a. Hence, emPAI data from both samples was combined, searched against a human tryptic peptide database, and an initial list of 181 hits was obtained. The antibodies used for the experiment co-eluted with immunoprecipitated proteins and appeared as a very strong band around the 50 kDa region in SDS-PAGE (Figure 5.3). However, as the antibodies used were of mouse and rabbit origin, they were not identified in the search against the human database.

Protein functions were determined according to information available on UniProt (www.uniprot.org). To eliminate background hits (proteins identified in mass spectrometry),
the following were removed according to a pre-determined criteria; proteins of keratin origin, redundant hits (incorrectly classified by the software as two unique hits under the same protein name), and proteins with emPAI value below 0.1 (too low in relative abundance).

The remaining proteins were classified according to their protein functions into the following categories; transcription and translation related, actin and tubulin related, mitochondria related. Where proteins were uncharacterized or had irrelevant functions which were too diverse to classify, they were classified as 'ungrouped proteins'.

A final list of 12 candidate chaperone/modification proteins was obtained. Of these, two proteins were found in both CD11a and CD18 samples, four proteins bound exclusively to CD18, and six bound exclusively to CD11a. Graphical summaries of the procedure used to remove uninteresting hits, protein classification and the distribution of the chaperone/modifying proteins are presented using Figure 5.4 to Figure 5.6 , while emPAI values for chaperone/modifying proteins are presented in Table 5.1 and Table 5.2. No repeats were performed for this experiment.


Table 5.1 Summary of the combined LCMSMS results. 44 hits were removed, excluding the CD18 and CD11a proteins.


Figure 5.4 Summary of the hits removed, relating to keratin, redundant and low adundance (emPAI below 0.1). Of the 181 hits identified, 135 were determined to be 'non background' hits using the procedure summarized in the right panel.


Figure 5.5 The 135 consistent hits were classified according to known functions, where proteins that had diverse functions that could not be classified meaningfully were classified as ungrouped proteins. 12 chaperone/modifying proteins were identified.


Figure 5.6 A Venn diagram illustrating the distribution of the hits across the CD11a (SKB2.7) and CD18 (JB2.7) samples. Of the 12 chaperone/modifying proteins, 6 were found in CD11a, 4 in CD18 and 2 were overlapping.

| Gene <br> Symbol | Protein Name | emPAI values |  |
| :--- | :--- | :---: | :---: |
|  |  | SKB2.7 | JB2.7 |
|  |  | (CD11a expressing) | (CD18 expressing) |
|  | CD11a binding proteins |  |  |
| CANX | Calnexin | 0.34 | -1 |
| HSPA8 | Isoform 1 of Heat shock cognate 71 kDa <br> protein | 0.27 | -1 |
| DNAJB11 | DnaJ homolog subfamily B member 11 | 0.23 | -1 |
| HSP90AB1 | Heat shock protein HSP 90-beta | 0.15 | -1 |
| HSPA1L | Heat shock 70kDa protein 1-like variant | 0.13 | -1 |
| YWHAZ | 14-3-3 protein zeta/delta | 0.11 | -1 |
|  |  |  |  |
|  | CD18 binding proteins | -1 | 0.43 |
| HSPA8 | HSPA8 54 kDa protein | -1 | 0.32 |
| S100A8 | Protein S100-A8 | -1 | 0.16 |
| UGGT1 | Isoform 1 of UDP-glucose:glycoprotein <br> glucosyltransferase 1 | -1 | 0.14 |
|  | Isoform 1 of Protein-L-isoaspartate(D- <br> aspartate) O-methyltransferase |  |  |
| PCMT1 |  | 0.27 | 0.05 |
|  |  | 0.15 | 0.17 |
|  | Found in both samples |  |  |
| HSPA5 | HSPA5 protein |  |  |
| HSPA7 | Putative heat shock 70 kDa protein 7 |  |  |

Table 5.2 emPAI values for the chaperone proteins identified, classified according to the respective samples.

### 5.2.3 HIS tag purification from human cell line lysate was unspecific

When co-transfected into HEK293 cells with the CD11a expression plasmid, the CD18_HIS protein could dimerize with CD11a, as detected at the cell surface using FACS (Figure 5.7). This indicated that the $6 \times$ HIS tag did not interfere in heterodimer formation.

Attempts to purify the CD18_HIS protein using nickel beads and cobalt beads resulted in unspecific protein binding to and eluting from the beads. Various conditions were tested in FPLC, including different concentrations and gradients of immidazole and NP40 during column washing and elution. The UV280 profile and Western blot analysis indicated that peak elution occurred when immidazole concentration reached between 150 mM to 200 mM .

Later on, MonoQ columns were also used to purify the HIS tagged CD18, either separately or in tandem with nickel or cobalt bead purification. Eight conditions were tested, but CD18_HIS protein eluted from the beads were consistently masked by a high amount of contaminating proteins across a large range of molecular weights (Figure 5.8 shown here for reference).


Figure 5.7 FACS analysis of HEK293 cells transfected with CD11a / CD18_HIS or CD18_HIS individually.


Figure 5.8 Silver stain of SDS-PAGE loaded with eluates from tandem purification using monoQ followed by nickel beads. Fractions 18, 19 and 20 from the monoQ elution had the highest amount of CD18_HIS detected in Western blot (not shown), and so were pooled and further purified using the nickel column. Tandem purified eluates were resolved in lanes represented by fractions 4 through 11. However, no improved resolution of the CD18_HIS protein was observed with the second purification step.

### 5.2.4 FACS analysis of CD18F demonstrated that the FLAG tag did not interfere with the dimer formation

Since HIS tag purification was unsatisfactory, a FLAG tag was cloned behind the CD18 protein instead. FLAG tagged proteins could be specifically bound using M2 antibody conjugated - protein A sepharose beads (M2 beads). A high concentration of FLAG peptide could be used to compete with the antibody bound FLAG tagged proteins during elution.

To ensure that the FLAG tag did not interfere in normal expression of the CD18 protein, HEK293 cells were co-transfected with CD11a and CD18F. Cells transfected with CD11a/CD18F expressed the integrin dimer on the cell surface, as detected by MHM23, MHM24 and H52 (Figure 5.9).


Figure 5.9 FACS analysis profile of CD11a/CD18F transfected HEK293 cells. The background antibody used was OX68.

### 5.2.5 FLAG tag purification was superior to HIS tag purification

CD11a/CD18F was transfected into HEK293. After IP, CD18F was eluted. The eluate was concentrated using the Amicon 3K micro filter and resolved in SDS-PAGE. Silver stain analysis detected two main proteins bands at 95 kDa and 140 kDa . Purification of a FLAG tagged protein was much cleaner compared to HIS elution. Furthermore, elution was complete as M2 beads boiled in 2X loading dye did not contain any remnant CD18F (Figure 5.10).

Eluted samples were also analyzed in Western blot using the M2 anti-FLAG antibody (Sigma), and the anti-CD11a antibody (AbD serotec). Molecular weights of the detected protein bands corresponded to proteins detected in silver stain analysis (Figure 5.11). Hence, Western blot analysis confirmed that CD11a and CD18F were present in both the crude cell lysate of transfected cells and the eluates.


Figure 5.10 Silver stain analysis of purified CD18F from co-transfected HEK293 cells. For comparison of purification efficiency, cell lysate from transfectants was resolved. To evaluate the efficiency of FLAG protein elution, the remaining M2 affinity gel (M2 beads) after elution of FLAG tagged CD18, were boiled in SDS loading buffer and analyzed in the same gel.


Figure 5.11 Western blot analysis of FLAG purified proteins from HEK293 transiently co-transfected with CD11a and CD18F. The CD11a and CD18F protein bands are clearly seen in the lanes loaded with eluate, and crude cell lysate.

### 5.2.6 Silver staining showed that unspecific proteins could bind to the M2 beads

In preparation for LC-MSMS analysis, all conditions were scaled up by five fold. CD18F was transfected by itself or co-transfected with CD11a. As a background control, HEK293 cells were mock transfected and subjected to the same procedure of wash, lysis and FLAG purification.
$4 \mu \mathrm{l}$ of the concentrated samples were resolved in SDS-PAGE and silver stained. Developed for three minutes, protein bands corresponding to CD11a and CD18F were detected in the cotransfected sample, while a protein band corresponding to CD18F was detected in the CD18F transfected sample (Figure 5.12). When the silver stain was developed for an additional five minutes, other protein bands could be detected in all three samples, including the background control (Figure 5.13). This indicated that some proteins could bind to the M2 beads and that a background (negative control) sample was essential to remove false positive results in mass spectrometry.


Figure 5.12 SDS-PAGE of eluates, silver stained. Developed for three minutes, the CD18F and CD11a proteins can be clearly discerned in the gel.


Figure 5.13 SDS-PAGE of eluates, silver stained. This was the same gel as above, developed for a total of eight minutes. Background proteins were visible in the untransfected HEK293 background control.

### 5.2.7 LC-MSMS consistently identified 72 integrin binding proteins over 3

## biological repetitions

For mass spec analysis, concentrated eluates were resolved in SDS-PAGE. The polyacrylamide gel was stained with Commassie blue and portioned into five segments for LC-MSMS analysis. One of the stained gels is shown here for reference (Figure 5.14). Three biological repeats of the experiment were performed and 512 proteins were identified in total. 441 hits were excluded, and the remaining hits were grouped according to the respective sample sets (i.e. the three experimental repetitions for CD18F were considered as one set while those for CD11a/CD18F were considered as another set) (Table 5.3). For clearer presentation, the numerical data is summarized in a series of charts from Figure 5.15 to

Figure 5.18.


Figure 5.14 Commassie blue stain of concentrated eluates from CD11a/CD18F, CD18F, and mock transfected HEK293 cells. Each sample was loaded into 2 lanes to avoid overloading. After destaining, the gel was separated into 5 portions and diced for LCMSMS analysis. The main bands observed correspond to the CD18F protein. Due to the lower sensitivity of Commassie blue staining compared to silver stain, only faint bands corresponding to CD11a were detected in purified CD11a/CD18F.

Due to the nature of software analysis, 21 proteins were redundant hits which were identified twice by the software during data analysis. Proteins with diverse functions that could not be classified meaningfully were classified as ungrouped proteins. Hits identified at least twice in each sample set (further explained in Figure 5.16) were considered as consistent. Sorted according to the known protein functions, the relevant proteins in with chaperone, protein modification or protein transport functions are listed in Table 5.4. Each experiment is represented according to the date on which the experiment was performed. For example 110216 refers to the experiment performed on the $16^{\text {th }}$ of February, 2011.

Interestingly, the majority of proteins found in CD18F overlapped with CD11a/CD18F, consistent with the expectation that a proportion of nascent CD18F in the latter was still in the monomeric form.

| Proteins (Hits) Removed due to various reasons |  |
| :---: | :---: |
| Protein Category | No of proteins in the category |
| Background hits (found in negative control) | 192 |
| Redundant hits (the same protein classified erroneously by the software under different names) | 21 |
| Keratin related (likely to be dust contamination) | 5 |
| Tubulin, Actin related | 7 |
| Inconsistent hits (detected less than twice out of three repeats) | 214 |
| CD18 / CD11a protein | 2 |
| Consistent hits |  |
| Protein Category | No of proteins in the category |
| 1.Found in both CD18F and CD11a/CD18F samples |  |
| Chaperone proteins | 8 |
| Protein modification related | 5 |
| ER related | 3 |
| Proteasome related | 7 |
| Ribosome, translation related | 5 |
| Mitochondria related | 9 |
| Ungrouped proteins | 19 |
|  |  |
| 2.Found in CD18F samples only (too few to categorize) | 3 |
|  |  |
| 3. Found in CD11a/CD18F samples only |  |
| Chaperone proteins | 1 |
| Protein modification | 1 |
| ER retention | 1 |
| Ungrouped | 9 |

Table 5.3 Summary of the hits obtained in three biological repeats of IP-LCMS from CD18F and CD11a/CD18F transfectants.


Figure 5.15 A pie chart summarizing the hits removed from the 512 hits obtained in total. The procedure to remove the uninteresting hits is summarized in the right panel.


Figure 5.16 A Venn diagram depicting that hits identified at least twice out of the three experiments were considered as consistent hits. NB: 'EX' is used as a short form for 'experiment'; i.e. ${ }^{\text {st }} \mathbf{E X}$ refers to the first experiment. Hits identified in at least two out of three repeats were considered as consistent hits (represented by the overlapping regions of the chart). 1n 2: detected in 2 out of 3 experiments; All 3: detected in all three experiments.


Figure 5.17 A Venn diagram depicting that most of the consistent hits identified in CD18F were also identified in CD11a/CD18F, consistent with expectations.


Figure 5.18 A pie chart summarizing the classification of the 57 overlapping consistent hits according to known protein functions. Of main interest to this experiment were proteins identified with chaperone or protein modifying functions.

| GS |  | Pro_Mass | emPAI values obtained for each protein at each experimental repeat <br> CD11a/CD18F <br> CD18F |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 110216 | 110119 | 101222 | 110216 | 110119 | 101222 |
|  | CD11a/CD18F and CD18F |  |  |  |  |  |  |  |
| ITGB2 | (Identified in both sets of data at least twice) |  |  |  |  |  |  |  |
|  | Integrin beta | 86856 | 25.08 | 2.49 | 6.68 | 10.54 | 6.68 | 9.33 |
|  | Chaperones |  |  |  |  |  |  |  |
| CANX | Calnexin | 67982 | 1.14 | 0.13 | 0.14 | 0.71 | 0.68 | 1.13 |
| B3KY95 | Protein disulfide-isomerase A6 (EC 5.3.4.1) | 53655 | 0.2 | 0.11 | -1 | 0.2 | 0.2 | 0.06 |
| DNAJB11 | DnaJ homolog subfamily B member 11 | 20994 | 0.37 | 0.14 | 0.17 | 0.16 | 0.16 | -1 |
| BAT3 | HLA-B associated transcript 3 | 119085 | -1 | -1 | 0.03 | 0.15 | 0.15 | -1 |
| B2R774 | Lectin, mannose-binding, 1 (LMAN1) | 57768 | 0.12 | -1 | 0.06 | 0.12 | 0.18 | 0.12 |
| HSP90AA1 | HSP90AA1 protein | 68614 | 0.15 | -1 | -1 | 0.05 | -1 | 0.1 |
| SRPRB | Signal recognition particle receptor subunit beta | 29912 | 0.88 | -1 | 0.37 | -1 | 0.37 | -1 |
| Q53HF2 | Heat shock 70kDa protein 8 isoform 2 variant | 53580 | 1.17 | 0.45 | -1 | -1 | -1 | 0.71 |
|  | Protein Modification |  |  |  |  |  |  |  |
| A8KAK1 | UDP-glucose ceramide glucosyltransferase-like 1, transcript variant 2 | 175467 | 0.5 | 0.14 | 0.1 | 0.22 | 0.18 | 0.12 |
| OSTC | Oligosaccharyltransferase complex subunit | 16932 | 0.2 | -1 | -1 | 0.2 | 0.2 | -1 |
| CDIPT | CDP-diacylglycerol--inositol3-phosphatidyltransferase (EC 2.7.8.11) | 20575 | 0.58 | -1 | -1 | 0.16 | 0.16 | -1 |
| B4DJE3 | Dolichyl-diphosphooligosaccharide--proteinglycosyltransferase 48 kDa subunit (EC2.4.1.119) | 46620 | -1 | 0.2 | -1 | -1 | 0.07 | 0.07 |
| TMX1 | Thioredoxin-related transmembrane protein 1 | 32170 | -1 | -1 | 0.1 | -1 | 0.1 | 0.1 |
| B3KQT9 | Protein disulfide-isomerase A3 (EC 5.3.4.1) | 54468 | 0.19 | 0.05 | 0.06 | -1 | 0.06 | -1 |

## CD18F only

(Identified in CD18F samples at least twice)
Protein modification and transport

| A8K9K4 | glucosidase I | 62314 | -1 | -1 | -1 | 0.05 | 0.05 | -1 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| CLTC | Clathrin heavy chain 1 | 193260 | -1 | -1 | -1 | 0.02 | 0.02 | -1 |
|  | CD11a/CD18F only <br> (Identified in CD11a/CD18F transfected samples at least twice) |  |  |  |  |  |  |  |
| ITGAL | Integrin alpha-L | 129942 | 3.55 | 0.7 | 0.68 | -1 | -1 | -1 |
|  | Chaperones |  |  |  |  |  |  |  |
| HSPH1 | Heat shock protein 105 kDa | 97716 | 0.03 | 0.03 | -1 | -1 | -1 | -1 |
|  | Protein Modification |  |  |  |  |  |  |  |
| A8K4K6 | protein disulfide isomerase family A , member 4 | 73157 | -1 | 0.04 | 0.19 | -1 | -1 | -1 |
|  | ER retention |  |  |  |  |  |  |  |
| ERP44 | Endoplasmic reticulum resident protein 44 | 47341 | -1 | 0.06 | 0.07 | -1 | -1 | -1 |

Table 5.4 Shortlisted proteins with relevant functions, from 3 repetitions of the LC-MSMS experiment, sorted according to experimental date and sample type (CD18F or CD11a/CD18F ). The emPAI values for proteins identified at least twice in each group are presented here. Where the $\mathbf{- 1}$ value is shown for a protein, the protein was not present in the sample analyzed.

### 5.2.8 FACS analysis of t-CD18 demonstrated that the $\beta I$ domain was essential for intracellular retention of CD18

While consistent hits were identified in the previous experiments, it was difficult to conclusively identify integrin interacting proteins that caused retention of CD18, since monomeric CD18F overlapped with nascent CD18F in the doubly transfected samples. To overcome this limitation, it was hypothesized that a truncated-CD18 protein with the $\beta \mathrm{I}$ domain removed (t-CD18) would not be retained.

In FACS analysis, t -CD18 was detected on the cell surface as a monomer, confirming that the $\beta I$ domain was critical for CD18 retention. Dimerization at the cell surface was not detected in t-CD18/CD11a co-transfected samples.


Figure 5.19 FACS analysis of $t$-CD18 transfectants. No membrane expression of $t$-CD18/CD11a was detected, but t-CD18 was detected on the cell surface as a monomer. In contrast, CD18F was able to dimerize with CD11a, but was not detected on the cell surface as a monomer.

### 5.2.9 Analysis of known protein functions suggested calnexin as a candidate integrin retention protein

Having confirmed that the $\beta \mathrm{I}$ domain caused retention of CD18, the list of proteins which were consistently identified in IP-LCMS was searched for candidate proteins with retention functions. In this manner, calnexin was identified as a strong candidate for retention of the nascent integrin monomers. UDP-glucose ceramide glucosyltransferase 1 (UGGT1) works together with calnexin by glucosylating improperly folded proteins, enabling calnexin to retain the unfolded or misfolded proteins (Williams, 2006). UGGT1 was also identified consistently in the experiments.

### 5.2.10 Co-IP of calnexin from CD18F and t-CD18 demonstrated that calnexin associated with other domains of CD18, including the $\beta I$ domain

To demonstrate that the calnexin protein was retaining the CD18 protein by binding the $\beta I$ domain, one might aim to demonstrate that t-CD18 did not associate physically with calnexin. CD18F and t-CD18 were immunoprecipitated using the mAb H52, alongside a mock transfected sample (Mock TF).

Eluates were resolved in SDS-PAGE, and the gel was analyzed using the silver stain method. Calnexin ( 75 kDa ) co-eluted with both CD18F ( 95 kDa ) and t-CD18 ( 60 kDa ) (Figure 5.20 ).

Cell lysates were resolved alongside eluates as a control for protein size in Western blot. From the Western blot, calnexin was detected in eluates from both samples. In all, the results indicated that both CD18F and t-CD18 associated physically with calnexin.


Figure 5.20 Silver stain of SDS-PAGE used to resolve the IP samples and the crude cell lysates. Calnexin was detected in both the purified CD18F and t-CD18 samples.


Figure 5.21 Western blot of cell lysates and purified FLAG proteins from CD18F or t-CD18F transfectants, probed for the calnexin protein. Calnexin was detected in both the purified CD18F and tCD18 samples.

### 5.2.11 FACS analysis of t-CD18F and CD11aF showed that the $t$-CD18F protein

 could be expressed as a monomerThe earlier t-CD18 plasmid was constructed for a 'quick and dirty' experiment, and contained an 8 bp NotI sequence in place of the $\beta \mathrm{I}$ domain. Nonetheless, earlier results with t-CD18 were sufficient to provide an understanding of t-CD18 membrane expression and demonstrate protein interaction with calnexin.

Since calnexin bound to t-CD18, it was hypothesized that an additional chaperone protein may be responsible for CD18 retention. To find this protein, an expression vector for t-CD18 containing a FLAG tag (t-CD18) was constructed for IP-LCMS. CD11aF (CD11a sequence containing a FLAG tag at the C-terminal) was included in the experiment to study CD11a interacting proteins.

Analyzed in FACS, CD11aF dimerized with co-transfected CD18, and was not expressed on the cell surface as a monomer. In contrast, t -CD18F was not detected with co-transfected CD11a, but was expressed on the cell surface as a monomer. Both results were as expected.

A higher proportion of cells expressed monomeric t -CD18F when CD11a was co-transfected, suggesting that retention of t -CD18F was occurring, and may be further impaired by simultaneous over-expression of two plasmids (Figure 5.22).


Figure 5.22 FACS analysis of CD11aF and t-CD18F. CD18F was used as a positive control in this analysis. This figure is different from that in Figure 5.19 as the $t$-CD18F sequence used here corresponded to the $\beta I$ domain deletion described in Shi et al [2007].

### 5.2.12 Combining all FLAG purification data, analysis yielded 127 consistent hits

Three experiments (repeats) to analyze the $\mathrm{t}-\mathrm{CD} 18 \mathrm{~F}$ interacting proteins were performed. In addition, one repeat for CD11a/CD18F, CD11a/t-CD18F, CD11aF/CD18 and CD11aF were performed (Table 5.5). The IP-LCMS results obtained were combined with results described in the previous section and analyzed together. A total of 839 proteins were identified in total, as summarized in Table 5.6. For clarity, the numerical data is summarized in a series of charts (Figure 5.23 and Figure 5.24).

CD18F, t-CD18F were separately grouped as two sets, while samples containing CD11a (i.e. CD11a/CD18F, CD11aF and CD11a) were grouped as one set. However, hits identified in the mixed grouping under CD11a at least twice were considered as consistent hits, but may present as false positive results since the group is not homogenous The list of all identified consistent hits is presented in Appendix II (Table II.2) according to the experiment dates and plasmids transfected.

Sorting of the data according to these pre-determined groups yielded 127 consistent hits, of which the majority were found in CD11a and CD18F. 28 were identified in CD18F, 43 in CD11a, and 38 were overlapping between the two groups (Figure 5.23). Interestingly, while calnexin was found consistently in all three sets, the LMAN1 protein was consistently found in CD18F, but was absent in t-CD18F. The HSP5A protein bound in high abundance where detected, and unexpectedly, CD11a was pulled down with t-CD18F in the CD11a/t-CD18F sample (data presented in Appendix II).

| Number of experimental repetitions performed for each sample type |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| CD18F | t -CD18F | $\mathrm{t}-\mathrm{CD} 18 \mathrm{~F} /$ CD11a | CD11a/CD18F | CD11aF / CD18 | CD11aF |
| 5 | 3 | 1 | 4 | 1 | 1 |

Table 5.5 A summary of the number of repetitions performed for each sample. The data obtained from all of these experiments were pooled for analysis. The pooled data is presented in Appendix II.

| Proteins (hits) removed |  |
| :---: | :---: |
| Category | No of proteins |
| Background | 204 |
| Keratin related | 88 |
| Inconsistent hits | 371 |
| Redundant hits | 49 |
|  |  |
| Consistent hits |  |
| Category (according to known protein functions) | No of proteins |
| 1. CD18F only |  |
| - Chaperone | 3 |
| - Protein modification | 5 |
| - Protein binding | 2 |
| - Proteasome | 4 |
| - Mitochondria | 3 |
| - Translation | 2 |
| - Ungrouped | 7 |
|  |  |
| 2. CD11a only |  |
| - Protein modification | 6 |
| - Chaperone | 7 |
| - Translation | 4 |


| - Actin tubulin | 4 |
| :---: | :---: |
| - Ungrouped | 22 |
| 3. CD11a and CD18F (two groups) |  |
| - Chaperone | 2 |
| - Protein modification | 4 |
| - Protein binding | 3 |
| - Mitochondria | 10 |
| - Proteasome | 5 |
| - Translation | 7 |
| - Actin tubulin | 3 |
| - Ungrouped | 4 |
|  |  |
| 4. tCD18 and CD11a (two groups) |  |
| - Chaperone | 1 |
|  |  |
| 5. tCD18F and CD18F (two groups) |  |
| - Protein binding | 2 |
|  |  |
| 6. CD11a, CD18F and tCD18F (all three groups) |  |
| - Chaperone | 5 |
| - Protein modification | 4 |
| - Actin tubulin | 1 |
| - Ungrouped | 2 |

Table 5.6 Summary of the IP-LCMS results from all the FLAG purification experiments.


Figure 5.23 A pie chart showing the breakdown of the 839 hits identified when all the FLAG purification results were combined. There were 127 consistent hits after the uninteresting hits corresponding to background, keratin, redundant and inconsistent hits were removed.


Figure 5.24 A Venn diagram summarizing the 127 hits obtained and the samples from which the hits were identified. * The CD11a sample set consisted of CD11aF/CD18, CD11a/CD18F and CD11aF. Hits identified in this mixed grouping at least twice were considered as consistent hits, but may present as false positive results since the group is not homogenous.

### 5.2.13 Comparing t-CD18F and CD18F LCMS data, LMAN1 did not bind the $\beta$ I domain

The t -CD18F plasmid was created to identify proteins that bind to the CD18 $\beta \mathrm{I}$ domain. Hence, LCMS data obtained from the five CD18F repeats were compared against the three tCD18F repeats. In this analysis, a total of 637 hits were identified. 92 hits were determined to be consistent (Figure 5.25), the majority of which were present in CD18F (Figure 5.26). When the 92 hits were grouped according to protein function, 6,12 and 7 hits were found to have protein binding, protein modification and chaperone functions, respectively (Figure 5.27). Consistent hits from both sets of data found in these three categories were determined to be interesting, as they may be involved in retention of CD18. emPAI values of the interesting hits are presented in Table 5.7, with the CD11a/t-CD18F data appended to the same table for reference.


Figure 5.25637 hits were identified when the data from five repeats of CD18F and three repeats of $t$ CD18F were combined. Of these, 92 hits were determined to be consistent hits.


Figure 5.26 Grouping of the 92 consistent hits. When grouped according to the samples from which the 92 consistent hits were identified, the majority of 72 was found in CD18F.


Figure 5.27 The 92 consistent hits were grouped according to known functions and proteins found to have protein binding, protein modifying and chaperone functions were considered to be interesting hits.


Figure 5.28 Distribution of the three groups of proteins determined to be interesting hits between the two sets of data compared.

| Protein Functions of consistent hits identified | emPAI values obtained for each protein at each experimental repeat |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| (Comparing CD18F and t-CD18F only) | CD18F |  |  |  |  | t-CD18F |  |  | CD11a/tCD18F |
|  | 101222 | 110119 | 110216 | 110531 | 110701 | 110531 | 110701 | 110701 | 110531 |
| Protein binding (CD18F) |  |  |  |  |  |  |  |  |  |
| Clathrin heavy chain 1 (CLTC) | -1 | 0.02 | 0.02 | -1 | 0.03 | -1 | -1 | 0.03 | -1 |
| SCAMP3 protein (SCAMP3) | -1 | 0.09 | -1 | -1 | 0.08 | -1 | -1 | -1 | -1 |
| Karyopherin alpha 2 (Importin alpha 1, KPNA2) | 0.06 | 0.12 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| Exportin-1 (XPO1) | -1 | 0.08 | 0.08 | -1 | -1 | -1 | -1 | 0.02 | 0.11 |
| Lectin, mannose-binding, 1 (LMAN1) | 0.12 | 0.18 | 0.12 | 0.39 | -1 | -1 | -1 | -1 | -1 |
| Protein binding (CD18F and t-CD18F) |  |  |  |  |  |  |  |  |  |
| Karyopherin (importin) beta 1 (KPNB1) | -1 | 0.03 | -1 | 0.07 | -1 | 0.07 | 0.07 | 0.06 | 0.03 |
| Protein Modifying (CD18F) |  |  |  |  |  |  |  |  |  |
| CDP-diacylglycerol--inositol3phosphatidyltransferase (CDIPT) | -1 | 0.16 | 0.35 | -1 | 0.14 | -1 | -1 | -1 | -1 |
| Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit (STT3A) | -1 | 0.1 | 0.22 | -1 | -1 | -1 | -1 | -1 | -1 |
| Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit (STT3B) | -1 | 0.03 | 0.07 | -1 | -1 | -1 | -1 | -1 | -1 |
| Oligosaccharyltransferase complex subunit (OSTC) | -1 | 0.2 | 0.2 | -1 | -1 | -1 | -1 | -1 | -1 |
| Ribophorin II (RPN2) | -1 | -1 | 0.26 | 0.05 | -1 | -1 | -1 | -1 | -1 |
| Glucosidase I | -1 | 0.05 | 0.05 | -1 | -1 | -1 | -1 | -1 | -1 |
| UDP-glucose ceramide glucosyltransferase-like 1 | 0.12 | 0.18 | 0.22 | 0.02 | 0.07 | -1 | 0.04 | -1 | -1 |
| Neutral alpha-glucosidase AB (Glucosidase II $\alpha$ subunit) | 0.03 | 0.07 | 0.06 | 0.12 | -1 | -1 | -1 | 0.22 | 0.07 |
| Protein Modifying (t-CD18) |  |  |  |  |  |  |  |  |  |
| DLST protein | -1 | -1 | -1 | -1 | -1 | -1 | 0.07 | 0.06 | -1 |


| Protein Modifying (CD18F and tCD18F) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Dolichyl-diphosphooligosaccharide-proteinglycosyltransferase 48 kDa subunit | 0.07 | 0.07 | -1 | 0.15 | -1 | -1 | 0.15 | 0.2 | -1 |
| Protein disulfide-isomerase A6 | 0.06 | 0.2 | 0.2 | 0.13 | -1 | 0.13 | -1 | 0.24 | 0.13 |
| Protein disulfide-isomerase A3 | -1 | 0.06 | -1 | -1 | 0.17 | 0.94 | 0.19 | 0.43 | 0.12 |
| Chaperone (CD18F) |  |  |  |  |  |  |  |  |  |
| HSP90AA1 protein | 0.1 | -1 | 0.05 | -1 | 0.13 | -1 | -1 | 0.23 | -1 |
| DnaJ (Hsp40) homolog, subfamily A, member 1 (DNAJA1) | 0.07 | -1 | -1 | -1 | 0.13 | -1 | 0.15 | -1 | -1 |
| 78 kDa glucose-regulated protein (HSPA5) | -1 | -1 | 6.69 | 15.36 | -1 | 2.17 | -1 | -1 | 2.78 |
| Chaperone (CD18F t-CD18F) |  |  |  |  |  |  |  |  |  |
| Heat shock 70kDa protein 8 isoform 1 variant | -1 | 0.39 | -1 | 0.6 | 0.43 | 0.15 | -1 | 0.18 | -1 |
| DnaJ homolog subfamily B member 11 (DNAJB11) | -1 | 0.39 | 0.16 | 0.6 | 0.43 | 0.16 | 0.16 | 0.18 | -1 |
| Calnexin (CANX) | 0.94 | 0.33 | 0.87 | 1.45 | 0.32 | 0.12 | 0.21 | 0.04 | 0.21 |
| Chaperone ( tCD18) |  |  |  |  |  |  |  |  |  |
| Endoplasmic reticulum resident protein 44 (ERP44) | -1 | -1 | -1 | -1 | -1 | -1 | 0.07 | 0.06 | -1 |

Table 5.7 Consistent hits with protein binding, modifying and chaperone functions identified during analysis of CD18F and t-CD18F LCMS data.

### 5.3 Discussion

### 5.3.1 Integrin IP using antibodies

Twelve integrin interacting chaperone/modification proteins were identified using the antibody-based IP-LCMS method. These proteins may be responsible for retention of the partially unfolded monomers, or may assist in formation of the integrin dimer. However, since this was a trial experiment, no background or negative controls were used. Even though a pre-clearing step was used, the identified proteins may be background proteins which bound to the MHM24 or H52 PAS-RAM conjugated beads unspecifically. Indeed, only the hit for HSPA5 (75 kDa glucose regulated protein) corroborated with samples from FLAG purification, indicating that the antibody-PAS-RAM system generated a high proportion of false positive results in LCMSMS. Since the PAS-RAM system consisted of a rabbit antimouse antibody and a mouse anti-human antibody, there was probably much unspecific protein binding to the antibodies used.

Adding to that, as observed in the Commassie blue stained SDS-PAGE (Figure 5.3), a large protein band was present around 50 kDa . This resulted from co-elution of the integrin antibody IgG chain and the mouse IgG chain from conjugated PAS-RAM. This strong protein band masks other proteins which resolve around the same molecular weight in mass spectrometry. Hence, to improve on this technique, a HIS tag was cloned behind the CD18 protein.

### 5.3.2 HIS tag purification

Protein yield from HIS tag purification was poor and many unspecific (i.e. untagged) proteins were able to bind to the beads, both nickel and cobalt, and were subsequently eluted. These
contaminating proteins masked the CD18_HIS protein as well as the desired interacting proteins. These results were in contrast to previous results (Shi, et al., 2007a).

To explain this discrepancy, one might argue that the HIS tagged PHE1 and PHE2 protein were secreted into the culture media. In addition, these proteins were purified from cell culture supernatant that did not contain FBS. As a result, less contaminating proteins were present during bead binding (loading) and elution.

In contrast, the CD18_HIS experiments here were performed on cell lysate obtained from a transfected human cell line. Possibly, many proteins in the human proteome are able to bind strongly to nickel and cobalt beads. As a result, despite strong wash conditions in the presence of detergents, contaminating proteins could not be removed. Nonetheless, FACS analysis of CD18_HIS/CD11a expression was useful in indicating that the HIS sequence did not interfere with normal expression of LFA-1. This culminated in the cloning of a FLAG tag behind the CD18 protein instead.

### 5.3.3 FLAG tag purification

LCMS results from all the FLAG tag purification experiments were pooled and analyzed (presented in Appendix II). The hits were classified according to known protein functions which may be useful to provide an overview of cellular processes relating to the integrins in transfectants, and each category is analyzed in turn in the following sections.

FLAG tag purification resulted in relatively clean samples which were suitable for LCMSMS analysis, and is superior over the traditional mAb based immunoprecipitation in that the FLAG peptide could be used to elute the immunoprecipitated protein. Furthermore,

FLAG purification avoids the 50 kDa IgG band that typically contaminates the eluate, and has the advantage in that a negative control could be easily performed.

### 5.3.3.1 CD18F vs CD11a/CD18F

Individually transfected CD18F samples were initially compared against CD11a/CD18F samples. The majority of consistent hits were overlapping, present in both CD11a/CD18F and CD18F transfectants (Figure 5.17) This is consistent with the expectation that while CD18F folds together with CD11a intracellularly, each nascent integrin protein is a monomer before heterodimer formation occurs.

In this study, many of the known intergrin interacting cytoplasmic proteins important for inside-out activation were not identified. However, a human embryonic kidney cell line was used instead here of a cell line with a leukocyte lineage. This may explain the absence of the interacting cytoplasmic proteins since they are known to bind activated leukocyte membrane expressed LFA-1 upon ligand binding, or endogenous activation downstream of chemokine, B and T-cell receptor activation (Alon and Feigelson, 2009; Arana, et al., 2008; Laudanna, et al., 2002; Zell, et al., 1999). For example, 14-3-3 proteins were shown to bind to the cytoplasmic domain of membrane expressed LFA-1 in activated T-cells, facilitating LFA-1 mediated cell spreading (Nurmi, et al., 2006).

Nonetheless, the objective of the current study was to identify proteins that mediate retention of the partially unfolded CD18 protein. Since CD11a and CD18 monomer retention was consistently observed in cells from LAD-I patients (Anderson, et al., 1985; Kohl, et al., 1984; Marlin, et al., 1986), leukocyte cell lines where either subunit was deleted (Weber, et al., 1997) and individually transfected HEK cells, it was reasoned that the retention mechanism
for partially folded CD18 and CD11a is ubiquitous. Furthermore, CD11a and CD18 formed heterodimers when co-expressed in HEK293 cells, indicating that the integrin expression and dimerization processes that occur endogenously in leukocytes do occur in the transfectants as well.

Furthermore, the transfection system was useful in a two fold manner. Firstly, HEK293 cells were much easier to transfect than leukocytes. Secondly, this system could be used to express a FLAG-tagged integrin monomer, which facilitated purification of relatively 'clean' samples suitable for mass spectrometry. FLAG tagged integrin purification would be difficult to achieve in a leukocyte cell line since endogenous LFA-1 would compete with over-expressed, FLAG tagged integrin, which is in turn difficult to transfect. Given that the generation of a stable leukocyte cell line expressing FLAG tagged integrin monomers would create time constraints on the project, the HEK293 transfection system was probably the more efficient system for this study.

This segment of the study provided an opportunity to identify proteins responsible for integrin dimerization. To this end, fourteen chaperone and modifying proteins were identified consistently in both sets of FLAG tag purified samples. Together with HSPH1 (HSP 105 kDa ) and PDIA4 (identified exclusively in CD11a/CD18F), these results represent chaperone/modifying proteins that have relatively stable interactions with the CD18 protein.

An alternative view of HSP105 and PDIA4 may also be valid. Both were not present in CD18F and as such, may be candidate proteins for mediating integrin heterodimer formation, binding only during the dimerization process. However, these proteins may simply be CD11a
binding proteins. The latter possibility can only be eliminated by analyzing a minimum of three biological repetitions for CD11aF.

Nonetheless, the list of 14 proteins presented in Table 5.4 may serve as a good starting point to investigate the process of integrin dimerization.

### 5.3.3.2 Analysis of CD11a/CD18F vs CD11a/t-CD18F showed that CD11a associated intracellularly with t-CD18F

CD11a was pulled down together with t -CD18F when co-transfected. Using emPAI values obtained in the experiment, protein content was calculated based on an established mathematical formula (Ishihama, et al., 2005). The ratio of CD11a pulled down together with t -CD18F (i.e. protein content of CD11a divided by that for $\mathrm{t}-\mathrm{CD} 18 \mathrm{~F}$ ) was 0.42 while ratios for CD11a/CD18F and CD11aF/CD18 ranged from 0.11-0.33 (Table 5.8). Albeit that only one sample was analyzed in IP-LCMS, this demonstrated that t -CD18F associated with CD11a intracellularly, even though only monomeric t-CD18F was detected at the cell surface, and suggested stoichiometrically that more CD11a was bound to t -CD18F compared to the other doubly transfected samples.

Based on the hypothesis that the $\beta \mathrm{I}$ domain folds together with CD11a, the mutant dimer should not form. However, results here showed an intracellular association of t-CD18F and CD11a, suggesting that the $\beta$ I domain flanking regions may help to anchor the two nascent proteins together, and may fold with the CD11a beta-propeller to form a defective dimer, in the absence of the $\beta I$ domain. This is consistent with a previous study indicating that removal of the $\beta I$ domain does not affect folding of the CD18 NTR/mid-region complex (Tan, et al., 2001a). Since it is retained, the CD11a/t-CD18F system may be useful in comparison with completed integrin dimers in the study of integrin exocytosis.

| Sample | CD18F |  |  |  |  | t-CD18F |  |  | $\begin{aligned} & \mathrm{t}-\mathrm{CD} 18 \mathrm{~F} \\ & \text { /CD11a } \end{aligned}$ | CD11a/CD18F |  |  |  | $\begin{aligned} & \text { CD11aF } \\ & \text { /CD18 } \end{aligned}$ | CD11aF |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Date | 101222 | 110119 | 110216 | 110531 | 110701 | 110531 | 110701 | 110701 | 110531 | 101222 | 110119 | 110216 | 110531 | 110531 | 110531 |
| Total emPAI | 39.96 | 53.62 | 98.46 | 83.27 | 111.7 | 61.6 | 50.05 | 37.19 | 71.17 | 42.99 | 42.55 | 156.78 | 52.05 | 46.55 | 131.33 |
| CD18 emPAI | 9.33 | 6.4 | 10.98 | 15.72 | 5.32 | 3.64 | 2.27 | 1.6 | 5.25 | 6.68 | 2.61 | 23.22 | 13.42 | 7.49 | -1 |
| CD11a emPAI | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 2.01 | 0.73 | 0.85 | 3.55 | 2.64 | 1.95 | 5.28 |
| CD18 protein content (mol\%) | 23.35 | 11.94 | 11.15 | 18.88 | 4.76 | 5.91 | 4.54 | 4.30 | 0.39 | 15.54 | 6.13 | 14.81 | 25.78 | 16.09 | - |
| CD11a protein content (mol\%) | - | - | - | - | - | - | - | - | 0.16 | 1.70 | 2.00 | 2.26 | 5.07 | 4.19 | 4.02 |
| CD11a/CD18 |  |  |  |  |  |  |  |  | 0.42 | 0.11 | 0.33 | 0.15 | 0.20 | 0.26 | - |

Table 5.8 Calculations of the protein content (mol \%) and ratio of CD18F and CD11a in each of the samples analyzed in IP-LCMS.

### 5.3.3.3 Chaperone and Protein modification proteins

Of main interest to the current study, are proteins grouped under the headings 'Chaperones' or 'Protein Modification' (chaperone/modification), based on currently known functions. The former have functional roles in sorting of proteins and protein folding, and the latter are responsible for protein modification. A summary of the known protein functions of the chaperone/modification proteins identified is presented in Appendix II of this thesis.

It is difficult to discern which of the chaperone proteins found in both CD 18 F and CD11a/CD18F, is solely responsible for integrin heterodimer formation. While calnexin may retain unfolded integrin proteins in the ER (discussed later in this chapter), it is also possible that folding of CD18 with the CD11a protein may not be dependent on a single chaperone protein, but may instead rely on an array of chaperone proteins in tandem or simultaneously, to attain the required conformation.

When compared across CD18F, t-CD18 and CD11a, calnexin was found consistently in all three sample sets. Removal of the $\beta \mathrm{I}$ domain resulted in monomeric expression of t -CD18F. In comparing CD18F and t -CD18F, several chaperone/modifying proteins which bound to the $\beta$ domain to mediate CD18 retention were identified. The most interesting hit was for Lectin, Mannose Binding protein 1 (LMAN1), which is discussed in more detail later on.

### 5.3.3.4 Proteasome and ribosomal proteins

Proteasome related proteins were found consistently in all the FLAG purified samples. This is reasonable as an over-expression system was used in this analysis; a proportion of the overexpressed and partially unfolded integrins may be in the process of degradation in the proteasome.

Similarly, many ribosomal proteins were consistently identified. While one might argue that a proportion of immunoprecipitated integrins are nascent proteins emerging from protein synthesis, the FLAG tag is present in the C terminus of the protein. Thus, the FLAG domain is the last portion to be synthesized, after which the ribosomes would have been released. It is unlikely that some of the CD18F obtained during IP was still in the process of synthesis.

It may be possible to explain the consistent, non-background hits corresponding to ribosomal proteins in the following manner. While it has not been demonstrated conclusively for the leukocyte integrins, ribosomal and translational initiation proteins have been shown to colocalize with talin/ $\beta 3$ in migrating or spreading mammalian cells (Willett, et al., 2010). Colocalization at the adhesion complexes of the cell membrane was demonstrated with microscopy. As Willett et al [2010] have proposed, ribosomal proteins associate with the integrin proteins outside of the endoplasmic reticulum, at subcellular locations where the integrin proteins are required. This would enable synthesis of the integrin proteins to be upregulated at sublocations such as the cell membrane. LC-MSMS results obtained here suggest a physical association between ribosomal proteins and the leukocyte integrins and seem to support this hypothesis.

### 5.3.3.5 Calnexin as an integrin retaining protein

Calnexin is known to function in synergy with UDP-glucosytransferase in binding and retaining nascent or misfolded proteins (Kim and Arvan, 1995; Tsukamoto, et al., 2002; Wada, et al., 1997; Williams, 2006) (see Figure 5.29). Based on the LC-MSMS results and the known protein functions, calnexin was identified as a strong candidate to retain the partially unfolded integrin proteins intracellularly. While it was hypothesized initially that calnexin would not bind t-CD18, IP results demonstrated that calnexin associates physically with both CD18 and t-CD18.

The CD18 protein folds independently of CD11a, with the exception of the $\beta \mathrm{I}$ domain (Huang, et al., 1997; Huang and Springer, 1997). FACS analysis demonstrated that removal of the $\beta$ d domain can allow some of the nascent t -CD18 protein to escape retention. It is plausible that t-CD18 association with calnexin is a result of calnexin association with other domains of the CD1 8 protein, apart from the $\beta$ domain.

The absence of the unfolded $\beta I$ domain in the largely folded t -CD18 protein may have facilitated escape from calnexin retention after proper folding of all the other domains, resulting in t-CD18 exocytosis to the cell surface as a monomer. Flow cytometric analysis also indicated that relative to wild type CD11a/CD18F transfectants, t -CD18 transfectants are less abundantly expressed on the cell surface. Hence it is likely that t-CD18, despite the lack of the $\beta \mathrm{I}$ domain, is subject to intracellular retention that could be overwhelmed by the transfection of twice as much DNA plasmid in doubly transfected (CD11a/t-CD18) samples.


Figure 5.29 A schematic diagram of the calnexin cycle. Calnexin (cnx) binds to nascent and misfolded proteins in the endoplasmic reticulum by recognizing and binding to a glucose residue. Calnexin mediates refolding of the improperly folded protein. Glucosidase II (GII) removes the glucose residue such that calnexin can no longer bind to the protein and that properly folded proteins can then be exocytosed. UGGT1 recognizes proteins that are still misfolded, adds a glucose residue on misfolded proteins, which are then bound by calnexin for further cycles of refolding.

### 5.3.3.6 Mitochondria proteins

It is harder to explain some of the other hits. Mitochondria proteins were consistently pulled down in all the FLAG purification samples. In an attempt to explain this observation, one might speculate about the relationship between integrins and mitochondria proteins in the following manner.

Several studies indicate a physical association between the mitochondria and the ER (Mannella, et al., 1994; Soltys and Gupta, 1992). The smooth and rough ER are tethered to the mitochondria outer membrane at a distance of 10 nm and 25 nm respectively, a distance close enough such that ER proteins can directly associate with proteins and lipids of the outer mitochondria membrane (Csordas, et al., 2006). Molecular chaperones such as calnexin, calreticulin and HSPA5 (also known as BiP) have been shown to be localized at the Mitochondria Associated ER Membranes (MEM) where they function as high-capacity $\mathrm{Ca}^{2+}$ binding proteins that serve as intracellular $\mathrm{Ca}^{2+}$ buffering pools. In particular, $80 \%$ of intracellular calnexin occurred in the ER, where the majority of calnexin protein was observed to be in the MEM (Hayashi and Su, 2007; Myhill, et al., 2008).

Calnexin was consistently identified in this study to be an integrin binding protein. Hence, one might speculate that the integrins are localized to the MEM together with calnexin, where they are retained by calnexin during the process of integrin folding. Furthermore, it was recently demonstrated that $\beta 1$ integrin folding and assembly are completely dependent on $\mathrm{Ca}^{2+}$ binding, and that $\mathrm{Ca}^{2+}$ ion binding maintains the integrin in an inactive state until activation at the cell surface, mediated by the displacement of $\mathrm{Ca}^{2+}$ with $\mathrm{Mg}^{2+}$ or $\mathrm{Mn}^{2+}$ (Tiwari, et al., 2011). Given the coinciding role of calnexin as a molecular chaperone and as
an intracellular $\mathrm{Ca}^{2+}$ buffer at the MEM, it is possible that nascent integrins are localized to the MEM, close to the outer mitochondria membrane.

Thus, localization of integrins to the MEM may allow the integrins to interact with several mitochondria proteins in a yet uncharacterized manner.

### 5.3.4 LMAN1 may mediate retention of CD18 by binding the $\beta$ I domain

ER stress occurs as a result of accumulating misfolded/unfolded proteins. This is a stress response that triggers upregulation of LMAN1 expression and the binding of HSP5A (BiP) to misfolded/unfolded proteins (Gething, et al., 1986; Kozutsumi, et al., 1988; Renna, et al., 2007). In turn, LMAN1 is known to have a cargo receptor function, transporting specific proteins from the ER to the Golgi apparatus. This is mediated through an adaptor protein. For example, MCFD2 is an adaptor protein required to recruit the factor V and factor VIII proteins to sites of transport vesicle budding, allowing LMAN1 to mediate ER to Golgi transport of these blood clotting proteins (Zhang, et al., 2003; Zhang, et al., 2005; Zheng, et al., 2010).

In this study, LMAN1 was found to bind CD18F but not to $\mathrm{t}-\mathrm{CD} 18 \mathrm{~F}$. This suggests a possible role for LMAN1 in retention of monomeric CD18 by binding to the $\beta I$ domain. On the other hand, given that LMAN1 is upregulated during ER stress, and that monomeric CD18 cannot fold correctly by itself and is retained in the ER as a consequence, it is also possible that LMAN1 detection is a result of stalled ER to Golgi transport of partially unfolded CD18. In such a scenario, LMAN1 was unable to transport partially unfolded CD18 to the Golgi, thus remaining bound to CD18 which was in turn retained in the ER.

A separate observation lends credibility to the former theory. A proportion of the CD11a/tCD18F sample occurred in the (probably partially unfolded) dimeric form, which was retained intracellularly. Since the $\beta I$ domain was absent, retention of the defective dimer was probably not mediated by LMAN1. Indeed, LMAN1 was not detected in this sample.

### 5.4 A brief conclusion on the studies of integrin monomer retention

This experimental series was initiated with the intention of identifying chaperone/modification proteins that mediate integrin heterodimer formation. Several candidate chaperone/modification proteins that probably interact with the integrins in a relatively consistent and stable manner were identified. This serves as a good starting point to identify mechanisms underlying integrin heterodimer formation.

To narrow the experimental objective further, it was hypothesized and demonstrated that that removal of the $\beta \mathrm{I}$ domain allows cell surface expression of t -CD18F. Furthermore, it was demonstrated that calnexin interacts with nascent CD18, and through an analysis of the IPLCMS data, CD11a. Furthermore, LMAN1 interacts with the $\beta$ I domain of CD18F. Taken together, the data suggests that LMAN1 may mediate retention of CD18F in conjunction with calnexin.

After removing the $\beta \mathrm{I}$ domain, heterodimer formation can occur intracellularly, but export of the integrins was not detected. Thus, intracellular retention of the mutant heterodimer may be mediated by calnexin, either alone or in conjunction with another uncharacterized retention protein. Last but not least, the observation that doubly transfected sample express a higher proportion of t-CD18F on the cell surface suggests that the truncated CD18 protein remains subject to intracellular retention.

## Chapter 6 : Miscellaneous Experiments

### 6.1 Analyzing the MEM148 epitope

### 6.1.1 Introduction

The mouse anti-human MEM148 antibody was previously obtained from V. Horejsi (Institute of Molecular Genetics, Prague, Czech Republic). This monoclonal antibody was useful for detecting a masked epitope in the CD18 hybrid domain, and reports hybrid domain displacement (leg separation) (Cheng, et al., 2007; Tang, et al., 2005).

Subsequently, the MEM148 antibody was commercialized. MEM148 and MEM148-FITC (FITC conjugated antibody) were purchased from Serotech after the aliquot from V. Horejsi was used up. However, the purchased antibodies were inefficient in reporting leg separation in MOLT-4 LFA-1, such that the FACS profile with $\mathrm{Mg} / E G T A(M / E)$ activation had lower fluorescent intensity compared to before.

To illustrate this, two publications in 2001 and 2005 which describe FACS profiles of MEM148 reporter activity in LFA-1 leg opening in MOLT-4, were compared to unpublished results obtained by Cheng, et al. in 2008. In the top two panels of Figure 6.1, MEM148 obtained from V. Horejsi was used to stain M/E activated LFA-1 compared to inactive LFA-1. An irrelevant antibody was used as background control (represented with a black line), as labeled.

In 2008 (bottom panel), the proportion of cells reported to be active in the presence of $\mathrm{M} / \mathrm{E}$ was lower than in 2002 and 2005 (Cheng, et al., 2008). Using MEM148-FITC (Serotech), reporter activity in M/E activated LFA-1 (black line) against inactive LFA-1 (black line) was poor. Note that the FACS profile of MEM148 (Serotech) used in conjunction with FITC
conjugated rabbit anti-mouse secondary antibody is similar to that obtained with MEM148-
FITC (data not shown).


Figure 6.1 FACS profiles of MEM148 reporting activity for MOLT-4 LFA-1 leg opening from two publications in 2001 and 2005, compared to that obtained in 2008. In the top two panels, the fluorescence intensity with MEM148 staining of active (M/E activated) or inactive (without M/E or additives) is shown against that of an irrelevant antibody as background control. The background fluorescence is represented with a black line in Tan SM et al. [2001] and in Tang RH et al. [2005]. The MEM148 antibody used was obtained from V. Horejsi.

In 2008 (bottom panel), the proportion of cells reported to be active in the presence of $\mathrm{M} / \mathrm{E}$ was lower than in 2002 and 2005. This experiment was performed by Cheng $M$ et al. (2008) comparing M/E activated LFA-1 (black line) against inactive LFA-1 (solid curve). Here, the MEM148 antibody was purchased from Serotech.

In order to publish results from the previous project [Cheng, et al., 2008, unpublished], we attempted to troubleshoot and optimize the 'leg separation' reporter activity of MEM148 in various activating conditions and reagent concentrations.

Later on, we mapped the MEM148 epitope and compared our results to those previously obtained by Tang, et al. in 2005. This was achieved using chimeric plasmids bearing either the mouse or human MEM148 epitope in a human or mouse background, respectively. The $\beta 2 \mathrm{Hu} / \mathrm{Mo}$ A construct in pcDNA3.0 was previously described, in which $\mathrm{Met}^{1}$ to $\mathrm{Asn}^{584}$ of the human integrin $\beta 2$ sequence was replaced with the corresponding region from mouse $\beta 2$
(Figure 6.2) (Tng, et al., 2004). This plasmid was used to generate alternative constructs in which the human MEM148 epitope (VTHRNQP) was partially or wholly swapped into the mouse $\beta 2 \mathrm{Hu} / \mathrm{Mo}$ A construct (Tang, et al., 2005). Similarly, the human pcDNA3.0 plasmid was used as a template into which the corresponding mouse epitope (ASSIGKS), identified by protein alignment, were partially or wholly swapped. As such, 17 plasmids were obtained in total. The protein alignment between the mouse and human CD18 sequence at the midregion is depicted in Figure 6.3.

For epitope mapping, each plasmid was transfected into COS7 cells, which supports the expression of monomeric CD18 on the cell surface. The transfectants were then analyzed for MEM148 epitope expression in FACS.


Figure 6.2 Schematic diagram of the $\mathbf{\beta 2} \mathbf{H u} / \mathrm{Mo} A$ construct in pcDNA3.0, taken from Tang, et al. [2005] Met $^{1}$ to Asn $^{584}$ of the human integrin $\beta 2$ sequence was replaced with the corresponding region from mouse $\boldsymbol{\beta 2}$


Figure 6.3 Protein alignment showing the matching sequences (highlighted with solid circles) between the human and mouse epitopes for MEM148 [Tang, et al., 2005]. This sequence is within the mid-region, which is now known to be defined as insertions of the hybrid/PSI domain (refer to Figure $\mathbf{1 . 3}$ in Chapter 1).

### 6.1.2 Results

### 6.1.2.1 Increasing M/E or MEM148 concentration in both MEM148 (Serotech) and

## MEM148 Ascites, did not improve reporter activity

The poor leg opening reporting by MEM148 (Serotech) in MOLT-4 LFA-1 may be due to a number of possibilities, including differences in antibody concentration used, M/E concentration, a mutation in the antibody paratope, or the use of a unreliable cell line (i.e. contaminated or mixed up with another cell line).

To investigate the possibility that the poor reporter activity of MEM148 (Serotech) in MOLT-4 LFA-1 was due to differences in mAb or $\mathrm{M} / \mathrm{E}$ concentration, different concentrations of $\mathrm{M} / \mathrm{E}$ ( 1 x concentration: $5 \mathrm{mM} \mathrm{MgCl} 2,1.5 \mathrm{mM}$ EGTA, ranging from 1x to 2 x ) and MEM148 (Serotech) ( 1 x concentration: $10 \mathrm{ng} / \mu \mathrm{l}$, ranging from, 1 x to 6 x ) were tested. In addition, we requested for an old stock of MEM148 from V. Horejsi and were kindly provided with MEM148 (Ascites) (MEM148 ascites fluid). This was used as a primary antibody and tested alongside MEM148 (Serotech) in a range from 1x (1:500 dilution) to 2 x (1:250 dilution).

In addition, to ensure that the MOLT-4 (laboratory stock) cell line used in our laboratory was reliable (i.e. not contaminated or mixed up with another cell line), MOLT-4 cells were newly purchased from ATCC (MOLT-4 ATCC) and tested alongside MOLT-4 (laboratory stock). KIM127 with M/E activation was used as a positive control to show that M/E activation was working.

However, our results showed that none of the conditions improved reporter efficiency in both MEM148 (Serotech) or MEM148 Ascites. Furthermore, there was no significant difference in MEM148 reporting of LFA-1 leg opening in MOLT-4 (laboratory stock) and MOLT-4 ATCC (Figure 6.4).


Figure 6.4 FACS analysis of MOLT-4 cells labeled using normal (1x) M/E and $2 x$ ME, or labeled using a range of MEM148 concentrations ranging from $1 x$ to $6 x$. The background mAb used was LPM19C.

### 6.1.2.2 MEM148 reporter activity was more efficient in SKW3

To determine if this problem was restricted to MOLT-4, MEM148 (Serotech) reporter activity in MOLT-4 (ATCC) and SKW3 was compared. An additional comparison was made between $\mathrm{M} / \mathrm{E}$ and Mn activation; 2 mM manganese chloride ( Mn ) was used to activate LFA-1 in MOLT-4 and SKW3 cells. The results showed that M/E and Mn activation of SKW3 LFA1 could be detected by MEM148 (Serotech) clearly. However, MEM148 (Serotech) reporting of MOLT-4 LFA-1 leg opening was inefficient (Figure 6.5).


Figure 6.5 FACS analysis of MOLT-4 and SKW3 cells, using both M/E and Mn as activating agents. While reporter activity of MOLT-4 LFA-1 leg opening was poor, MEM148 could efficiently report leg opening in SKW3. The background mAb used was LPM19C.

### 6.1.2.3 Results from epitope mapping of MEM148 were similar to previous results

The epitope swapping strategy was described earlier (Tang, et al., 2005). In total, 17 hybrid CD18 clones were obtained and transfected into COS7 cells as these cells were able to support CD18 monomer expression. Each knock out plasmid contained the entire or part of the mouse MEM148 epitope swapped into the human CD18 sequence, while each knock in plasmid contained the entire or part of the human MEM148 epitope, swapped into the mouse CD18 sequence (Tang, et al., 2005). The expression of the monomeric mutant CD18 from each clone was confirmed according to KIM127 expression, which bound to an epitope in CD18 away from that of MEM148. While the KIM127 epitope is masked in resting LFA-1, this epitope is exposed in monomeric CD18 (data not shown).

Each transfectant was tested for MEM148 epitope expression levels by staining with MEM148-FITC (Serotech) and analyzed in FACS as previously described (Tang, et al., 2005). The proportion of transfectants stained by MEM148-FITC was compared to that of COS7 cell transfected with CD18 (used as a positive control). These results were converted to Mean Fluorescence Index (\% gated positive multiplied by mean fluorescence index) and presented in Table 6.1 as a percentage of the positive control.

| Knock out Expression Plasmids |  |  |  | Mean Fluorescence Index (\%) |  |
| :--- | :--- | :--- | :---: | :---: | :---: |
| Original residues <br> on human $\beta 2$ | Replacing <br> residues from <br> mouse epitope | Final epitope <br> sequence | MEM148-2005 |  |  |$\quad$ MEM148-2009

Table 6.1 Summary of the Mean Fluoresence Index (MFI) obtained from MEM148 epitope expression in transfected COS7 cells after reciprocal swapping of the residues. Residues from the mouse epitope are depicted in red.

### 6.1.3 Discussion

The poor reporter activity of MOLT-4 LFA-1 leg opening was independent of M/E concentration. Furthermore, no improvement in reporter activity could be achieved with higher MEM148 concentrations, suggesting that the amount of antibody at 1 x concentration was sufficient to bind available VTHRNQP epitopes in activated MOLT-4 LFA-1, and that the limiting factor for reporting of leg opening was the number of epitopes available.

In addition, the FACS profile of MOLT-4 ATCC and MOLT-4 (laboratory stock) was similar, indicating that the low reporter activity of MEM148 (Serotech) was not due to a problem with the cell line.

A high proportion of SKW3 cells bind to ICAM-1 without activation (data not shown), indicating that SKW3 cells expressed LFA-1 that were more active. It is probably likely that a high proportion of SKW3 LFA-1 exists endogenously in the activated, leg-open conformation.

FACS analysis comparing M/E and Mn activated LFA-1 in both cell lines suggested that MEM148 (Serotech) could detect LFA-1 leg opening. However, since the reporting of MOLT-4 LFA-1 leg opening was poor, one might suggest that MEM148 (Serotech) required a higher stoichiometric ratio of the exposed VTHRNQP epitope to bind activated LFA-1 efficiently (such as in SKW3), compared to MEM148 (V. Horejsi).

To investigate this possibility, epitope mapping was performed on mutant CD18 plasmids using MEM148-FITC (Serotech). Epitope mapping showed that MEM148-FITC (Serotech) (labeled as MEM148 2009) mapped to the VTHRNQP sequence. Up to $99 \%$ of MEM148-

FITC (Serotech) binding could be rescued when VTHRNQP was swapped into the chimeric $\beta 2 \mathrm{Hu} / \mathrm{Mo}$ (Table 6.1). On the whole, mapping results from all plasmids were similar to earlier results obtained using MEM148 (V. Horejsi) (Tang, et al., 2005), indicating that the antibody paratope was intact in MEM148-FITC (Serotech). This was in contrary to our earlier postulation that the MEM148 paratope had mutated over the course of several years.

Thus, the observed loss of MEM148 efficiency in reporting LFA-1 activation in MOLT-4 was not due to any of the conditions tested.

## Chapter 7 : Conclusion

In this thesis, three disparate studies relating to the CD18 integrin were performed. First, a $27,703 \mathrm{bp}$ deletion resulting in a fusion gene between $I T G B 2$ and PTTG1IP was characterized in an LAD-I patient. Secondly, studies of the intron 2 sequence demonstrated that the presence of an AG dinucleotide located 11 nt upstream represses downstream 3'ss activation, even in the absence of competing 3'ss. Thirdly, the process of integrin monomer retention was studied. Calnexin and LMAN1 were identified to be strong candidates for retaining the partially unfolded CD18 protein.

Further studies can be performed following this work. In Chapter 4, 3'ss scores for c .500 $12 \mathrm{~T}>\mathrm{G}$ predicted activation of the -973 'ss, but analysis of the cDNA sequence showed activation of a cryptic 3 'ss at +149 (Roos and Law, 2001; Roos, et al., 2002), which had scores below background or low scores.

In addition, c. $742-14 \mathrm{C}>\mathrm{A}$ resulted in use of a cryptic 3 'ss at -12 instead of -57 , despite low scores at -12 and scores at -57 which were higher than that of the authentic 3 'ss. Both cases warrant an in depth study of cryptic 3'ss activation, and may provide additional novel information related to cryptic 3'ss activation.

While the first two studies presented in Chapter 3 and 4 are complete studies, the third study presented in Chapter 5 is incomplete. Further experiments are required to determine the exact roles of calnexin and LMAN1 in nascent CD18 processing, modification and exocytosis. Nonetheless, mass spectrometry-based studies on the integrin interactome has not been performed elsewhere. Hence, the CD18 interacting proteins identified in transfectants represent a novel set of information.

Furthermore, the third study was conceived with the intention of identifying mechanisms involved in integrin dimerization. The current results are insufficient to provide a good insight into integrin dimerization. More information on CD11a chaperone/modifying proteins is required. This project can be extended to encompass CD11b and CD11c to provide a mass spectrometry based overview of the cellular processes that nascent integrin proteins undergo.

## References

Abram CL, Lowell CA. 2009. Leukocyte adhesion deficiency syndrome: a controversy solved. Immunol Cell Biol 87(6):440-2.
Agace WW, Patarroyo M, Svensson M, Carlemalm E, Svanborg C. 1995. Escherichia coli induces transuroepithelial neutrophil migration by an intercellular adhesion molecule-1-dependent mechanism. Infection and Immunity 63(10):4054-4062.
Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2007. Molecular Biology of the Cell: Garland Science.
Alon R, Aker M, Feigelson S, Sokolovsky-Eisenberg M, Staunton DE, Cinamon G, Grabovsky V, Shamri R, Etzioni A. 2003. A novel genetic leukocyte adhesion deficiency in subsecond triggering of integrin avidity by endothelial chemokines results in impaired leukocyte arrest on vascular endothelium under shear flow. Blood 101(11):4437-45.
Alon R, Etzioni A. 2003. LAD-III, a novel group of leukocyte integrin activation deficiencies. Trends. Immunol. 24(10):561-6.
Alon R, Feigelson SW. 2009. Chemokine signaling to lymphocyte integrins under shear flow. Microcirculation 16(1):3-16.
Alon R, Hammer DA, Springer TA. 1995. Lifetime of the P-selectin-carbohydrate bond and its response to tensile force in hydrodynamic flow. Nature 374(6522):539-42.
Anderson DC, Schmalsteig FC, Finegold MJ, Hughes BJ, Rothlein R, Miller LJ, Kohl S, Tosi MF, Jacobs RL, Waldrop TC and others. 1985. The severe and moderate phenotypes of heritable Mac-1, LFA-1 deficiency: their quantitative definition and relation to leukocyte dysfunction and clinical features. J Infect Dis 152(4):668-89.
Andoniou CE, van Dommelen SL, Voigt V, Andrews DM, Brizard G, Asselin-Paturel C, Delale T, Stacey KJ, Trinchieri G, Degli-Esposti MA. 2005. Interaction between conventional dendritic cells and natural killer cells is integral to the activation of effective antiviral immunity. Nat Immunol 6(10):1011-9.
Arana E, Harwood NE, Batista FD. 2008. Regulation of integrin activation through the B-cell receptor. Journal of Cell Science 121(14):2279-2286.
Armulik A, Nilsson I, von Heijne G, Johansson S. 1999. Determination of the border between the transmembrane and cytoplasmic domains of human integrin subunits. J Biol Chem 274(52):37030-4.
Arnaout MA. 1990. Leukocyte adhesion molecules deficiency: its structural basis, pathophysiology and implications for modulating the inflammatory response. Immunol Rev 114:145-80.
Arnaout MA, Mahalingam B, Xiong JP. 2005. Integrin structure, allostery, and bidirectional signaling. Annu Rev Cell Dev Biol 21:381-410.
Arnaout MA, Spits H, Terhorst C, Pitt J, Todd RF. 1984. Deficiency of a leukocyte surface glycoprotein (LFA-1) in two patients with Mol deficiency. Effects of cell activation on Mo1/LFA-1 surface expression in normal and deficient leukocytes. J Clin Invest 74(4):1291-300.
Back AL, Kwok WW, Hickstein DD. 1992. Identification of two molecular defects in a child with leukocyte adherence deficiency. J Biol Chem 267(8):5482-7.
Bajt ML, Farhood A, Jaeschke H. 2001. Effects of CXC chemokines on neutrophil activation and sequestration in hepatic vasculature. American Journal of Physiology Gastrointestinal and Liver Physiology 281(5):G1188-G1195.
Barczyk M, Carracedo S, Gullberg D. 2010. Integrins. Cell Tissue Res 339(1):269-80.

Beglova N, Blacklow SC, Takagi J, Springer TA. 2002. Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. Nat Struct Biol 9(4):282-7.
Bell M, Schreiner S, Damianov A, Reddy R, Bindereif A. 2002. p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. EMBO J 21(11):2724-2735.
Berglund JA, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes \& Development 12(6):858-867.
Bork P, Doerks T, Springer TA, Snel B. 1999. Domains in plexins: links to integrins and transcription factors. Trends Biochem Sci 24(7):261-3.
Bretscher MS. 1992. Circulating integrins: alpha 5 beta 1, alpha 6 beta 4 and Mac-1, but not alpha 3 beta 1, alpha 4 beta 1 or LFA-1. Embo J 11(2):405-10.
Brow DA. 2002. Allosteric cascade of spliceosome activation. Annual Review of Genetics. p 333-360.
Burset M, Seledtsov IA, Solovyev VV. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Research 28(21):4364-4375.
Butta N, Arias-Salgado EG, Gonzalez-Manchon C, Ferrer M, Larrucea S, Ayuso MS, Parrilla R. 2003. Disruption of the beta3 663-687 disulfide bridge confers constitutive activity to beta3 integrins. Blood 102(7):2491-7.
Calderwood DA. 2003. Integrin [beta] cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling. Proc. Natl Acad. Sci. USA 100:2272-2277.
Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3(4):285-98.
Cary LA, Han DC, Guan JL. 1999. Integrin-mediated signal transduction pathways. Histol. Histopathol.(14):1001-1009.
Caswell PT, Norman JC. 2006. Integrin trafficking and the control of cell migration. Traffic 7(1):14-21.
Chen S, Alon R, Fuhlbrigge RC, Springer TA. 1997. Rolling and transient tethering of leukocytes on antibodies reveal specializations of selectins. Proc Natl Acad Sci U S A 94(7):3172-7.
Cheng M, Foo SY, Shi ML, Law SA. 2008. Conformational States of LFA-1 and Its Adhesion to the Intercellular Molecules (ICAM)-1 and ICAM-3 (unpublished).
Cheng M, Foo SY, Shi ML, Tang RH, Kong LS, Law SK, Tan SM. 2007. Mutation of a conserved asparagine in the I-like domain promotes constitutively active integrins alphaLbeta 2 and alphaIIbbeta3. J Biol Chem 282(25):18225-32.
Cher THB, Chan HS, Klein GF, Jabkowski J, Schadenbock-Kranzl G, Zach O, Roca X, Law SK. 2011. A novel 3' splice-site mutation and a novel gross deletion in leukocyte adhesion deficiency (LAD)-1. Biochem Biophys Res Commun 404(4):1099-104.
Chien W, Pei L. 2000. A novel binding factor facilitates nuclear translocation and transcriptional activation function of the pituitary tumor-transforming gene product. J Biol Chem 275(25): 19422-7.
Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, Ghosh G. 2011. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc Natl Acad Sci U S A 108(20):8233-8.
Chua K, Reed R. 2001. An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol Cell Biol 21(5):1509-14.
Clark EA, Brugge JS. 1995. Integrins and signal transduction pathways: the road taken. Science 268(5208):233-239.

Collins L, Penny D. 2005. Complex Spliceosomal Organization Ancestral to Extant Eukaryotes. Molecular Biology and Evolution 22(4):1053-1066.
Corbi AL, Larson RS, Kishimoto TK, Springer TA, Morton CC. 1988. Chromosomal location of the genes encoding the leukocyte adhesion receptors LFA-1, Mac-1 and p150,95. Identification of a gene cluster involved in cell adhesion. J Exp Med 167(5):1597-607.
Corrionero A, Raker VA, Izquierdo JM, Valcárcel J. 2011. Strict $3^{\prime}$ splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease. RNA 17(3):401-411.
Crowley CA, Curnutte JT, Rosin RE, Andre-Schwartz J, Gallin JI, Klempner M, Snyderman R, Southwick FS, Stossel TP, Babior BM. 1980. An inherited abnormality of neutrophil adhesion. Its genetic transmission and its association with a missing protein. N Engl J Med 302(21):1163-8.
Csordas G, Renken C, Varnai P, Walter L, Weaver D, Buttle KF, Balla T, Mannella CA, Hajnoczky G. 2006. Structural and functional features and significance of the physical linkage between ER and mitochondria. J Cell Biol 174(7):915-21.
Das R, Zhou Z, Reed R. 2000. Functional association of U2 snRNP with the ATPindependent spliceosomal complex E. Mol Cell 5(5):779-87.
den_Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. Hum Mutat 15(1):7-12.
Dong Kim K, Zhao J, Auh S, Yang X, Du P, Tang H, Fu Y-X. 2007. Adaptive immune cells temper initial innate responses. Nat Med 13(10):1248-1252.
Etzioni A, Frydman M, Pollack S, Avidor I, Phillips ML, Paulson JC, Gershoni-Baruch R. 1992. Brief report: recurrent severe infections caused by a novel leukocyte adhesion deficiency. N Engl J Med 327(25):1789-92.
Etzioni A, Gershoni-Baruch R, Pollack S, Shehadeh N. 1998. Leukocyte adhesion deficiency type II: long-term follow-up. J Allergy Clin Immunol 102(2):323-4.
Fabbri M, Di Meglio S, Gagliani MC, Consonni E, Molteni R, Bender JR, Tacchetti C, Pardi R. 2005. Dynamic Partitioning into Lipid Rafts Controls the Endo-Exocytic Cycle of the \{alpha\}L/\{beta\}2 Integrin (LFA-1) during Leukocyte Chemotaxis. Mol. Biol. Cell:E05-05-0413.
Fagerholm SC, Varis M, Stefanidakis M, Hilden TJ, Gahmberg CG. 2006. alpha-Chain phosphorylation of the human leukocyte CD11b/CD18 (Mac-1) integrin is pivotal for integrin activation to bind ICAMs and leukocyte extravasation. Blood 108(10):337986.

Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. Genes \& Development 17(4):419-437.
Feizi T. 2000. Carbohydrate-mediated recognition systems in innate immunity. Immunol Rev 173:79-88.
Fiorini M, Piovani G, Schumacher RF, Magri C, Bertini V, Mazzolari E, Notarangelo L, Notarangelo LD, Barlati S. 2009. ITGB2 mutation combined with deleted ring 21 chromosome in a child with leukocyte adhesion deficiency. Journal of Allergy and Clinical Immunology 124(6):1356-1358.
Fischer A, Haddad E, Jabado N, Casanova J, Blanche S, Le Deist F, Cavazzana-Calvo M. 1998. Stem cell transplantation for immunodeficiency. Springer Seminars in Immunopathology 19(4):479-492.
Friend K, Kolev NG, Shu MD, Steitz JA. 2008. Minor-class splicing occurs in the nucleus of the Xenopus oocyte. Rna 14(8):1459-62.
Gahmberg CG, Fagerholm SC, Nurmi SM, Chavakis T, Marchesan S, Gronholm M. 2009. Regulation of integrin activity and signalling. Biochim Biophys Acta 1790(6):431-44.

Garcia-Alvarez B. 2003. Structural determinants of integrin recognition by talin. Mol. Cell 11:49-58.
Gething MJ, McCammon K, Sambrook J. 1986. Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. Cell 46(6):93950.

Giancotti FG. 1997. Integrin signaling: specificity and control of cell survival and cell cycle progression. Curr Opin Cell Biol 9(5):691-700.
Gingras AR. 2008. The structure of the C-terminal actin-binding domain of talin. EMBO J. 27:458-469.
Gompertz S, Stockley RA. 2000. Inflammation--role of the neutrophil and the eosinophil. Semin Respir Infect 15(1):14-23.
Gozani O, Feld R, Reed R. 1996. Evidence that sequence-independent binding of highly conserved U 2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes and Development 10(2):233-243.
Guth S, Tange TO, Kellenberger E, Valcarcel J. 2001. Dual Function for U2AF35 in AGDependent Pre-mRNA Splicing. Mol. Cell. Biol. 21(22):7673-7681.
Guth S, Valcarcel J. 2000. Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. J Biol Chem 275(48):38059-66.
Hall SL, Padgett RA. 1996. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science 271(5256):1716-8.
Harris ES, McIntyre TM, Prescott SM, Zimmerman GA. 2000. The leukocyte integrins. J Biol Chem 275(31):23409-12.
Hayashi T, Su TP. 2007. Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate $\mathrm{Ca}(2+)$ signaling and cell survival. Cell 131 (3):596-610.
Hayward AR, Leonard J, Wood CBS, Harvey BAM, Greenwood MC, Soothill JF. 1979. Delayed separation of the umbilical cord, widespread infections and defective neutrophil mobility. The Lancet 313(8126):1099-1101.
Hidalgo A, Ma S, Peired AJ, Weiss LA, Cunningham-Rundles C, Frenette PS. 2003. Insights into leukocyte adhesion deficiency type 2 from a novel mutation in the GDP-fucose transporter gene. Blood 101(5):1705-12.
Hixson P, Smith CW, Shurin SB, Tosi MF. 2004. Unique CD18 mutations involving a deletion in the extracellular stalk region and a major truncation of the cytoplasmic domain in a patient with leukocyte adhesion deficiency type 1. Blood 103(3):11051113.

Hogg N, Stewart MP, Scarth SL, Newton R, Shaw JM, Law SA, Klein N. 1999. A novel leukocyte adhesion deficiency caused by expressed but nonfunctional $\hat{I}^{2} 2$ integrins Mac-1 and LFA-1. J Clin Invest 103(1):97-106.
Horwitz A, Duggan K, Buck C, Beckerle MC, Burridge K. 1986. Interaction of plasma membrane fibronectin receptor with talin--a transmembrane linkage. Nature 320(6062):531-3.
Huang C, Lu C, Springer TA. 1997. Folding of the conserved domain but not of flanking regions in the integrin beta 2 subunit requires association with the alpha subunit. Proc Natl Acad Sci U S A 94(7):3156-61.
Huang C, Springer TA. 1997. Folding of the beta-propeller domain of the integrin alphaL subunit is independent of the I domain and dependent on the beta2 subunit. Proc Natl Acad Sci U S A 94(7):3162-7.
Huth JR, Olejniczak ET, Mendoza R, Liang H, Harris EA, Lupher ML, Jr., Wilson AE, Fesik SW, Staunton DE. 2000. NMR and mutagenesis evidence for an I domain allosteric
site that regulates lymphocyte function-associated antigen 1 ligand binding. Proc Natl Acad Sci U S A 97(10):5231-6.
Hynes RO. 1992. Integrins: Versatility, modulation, and signaling in cell adhesion Cell 69(1):11-25.
Hynes RO. 2002. Integrins: bidirectional, allosteric signaling machines. Cell 110(6):673-87.
Hyun YM, Lefort CT, Kim M. 2009. Leukocyte integrins and their ligand interactions. Immunol Res.
Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. 2005. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. Molecular \& Cellular Proteomics 4(9):1265-1272.
Ishikawa H, Heaney AP, Yu R, Horwitz GA, Melmed S. 2001. Human Pituitary TumorTransforming Gene Induces Angiogenesis. J Clin Endocrinol Metab 86(2):867-874.
Isken O, Maquat LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev 21(15):1833-56.
Issekutz TB. 1995. In vivo blood monocyte migration to acute inflammatory reactions, IL-1 alpha, TNF-alpha, IFN-gamma, and C5a utilizes LFA-1, Mac-1, and VLA-4. The relative importance of each integrin. The Journal of Immunology 154(12):6533-6540.
Katagiri K, Hattori M, Minato N, Irie S-k, Takatsu K, Kinashi T. 2000. Rap1 Is a Potent Activation Signal for Leukocyte Function-Associated Antigen 1 Distinct from Protein Kinase C and Phosphatidylinositol-3-OH Kinase. Molecular and Cellular Biology 20(6):1956-1969.
Katagiri K, Ohnishi N, Kabashima K, Iyoda T, Takeda N, Shinkai Y, Inaba K, Kinashi T. 2004. Crucial functions of the Rap1 effector molecule RAPL in lymphocyte and dendritic cell trafficking. Nat Immunol 5(10):1045-51.
Kent OA, Reayi A, Foong L, Chilibeck KA, MacMillan AM. 2003. Structuring of the 3' splice site by U2AF65. J Biol Chem 278(50):50572-7.
Kieffer JD, Plopper G, Ingber DE, Hartwig JH, Kupper TS. 1995. Direct binding of F actin to the cytoplasmic domain of the alpha 2 integrin chain in vitro. Biochem Biophys Res Commun 217(2):466-74.
Kim M, Carman CV, Springer TA. 2003. Bidirectional Transmembrane Signaling by Cytoplasmic Domain Separation in Integrins. Science 301(5640):1720-1725.
Kim PS, Arvan P. 1995. Calnexin and BiP act as sequential molecular chaperones during thyroglobulin folding in the endoplasmic reticulum. The Journal of cell biology 128(1-2):29-38.
Kishimoto TK, O'Conner K, Springer TA. 1989. Leukocyte adhesion deficiency. Aberrant splicing of a conserved integrin sequence causes a moderate deficiency phenotype. J Biol Chem 264(6):3588-95.
Knezevic I, Leisner TM, Lam SCT. 1996. Direct Binding of the Platelet Integrin aIIbB3 (GPIIb-IIIa) to Talin. Journal of Biological Chemistry 271(27):16416-16421.
Kohl S, Springer TA, Schmalstieg FC, Loo LS, Anderson DC. 1984. Defective natural killer cytotoxicity and polymorphonuclear leukocyte antibody-dependent cellular cytotoxicity in patients with LFA-1/OKM-1 deficiency. J Immunol 133(6):2972-8.
Konarska MM, Vilardell J, Query CC. 2006. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Molecular Cell 21(4):543-553.
Kozutsumi Y, Segal M, Normington K, Gething M-J, Sambrook J. 1988. The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucoseregulated proteins. Nature 332(6163):462-464.
Královicová J, Christensen MB, Vorechovsky I. 2005. Biased exon/intron distribution of cryptic and de novo $3^{\prime}$ splice sites. Nucleic acids research 33:4882-4898.

Královicová J, Vorechovsky I. 2007. Global control of aberrant splice-site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. Nucl. Acids Res. 35(19):6399-6413.
Kuijpers TW, van de Vijver E, Weterman MA, de Boer M, Tool AT, van den Berg TK, Moser M, Jakobs ME, Seeger K, Sanal O and others. 2009. LAD-1/variant syndrome is caused by mutations in FERMT3. Blood 113(19):4740-6.
Kunkel EJ, Butcher EC. 2002. Chemokines and the tissue-specific migration of lymphocytes. Immunity 16(1):1-4.
Kurzinger K, Reynolds T, Germain RN, Davignon D, Martz E, Springer TA. 1981. A novel lymphocyte function-associated antigen (LFA-1): cellular distribution, quantitative expression, and structure. The Journal of Immunology 127(2):596-602.
Lafuente E, Boussiotis VA. 2006. Rap1 regulation of RIAM and cell adhesion. Methods Enzymol 407:345-58.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W and others. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860-921.
Larsson BM, Larsson K, Malmberg P, Palmberg L. 1999. Gram positive bacteria induce IL-6 and IL-8 production in human alveolar macrophages and epithelial cells. Inflammation 23(3):217-30.
Laudanna C, Kim JY, Constantin G, Butcher E. 2002. Rapid leukocyte integrin activation by chemokines. Immunol Rev 186:37-46.
Lee H-S, Lim CJ, Puzon-McLaughlin W, Shattil SJ, Ginsberg MH. 2009. RIAM Activates Integrins by Linking Talin to Ras GTPase Membrane-targeting Sequences. Journal of Biological Chemistry 284(8):5119-5127.
Lee JO, Rieu P, Arnaout MA, Liddington R. 1995. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell 80(4):631-8.
Leitinger B, Hogg N. 2000. Effects of I domain deletion on the function of the beta2 integrin lymphocyte function-associated antigen-1. Mol Biol Cell 11(2):677-90.
Ley K, Laudanna C, Cybulsky MI, Nourshargh S. 2007. Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nature Reviews Immunology 7(9):678-689.
Liu S, Calderwood DA, Ginsberg MH. 2000. Integrin cytoplasmic domain-binding proteins. J Cell Sci 113(20):3563-3571.
Long EO. 2011. ICAM-1: Getting a Grip on Leukocyte Adhesion. The Journal of Immunology 186(9):5021-5023.
Lopez-Bigas N, Audit B, Ouzounis C, Parra G, Guigo R. 2005. Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett 579(9):1900-3.
Lu C, Ferzly M, Takagi J, Springer TA. 2001. Epitope mapping of antibodies to the Cterminal region of the integrin beta 2 subunit reveals regions that become exposed upon receptor activation. J Immunol 166(9):5629-37.
Luhn K, Marquardt T, Harms E, Vestweber D. 2001. Discontinuation of fucose therapy in LADII causes rapid loss of selectin ligands and rise of leukocyte counts. Blood 97(1):330-332.
Luo BH, Carman CV, Springer TA. 2007. Structural basis of integrin regulation and signaling. Annu Rev Immunol 25:619-47.
Luo BH, Takagi J, Springer TA. 2004. Locking the beta3 integrin I-like domain into high and low affinity conformations with disulfides. J Biol Chem 279(11):10215-21.
Ma YQ, Qin J, Wu C, Plow EF. 2008. Kindlin-2 (Mig-2): a co-activator of [beta]3 integrins. J. Cell Biol. 181:439-446.
Malech HL, Nauseef WM. 1997. Primary inherited defects in neutrophil function: etiology and treatment. Semin Hematol 34(4):279-90.

Malinin NL, Zhang L, Choi J, Ciocea A, Razorenova O, Ma YQ, Podrez EA, Tosi M, Lennon DP, Caplan AI and others. 2009. A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans. Nat Med 15(3):313-8.
Mannella CA, Marko M, Penczek P, Barnard D, Frank J. 1994. The internal compartmentation of rat-liver mitochondria: tomographic study using the high-voltage transmission electron microscope. Microsc Res Tech 27(4):278-83.
Marlin SD, Morton CC, Anderson DC, Springer TA. 1986. LFA-1 immunodeficiency disease. Definition of the genetic defect and chromosomal mapping of alpha and beta subunits of the lymphocyte function-associated antigen 1 (LFA-1) by complementation in hybrid cells. J Exp Med 164(3):855-67.
Martz E. 1987. LFA-1 and other accessory molecules functioning in adhesions of T and B lymphocytes. Hum Immunol 18(1):3-37.
Mathew EC, Shaw JM, Bonilla FA, Law SK, Wright DA. 2000. A novel point mutation in CD18 causing the expression of dysfunctional CD11/CD18 leucocyte integrins in a patient with leucocyte adhesion deficiency (LAD). Clin Exp Immunol 121(1):133-8.
Matsuura S, Kishi F, Tsukahara M, Nunoi H, Matsuda I, Kobayashi K, Kajii T. 1992. Leukocyte adhesion deficiency: identification of novel mutations in two Japanese patients with a severe form. Biochem Biophys Res Commun 184(3):1460-7.
McCabe CJ, Boelaert K, Tannahill LA, Heaney AP, Stratford AL, Khaira JS, Hussain S, Sheppard MC, Franklyn JA, Gittoes NJL. 2002. Vascular Endothelial Growth Factor, Its Receptor KDR/Flk-1, and Pituitary Tumor Transforming Gene in Pituitary Tumors. J Clin Endocrinol Metab 87(9):4238-4244.
McDowall A, Inwald D, Leitinger B, Jones A, Liesner R, Klein N, Hogg N. 2003. A novel form of integrin dysfunction involving B1, B2, and B3 integrins. The Journal of Clinical Investigation 111(1):51-60.
McEver RP. 1997. Selectin-carbohydrate interactions during inflammation and metastasis. Glycoconj J 14(5):585-91.
Montanez E. 2008. Kindlin-2 controls bidirectional signaling of integrins. Genes Dev. 22:1325-1330.
Moore MJ. 2000. Intron recognition comes of AGe. Nat Struct Biol 7(1):14-6.
Moser M, Bauer M, Schmid S, Ruppert R, Schmidt S, Sixt M, Wang HV, Sperandio M, Fassler R. 2009. Kindlin-3 is required for beta2 integrin-mediated leukocyte adhesion to endothelial cells. Nat Med 15(3):300-5.
Mould AP, Symonds EJH, Buckley PA, Grossmann JG, McEwan PA, Barton SJ, Askari JA, Craig SE, Bella J, Humphries MJ. 2003. Structure of an Integrin-Ligand Complex Deduced from Solution X-ray Scattering and Site-directed Mutagenesis. J. Biol. Chem. 278(41):39993-39999.
Murphy KM, Travers P, Walport M. 2007. Janeway's Immunobiology: Garland Science.
Myhill N, Lynes EM, Nanji JA, Blagoveshchenskaya AD, Fei H, Carmine Simmen K, Cooper TJ, Thomas G, Simmen T. 2008. The subcellular distribution of calnexin is mediated by PACS-2. Mol Biol Cell 19(7):2777-88.
Navab M, Hough GP, Stevenson LW, Drinkwater DC, Laks H, Fogelman AM. 1988. Monocyte migration into the subendothelial space of a coculture of adult human aortic endothelial and smooth muscle cells. The Journal of Clinical Investigation 82(6):1853-1863.
Nelson C, Rabb H, Arnaout MA. 1992. Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing and a missense mutation in a conserved region of CD18 impair cell surface expression of beta 2 integrins. J. Biol. Chem. 267(5):33513357.

Nelson KK, Green MR. 1989. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev 3(10):1562-71.
Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature 463(7280):457-63.
Nurmi SM, Gahmberg CG, Fagerholm SC. 2006. 14-3-3 Proteins Bind Both Filamin and $\alpha \mathrm{L} \beta 2$ Integrin in Activated T Cells. Annals of the New York Academy of Sciences 1090(1):318-325.
Ochman H, Gerber AS, Hartl DL. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120(3):621-623.
Ostermann G, Weber KS, Zernecke A, Schroder A, Weber C. 2002. JAM-1 is a ligand of the beta(2) integrin LFA-1 involved in transendothelial migration of leukocytes. Nat Immunol 3(2):151-8.
Pagani F, Baralle FE. 2004. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5(5):389-396.
Pagani F, Raponi M, Baralle FE. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A 102(18):6368-72.
Pagenstecher C, Wehner M, Friedl W, Rahner N, Aretz S, Friedrichs N, Sengteller M, Henn W, Buettner R, Propping P and others. 2006. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. Hum Genet 119(1-2):9-22.
Parvaneh N, Mamishi S, Rezaei A, Rezaei N, Tamizifar B, Parvaneh L, Sherkat R, Ghalehbaghi B, Kashef S, Chavoshzadeh Z and others. 2010. Characterization of 11 new cases of leukocyte adhesion deficiency type 1 with seven novel mutations in the ITGB2 gene. J Clin Immunol 30(5):756-60.
Pasvolsky R, Feigelson SW, Kilic SS, Simon AJ, Tal-Lapidot G, Grabovsky V, Crittenden JR, Amariglio N, Safran M, Graybiel AM and others. 2007. A LAD-III syndrome is associated with defective expression of the Rap-1 activator CalDAG-GEFI in lymphocytes, neutrophils, and platelets. J Exp Med 204(7): 1571-82.
Pfeffer K. 2003. Biological functions of tumor necrosis factor cytokines and their receptors. Cytokine Growth Factor Rev 14(3-4):185-91.
Plow EF, Haas TA, Zhang L, Loftus J, Smith JW. 2000. Ligand binding to integrins. J Biol Chem 275(29):21785-8.
Pluskota E, Soloviev DA, Szpak D, Weber C, Plow EF. 2008. Neutrophil apoptosis: selective regulation by different ligands of integrin alphaMbeta2. J Immunol 181(5):3609-19.
Ponting CP, Schultz J, Copley RR, Andrade MA, Bork P. 2000. Evolution of domain families. Adv Protein Chem 54:185-244.
Qasim W, Cavazzana-Calvo M, Davies EG, Davis J, Duval M, Eames G, Farinha N, Filopovich A, Fischer A, Friedrich W and others. 2009. Allogeneic Hematopoietic Stem-Cell Transplantation for Leukocyte Adhesion Deficiency. Pediatrics 123(3):836-840.
Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4(3):311-23.
Renna M, Caporaso MG, Bonatti S, Kaufman RJ, Remondelli P. 2007. Regulation of ERGIC-53 Gene Transcription in Response to Endoplasmic Reticulum Stress. Journal of Biological Chemistry 282(31):22499-22512.
Roberts M, Barry S, Woods A, van der Sluijs P, Norman J. 2001. PDGF-regulated rab4dependent recycling of alphavbeta3 integrin from early endosomes is necessary for cell adhesion and spreading. Curr Biol 11(18):1392-402.
Roca X, Olson AJ, Rao AR, Enerly E, Kristensen VN, Burresen-Dale A-L, Andresen BS, Krainer AR, Sachidanandam R. 2008. Features of 5' splice-site efficiency derived
from disease-causing mutations and comparative genomics. Genome Research 18(1):77-87.
Roca X, Sachidanandam R, Krainer AR. 2003. Intrinsic differences between authentic and cryptic 5' splice sites. Nucl. Acids Res. 31(21):6321-6333.
Roos D, Law SKA. 2001. Hematologically important mutations: Leukocyte Adhesion Deficiency. Blood Cells, Molecules, and Diseases 27(6):1000-1004.
Roos D, Meischl C, de Boer M, Simsek S, Weening RS, Sanal Ö, Tezcan I, Güngör T, Law SKA. 2002. Genetic analysis of patients with leukocyte adhesion deficiency: Genomic sequencing reveals otherwise undetectable mutations. Experimental Hematology 30(3):252-261.
Ross GD, Thompson RA, Walport MJ, Springer TA, Watson JV, Ward RH, Lida J, Newman SL, Harrison RA, Lachmann PJ. 1985. Characterization of patients with an increased susceptibility to bacterial infections and a genetic deficiency of leukocyte membrane complement receptor type 3 and the related membrane antigen LFA-1. Blood 66(4):882-90.
Rothlein R, Dustin ML, Marlin SD, Springer TA. 1986. A human intercellular adhesion molecule (ICAM-1) distinct from LFA-1. J Immunol 137(4):1270-4.
Ruskin B, Zamore PD, Green MR. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52(2):207-19.
Sanchez-Madrid F, Nagy JA, Robbins E, Simon P, Springer TA. 1983. A human leukocyte differentiation antigen family with distinct alpha-subunits and a common beta-subunit: the lymphocyte function-associated antigen (LFA-1), the C3bi complement receptor (OKM1/Mac-1), and the p150,95 molecule. J Exp Med 158(6):1785-1803.
Schleiffenbaum B, Moser R, Patarroyo M, Fehr J. 1989. The cell surface glycoprotein Mac-1 (CD11b/CD18) mediates neutrophil adhesion and modulates degranulation independently of its quantitative cell surface expression. J Immunol 142(10):3537-45.
Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell 11(4):965-76.
Sharp PA, Burge CB. 1997. Classification of introns: U2-type or U12-type. Cell 91(7):875-9.
Shaw JM, Al-Shamkhani A, Boxer LA, Buckley CD, Dodds AW, Klein N, Nolan SM, Roberts I, Roos D, Scarth SL and others. 2001. Characterization of four CD18 mutants in leucocyte adhesion deficient (LAD) patients with differential capacities to support expression and function of the CD11/CD18 integrins LFA-1, Mac-1 and p150,95. Clin Exp Immunol 126(2):311-8.
Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. 2006. Comprehensive splice-site analysis using comparative genomics. Nucl. Acids Res. 34(14):3955-3967.
Shi M, Foo SY, Tan SM, Mitchell EP, Law SK, Lescar J. 2007a. A Structural Hypothesis for the Transition between Bent and Extended Conformations of the Leukocyte beta2 Integrins. J Biol Chem 282(41):30198-206.
Shi M, Sundramurthy K, Liu B, Tan S-M, Law SKA, Lescar J. 2005. The Crystal Structure of the Plexin-Semaphorin-Integrin Domain/Hybrid Domain/I-EGF1 Segment from the Human Integrin \{beta\}2 Subunit at 1.8-A Resolution. J. Biol. Chem. 280(34):3058630593.

Shi X, Ma Y-Q, Tu Y, Chen K, Wu S, Fukuda K, Qin J, Plow EF, Wu C. 2007b. The MIG2/Integrin Interaction Strengthens Cell-Matrix Adhesion and Modulates Cell Motility. Journal of Biological Chemistry 282(28):20455-20466.

Sigal A, Bleijs DA, Grabovsky V, van Vliet SJ, Dwir O, Figdor CG, van Kooyk Y, Alon R. 2000. The LFA-1 integrin supports rolling adhesions on ICAM-1 under physiological shear flow in a permissive cellular environment. J Immunol 165(1):442-52.
Silva TA, Lara VS, Silva JS, Oliveira SHP, Butler WT, Cunha FQ. 2005. Macrophages and Mast Cells Control the Neutrophil Migration Induced by Dentin Proteins. Journal of Dental Research 84(1):79-83.
Smith SJ, McCann RO. 2007. A C-terminal dimerization motif is required for focal adhesion targeting of Talin1 and the interaction of the Talin1 I/LWEQ module with F-actin. Biochemistry 46:10886-10898.
Soares LMM, Zanier K, Mackereth C, Sattler M, Valcorcel J. 2006. Intron Removal Requires Proofreading of U2AF/3' Splice Site Recognition by DEK. Science 312(5782):19611965.

Solomon E, Palmer RW, Hing S, Law SKA. 1988. Regional localization of CD18, the $\beta$ subunit of the cell surface adhesion molecule LFA-1, on human chromosome 21 by in situ hybridization. Annals of Human Genetics 52(2):123-128.
Soltys BJ, Gupta RS. 1992. Interrelationships of endoplasmic reticulum, mitochondria, intermediate filaments, and microtubules--a quadruple fluorescence labeling study. Biochem Cell Biol 70(10-11):1174-86.
Springer T, Galfre G, Secher DS, Milstein C. 1979. Mac-1: a macrophage differentiation antigen identified by monoclonal antibody. Eur J Immunol 9(4):301-6.
Springer TA. 1990. Adhesion receptors of the immune system. Nature 346(6283):425-434.
Springer TA, Miller LJ, Anderson DC. 1986. p150,95, the third member of the Mac-1, LFA-1 human leukocyte adhesion glycoprotein family. J Immunol 136(1):240-5.
Springer TA, Thompson WS, Miller LJ, Schmalstieg FC, Anderson DC. 1984. Inherited deficiency of the Mac-1, LFA-1, p150,95 glycoprotein family and its molecular basis. J Exp Med 160(6):1901-18.
Springer TA, Wang JH. 2004. The three-dimensional structure of integrins and their ligands, and conformational regulation of cell adhesion. Adv Protein Chem 68:29-63.
Stein JV, Rot A, Luo Y, Narasimhaswamy M, Nakano H, Gunn MD, Matsuzawa A, Quackenbush EJ, Dorf ME, von Andrian UH. 2000. The CC chemokine thymusderived chemotactic agent 4 (TCA-4, secondary lymphoid tissue chemokine, 6Ckine, exodus-2) triggers lymphocyte function-associated antigen 1-mediated arrest of rolling T lymphocytes in peripheral lymph node high endothelial venules. J Exp Med 191(1):61-76.
Svensson L, Howarth K, McDowall A, Patzak I, Evans R, Ussar S, Moser M, Metin A, Fried M, Tomlinson I and others. 2009. Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation. Nat Med 15(3):306-312.
Tadokoro S. 2003. Talin binding to integrin [beta] tails: a final common step in integrin activation. Science 302:103-106.
Takagi J, Beglova N, Yalamanchili P, Blacklow SC, Springer TA. 2001. Definition of EGFlike, closely interacting modules that bear activation epitopes in integrin beta subunits. PNAS 98(20):11175-11180.
Takala H, Nurminen E, Nurmi SM, Aatonen M, Strandin T, Takatalo M, Kiema T, Gahmberg CG, Ylanne J, Fagerholm SC. 2008. Beta2 integrin phosphorylation on Thr758 acts as a molecular switch to regulate 14-3-3 and filamin binding. Blood 112(5):1853-62.
Tan SM, Hyland RH, Al-Shamkhani A, Douglass WA, Shaw JM, Law SK. 2000. Effect of integrin beta 2 subunit truncations on LFA-1 (CD11a/CD18) and Mac-1 (CD11b/CD18) assembly, surface expression, and function. J Immunol 165(5):257481.

Tan SM, Robinson MK, Drbal K, van Kooyk Y, Shaw JM, Law SK. 2001a. The N-terminal region and the mid-region complex of the integrin beta 2 subunit. J Biol Chem 276(39):36370-6.
Tan SM, Walters SE, Mathew EC, Robinson MK, Drbal K, Shaw JM, Law SK. 2001b. Defining the repeating elements in the cysteine-rich region (CRR) of the CD18 integrin beta 2 subunit. FEBS Lett 505(1):27-30.
Tanentzapf G, Brown NH. 2006. An interaction between integrin and the talin FERM domain mediates integrin activation but not linkage to the cytoskeleton. Nature Cell Biol. 8:601-606.
Tang RH, Tng E, Law SK, Tan SM. 2005. Epitope mapping of monoclonal antibody to integrin alphaL beta2 hybrid domain suggests different requirements of affinity states for intercellular adhesion molecules (ICAM)-1 and ICAM-3 binding. J Biol Chem 280(32):29208-16.
Tanriver Y, Martin-Fontecha A, Ratnasothy K, Lombardi G, Lechler R. 2009. SuperantigenActivated Regulatory T Cells Inhibit the Migration of Innate Immune Cells and the Differentiation of Naive T Cells. The Journal of Immunology 183(5):2946-2956.
Tarn WY, Steitz JA. 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell 84(5):801-11.
Thanaraj TA, Clark F. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res 29(12):2581-93.
Tiwari S, Askari JA, Humphries MJ, Bulleid NJ. 2011. Divalent cations regulate the folding and activation status of integrins during their intracellular trafficking. Journal of Cell Science 124(10):1672-1680.
Tng E, Tan SM, Ranganathan S, Cheng M, Law SK. 2004. The integrin alpha L beta 2 hybrid domain serves as a link for the propagation of activation signal from its stalk regions to the I-like domain. J Biol Chem 279(52):54334-9.
Tone Y, Wada T, Shibata F, Toma T, Hashida Y, Kasahara Y, Koizumi S, Yachie A. 2007. Somatic revertant mosaicism in a patient with leukocyte adhesion deficiency type 1. Blood 109(3):1182-1184.
Tsai YC, Lee WI, Huang JL, Hung IJ, Jaing TH, Yao TC, Chen MT, Kuo ML. 2008. Neutrophil function and molecular analysis in severe leukocyte adhesion deficiency type I without separation delay of the umbilical cord. Pediatr Allergy Immunol 19(1):25-32.
Tsukamoto H, Tousson A, Circolo A, Marchase RB, Volanakis JE. 2002. Calnexin is associated with and induced by overexpressed human complement protein C2. Anat Rec 267(1):7-16.
Uzel G, Tng E, Rosenzweig SD, Hsu AP, Shaw JM, Horwitz ME, Linton GF, Anderson SM, Kirby MR, Oliveira JB and others. 2008. Reversion mutations in patients with leukocyte adhesion deficiency type-1 (LAD-1). Blood 111(1):209-218.
Valcarcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with premRNA of branch point and promotion base pairing with U2 snRNA. Science 273(5282):1706-1709.
van de Stolpe A, van der Saag PT. 1996. Intercellular adhesion molecule-1. J Mol Med (Berl) 74(1):13-33.
van de Vijver E, Maddalena A, Sanal O, Holland SM, Uzel G, Madkaikar M, de Boer M, van Leeuwen K, Koker MY, Parvaneh N and others. 2012. Hematologically important mutations: Leukocyte adhesion deficiency (first update). Blood Cells, Molecules, and Diseases 48(1):53-61.

Van der Vieren M, Le Trong H, Wood CL, Moore PF, St John T, Staunton DE, Gallatin WM. 1995. A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3. Immunity 3(6):683-90.
Vararattanavech A, Lin X, Torres J, Tan S-M. 2009. Disruption of the Integrin aLB2 Transmembrane Domain Interface by B2 Thr-686 Mutation Activates aLB2 and Promotes Micro-clustering of the aL Subunits. Journal of Biological Chemistry 284(5):3239-3249.
Vorechovsky I. 2006. Aberrant $3^{\prime}$ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucl. Acids Res. 34(16):4630-4641.
Wada I, Kai M, Imai S, Sakane F, Kanoh H. 1997. Promotion of transferrin folding by cyclic interactions with calnexin and calreticulin. EMBO J 16(17):5420-5432.
Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 8(10):749-61.
Wang Z, Yu R, Melmed S. 2001. Mice lacking pituitary tumor transforming gene show testicular and splenic hypoplasia, thymic hyperplasia, thrombocytopenia, aberrant cell cycle progression, and premature centromere division. Mol Endocrinol 15(11):1870-9.
Weber KS, York MR, Springer TA, Klickstein LB. 1997. Characterization of lymphocyte function-associated antigen 1 (LFA-1)-deficient T cell lines: the alphaL and beta2 subunits are interdependent for cell surface expression. Journal of immunology (Baltimore, Md. : 1950) 158(1):273-9.
Weidenhammer EF, M Ruiz-Noriega, Woolford JL. 1997. Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae. Molecular and Cellular Biology 17(7):3580-3588.
Willett M, Pollard HJ, Vlasak M, Morley SJ. 2010. Localization of ribosomes and translation initiation factors to talin/beta3-integrin-enriched adhesion complexes in spreading and migrating mammalian cells. Biol Cell 102(5):265-76.
Williams DB. 2006. Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. J Cell Sci 119(4):615-623.
Wimmer K, Roca X, Beiglbock H, Callens T, Etzler J, Rao AR, Krainer AR, Fonatsch C, Messiaen L. 2007. Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5 ' splice-site disruption. Hum Mutat 28(6):599-612.
Wong DA, Davis EM, LeBeau M, Springer TA. 1996. Cloning and chromosomal localization of a novel gene-encoding a human beta 2-integrin alpha subunit. Gene 171(2):291-4.
Wright AH, Douglass WA, Taylor GM, Lau YL, Higgins D, Davies KA, Law SK. 1995. Molecular characterization of leukocyte adhesion deficiency in six patients. Eur J Immunol 25(3):717-22.
Wright SD, Rao PE, Van Voorhis WC, Craigmyle LS, Iida K, Talle MA, Westberg EF, Goldstein G, Silverstein SC. 1983. Identification of the C3bi receptor of human monocytes and macrophages by using monoclonal antibodies. Proc Natl Acad Sci U S A 80(18):5699-703.
Wu S, Romfo CM, Nilsen TW, Green MR. 1999. Functional recognition of the 3 ' splice site AG by the splicing factor U2AF35. Nature 402(6763):832-5.
Xiao T, Takagi J, Coller BS, Wang JH, Springer TA. 2004. Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. Nature 432(7013):59-67.
Xie C, Shimaoka M, Xiao T, Schwab P, Klickstein LB, Springer TA. 2004. The integrin alpha-subunit leg extends at a Ca2+-dependent epitope in the thigh/genu interface upon activation. Proc Natl Acad Sci U S A 101(43):15422-7.

Xie C, Zhu J, Chen X, Mi L, Nishida N, Springer TA. 2010. Structure of an integrin with an [alpha]I domain, complement receptor type 4. Embo J 29(3):666-679.
Xie J, Beickman K, Otte E, Rymond BC. 1998. Progression through the spliceosome cycle requires $\operatorname{Prp} 38$ p function for U4/U6 snRNA dissociation. EMBO J 17(10):2938-2946.
Xiong J-P, Stehle T, Goodman SL, Arnaout MA. 2003a. New insights into the structural basis of integrin activation. Blood 102(4):1155-1159.
Xiong J-P, Stehle T, Goodman SL, Arnaout MA. 2004. A Novel Adaptation of the Integrin PSI Domain Revealed from Its Crystal Structure. J. Biol. Chem. 279(39):4025240254.

Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL, Joachimiak A, Goodman SL, Arnaout MA. 2001. Crystal structure of the extracellular segment of integrin alpha Vbeta3. Science 294(5541):339-45.
Xiong JP, Stehle T, Goodman SL, Arnaout MA. 2003b. Integrins, cations and ligands: making the connection. J Thromb Haemost 1(7):1642-54.
Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA. 2002. Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science 296(5565):151-5.
Yalamanchili P, Lu C, Oxvig C, Springer TA. 2000. Folding and function of I domaindeleted Mac-1 and lymphocyte function-associated antigen-1. J Biol Chem 275(29):21877-82.
Yang W, Shimaoka M, Salas A, Takagi J, Springer TA. 2004. Intersubunit signal transmission in integrins by a receptor-like interaction with a pull spring. Proc Natl Acad Sci U S A 101(9):2906-11.
Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11(2-3):377-94.
Yoshie O. 2000. Role of chemokines in trafficking of lymphocytes and dendritic cells. Int J Hematol 72(4):399-407.
Yoshinaga IG, Dekker SK, Mihm MC, Jr., Byers HR. 1994. Differential effect of magnesium and calcium on integrin-mediated melanoma cell migration on type IV collagen and fibronectin. Melanoma Res 4(6):371-8.
Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature 355(6361):609-614.
Zell T, Kivens WJ, Kellermann SA, Shimizu Y. 1999. Regulation of integrin function by T cell activation: points of convergence and divergence. Immunol Res 20(2):127-45.
Zhang B, Cunningham MA, Nichols WC, Bernat JA, Seligsohn U, Pipe SW, McVey JH, Schulte-Overberg U, de Bosch NB, Ruiz-Saez A and others. 2003. Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex. Nat Genet 34(2):220225.

Zhang B, Kaufman RJ, Ginsburg D. 2005. LMAN1 and MCFD2 Form a Cargo Receptor Complex and Interact with Coagulation Factor VIII in the Early Secretory Pathway. Journal of Biological Chemistry 280(27):25881-25886.
Zhang MQ. 1998. Statistical features of human exons and their flanking regions. Hum Mol Genet 7(5):919-32.
Zheng C, Liu H-h, Yuan S, Zhou J, Zhang B. 2010. Molecular basis of LMAN1 in coordinating LMAN1-MCFD2 cargo receptor formation and ER-to-Golgi transport of FV/FVIII. Blood 116(25):5698-5706.
Zhu J, Luo B-H, Xiao T, Zhang C, Nishida N, Springer TA. 2008. Structure of a Complete Integrin Ectodomain in a Physiologic Resting State and Activation and Deactivation by Applied Forces. Molecular cell 32(6):849-861.

Zou H, McGarry TJ, Bernal T, Kirschner MW. 1999. Identification of a Vertebrate SisterChromatid Separation Inhibitor Involved in Transformation and Tumorigenesis. Science 285(5426):418-422.

## Appendix I: Experiments repeated for the LAD-I publication

## I. 1 Sequencing of the intron 2 cDNA product from patient cDNA

As described in Chapter 3, a 245 bp cDNA fragment was obtained together the expected 202 bp splicing product from PCR amplification of patient cDNA, using the primers F23051/R26792. Both PCR products were excised and sequenced using R26792. The 245 and 202 bp products corresponded to splicing at the -43 3'ss and the authentic 3 'ss (Figure I.1).


Figure I.1Reverse sequencing chromatograms of the 200 and 240 bp PCR products obtained from patient cDNA. (a) The 245 bp product contained a splicing mutation where 43 bp of the 3 ' end of intron 2 was spliced into the cDNA. The $\mathbf{C}$ to $A$ point mutation was observed at the $\mathbf{- 1 0}$ position. This CD18 allele coincided with the CTG codon at Leu ${ }^{8}$ as indicated in the diagram (b) The 202 bp product was the expected product, which coincided with the CTT codon at Leu ${ }^{8}$.

## I. 2 The $C$ to $A$ point mutation in intron 2 resulted in a splicing defect

Intron 2 of the patient's $I T G B 2$ gene was amplified and sequenced. Except for the C and A difference at position -10 , no other difference was found. The wildtype and mutant intron 2 were introduced into the CD18 cDNA expression plasmid and the two minigenes were annotated as CD18 I2-wt and CD18 I2-mut respectively. The plasmids with the CD18 minigenes were transfected into HEK293 cells. cDNA was prepared from the mRNA and analyzed (Figure I.2a). Most of the cDNA contain the unspliced intron although faint PCR products of spliced cDNA were detected. When the spliced fragments were purified and reamplified, the two fragments of the expected sizes of 245 bp and 202 bp for the mutant and wildtype minigenes were obtained (Figure I.2b). The larger fragment is generated by activation of a cryptic 3 'ss 43 nts upstream of the authentic 3 'ss. Cryptic splice sites are defined as splice sites that are only used when use of the authentic splice site is disrupted by mutation (Královicová, et al., 2005; Roca, et al., 2003). Use of the -43 cryptic 3'ss in CD18 intron 2 is not detected in the wild-type (CD18 I2-wt) minigene. In contrast, the -43 cryptic 3'ss is used exclusively in the -10 C to A mutant (CD18 I2-mut) minigene.

The two CD18 I2 plasmids were co-transfected with the CD11a, b, or cexpression plasmids into HEK293 cells. Expression of the three CD11/ CD18 integrin dimers can be detected by flow cytometry using the heterodimeric specific mAb IB4 (Wright, et al., 1983) on transfectants with the CD18 I2-wt plasmid, but not on those with the CD18 I2-mut plasmid (Figure I.3). These results showed that the HEK293 transfectants can splice the wildtype intron correctly, and that the C to A mutation at position -10 caused insertion of the -43 cryptic 3'ss.


Figure I. 2 Analysis of the $\mathbf{C}$ to A polymorphic mutation in intron 2 of CD18. Using a PCR based method, the entire mutant intron 2 sequence was inserted between the exon 2 to exon 3 junction of a CD18 expression plasmid to create the CD18-i2mut expression plasmid. Similarly, the entire wild type intron 2 sequence of CD18 was inserted into the corresponding region of a CD18 expression plasmid to create CD18-i2wt. These expression plasmids were transfected into HEK293 cells for RT-PCR analysis of the mRNA region between exon 2 and exon 3. (a) Initial amplification resulted in a 550 bp product corresponding to the unspliced exon 2 - intron 2 - exon 3 region. A faint 245 bp and 202 bp product were observed for the CD18-i2wt and CD18-i2mut. (b) Further amplification of the excised 245/202 bp product was performed. PCR products obtained corresponded to the expected 202 bp PCR product in CD18-i2wt, and a 245 bp PCR product containing an extra 43 bp from intron 2 in CD18-i2mut transfected cells, demonstrating conclusively that the 202 bp product was not present in CD18-i2mut.


Fluorescence Intensity

Figure I. 3 FACS analysis of the intron 2 mutation in transfectants. The CD18-i2mut and CD18-i2wt expression plasmids were co-transfected into HEK293 cells together with the CD11a, CD11b or CD11c expression plasmids for analysis of cell surface expression. Separately, a wild type CD18 expression plasmid (containing no inserted intronic region) was co-transfected with each $\alpha$-subunit expression plasmid as a positive control, while each $\alpha$-subunit expression plasmid was transfected individually as a negative control. FACS analysis was performed using CD11a, CD11b or CD11c specific antibodies (MHM24, LPM19C and KB43 respectively) and a CD18 dimer specific antibody (1B4). The results showed that the expression plasmid containing the wild type intronic region (CD18-i2wt) could be expressed as a dimer together with each $\alpha$-subunit, and the expression profile of each integrin dimer was similar to that of the positive control. In contrast, the CD18-i2mut expression plasmid could not support cell surface expression of integrin dimers in all cases.

## I. 3 Genomic DNA sequencing demonstrated that the patient was polymorphic at

 $V a l^{367}$ and Val ${ }^{441}$Using the primers F10 and R10 to amplify the genomic DNA region flanking exon 10, a 473 bp product was obtained in the patient (Figure I.4). Sequencing showed that the patient was polymorphic A/C polymorphism at $\mathrm{Val}^{367}$ (Figure I.5). This result was confirmed and further extended using the primers F10 and A8913C to amplify the genomic DNA region flanking exon 10 to exon 11 (Figure I.6); the patient had a $\mathrm{A} / \mathrm{C}$ polymorphism at $\mathrm{Val}^{367}$ and a $\mathrm{C} / \mathrm{T}$ polymorphism at Val ${ }^{441}$ (Figure I.7).

In contrast, the father was polymorphic $(\mathrm{A} / \mathrm{C})$ at $\mathrm{Val}^{367}$ and homozygous $(\mathrm{C})$ at $\mathrm{Val}^{441}$, while the mother was homozygous (C) at $\mathrm{Val}^{367}$ and polymorphic (C/T) at $\mathrm{Val}^{441}$. Using these results, it was possible to deduce that the parents and patient each possessed 2 CD18 alleles corresponding to these polymorphic markers.


Figure I. 4 Gel electrophoresis of the 473 bp PCR product obtained from genomic DNA using F10/R10. This product spanned exon 10 of CD18.


Figure I. 5 Reverse sequencing chromatogram of the F10/R10 PCR product obtained from genomic DNA. At $\mathrm{Val}^{367}$, the patient was polymorphic ( $\mathrm{A} / \mathrm{C}$ ), the father was polymorphic ( $\mathrm{A} / \mathrm{C}$ ), while the mother was not polymorphic (C).


Figure I. 6 Gel electrophoresis of the 1809 bp PCR product obtained from genomic DNA using F10/A8913C. This product spanned exon 10 to exon 11 of CD18.


Father Reverse Sequencing - A8913C


Figure I. 7 Forward and reverse sequencing chromatograms of the F10/A8913C PCR product obtained from genomic DNA. The patient had a $\mathbf{A} / \mathrm{C}$ polymorphism at $\mathrm{Val}^{367}$ and a C/T polymorphism at Val ${ }^{441}$. The father was polymorphic ( $\mathrm{A} / \mathrm{C}$ ) at $\mathrm{Val}^{367}$ and homozygous ( C ) at $\mathbf{V a l}^{441}$, while the mother was homozygous (C) at $\mathbf{V a l}^{367}$ and polymorphic (C/T) at Val ${ }^{41}$.

## I. 4 cDNA and genomic DNA PCR products extending beyond exon 11 in the father and patient were not polymorphic

The primers H9025/R37162 were used to amplify a region spanning exon 9 to exon 16 from cDNA. The PCR product from the patient was not polymorphic at both positions but instead had a C and T at $\mathrm{Val}^{367}$ and $\mathrm{Val}^{441}$ respectively (Figure I.11). However, when the cDNA PCR product flanked exon 6 to exon 11 (F9005/A8913C) (Figure I.12), the patient was polymorphic (Figure I.13). Based on the small dye peaks in the sequencing chromatogram, the patient $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele PCR product was weak, suggesting that the $\mathrm{C}\left(\mathrm{Val}^{367}\right)$ and $\mathrm{T}\left(\mathrm{Val}^{441}\right)$ allele was more abundant than that $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele at the mRNA level. As the H9025/R37162 PCR product was not polymorphic; a sizable genomic mutation may have occurred after exon 11 in the patient $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele. As discussed in Chapter 3, a large deletion was the mostly likely cause of the mutation.

As expected, the H9025/R37162 cDNA PCR product from the mother was polymorphic at Val $^{441}$ (Figure I.10); both alleles were amplified from cDNA spanning exon 9 to exon 16 in the mother.

With H9025/R37162, the father had a C at $\mathrm{Val}^{367}$ (Figure I.9) instead of the expected $\mathrm{A} / \mathrm{C}$ polymorphism, even though the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele was present in father genomic DNA (Figure I.13). Possibly, the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele was less abundant at the mRNA level, and was masked by the $\mathrm{C}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele.


Figure I. 8 Gel electrophoresis of the 1682 bp PCR product obtained from cDNA using H9025/R37162. This product spanned exon 9 to exon 16 of CD18.

Father Forward Sequencing - H9025


Figure I. 9 Forward and reverse sequencing chromatogram of the H9025/R37162 PCR product obtained from father cDNA. In the father, there was a $\mathbf{C}$ at $\mathrm{Val}^{367}$ and a C at $\mathrm{Val}^{441}$ in the PCR product spanning exon 9 to exon 16. As expected, the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele was not amplified.


Figure I.10 Forward and reverse sequencing chromatograms of the H9025/R37162 PCR product obtained from mother cDNA. In the mother, the PCR product spanning exon 9 to exon 16 was polymorphic (C/T) at $\mathrm{Val}^{441}$ and homozygous for C at $\mathrm{Val}^{\mathbf{3 6 7}}$. The polymorphism at $\mathrm{Val}^{441}$ demonstrated that both CD18 alleles in the mother were amplified.


Figure I.11 Forward and Reverse sequencing chromatograms of the H9025/R37162 PCR product obtained from patient cDNA. In the patient, there was a C at $\mathrm{Val}^{367}$ and a T at $\mathrm{Val}^{441}$ in the PCR product spanning exon 9 to exon 16. The polymorphisms detected in the F9005B/A8913C PCR product were not detected, suggesting that the cDNA product from the $A\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele contained a deletion somewhere after exon 11.


Figure I. 12 Gel electrophoresis of the 862 bp PCR product obtained from cDNA using F9005/A8913C. This product spanned exon 6 to exon 11 of CD18.


Patient Forward Sequencing - F9005B


Figure I. 13 Forward and reverse sequencing chromatograms of the $\mathrm{F9005B}$ /A8913C PCR product obtained from cDNA. The genomic DNA polymorphism detected in father genomic DNA could not be detected at the cDNA level. Instead, the father had a $\mathbf{C}$ at $\mathrm{Val}^{367}$ and a $\mathbf{C}$ at $\mathbf{V a l}^{441}$, suggesting that these 2 polymorphic markers were present in a normal CD18 allele, and that the $\mathrm{A}\left(\mathrm{Val}^{367}\right) \mathrm{C}\left(\mathrm{Val}^{441}\right)$ allele was masked. In the mother, there was a C at $\mathrm{Val}^{367}$ and a $\mathrm{C} / \mathrm{T}$ polymorphism at Val ${ }^{441}$, demonstrating that both CD18 alleles were amplified. In the patient, there was a $A / C$ and C/T polymorphism at each respective position. Hence, both copies of CD18 alleles were amplified in the patient.

## Appendix II: Protein functions of all proteins consistently identified in IP-LCMS

Fron 839 proteins identified in the IP-LCMS experiments combined, 127 hits were determined to be consistent hits, identified at least twice in each of the three groups of samples; t-CD18F (including t-CD18F/CD11a), CD18F, and CD11a (all CD18F/CD11a and CD11aF/CD18 co-transfected samples and CD11aF).

The proteins were individually analyzed and categorized into protein functions as described earlier. Protein functions were according to the annotations available on UniProt. The list of proteins and the known functions are presented in Table II.1. Furthermore, the emPAI scores for all the consistent hits identified is presented in Table II.2.

| Protein Name | Available Uniprot Annotated Functions |
| :---: | :---: |
| CD18F only |  |
| Chaperones |  |
| SCAMP3 protein | Functions in post-Golgi recycling pathways. Acts as a recycling carrier to the cell surface |
| DnaJ (Hsp40) homolog, subfamily A, member 1 (DNAJA1) | Co-chaperone of Hsc70. Seems to play a role in protein import into mitochondria. |
| HSP90AA1 protein | Molecular chaperone that promotes the maturation, structural maintenance and proper regulation of specific target proteins involved for instance in cell cycle control and signal transduction. Undergoes a functional cycle that is linked to its ATPase activity. This cycle probably induces conformational changes in the client proteins, thereby causing their activation. Interacts dynamically with various co-chaperones that modulate its substrate recognition, ATPase cycle and chaperone function. |
| Protein modification |  |
| Oligosaccharyltransferase complex subunit (OSTC) | Essential subunit of N -oligosaccharyl transferase enzyme which catalyzes the transfer of a high mannose oligosaccharide from a lipid-linked oligosaccharide donor to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains. Essential subunit of N -oligosaccharyl transferase enzyme which catalyzes the transfer of a high mannose oligosaccharide from a lipid-linked oligosaccharide donor to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains. |
| Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit (STT3B) |  |
| Glucosidase I | mannosyl-oligosaccharide glucosidase activity |
| Triosephosphate isomerase | Catalytic function D-glyceraldehyde 3-phosphate = glycerone phosphate. |
| CDP-diacylglycerol--inositol3phosphatidyltransferase (CDIPT) | Catalyzes the biosynthesis of phosphatidylinositol (Ptdlns) as well as PtdIns:inositol exchange reaction. May thus act to reduce an excessive cellular PtdIns content. The exchange activity is due to the reverse reaction of Ptdlns synthase and is dependent on CMP, which is tightly bound to the enzyme |
| Protein Binding |  |
| Ras-related protein Rab-35 | In the process of endocytosis, essential rate-limiting regulator of a fast recycling pathway back to the plasma membrane. During cytokinesis, required for the postfurrowing terminal steps, namely for intercellular bridge stability and abscission, possibly by controlling phosphatidylinositol 4,5-bis phosphate (PIP2) and SEPT2 localization at the intercellular bridge |
| Ras-related protein Rab-15 | May act in concert with RAB3A in regulating aspects of synaptic vesicle membrane flow within the nerve terminal |
| Proteasome |  |


| Proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog) (PSMD7) |  |
| :---: | :---: |
| 26S PROTEASOME REGULATORY SUBUNIT S3 |  |
| Proteasome (prosome, macropain) 26S subunit, non-ATPase, 1 (PSMD1) |  |
| Proteasome subunit beta type-5 (PSMB5) |  |
|  |  |
| Mitochondria |  |
| Solute carrier family 16 (monocarboxylic acid transporters), member 1 (SLC16A1) |  |
| MTCH1 protein | Potential mitochondrial transporter. May play a role in apoptosis. Located in Mitochondrion inner membrane; Multi-pass membrane protein |
| ATP synthase subunit alpha | Mitochondrial membrane ATP synthase ( $F_{1} F_{0}$ ATP synthase or Complex $V$ ) produces ATP from ADP in the presence of a proton gradient across the membrane which is generated by electron transport complexes of the respiratory chain. F-type ATPases consist of two structural domains, $\mathrm{F}_{1}$ - containing the extramembraneous catalytic core, and $\mathrm{F}_{0}$ - containing the membrane proton channel, linked together by a central stalk and a peripheral stalk. During catalysis, ATP synthesis in the catalytic domain of $F_{1}$ is coupled via a rotary mechanism of the central stalk subunits to proton translocation. Subunits alpha and beta form the catalytic core in $F_{1}$. Rotation of the central stalk against the surrounding alpha beta $_{3}$ subunits leads to hydrolysis of ATP in three separate catalytic sites on the beta subunits. Subunit alpha does not bear the catalytic high-affinity ATP-binding sites |
|  |  |
| Translation related |  |
| RPLPO protein | Ribosomal protein PO is the functional equivalent of E.coli protein L10. P0 forms a pentameric complex by interaction with dimers of P1 and P2. Identified in a mRNP granule complex, at least composed of ACTB, ACTN4, DHX9, ERG, HNRNPA1, HNRNPA2B1, HNRNPAB, HNRNPD, HNRNPL, HNRNPR, HNRNPU, HSPA1, HSPA8, IGF2BP1, ILF2, ILF3, NCBP1, NCL, PABPC1, PABPC4, PABPN1, RPLPO, RPS3, RPS3A, RPS4X, RPS8, RPS9, SYNCRIP, TROVE2, YBX1 and untranslated mRNAs. Interacts with Lassa virus Z protein. Interacts with APEX1. Localized to Nucleus. Cytoplasm. Note: Localized in cytoplasmic mRNP granules containing untranslated mRNAs |
| Ribosomal protein L12 variant |  |
|  |  |
| Ungrouped proteins |  |
| Paraoxonase 2 (PON2) | Capable of hydrolyzing lactones and a number of aromatic carboxylic acid esters. Has antioxidant activity. Is not associated with high density lipoprotein. Prevents LDL lipid peroxidation, reverses the oxidation of mildly oxidized LDL, and inhibits the ability of MM-LDL to induce monocyte chemotaxis |
| 7-dehydrocholesterol reductase (DHCR7) | Production of cholesterol by reduction of C7-C8 double bond of 7-dehydrocholesterol. Located in Endoplasmic |


|  | reticulum membrane; Multi-pass membrane protein |
| :---: | :---: |
| Ubiquitin associated domain containing 2 (Ubac2, PHGDHL1) |  |
| 4F2 cell-surface antigen heavy chain | Required for the function of light chain amino-acid transporters. Involved in sodium-independent, high-affinity transport of large neutral amino acids such as phenylalanine, tyrosine, leucine, arginine and tryptophan. Involved in guiding and targeting of LAT1 and LAT2 to the plasma membrane. When associated with SLC7A6 or SLC7A7 acts as an arginine/glutamine exchanger, following an antiport mechanism for amino acid transport, influencing arginine release in exchange for extracellular amino acids. Plays a role in nitric oxide synthesis in human umbilical vein endothelial cells (HUVECs) via transport of L-arginine. Required for normal and neoplastic cell growth. When associated with SLC7A5/LAT1, is also involved in the transport of L-DOPA across the blood-brain barrier, and that of thyroid hormones triiodothyronine ( $T 3$ ) and thyroxine ( $T 4$ ) across the cell membrane in tissues such as placenta. Involved in the uptake of methylmercury ( MeHg ) when administered as the L-cysteine or D,Lhomocysteine complexes, and hence plays a role in metal ion homeostasis and toxicity. When associated with SLC7A5 or SLC7A8, involved in the cellular activity of small molecular weight nitrosothiols, via the stereoselective transport of L-nitrosocysteine (L-CNSO) across the transmembrane. Together with ICAM1, regulates the transport activity LAT2 in polarized intestinal cells, by generating and delivering intracellular signals. When associated with SLC7A5, plays an important role in transporting L-leucine from the circulating blood to the retina across the inner blood-retinal barrier |
| Nbla03646 |  |
| Implantation-associated protein |  |
| HLA-B associated transcript 1 (BAT1) | Component of the THO subcomplex of the TREX complex. The TREX complex specifically associates with spliced mRNA and not with unspliced pre-mRNA. It is recruited to spliced mRNAs by a transcription-independent mechanism. Binds to mRNA upstream of the exon-junction complex (EJC) and is recruited in a splicing- and capdependent manner to a region near the 5' end of the mRNA where it functions in mRNA export. The recruitment occurs via an interaction between THOC4 and the cap-binding protein NCBP1. DDX39B functions as a bridge between THOC4 and the THO complex. The TREX complex is essential for the export of Kaposi's sarcomaassociated herpesvirus (KSHV) intronless mRNAs and infectious virus production. The recruitment of the TREX complex to the intronless viral mRNA occurs via an interaction between KSHV ORF57 protein and THOC4. Splice factor that is required for the first ATP-dependent step in spliceosome assembly and for the interaction of U2 snRNP with the branchpoint. Has both RNA-stimulated ATP binding/hydrolysis activity and ATP-dependent RNA unwinding activity. Even with the stimulation of RNA, the ATPase activity is weak. Can only hydrolyze ATP but not other NTPs. The RNA stimulation of ATPase activity does not have a strong preference for the sequence and length of the RNA. However, ssRNA stimulates the ATPase activity much more strongly than dsRNA. Can unwind 5' or 3' overhangs or blunt end RNA duplexes in vitro. The ATPase and helicase activities are not influenced by U2AF2 and THOC4. Spliceosome RNA helicase DDX39B |


| tCD18F only |  |
| :---: | :---: |
| Mitochondria |  |
| DLST protein | The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and $\mathrm{CO}_{2}$. It contains multiple copies of 3 enzymatic components: 2-oxoglutarate dehydrogenase ( E 1 ), dihydrolipoamide succinyltransferase (E2) and lipoamide dehydrogenase (E3). Full name Dihydrolipoyllysineresidue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial |
| Chaperone |  |
| DnaJ (Hsp40) homolog, subfamily C, member 10 (DNAJC10) | This endoplasmic reticulum co-chaperone may play a role in protein folding and translocation across the endoplasmic reticulum membrane. May act as a co-chaperone for HSPA5. |
| Ungrouped |  |
| Chondroitin sulfate proteoglycan 2 (Versican) variant | May play a role in intercellular signaling and in connecting cells with the extracellular matrix. May take part in the regulation of cell motility, growth and differentiation. Binds hyaluronic acid. Cerebral white matter and plasma. Isoform V0 and isoform V1 are expressed in normal brain, gliomas, medulloblastomas, schwannomas, neurofibromas, and meningiomas. Isoform V2 is restricted to normal brain and gliomas. Isoform V3 is found in all these tissues except medulloblastomas |
| Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) |  |
| Uncharacterized protein ENSP00000408518 | Zinc finger protein 878 |
| CD11a only |  |
| Protein modification |  |
| Dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit (DPM1) |  |
| Peptidyl-prolyl cis-trans isomerase | PPlases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides. May be implicated in the folding, transport, and assembly of proteins. May play an important role in the regulation of pre-mRNA splicing. |
| Peptidyl-prolyl cis-trans isomerase B (PPIB) | PPlases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides. May be implicated in the folding, transport, and assembly of proteins. May play an important role in the regulation of pre-mRNA splicing. |
| Procollagen galactosyltransferase 1 (GLT25D1) | Has a beta-galactosyltransferase activity; transfers beta-galactose to hydroxylysine residues of collagen. Localized to Endoplasmic reticulum lumen. Note: Colocalized with PLOD3 and mannose binding lectin. |
| Protein disulfide-isomerase (PDIA1) |  |


| Protein disulfide isomerase family A, member 4 |  |
| :---: | :---: |
| Chaperone |  |
| Heat shock protein 105 kDa (HSPH1) |  |
| Tumor rejection antigen (Gp96, TRA1) | Molecular chaperone that functions in the processing and transport of secreted proteins. Functions in endoplasmic reticulum associated degradation (ERAD). Has ATPase activity |
| Signal recognition particle receptor subunit beta (SRPRB) | Component of the SRP (signal recognition particle) receptor. Ensures, in conjunction with the signal recognition particle, the correct targeting of the nascent secretory proteins to the endoplasmic reticulum membrane system. Has GTPase activity. May mediate the membrane association of SRPR |
| FAS-associated factor 2 (FAF2) | May play a role in the translocation of terminally misfolded proteins from the endoplasmic reticulum lumen to the cytoplasm and their degradation by the proteasome |
| T-complex protein 1 subunit delta (CCT4) | Molecular chaperone; assists the folding of proteins upon ATP hydrolysis. As part of the BBS/CCT complex may play a role in the assembly of BBSome, a complex involved in ciliogenesis regulating transports vesicles to the cilia. Known to play a role, in vitro, in the folding of actin and tubulin. |
| Translocon-associated protein subunit delta (SSR4) | TRAP proteins are part of a complex whose function is to bind calcium to the ER membrane and thereby regulate the retention of ER resident proteins. |
| HLA-B associated transcript 3 (BAT3) | haperone that plays a key role in various processes such as apoptosis, insertion of tail-anchored (TA) membrane proteins to the endoplasmic reticulum membrane and regulation of chromatin. Acts in part by regulating stability of proteins and their degradation by the proteasome. Participates in endoplasmic reticulum stress-induced apoptosis via its interaction with AIFM1/AIF by regulating AIFM1/AIF stability and preventing its degradation. Also required during spermatogenesis for synaptonemal complex assembly via its interaction with HSPA2, by inhibiting polyubiquitination and subsequent proteasomal degradation of HSPA2. Required for selective ubiquitin-mediated degradation of defective nascent chain polypeptides by the proteasome. In this context, may play a role in immuno-proteasomes to generate antigenic peptides via targeted degradation, thereby playing a role in antigen presentation in immune response. Key component of the BAG6/BAT3 complex, a cytosolic multiprotein complex involved in the post-translational delivery of tail-anchored (TA) membrane proteins to the endoplasmic reticulum membrane. TA membrane proteins, also named type II transmembrane proteins, contain a single C-terminal transmembrane region. BAG6/BAT3 acts by facilitating TA membrane proteins capture by ASNA1/TRC40: it is recruited to ribosomes synthesizing membrane proteins, interacts with the transmembrane region of newly released TA proteins and transfers them to ASNA1/TRC40 for targeting to the endoplasmic reticulum membrane. Also involved in DNA damage-induced apoptosis: following DNA damage, accumulates in the nucleus and forms a complex with p300/EP300, enhancing p300/EP300-mediated p53/TP53 acetylation leading to increase p53/TP53 transcriptional activity. When nuclear, may also act as a component of some chromatin regulator complex that regulates histone 3 'Lys-4' dimethylation (H3K4me2). |
| Translation related |  |


| Poly(rC)-binding protein 2 |  |
| :--- | :--- |
| Poly(RC) binding protein 1 (PCBP1) |  |
| RuvB-like 2 (RUVBL2) | Prohibitin inhibits DNA synthesis. It has a role in regulating proliferation. As yet it is unclear if the protein or the <br> mRNA exhibits this effect. May play a role in regulating mitochondrial respiration activity and in aging. Located in <br> mitochondrion inner membrane |
| PHB protein |  |
|  |  |
| Actin tubulin related |  |
| Yubulin beta-6 chain |  |
| Tubulin alpha-1A chain (TUBA1A) |  |
| Tubulin alpha-8 chain (TUBA8) | Receptor for the cytotoxic ligand TNFSF10/TRAIL. The adapter molecule FADD recruits caspase-8 to the activated <br> receptor. The resulting death-inducing signaling complex (DISC) performs caspase-8 proteolytic activation which <br> initiates the subsequent cascade of caspases (aspartate-specific cysteine proteases) mediating apoptosis. <br> Promotes the activation of NF-kappa-B. |
| Tubulin beta-2C chain | Tumor necrosis factor receptor superfamily member 10B. Receptor for the cytotoxic ligand TNFSF10/TRAIL. The <br> adapter molecule FADD recruits caspase-8 to the activated receptor. The resulting death-inducing signaling <br> complex (DISC) performs caspase-8 proteolytic activation which initiates the subsequent cascade of caspases <br> (aspartate-specific cysteine proteases) mediating apoptosis. Promotes the activation of NF-kappa-B. |
| Ungrouped | Cancer-related nucleoside-triphosphatase. Has nucleotide phosphatase activity towards ATP, GTP, CTP, TTP and <br> UTP. Hydrolyzes nucleoside diphosphates with lower efficiency |
| Glutathione peroxidase | Apoptosis-inducing factor short isoform 3 <br> (PDCD8) |
| Coiled-coil domain-containing protein <br> 51(CCDC51) | Receptor for progesterone. Membrane-associated progesterone receptor component 1. |
| Chromosome 1 open reading frame 57 (C1orf57) | Component of intercellular desmosome junctions. Involved in the interaction of plaque proteins and intermediate <br> filaments mediating cell-cell adhesion. May contribute to epidermal cell positioning (stratification) by mediating <br> differential adhesiveness between cells that express different isoforms. Linked to the keratinization of epithelial <br> tissues |
| Putative uncharacterized protein | Acts as a regulatory subunit of the 26S proteasome which is involved in the ATP-dependent degradation of <br> ubiquitinated proteins. 26S proteasome non-ATPase regulatory subunit 11 |
| PGRMC1 protein | Glutathione peroxidase 8 (GPX8) |
| Desmocollin 1, isoform CRA_b (DSC1) |  |
| Monocarboxylate transporter 1 |  |


| PRDX4 | Probably involved in redox regulation of the cell. Regulates the activation of NF-kappa-B in the cytosol by a modulation of I-kappa-B-alpha phosphorylation |
| :---: | :---: |
| Transmembrane and tetratricopeptide repeat containing 1 (TMTC1) | Transmembrane and TPR repeat-containing protein 1 |
| Nodal modulator 2 (NOMO2) | May antagonize Nodal signaling and subsequent organization of axial structures during mesodermal patterning, via its interaction with NCLN |
| Replication factor C subunit 3 | The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins PCNA and activator 1 . This subunit binds to the primer-template junction. Binds the PO-B transcription element as well as other GA rich DNA sequences. Could play a role in DNA transcription regulation as well as DNA replication and/or repair. Can bind single- or double-stranded DNA. Interacts with C-terminus of PCNA. 5' phosphate residue is required for binding of the N-terminal DNA-binding domain to duplex DNA, suggesting a role in recognition of non-primer template DNA structures during replication and/or repair. |
| Thymopoietin zeta isoform (TMPO) | May help direct the assembly of the nuclear lamina and thereby help maintain the structural organization of the nuclear envelope. Possible receptor for attachment of lamin filaments to the inner nuclear membrane. May be involved in the control of initiation of DNA replication through its interaction with NAKAP95 |
| Torsin-1A-interacting protein 2 (TOR1AIP2) | inds to A- and B-type lamins. Possible role in membrane attachment and assembly of the nuclear lamina |
| Pyruvate dehydrogenase (lipoamide) alpha 1 (PDHA1) | The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and $\mathrm{CO}_{2}$. It contains multiple copies of three enzymatic components: pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and lipoamide dehydrogenase (E3). |
| Pyruvate kinase isozymes M1/M2 (PKM2) |  |
| Collagen-binding protein 2 |  |
| D-3-phosphoglycerate dehydrogenase | 3-phospho-D-glycerate + NAD $+=3$-phosphonooxypyruvate + NADH. 2-hydroxyglutarate + NAD $+=$ 2-oxoglutarate + NADH |
| Bid | The major proteolytic product p15 BID allows the release of cytochrome c By similarity. Isoform 1, isoform 2 and isoform 4 induce ICE-like proteases and apoptosis. Isoform 3 does not induce apoptosis. Counters the protective effect of Bcl-2 |
| CD11a and CD18F |  |
| Chaperones |  |
| FK506-binding protein 10 | Peptidyl-prolyl cis-trans isomerase FKBP10 PPlases accelerate the folding of proteins during protein synthesis. |
| Lectin, mannose-binding, 1 (LMAN1) | Mannose-specific lectin. May recognize sugar residues of glycoproteins, glycolipids, or glycosylphosphatidyl inositol anchors and may be involved in the sorting or recycling of proteins, lipids, or both. The LMAN1-MCFD2 complex forms a specific cargo receptor for the ER-to-Golgi transport of selected proteins. |
| Protein modification |  |
| STT3A_HUMAN Dolichyl- |  |


| diphosphooligosaccharide--protein <br> glycosyltransferase subunit STT3A OS=Homo sapiens GN=STT3A PE=1 SV=2 |  |
| :---: | :---: |
| Ribophorin II (RPN2) |  |
| UDP-glucose ceramide glucosyltransferase-like 1, transcript variant 2 | Recognizes glycoproteins with minor folding defects. Reglucosylates single N -glycans near the misfolded part of the protein, thus providing quality control for protein folding in the endoplasmic reticulum. Reglucosylated proteins are recognized by calreticulin for recycling to the endoplasmic reticulum and refolding or degradation.Endoplasmic reticulum lumen. Endoplasmic reticulum-Golgi intermediate compartment |
| RPN1 protein | Essential subunit of N -oligosaccharyl transferase enzyme which catalyzes the transfer of a high mannose oligosaccharide from a lipid-linked oligosaccharide donor to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains. Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 |
| Protein binding |  |
| hioredoxin-related transmembrane protein 1 (TMX1) | May participate in various redox reactions through the reversible oxidation of its active center dithiol to a disulfide and catalyze dithiol-disulfide exchange reactions. |
| ER lipid raft associated 1 (ERLIN1) | Component of the ERLIN1/ERLIN2 complex which mediates the endoplasmic reticulum-associated degradation (ERAD) of inositol 1,4,5-trisphosphate receptors (IP3Rs). |
| Karyopherin alpha 2 (RAG cohort 1, importin alpha 1) (KPNA2) | Functions in nuclear protein import as an adapter protein for nuclear receptor KPNB1. Binds specifically and directly to substrates containing either a simple or bipartite NLS motif. Docking of the importin/substrate complex to the nuclear pore complex (NPC) is mediated by KPNB1 through binding to nucleoporin FxFG repeats and the complex is subsequently translocated through the pore by an energy requiring, Ran-dependent mechanism. At the nucleoplasmic side of the NPC, Ran binds to importin-beta and the three components separate and importinalpha and -beta are re-exported from the nucleus to the cytoplasm where GTP hydrolysis releases Ran from importin. The directionality of nuclear import is thought to be conferred by an asymmetric distribution of the GTPand GDP-bound forms of Ran between the cytoplasm and nucleus. In vitro, mediates the nuclear import of human cytomegalovirus UL84 by recognizing a non-classical NLS. |
| Mitochondria proteins |  |
| Voltage-dependent anion-selective channel protein 2 |  |
| Calcium-binding mitochondrial carrier protein Aralar1 |  |
| Inner membrane protein, mitochondrial (mitofilin) (IMMT) |  |
| Mitochondrial carrier homolog 2 (MTCH2) |  |


| Sideroflexin-1 (SFXN1) | Might be involved in the transport of a component required for iron utilization into or out of the mitochondria. <br> Localized in Mitochondrion membrane; Multi-pass membrane protein |
| :--- | :--- |
| SLC25A11 protein |  |
| Solute carrier family 2 (Facilitated glucose <br> transporter), member 1 (SLC2A1) |  |
| Coiled-coil-helix-coiled-coil-helix domain <br> containing 3, isoform CRA_d (CHCHD3) | May be a scaffolding protein that stabilizes protein complexes involved in maintaining mitochondrial crista <br> architecture and protein import |
| ATP2A2 |  |
| Pyruvate kinase |  |
|  |  |
| Proteasome proteins | Metalloprotease component of the 26S proteasome that specifically cleaves 'Lys-63'-linked polyubiquitin chains. <br> The 26S proteasome is involved in the ATP-dependent degradation of ubiquitinated proteins. The function of the |
| Xenopus laevis proteasome (prosome, <br> macropain) 26S subunit, ATPase 3 |  |
| PSMD14 |  |
| Proteasome subunit alpha type 6 (PSMA6) |  |
| Proteasome subunit beta type (PSMB6) |  |
| 26S protease regulatory subunit S10B (PSMC6) |  |
|  | unknown |
| Translation related proteins |  |
| FWP004 | Acts as a translation activator that mediates translational control and perform an EF3-related function on on the <br> ribosome by regulating GCN2 protein kinase (EIF2AK1-4) activity |
| Ribosomal protein S27(LOC392748) |  |
| 40S ribosomal protein S20) |  |
| Translational activator GCN1 (GCN1L1) |  |
| ADP-ribosylation factor 5 (ARF5) |  |
| Elongation factor 1-gamma |  |
| Lamin-B receptor (LBR) |  |
| Actin tubulin related proteins |  |
| Beta-actin-like protein 2 (ACTBL2) |  |
| Tubulin beta-2B chain(TUBB2B) |  |
| Cofilin-1 |  |
|  |  |


| Ungrouped |  |
| :---: | :---: |
| Butyrate-induced transcript 1 (HSPC121) |  |
| Nuclear pore membrane glycoprotein 210 (NUP210) | Nucleoporin essential for nuclear pore assembly and fusion, nuclear pore spacing, as well as structural integrity |
| Major histocompatibility complex, class I, C (HLAC) |  |
| MHC class I antigen (HLA-B) |  |
|  |  |
| CD18F, tCD18F and CD11a (all three groups) |  |
|  |  |
| Chaperones |  |
| Calnexin (CANX) | Calcium-binding protein that interacts with newly synthesized glycoproteins in the endoplasmic reticulum. It may act in assisting protein assembly and/or in the retention within the ER of unassembled protein subunits. It seems to play a major role in the quality control apparatus of the ER by the retention of incorrectly folded proteins |
| Heat shock 70kDa protein 8 isoform 2 variant | Acts as a repressor of transcriptional activation. Inhibits the transcriptional coactivator activity of CITED1 on Smadmediated transcription. Chaperone. Isoform 2 may function as an endogenous inhibitory regulator of HSC70 by competing the co-chaperones. |
| DnaJ homolog subfamily B member 11 (DNAJB11) | Serves as a co-chaperone for HSPA5. Binds directly to both unfolded proteins that are substrates for ERAD and nascent unfolded peptide chains, but dissociates from the HSPA5-unfolded protein complex before folding is completed. May help recruiting HSPA5 and other chaperones to the substrate. Stimulates HSPA5 ATPase activity |
| 78 kDa glucose-regulated protein (HSPA5) | Probably plays a role in facilitating the assembly of multimeric protein complexes inside the ER. Interacts with DNAJC1 (via J domain) By similarity. Component of an EIF2 complex at least composed of CELF1/CUGBP1, CALR, CALR3, EIF2S1, EIF2S2, HSP90B1 and HSPA5. Part a large chaperone multiprotein complex comprising DNAJB11, HSP90B1, HSPA5, HYOU, PDIA2, PDIA4, PDIA6, PPIB, SDF2L1, UGT1A1 and very small amounts of ERP29, but not, or at very low levels, CALR nor CANX. Interacts with TMEM132A and TRIM21. May form a complex with ERLEC1, OS9, SEL1L and SYVN1 |
| ERLIN2 | Component of the ERLIN1/ERLIN2 complex which mediates the endoplasmic reticulum-associated degradation (ERAD) of inositol 1,4,5-trisphosphate receptors (IP3Rs). Also involved in ITPR1 degradation by the ERAD pathway |
|  |  |
| Protein modification |  |
| Dolichyl-diphosphooligosaccharide-proteinglycosyltransferase 48 kDa subunit |  |
| Neutral alpha-glucosidase AB | Cleaves sequentially the 2 innermost alpha-1,3-linked glucose residues from the $\mathrm{Glc}_{2} \mathrm{Man}_{9} \mathrm{GlcNAc}_{2}$ oligosaccharide precursor of immature glycoproteins. |
| Protein disulfide-isomerase A6 | May function as a chaperone that inhibits aggregation of misfolded proteins. Plays a role in platelet aggregation and activation by agonists such as convulxin, collagen and thrombin |


| Protein disulfide-isomerase A3 | Catalyzes the rearrangement of -S-S- bonds in proteins |
| :---: | :---: |
| Ungrouped |  |
| Sodium/potassium-transporting ATPase alpha-1 chain |  |
| RBM10 protein | May be involved in post-transcriptional processing, most probably in mRNA splicing. Binds to RNA homopolymers, with a preference for $\operatorname{poly}(\mathrm{G})$ and poly(U) and little for poly(A) |
|  |  |
| Actin tubulin |  |
| Tubulin beta-7 chain |  |
|  |  |
| tCD18 CD11a |  |
|  |  |
| Chaperone |  |
| Endoplasmic reticulum resident protein 44 (ERP44) | Mediates thiol-dependent retention in the early secretory pathway, forming mixed disulfides with substrate proteins through its conserved CRFS motif. Inhibits the calcium channel activity of ITPR1. May have a role in the control of oxidative protein folding in the endoplasmic reticulum. Required to retain ERO1L and ERO1LB in the endoplasmic reticulum. Located in ER lumen |
|  |  |
| tCD18F CD18F |  |
|  |  |
| Protein Binding |  |
| Transitional endoplasmic reticulum ATPase (VCP) | Necessary for the fragmentation of Golgi stacks during mitosis and for their reassembly after mitosis. Involved in the formation of the transitional endoplasmic reticulum (tER). The transfer of membranes from the endoplasmic reticulum to the Golgi apparatus occurs via $50-70 \mathrm{~nm}$ transition vesicles which derive from part-rough, partsmooth transitional elements of the endoplasmic reticulum (tER). Vesicle budding from the tER is an ATPdependent process. The ternary complex containing UFD1L, VCP and NPLOC4 binds ubiquitinated proteins and is necessary for the export of misfolded proteins from the ER to the cytoplasm, where they are degraded by the proteasome. The NPLOC4-UFD1L-VCP complex regulates spindle disassembly at the end of mitosis and is necessary for the formation of a closed nuclear envelope By similarity. Regulates E3 ubiquitin-protein ligase activity of RNF19A |
| Karyopherin (importin) beta 1 (KPNB1) | Functions in nuclear protein import, either in association with an adapter protein, like an importin-alpha subunit, which binds to nuclear localization signals (NLS) in cargo substrates, or by acting as autonomous nuclear transport receptor. Acting autonomously, serves itself as NLS receptor. Docking of the importin/substrate complex to the nuclear pore complex (NPC) is mediated by KPNB1 through binding to nucleoporin FxFG repeats and the complex is subsequently translocated through the pore by an energy requiring, Ran-dependent mechanism. At the |


|  | nucleoplasmic side of the NPC, Ran binds to importin-beta and the three components separate and importin- <br> alpha and -beta are re-exported from the nucleus to the cytoplasm where GTP hydrolysis releases Ran from <br> importin. The directionality of nuclear import is thought to be conferred by an asymmetric distribution of the GTP- <br> and GDP-bound forms of Ran between the cytoplasm and nucleus. Mediates autonomously the nuclear import of <br> ribosomal proteins RPL23A, RPS7 and RPL5. Binds to a beta-like import receptor binding (BIB) domain of RPL23A. <br> In association with IPO7 mediates the nuclear import of H1 histone. In vitro, mediates nuclear import of H2A, H2B, <br> H3 and H4 histones. In case of HIV-1 infection, binds and mediates the nuclear import of HIV-1 Rev. Imports PRKCI <br> into the nucleus |
| :--- | :--- | :--- |

Table II. 1 Known functions of chaperone/modification proteins identified in the combined FLAG LC-MS results, assigned according to the groups in which the hits were consistently identified

| Protein identified | emPAI values of all proteins identified consistently in LCMS (arranged according to sample type and experiment date) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | CD18F |  |  |  |  | tCD18F |  |  | $\begin{aligned} & \text { CD11a/ } \\ & \text { t-CD18F } \end{aligned}$ | CD11a/CD18F |  |  |  | $\begin{aligned} & \text { CD11aF/ } \\ & \text { CD18 } \end{aligned}$ | CD11aF |
| Name | 110701 | 110119 | 101222 | 110216 | 110531 | 110701 | 110701 | 110531 | 110531 | 110119 | 101222 | 110531 | 110216 | 110531 | 110531 |
| Integrin beta (ITGB2) | 5.32 | 6.4 | 9.33 | 10.98 | 15.72 | 1.6 | 2.27 | 3.64 | 3.74 | 2.61 | 6.68 | 13.42 | 23.22 | 7.49 | -1 |
| Integrin alpha-L (ITGAL) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1.57 | 0.85 | 0.73 | 2.64 | 3.55 | 1.95 | 5.28 |
| CD18F only |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Chaperones |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SCAMP3 protein DnaJ (Hsp40) homolog, subfamily A, member 1 | 0.08 | 0.09 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.09 | -1 | -1 | -1 | -1 |
| (DNAJA1) | 0.13 | -1 | 0.07 | -1 | -1 | -1 | 0.15 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.15 |
| HSP90AA1 protein | 0.13 | -1 | 0.1 | 0.05 | -1 | 0.23 | -1 | -1 | -1 | -1 | -1 | -1 | 0.15 | -1 | -1 |
| Protein modification |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Oligosaccharyltransferase complex subunit (OSTC) | -1 | 0.2 | -1 | 0.2 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.2 | -1 | -1 |
| Dolichyl- <br> diphosphooligosaccharide-protein glycosyltransferase subunit (STT3B) | -1 | 0.03 | -1 | 0.07 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.03 | -1 | -1 |
| Glucosidase I | -1 | 0.05 | -1 | 0.05 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| Triosephosphate isomerase | 0.11 | -1 | 0.12 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| CDP-diacylglycerol--inositol3phosphatidyltransferase (CDIPT) | 0.14 | 0.16 | -1 | 0.35 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.58 | -1 | -1 |
| Protein Binding |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Ras-related protein Rab-35 | 0.36 | -1 | -1 | 0.42 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| Ras-related protein Rab-15 | 0.25 | -1 | -1 | 0.29 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| Proteasome |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Proteasome (prosome, macropain) 26 S subunit, nonATPase, 7 (Mov34 homolog) | 0.16 | -1 | 0.09 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |


| (PSMD7) |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 26S PROTEASOME |  |  |  |  |  |  |  |  |
| REGULATORY SUBUNIT S3 | -1 | 0.11 | -1 | 0.05 | -1 | -1 | -1 | -1 |
| Proteasome (prosome, macropain) 26S subunit, non- |  |  |  |  |  |  |  |  |
| ATPase, 1 (PSMD1) | -1 | -1 | -1 | 0.06 | 0.06 | -1 | -1 | -1 |
| Proteasome subunit beta type-5 (PSMB5) | -1 | 0.25 | 0.39 | -1 | -1 | -1 | -1 | -1 |
| Mitochondria |  |  |  |  |  |  |  |  |
| Solute carrier family 16 (monocarboxylic acid transporters), member 1 |  |  |  |  |  |  |  |  |
| (SLC16A1) | -1 | 0.06 | 0.06 | -1 | -1 | -1 | -1 | -1 |
| MTCH1 protein | 0.15 | -1 | -1 | 0.08 | -1 | -1 | -1 | -1 |
| ATP synthase subunit alpha | 0.17 | 0.34 | -1 | -1 | -1 | -1 | -1 | -1 |
| Translation related |  |  |  |  |  |  |  |  |
| RPLPO protein | -1 | -1 | 0.12 | 0.41 | -1 | -1 | -1 | -1 |
| Ribosomal protein L12 variant | 0.14 | -1 | 0.33 | -1 | -1 | -1 | -1 | -1 |
| Ungrouped proteins |  |  |  |  |  |  |  |  |
| Paraoxonase 2 (PON2) | -1 | -1 | 0.08 | 0.08 | 0.08 | -1 | -1 | -1 |
| 7-dehydrocholesterol reductase (DHCR7) | 0.05 | 0.06 | -1 | 0.12 | -1 | -1 | -1 | -1 |
| Ubiquitin associated domain containing 2 (Ubac2, |  |  |  |  |  |  |  |  |
| PHGDHL1) | 0.53 | 0.27 | -1 | -1 | -1 | -1 | -1 | -1 |
| 4F2 cell-surface antigen heavy chain | -1 | 0.12 | -1 | 0.12 | -1 | -1 | -1 | -1 |
| Nbla03646 | -1 | 0.07 | -1 | 0.07 | -1 | -1 | -1 | -1 |
| Implantation-associated protein | -1 | -1 | -1 | 0.08 | 0.08 | -1 | -1 | -1 |
| HLA-B associated transcript 1 (BAT1) | -1 | 0.15 | -1 | 0.15 | -1 | 0.14 | -1 | -1 |
| tCD18F only |  |  |  |  |  |  |  |  |


| -1 | -1 | -1 | -1 | 0.11 |
| :---: | :---: | :---: | :---: | :---: |
| -1 | -1 | -1 | -1 | 0.13 |
| -1 | -1 | -1 | -1 | 0.25 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | 0.26 |
| -1 | -1 | 0.15 | -1 | -1 |
| -1 | -1 | -1 | -1 | 0.08 |
| -1 | -1 | -1 | -1 | 0.06 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | 0.19 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | -1 |
| -1 | 0.14 | -1 | -1 | -1 |



| Tumor rejection antigen (Gp96, TRA1) | -1 | -1 | -1 | -1 | -1 | 0.2 | -1 | -1 | -1 | -1 | -1 | -1 | 0.11 | -1 | 0.32 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Signal recognition particle receptor subunit beta (SRPRB) | -1 | 0.37 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.24 | -1 | 1.09 | -1 | -1 |
| FAS-associated factor 2 (FAF2) | -1 | 0.2 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.06 | -1 | -1 | 0.2 | -1 | -1 |
| T-complex protein 1 subunit delta (CCT4) | -1 | 0.06 | -1 | -1 | -1 | 0.05 | -1 | -1 | -1 | -1 | 0.06 | -1 | -1 | -1 | 0.06 |
| Translocon-associated protein subunit delta (SSR4) HLA-B associated transcript 3 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.18 | -1 | 0.62 | -1 | -1 |
| (BAT3) | -1 | 0.06 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.03 | -1 | -1 | -1 | 0.11 |
| Translation related |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Poly(rC)-binding protein 2 <br> Poly $(\mathrm{RC})$ binding protein 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.09 | 0.1 | -1 | 0.1 | -1 | 0.21 |
| Poly(RC) binding protein 1 (PCBP1) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.08 | -1 | -1 | 0.18 | -1 | -1 |
| RuvB-like 2 (RUVBL2) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.12 | -1 | -1 | 0.14 | -1 | -1 |
| PHB protein | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.1 | -1 | -1 | 0.24 | -1 | -1 |
| Actin tubulin related |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Yubulin beta-6 chain | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.62 | 0.4 | -1 | -1 | -1 | 0.61 |
| Tubulin alpha-1A chain (TUBA1A) | -1 | -1 | -1 | 1.73 | -1 | -1 | -1 | -1 | -1 | 1.06 | -1 | -1 | 2.98 | -1 | -1 |
| Tubulin alpha-8 chain (TUBA8) | -1 | -1 | -1 | 1.24 | -1 | -1 | -1 | -1 | -1 | 0.38 | -1 | -1 | 1.79 | -1 | -1 |
| Tubulin beta-2C chain | 0.76 | -1 | -1 | -1 | -1 | -1 | 1.03 | -1 | -1 | -1 | 1.18 | 0.33 | -1 | 0.65 | -1 |
| Ungrouped |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MHC class I antigen (HLA-A) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.09 | 0.18 | 0.28 | 0.1 | 0.09 | 0.49 |
| Glutathione peroxidase | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.35 | 0.18 | -1 | -1 | 0.16 | -1 |
| Apoptosis-inducing factor short isoform 3 (PDCD8) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.13 | -1 | 0.27 | -1 | -1 |
| Coiled-coil domaincontaining protein |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 51(CCDC51) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.06 | -1 | -1 | 0.07 | -1 | -1 |
| Chromosome 1 open reading frame 57 (C1orf57) | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0.14 | -1 | -1 | 0.56 | -1 | 0.35 |






| MHC class I antigen (HLA-B) | -1 | 0.35 | -1 | 0.22 | -1 | 0.09 | -1 | -1 | -1 | 0.42 | -1 | 0.35 | 0.1 | -1 | 0.35 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| CD18F, tCD18F and CD11a (all three groups) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Chaperones |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Calnexin (CANX) | 0.32 | 0.33 | 0.94 | 0.87 | 1.45 | 0.04 | 0.21 | 0.12 | 0.21 | 0.13 | 0.05 | 0.39 | 1.04 | -1 | 0.14 |
| Heat shock 70kDa protein 8 isoform 2 variant | -1 | 0.39 | 0.71 | -1 | 0.6 | 0.18 | -1 | 0.15 | -1 | 0.45 | -1 | 0.35 | 1.3 | 0.24 | -1 |
| DnaJ homolog subfamily B member 11 (DNAJB11) | 0.14 | -1 | -1 | 0.16 | 0.16 | -1 | 0.16 | 0.16 | -1 | 0.14 | 0.17 | 0.08 | 0.37 | 0.15 | 0.48 |
| 78 kDa glucose-regulated protein (HSPA5) | -1 | -1 | -1 | 6.69 | 15.36 | -1 | -1 | 2.17 | 2.78 | -1 | -1 | 2.78 | 6.69 | 1.38 | 21.31 |
| ERLIN2 | 0.16 | -1 | 0.28 | -1 | 0.18 | 0.16 | 0.18 | -1 | -1 | -1 | 0.18 | 0.18 | 0.28 | -1 | -1 |
| Protein modification |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Dolichyl-diphosphooligosaccharide-proteinglycosyltransferase 48 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| kDa subunit | 0.12 | 0.07 | 0.07 | 0.3 | 0.15 | 0.2 | 0.15 | -1 | -1 | 0.2 | -1 | 0.15 | 0.21 | -1 | 0.3 |
| Neutral alpha-glucosidase AB Protein disulfide-isomerase | -1 | 0.07 | 0.03 | 0.06 | 0.12 | 0.22 | -1 | -1 | 0.07 | 0.06 | 0.07 | 0.09 | 0.12 | -1 | 0.26 |
| A6 | -1 | 0.2 | 0.06 | 0.2 | 0.13 | 0.24 | -1 | 0.13 | 0.13 | 0.05 | -1 | 0.13 | 0.2 | 0.11 | 0.27 |
| Protein disulfide-isomerase |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A3 | 0.17 | 0.06 | -1 | -1 | -1 | 0.43 | 0.19 | 0.94 | 0.12 | 0.11 | 0.06 | -1 | 0.56 | -1 | 0.94 |
| Ungrouped |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Sodium/potassiumtransporting ATPase alpha-1 chain | 0.08 | 0.29 | -1 | 0.29 | -1 | 0.05 | 0.12 | -1 | -1 | 0.11 | 0.09 | -1 | 0.4 | -1 | -1 |
| RBM10 protein | 0.05 | 0.11 | -1 | -1 | -1 | 0.05 | 0.06 | -1 | -1 | 0.05 | 0.06 | -1 | -1 | -1 | 0.06 |
| Actin tubulin |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Tubulin beta-7 chain | 0.63 | -1 | -1 | -1 | -1 | 0.72 | 0.96 | -1 | -1 | -1 | 0.96 | 0.44 | -1 | 0.46 | -1 |
| tCD18 CD11a |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Chaperone |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Endoplasmic reticulum | -1 | -1 | -1 | -1 | -1 | 0.06 | 0.07 | -1 | -1 | 0.06 | 0.07 | -1 | -1 | -1 | 0.22 |


| resident protein 44 (ERP44) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| tCD18F CD18F |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Protein Binding |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Transitional endoplasmic reticulum ATPase (VCP) | 0.08 | 0.1 | -1 | 0.31 | 0.14 | 0.1 | 0.07 | -1 | -1 | -1 | -1 | -1 | 0.09 | -1 | -1 |
| Karyopherin (importin) beta 1 (KPNB1) | -1 | 0.03 | -1 | -1 | 0.07 | 0.06 | 0.07 | 0.07 | 0.03 | -1 | -1 | -1 | 0.07 | -1 | -1 |

Table II. 2 emPAI scores of consistent hits identified from a combined list of all the FLAG purification experiments performed.

## Appendix III: Genomic DNA sequence for ITGB2 variant 1

The human genomic DNA sequence of $I T G B 2$, variant 1 is appended here for reference. The first base pair of this sequence shown here is truncated by 10700 bp upsteam of exon 1 of variant 1 (NG_007270.2), i.e. compared to the database sequence, the first basepair depicted here is $10701 \overline{\mathrm{~b}}$ p. Highlighted in yellow, are the positions of the exon flanking primers used to sequence the exon sequence of the patient.

```
    1 ~ a t c t a a a a a g ~ t t c g a c c t g a ~ t t c a g g c t c a ~ a t a g t a a c a t ~ t t g t t c t t t c ~ t a t t t t t c t t ~
    tagatttttc aagctggact aagtccgagt tatcattgta aacaagaaag ataaaaagaa
    6 1 ~ g a t t t c t a t t ~ g t t a t t t c a t ~ c a a g a c t a a t ~ t c a t t a a a g g ~ t g t g t g c a t a ~ a c g t c t g c t g ~
        ctaaagataa caataaagta gttctgatta agtaatttcc acacacgtat tgcagacgac
1 2 1 ~ c g t c c g t g a c ~ g t t c g c a g c g ~ a g g g t c a t t g ~ c t t a c t t t a a ~ g t c a t c a g c g ~ g g t g t a g t a g
        gcaggcactg caagcgtcgc tcccagtaac gaatgaaatt cagtagtcgc ccacatcatc
1 8 1 ~ a a t g g t a t t c ~ t c t t t c t g t c ~ c t t c t t c a t t ~ a t t t t c t a g g ~ c t a a g t c t g t ~ a a t g a g a a g a ~
        ttaccataag agaaagacag gaagaagtaa taaaagatcc gattcagaca ttactcttct
2 4 1 ~ t a c c c t t g a c ~ a g a g a a c a t t ~ t c a t t g g a a c ~ t t g t t t t g t c ~ t c t t t a t t t c ~ t a a a a c c c a a ~
    atgggaactg tctcttgtaa agtaaccttg aacaaaacag agaaataaag attttgggtt
3 0 1 ~ c g c a a g a c c t ~ t g t t g a t t t a ~ a a a c a a g a a a ~ a t g t a a c c c a ~ t t c a c a g g t g ~ c t g t g a t t c c
    gcgttctgga acaactaaat tttgttcttt tacattgggt aagtgtccac gacactaagg
3 6 1 ~ t g a t g g g t t t ~ g a a a t t a t t c ~ c c a a t g t t t t ~ a t t g t t t g g t ~ t t t t a t t t a c ~ c t t t c t t g t t ~
    actacccaaa ctttaataag ggttacaaaa taacaaacca aaaataaatg gaaagaacaa
4 2 1 ~ t g t t c t a t t t ~ t c c c t t a c t t ~ t t t t c t t t t a ~ a a g g a t t t g c ~ c t t t t t t t t t ~ t t t t t t t t t t ~
    acaagataaa agggaatgaa aaaagaaaat ttcctaaacg gaaaaaaaaa aaaaaaaaaa
4 8 1 ~ c t c a g t t a c c ~ c a g a c t g g a g ~ t g c a g t g g t g ~ t a a g c a t g g t ~ t c a c c g t a g c ~ c t c a a c c t c t ~
    gagtcaatgg gtctgacctc acgtcaccac attcgtacca agtggcatcg gagttggaga
5 4 1 ~ t g g g c t c a a g ~ c a a t t c t c c c ~ a c c t c a g c c t ~ c c c a a g t a g c ~ c t g g a c c a t a ~ t c g c g t g c t a ~
    acccgagttc gttaagaggg tggagtcgga gggttcatcg gacctggtat agcgcacgat
6 0 1 ~ c c a t g c c t g g ~ c g a a t t c t t a ~ g a a a a t t t t t ~ t t g t a g a g a t ~ g a g g g t c t g g ~ g t c t c c a a a t ~
    ggtacggacc gcttaagaat cttttaaaaa aacatctcta ctcccagacc cagaggttta
6 6 1 ~ g t t g c c t a g g ~ c t g g t c t c g a ~ a c t c c t g g g c ~ t c a a g t g a t c ~ c t c c c a c c t c ~ g g c c t c c c a a ~
    caacggatcc gaccagagct tgaggacccg agttcactag gagggtggag ccggagggtt
7 2 1 ~ a g t g c t g g g a ~ t t a c a a g t t t ~ g a g c c a c c a t ~ g c c c a a a a t a ~ a a g a t t t g t c ~ a t a t t t t t t t ~
    tcacgaccct aatgttcaaa ctcggtggta cgggttttat ttctaaacag tataaaaaaa
7 8 1 ~ t c a g g t a t a t ~ a c a t g c t a t t ~ t t t a t t c t t a ~ a g g a g a c t c t ~ t g a a t t g t t a ~ g c a t a t a t a c
    agtccatata tgtacgataa aaataagaat tcctctgaga acttaacaat cgtatatatg
8 4 1 ~ t t a g a t a t t t ~ t c t a t c a a a a ~ t g c c c a c t g a ~ c a a g c a t c t a ~ t a t t c c t c c c ~ c a g a t a g g a t ~
    aatctataaa agatagtttt acgggtgact gttcgtagat ataaggaggg gtctatccta
901 ggaacggcaa tgtgctccct ggcccctccc aacccagtcc tctaggcgtt ttgtttcaca
    ccttgccgtt acacgaggga ccggggaggg ttgggtcagg agatccgcaa aacaaagtgt
961 gcgttaatat tcttaccact attttttaaa tcaacaggtt atgacatcat aatttttact
    cgcaattata agaatggtga taaaaaattt agttgtccaa tactgtagta ttaaaaatga
1 0 2 1 ~ t a t t t c t t t g ~ c c c a a c t t t g ~ t a t c a t c t g c ~ t c t t g t a t t t ~ c t c a t g g a t t ~ t c t t c a t g t c ~
    ataaagaaac gggttgaaac atagtagacg agaacataaa gagtacctaa agaagtacag
1 0 8 1 ~ t a t t t c t g c a ~ a g a c a t c t t t ~ a a a a c a g a g t ~ t t a g g a g c c g ~ g g a g c c a t g g ~ c t c a c g c c a g ~
    ataaagacgt tctgtagaaa ttttgtctca aatcctcggc cctcggtacc gagtgcggtc
1 1 4 1 ~ t a a t c c c a g c ~ a c t t t g g a a g ~ g c a g a g g t g g ~ a t g g a t c a c t ~ t g a a g c c a g g ~ a g t t c a a g a c
    attagggtcg tgaaaccttc cgtctccacc tacctagtga acttcggtcc tcaagttctg
1 2 0 1 ~ c a g c c t g g t c ~ a a c a t a g t g a ~ a a c c c t g t c t ~ c t a c t a a g a a ~ t a c a a a a a t t ~ a c c c a g g c g t ~
    gtcggaccag ttgtatcact ttgggacaga gatgattctt atgtttttaa tgggtccgca
1261 ggtggtacat gtgcctgtaa tcccagctac ttgggaggct gaggcaggag aattgcttga
    ccaccatgta cacggacatt agggtcgatg aaccctccga ctccgtcctc ttaacgaact
```

1321 acccgggagg cagaggttgc agtgagctga gatcacgcca ctgcactcca gcctgggtga tgggccctcc gtctccaacg tcactcgact ctagtgcggt gacgtgaggt cggacccact
1381 caaaagcaaa actctgtctc aaaaaaaact gaaaacataa aaataaataa ataaataaaa gttttcgttt tgagacagag ttttttttga cttttgtatt tttatttatt tatttatttt
1441 cagagtttag ggtagtaatg tgtctgagtc cttgctctga agatgtctga tttgggtttt gtctcaaatc ccatcattac acagactcag gaacgagact tctacagact aaacccaaaa
1501 ttttctaggt ataatctctt tccctcctct tcctcctcct cctcctctcc cttctctctc aaaagatcca tattagagaa agggaggaga aggaggagga ggaggagagg gaagagagag
1561 tctctgagcc tgtagtattt tccctttcgt gttctgtggt atcctcctaa tgtccctagt agagactcgg acatcataaa agggaaagca caagacacca taggaggatt acagggatca
1621 taacaggacc tcatccatcc atccctgggg tgttcagtgg cctcgcaatc cacgtccccg attgtcctgg agtaggtagg tagggacccc acaagtcacc ggagcgttag gtgcaggggc
1681 cccagatccg cagccttcct cctgtgacct gagtgtctcc tcctcgctgt ccctgttctg gggtctaggc gtcggaagga ggacactgga ctcacagagg aggagcgaca gggacaagac
1741 tcccctcccg ccctcacatg caggtggatg tcaatgcgtg cacgcaccca ggttttcttg aggggagggc gggagtgtac gtccacctac agttacgcac gtgcgtgggt ccaaaagaac
1801 cctgttccca cctctttatc ctccttggct ggttcctccg ctccattctc tggctcagga ggacaagggt ggagaaatag gaggaaccga ccaaggaggc gaggtaagag accgagtcct
1861 atttgctccc cagggttttt cattccgctc tgtagcagaa tgtgctactg ttgctttcca taaacgaggg gtcccaaaaa gtaaggcgag acatcgtctt acacgatgac aacgaaaggt
1921 gcatcaaact tttcactccc atggcctctg tgttcttctt cctaatgggt gccctttctg cgtagtttga aaagtgaggg taccggagac acaagaagaa ggattaccca cgggaaagac
1981 cctggcgctc ctggaggccg ggcccctctc ccaagtgtga atctgatgga agcaggtcct ggaccgcgag gacctccggc ccggggagag ggttcacact tagactacct tcgtccagga $\rightarrow$
2041 tctcgccggg acctcatgtg agaaatgaca tgtcccacct gtctcaaggg gccgctctct agagcggccc tggagtacac tctttactgt acagggtgga cagagttccc cggcgagaga
2101 gacatcagag ctgctgtaga gcggagaggg gcaggggtga agggccacgg tggtgcaacc ctgtagtctc gacgacatct cgcctctccc cgtccccact tcccggtgcc accacgttgg $\rightarrow$ EXON 1
2161 caccacttcc tccaaggagg agctgagagg aACAGGAAGT GTCAGGACTT TACGACCCGC gtggtgaagg aggttcctcc tcgactctcc tTGTCCTTCA CAGTCCTGAA ATGCTGGGCG
2221 GCCTCCAGCT GAGGTTTCTA GACGTGACCC AGGGCAGACT GGTAGCAAAG CCCCCACGCC CGGAGGTCGA CTCCAAAGAT CTGCACTGGG TCCCGTCTGA CCATCGTTTC GGGGGTGCGG
2281 CAGCCAGGAG CACCGCCGAG GACTCCAGCA CACCGAGGgt gagtgtgcag aggccccgct GTCGGTCCTC GTGGCGGCTC CTGAGGTCGT GTGGCTCCca ctcacacgtc tccggggcga EXON $1 \leftarrow$
2341 atgtccacgg tgatttttgg agggaagcac tctcctgccc cagcctggag tgaccctggt tacaggtgcc actaaaaacc tcccttcgtg agaggacggg gtcggacctc actgggacca
2401 ggccctgctt atagccccgg cccaggggtg tctggcagga aggggtctcc agaccgtctg ccgggacgaa tatcggggcc gggtccccac agaccgtcct tccccagagg tctggcagac $\leftarrow$
2461 ctccagggtc ccctttgaga cccggcaagg ggaagtgcct gccaccaaga gcagagccac gaggtcccag gggaaactct gggccgttcc ccttcacgga cggtggttct cgtctcggtg
2521 atggtgtgct gggagcgctt gcggctgagg atggggccgg catgggagga caaggggctg taccacacga ccctcgcgaa cgccgactcc taccccggcc gtaccctcct gttccccgac
2581 tgaggccggc aaggccagga agaggctggg cccacaagcc cctaggggga cggctggggg actccggccg ttccggtcct tctccgaccc gggtgttcgg ggatccccct gccgaccccc
2641 gtcctggtgc tccgggctgg gcaggaggca ggttcccggg gcctcctctg tgaacctgct caggaccacg aggcccgacc cgtcctccgt ccaagggcce cggaggagac acttggacga
2701 gcgcctgcct gtgttaggcc ctcctgtctg catgctgccc cgcctggcta cagcgattcc cgcggacgga cacaatccgg gaggacagac gtacgacggg gcggaccgat gtcgctaagg
2761 atctccagat cgactcgcag tgtgtggtgt gtgatatggt gtggtgtggt tcgtggtgtg tagaggtcta gctgagcgtc acacaccaca cactatacca caccacacca agcaccacac
2821 tagtgtcgtg tgtgtagtgt ggcatggtgt gcatgtatgg tgcacgtgtg tggcatagcg atcacagcac acacatcaca ccgtaccaca cgtacatacc acgtgcacac accgtatcgc
2881 tagtgtggtg tgtgtgtgtg catgtatggt gtgcgagtgt gcgcatgtga tgtggggagt atcacaccac acacacacac gtacatacca cacgctcaca cgcgtacact acacccctca
2941 ctgttgtatg tacatgtatg gtgcacatgt gtaatgtagt gtggtgtggt gtggcacatg gacaacatac atgtacatac cacgtgtaca cattacatca caccacacca caccgtgtac
3001 ctccatgctc ctctgggacc ctcgagcttc cctgcctcct ccgtggggac ccctgggtga gaggtacgag gagaccctgg gagctcgaag ggacggagga ggcacccctg gggacccact
3061 cgtcttgctg taggggagcg ggcaccgctg tgcgattctg gtcgagttgg ttcagctggt gcagaacgac atcccctcgc ccgtggcgac acgctaagac cagctcaacc aagtcgacca
3121 gaccctggcc tccccacaac aggctcctgg aggccattcc caaggtcctg gtgggagggg ctgggaccgg aggggtgttg tccgaggacc tccggtaagg gttccaggac caccctcccc
3181 gattgaagcc acgcccctgg gtcccaggcc cacctggccc ccttccccca acactgcact ctaacttcgg tgcggggacc cagggtccgg gtggaccggg ggaagggggt tgtgacgtga
3241 cttggggcca gccctcccca tgcatcctcc acccctggca ccaccccttg caccaccgtc gaaccccggt cgggaggggt acgtaggagg tggggaccgt ggtggggaac gtggtggcag
3301 cccacaccca atcccccctc ccctggatct gggagctggg attctgccct gctgtccctg gggtgtgggt taggggggag gggacctaga ccctcgaccc taagacggga cgacagggac
3361 tccatccggg gcccctctcc tctgtcagtt cctgggtggg gagaaggagg ggtggcatgg aggtaggccc cggggagagg agacagtcaa ggacccaccc ctcttcctcc ccaccgtacc
3421 ctgcaatggg gaaggggggg gatggccgga gagctcagcc tgggagcgag ctgtgggggc gacgttaccc cttccccccc ctaccggcct ctcgagtcgg accctcgctc gacacccccg

3481 ataggggaca ctcgctgtca ctcgccttcc ccgtgacggt ggcgcaacgc ctgcgggttt tatcccctgt gagcgacagt gagcggaagg ggcactgcca ccgcgttgcg gacgcccaaa
3541 gggggctgag gccggattct ggaggcttca cagaaatgag ctgctagaag ctctgggctg cccccgactc cggcctaaga cctccgaagt gtctttactc gacgatcttc gagacccgac
3601 tctgccgtca ccccagcgtt tcaccatgag tgaagcccag ctcactgcag aatccatgcc agacggcagt ggggtcgcaa agtggtactc acttcgggtc gagtgacgtc ttaggtacgg
3661 cagctcactg cagaatcccc gggggccagc actttaccaa gagtggcctc cagactccac gtcgagtgac gtcttagggg cccccggtcg tgaaatggtt ctcaccggag gtctgaggtg
3721 attttaggaa ggaactcctt ccccatccag ctcacagcag aatccccagg gtccctgcca taaaatcctt ccttgaggaa ggggtaggtc gagtgtcgtc ttaggggtcc cagggacggt
3781 gcctggcacc cagcgtgccc ctcagtattc tggggatgag tgggggagtc cctctgctgg cggaccgtgg gtcgcacggg gagtcataag acccctactc accccctcag ggagacgacc
3841 gaagaggtcc ctggctacgt ccctgcaggg cggcggctga gcccatgcct cggccccagg cttctccagg gaccgatgca gggacgtccc gccgccgact cgggtacgga gccggggtcc
3901 gccagtcccc gccgggtgcc atcctgagca gggccacaga acagatgtcc cacatggcca cggtcagggg cggcccacgg taggactcgt cccggtgtct tgtctacagg gtgtaccggt
3961 ggacccctgt ggccacctga ccccgggggg gctgtggccc tcggtgaagc aggtgaccgg cctggggaca ccggtggact ggggcccccc cgacaccggg agccacttcg tccactggcc
4021 gcgggctgtg gggtttgtca gctccgcagc gtcactgccg ctcagcgtat gagtttccac cgcccgacac cccaaacagt cgaggcgtcg cagtgacggc gagtcgcata ctcaaaggtg
4081 tcccagcact caatgaacag ctgagcaact gcaggtgcct ccgagtgctg ggagagggga agggtcgtga gttacttgtc gactcgttga cgtccacgga ggctcacgac cctctcccct
4141 atgctcagct ccgagtgctg ggagagggga cgctcagctc cgagtgctgg gagaggggac tacgagtcga ggctcacgac cctctcccet gcgagtcgag gctcacgacc ctctcccctg
4201 gcttagccac ccaagggcct ttgcctcaag ggagtaaata ggggcatttc ctggtgggcg cgaatcggtg ggttcccgga aacggagttc cctcatttat ccccgtaaag gaccacccgc
4261 tgggctgggc tggagaagca gggcctcctg gcaggggctg gcatgagatc caaacactgc acccgacccg acctcttcgt cccggaggac cgtccccgac cgtactctag gtttgtgacg
4321 ccaggcagct ggacactcct gagcctggag cttctcacct gaaccacagg gtggtaaccc ggtccgtcga cctgtgagga ctcggacctc gaagagtgga cttggtgtcc caccattggg
4381 ctgcccctca ggcggcatcg acgtgacctg cctagtgcac agtgagcctg aggcaacacg gacggggagt ccgccgtagc tgcactggac ggatcacgtg tcactcggac tccgttgtgc
4441 gccaccacca ggagactccc tggggaagag gcccccagac agcgaggtgc ccatgaagga cggtggtggt cctctgaggg accccttctc cgggggtctg tcgctccacg ggtacttcct
4501 agttctgcac aggggccgcg gccatccagg agatgggagg ggccctatct gcaccctgtc tcaagacgtg tccccggcgc cggtaggtcc tctaccctcc ccgggataga cgtgggacag
4561 tggaaggggc atggggtgag caggctattt ggtcccccgt gactcagttt ccccatctgt accttccccg taccccactc gtccgataaa ccagggggca ctgagtcaaa ggggtagaca
4621 gaaacaggag gggtggtctg gccaattgtg cgcactccag ctgtctctat agtcaaccgt ctttgtcctc cccaccagac cggttaacac gcgtgaggtc gacagagata tcagttggca
4681 gtaacttgaa ctcacttccc cagggctcct gccaggccca ggtctgtcca ctctggccca cattgaactt gagtgaaggg gtcccgagga cggtccgggt ccagacaggt gagaccgggt
4741 ggccgggccc caccaggtac cagggctgtg ggacatttgc ctctgtggat ggggcagcgt ccggcccggg gtggtccatg gtcccgacac cctgtaaacg gagacaccta ccccgtcgca
4801 gacgcctgtg gcgattccat ggaacaggtc ctgtgtgggc cccccttgct gcactcgtgt ctgcggacac cgctaaggta ccttgtccag gacacacccg gggggaacga cgtgagcaca
4861 cctgtgtcag agaaccgtgg tcctgcctga ggccatgtca ccggaggctg gctcccgttc ggacacagtc tcttggcacc aggacggact ccggtacagt ggcctccgac cgagggcaag
4921 ccggaggaga atgccagctg ccacaggggt cagtgacaca ggcatttgcc cagcaggtac ggcctcctct tacggtcgac ggtgtcccca gtcactgtgt ccgtaaacgg gtcgtccatg
4981 aggccaggtc gctgggaagc cccaactcaa ggggctctgc agtgggcgcc tagcaattat tccggtccag cgacccttcg gggttgagtt ccccgagacg tcacccgcgg atcgttaata
5041 ggagctgtgg gtgataccca ccccaagccg gacctggtgc cagcaattat ccctcccatt cctcgacacc cactatgggt ggggttcggc ctggaccacg gtcgttaata gggagggtaa
5101 ctgcttggca caaactggca ctgacacggg gcattgatta ggtggttgca tttgtcgatt gacgaaccgt gtttgaccgt gactgtgccc cgtaactaat ccaccaacgt aaacagctaa
5161 ccgagcccat gtgtgtggag cccaggccct gagcaagcac agagggccct ggtgctgtcc ggctcgggta cacacacctc gggtccggga ctcgttcgtg tctcccggga ccacgacagg
5221 tgagctgagg cccctgccca ccccgggctg cgtgttctct ccggccattc agggtctgag actcgactcc ggggacgggt ggggcccgac gcacaagaga ggccggtaag tcccagactc
5281 agtgcagaag agttagcccc aggggaacac cttcctgccc agctgatttg taaacagcct tcacgtcttc tcaatcgggg tccccttgtg gaaggacggg tcgactaaac atttgtcgga
5341 cgctggggac acatgttggt gttgggagtg gtggaggttt ctgtgtgtga tggaaacttg gcgacccctg tgtacaacca caaccctcac cacctccaaa gacacacact acctttgaac
5401 gggtaggtgt tggggaaaag aggtgctgag agtctcagca gagggcagag tccaggagtc cccatccaca accccttttc tccacgactc tcagagtcgt ctcccgtctc aggtcctcag
5461 ctggagtgca gctctcagcc cctgcgccgt gcctagagga gaggctggca agcccctgtc gacctcacgt cgagagtcgg ggacgcggca cggatctcct ctccgaccgt tcggggacag
5521 ccatggatgg gtgaatgggt gagtgacagt ggcaacaagc ccatgccagc agggaggtcc ggtacctacc cacttaccca ctcactgtca ccgttgttcg ggtacggtcg tccctccagg
5581 cactgccatg aagcagcgtc ctcatcaacg tcacagtggg aacaagaaag ccagtagctt gtgacggtac ttcgtcgcag gagtagttgc agtgtcaccc ttgttctttc ggtcatcgaa
5641 tggaatccga aggagcaatt tcagcataat tcgtcggcca ccatgaaatt ccatttttaa accttaggct tcctcgttaa agtcgtatta agcagccggt ggtactttaa ggtaaaaatt
5701 aatgtccttt attcatttat cacaaaaagg gtggaaaata caaacaggaa taatgaaaat ttacaggaaa taagtaaata gtgtttttcc caccttttat gtttgtcctt attactttta

5761 tgaaaatact catagtgcca ccacccagag aaatgcagaa tattttggaa ctgtgtctct acttttatga gtatcacggt ggtgggtctc tttacgtctt ataaaacctt gacacagaga
5821 ctttttttcc ccataaattc acttttacca cgaagccaag ctgccctgag cagctgtttg gaaaaaaagg ggtatttaag tgaaaatggt gcttcggttc gacgggactc gtcgacaaac
5881 catgttttcg ttgctgtgtg ggactccata gcgcaggcac cgaaaccgcc aaagcatctt gtacaaaagc aacgacacac cctgaggtat cgcgtccgtg gctttggcgg tttcgtagaa 5941 acacgttgac ggctacacag ggaatcaaga tattttggct gaaatccctg gttctttcct tgtgcaactg ccgatgtgtc ccttagttct ataaaaccga ctttagggac caagaaagga
6001 ggggtgactc atatgtttct tagccagttt tcttccttct tttgtgaatt acctgctcag ccccactgag tatacaaaga atcggtcaaa agaaggaaga aaacacttaa tggacgagtc 6061 ggtgccacac ccttgaaaca tccttgaatg gagtatttac ctttttgtta ttaatttcta ccacggtgtg ggaactttgt aggaacttac ctcataaatg gaaaaacaat aattaaagat
6121 agatctcttt agagatgaag attaccattt accttttgca ctagagagat aatagctatt tctagagaaa tctctacttc taatggtaaa tggaaaacgt gatctctcta ttatcgataa
6181 tcctctatgc aatttgttag tacaacatgt gccattttta gagtctgtct tcttgctatt aggagatacg ttaaacaatc atgttgtaca cggtaaaaat ctcagacaga agaacgataa
6241 ttccccctta gctttttttt tttttttctt ttgagacagt ttcactcttg ttgcccaggc aagggggaat cgaaaaaaaa aaaaaaagaa aactctgtca aagtgagaac aacgggtccg
6301 tggagtgcag aggtgagatc tcagctcact gcaccctctg cctcccgggc tcaagtgatt acctcacgtc tccactctag agtcgagtga cgtgggagac ggagggcccg agttcactaa
6361 ctcctgcctc agcctcgcga gtagctggga ttacaggcgc ccgccaccaa gcccagctaa gaggacggag tcggagcgct catcgaccct aatgtccgcg ggcggtggtt cgggtcgatt
6421 tttatgtatt tttagtagag atggggtttc actatgttgc tcaggctggt ctcaaactcc aaatacataa aaatcatctc taccccaaag tgatacaacg agtccgacca gagtttgagg
6481 tgacctcagg tgatccgcct gcctcagcct cccaaagtgc tgggattaca ggcatgagcc actggagtcc actaggcgga cggagtcgga gggtttcacg accctaatgt ccgtactcgg
6541 accatgcccg gcctcccctt agcttttatg ctgaggtttc cttactctta gatacctatt tggtacgggc cggaggggaa tcgaaaatac gactccaaag gaatgagaat ctatggataa 6601 cttctatttc ctttattgtt ttattcccta aaactgaagt cctacacatt tgtggccatt gaagataaag gaaataacaa aataagggat tttgacttca ggatgtgtaa acaccggtaa 6661 cttttatgtc cacagcaaga gtgtccgggg cctcggcctc acgtcgccac ccaggaacag gaaaatacag gtgtcgttct cacaggcccc ggagccggag tgcagcggtg ggtccttgtc
6721 acagtctcct tcatgttaga tggaggagtg gtgtcccctt gtctaaaatg ctgtaaatat tgtcagagga agtacaatct acctcctcac cacaggggaa cagattttac gacatttata
6781 catttgattt ccctttgact tcaggtattt agcagagtac tttagcattc cttattaata gtaaactaaa gggaaactga agtccataaa tcgtctcatg aaatcgtaag gaataattat
6841 tgtgtgtgtc gggggggtgg ggtgtctcat tttatttcat ttggccagaa attgtggcct acacacacag cccccccacc ccacagagta aaataaagta aaccggtctt taacaccgga
6901 ttaaaatttg tgctatcagg ccgggcgcag tggctcacgc ctgtaatccc agcactttgg aattttaaac acgatagtcc ggcccgcgtc accgagtgcg gacattaggg tcgtgaaacc
6961 gaggccgagg caggcggatc acgaggtcag gagttcgaga ccagcctggc caatatggtg ctccggctcc gtccgcctag tgctccagtc ctcaagctct ggtcggaccg gttataccac
7021 aaactccatc tctactaaaa atacaaaaat tagctgggcc tagtggtgtg cacctatagt tttgaggtag agatgatttt tatgttttta atcgacccgg atcaccacac gtggatatca
7081 cccagctact caggaggctg aggcagaaga atttcttgaa catgggaggc ggaggttgca gggtcgatga gtcctccgac tccgtcttct taaagaactt gtaccctccg cctccaacgt
7141 gtgagccgag atctagccat tgcactccag cctggatgac agagtgagac tccatctcaa cactcggctc tagatcggta acgtgaggtc ggacctactg tctcactctg aggtagagtt
7201 aaaaacaaac aaacaaacaa aactgtgcta tcaaacttca cagatagtca aaccaacagt tttttgtttg tttgtttgtt ttgacacgat agtttgaagt gtctatcagt ttggttgtca
7261 aggaacacgt ctccctgacc cttgggaaaa tacgcttttg cagggtcgca ggttcagagg tccttgtgca gagggactgg gaaccctttt atgcgaaaac gtcccagcgt ccaagtctcc
7321 cgacatgggc tgggttttcc tgcatcctct ctgctccgac ttgttcctat ttgatctgac gctgtacccg acccaaaagg acgtaggaga gacgaggctg aacaaggata aactagactg
7381 ctgtttttgt ccctggattt cctggagcct ctgtgatcca aggacagtgt tccagtcccc gacaaaaaca gggacctaaa ggacctcgga gacactaggt tcctgtcaca aggtcagggg
7441 acctgacggc acttctctcc agttctttat cacattccta caaagcagaa atgagtccat tggactgccg tgaagagagg tcaagaaata gtgtaaggat gtttcgtctt tactcaggta
7501 cttgatgctc tggcttttgg cttaaaaggt ggatgcctag tttataatca tcccttgagc gaactacgag accgaaaacc gaattttcca cctacggatc aaatattagt agggaactcg
7561 cttcagttcc cagcgcaggg gttggctccc atgtcctctt ccttccaacc ttgtcccctg gaagtcaagg gtcgcgtccc caaccgaggg tacaggagaa ggaaggttgg aacaggggac
7621 gcggcatcgt ccatcacttc ccgcgcacct ccgagtcact ttgatgcgag tgcctctcct cgccgtagca ggtagtgaag ggcgcgtgga ggctcagtga aactacgctc acggagagga
7681 ctaggacagc gcaaggtggg tttcagcaca ggcaggagga agctctgatg caaagttcca gatcctgtcg cgttccaccc aaagtcgtgt ccgtcctcct tcgagactac gtttcaaggt
7741 atgggtctgc ggggcctctg gttcctcaag cagtctcgtc ttagggtttt acggggttgc tacccagacg ccccggagac caaggagttc gtcagagcag aatcccaaaa tgccccaacg
7801 aaaaaggaag cctgagactg ggcacggtgg ctcacgcctg taatcccagc actttgggag tttttccttc ggactctgac ccgtgccacc gagtgcggac attagggtcg tgaaaccctc
7861 gccgaggtgg gcggatcatg aggtcaggag atcgagacca tcctgatcaa cacgatgaaa cggctccacc cgcctagtac tccagtcctc tagctctggt aggactagtt gtgctacttt
7921 cctcgtctct actaaaaaca caaaaaaatt agctgggcgt ggtcgtggga gcctgtagtc ggagcagaga tgatttttgt gtttttttaa tcgacccgca ccagcaccct cggacatcag
7981 ccagctactc aggaagctga ggcaggagaa tggcatgaac ctgggaggcg gagcttgcag ggtcgatgag tccttcgact ccgtcctctt accgtacttg gaccctccgc ctcgaacgtc

8041 tgagctgaga tcgcgccact gcactccagg ccggtgacag cgcaactccg tctcaaaaaa actcgactct agcgcggtga cgtgaggtcc ggccactgtc gcgttgaggc agagtttttt
8101 ataaaaagga agcctgaccc cgctggattc ctctcccagc ataggctctg ccctctctct tatttttcct tcggactggg gcgacctaag gagagggtcg tatccgagac gggagagaga
8161 gctgtcttgc ttaaagtttc ctttcttttt cttttctttt ttttttgaga cagggcctcg cgacagaacg aatttcaaag gaaagaaaaa gaaaagaaaa aaaaaactct gtcccggagc 8221 ttctgtcaac taggctggag ttcagcagtg ccattagggt tcactgcagt ctctccctcc aagacagttg atccgacctc aagtcgtcac ggtaatccca agtgacgtca gagagggagg
8281 tgggcgcgag tgatcctcct gcctcagcct cctgagtaac tgggactaga gatgcatgcc acccgcgctc actaggagga cggagtcgga ggactcattg accctgatct ctacgtacgg 8341 accatgtctg gctaatttat tttctatttt tattttttgt agaggtagga gtctcactgt tggtacagac cgattaaata aaagataaaa ataaaaaaca tctccatcct cagagtgaca
8401 gttgcccagg ctgctctcaa actcctggcc tcaagtgatt ctcctgcctt ggcctctcaa caacgggtcc gacgagagtt tgaggaccgg agttcactaa gaggacggaa ccggagagtt
8461 aacactggga ttgcaggcgt cagccaccag gctggcctct tcctttcctg tgtttgctcc ttgtgaccct aacgtccgca gtcggtggtc cgaccggaga aggaaaggac acaaacgagg
8521 tgttctcctg tcctcgactt ctgtgaaagg gtgtgtgtga tttggcccgg atctgctgca acaagaggac aggagctgaa gacactttcc cacacacact aaaccgggcc tagacgacgt
8581 ggtctcttct gacgccgctt tcgtgtctca gagcctgaat gttcagagct ttgctcaccg ccagagaaga ctgcggcgaa agcacagagt ctcggactta caagtctcga aacgagtggc
8641 tgaagtgtgg ttcttgttag ctgtgtccat ttggtgcccg ggagggaagg gcgcccggcc acttcacacc aagaacaatc gacacaggta aaccacgggc cctcccttcc cgcgggccgg
8701 cctgaggggt tctccgggga aggatgacct cgtccccctg gtcctgcccc ctcagccccc ggactcccca agaggccect tcctactgga gcagggggac caggacgggg gagtcggggg
8761 tgcttcccct gtgcagggga cttggtgaca ggtaccactt cctgcagctg ggctggggtt acgaagggga cacgtcccct gaaccactgt ccatggtgaa ggacgtcgac ccgaccccaa
8821 gcaccagcac ttccttctcg ggcagggttg aggtcaaggg cgactctgat gcaaggttct cgtggtcgtg aaggaagagc ccgtcccaac tccagttccc gctgagacta cgttccaaga
8881 gagaggcctg tggggctgtg gattcctaag cagcctcagc tttgggtttt gcaccttagg ctctccggac accccgacac ctaaggattc gtcggagtcg aaacccaaaa cgtggaatcc
8941 tcttggtaag gggacccgcc cgcccttaga tgggggatgt ccagagctgg aggatgagga agaaccattc ccctgggcgg gcgggaatct accccctaca ggtctcgacc tcctactcct
9001 gggcacccgc ccctggtgtg gggcagcagg gagggccttg gagaatgagg agggcacctg cccgtgggcg gggaccacac cccgtcgtcc ctcccggaac ctcttactcc tcccgtggac
9061 cccccagtgt ggggcagcag ggagggcctc agggtgatca ggctccaggc cagagctctg gggggtcaca ccccgtcgtc cctcccggag tcccactagt ccgaggtccg gtctcgagac 9121 gggccagctc aggggctggg gtggggggtg tctgtgctct tcctcccctc ctgccctgca cccggtcgag tccccgaccc caccccccac agacacgaga aggaggggag gacgggacgt
9181 gctctgccgg cttagcctca gggcaccaga ggtgacagtg gccaggcacc tgtttgcccg cgagacggcc gaatcggagt cccgtggtct ccactgtcac cggtccgtgg acaaacgggc
9241 aatgctgccc tacatctgtg ccttacccat tcccatcccc cagggagctg gcacaggccc ttacgacggg atgtagacac ggaatgggta agggtagggg gtccctcgac cgtgtccggg
9301 ttcgaccacc tccaggcaca tggaggaaac tgaggcagat aggcctgatg gggtcttggt aagctggtgg aggtccgtgt acctcctttg actccgtcta tccggactac cccagaacca
9361 ctctcaggcc tcaggtgagg ccctggacct ccgtgggcca tcctggctgc tgggcagagg gagagtccgg agtccactcc gggacctgga ggcacccggt aggaccgacg acccgtctcc
9421 ggctgccctg ccctgttctc ccacagaagg gaggctggct gccaggctca gtgcactgcc ccgacgggac gggacaagag ggtgtcttcc ctccgaccga cggtccgagt cacgtgacgg
9481 ttacctcctg gagcggacct tcatccccag ctgtttgtct gggctcagcc actggcattg aatggaggac ctcgcctgga agtaggggtc gacaaacaga cccgagtcgg tgaccgtaac 9541 gggggtgggg cagggcctgg cttctcttgt gcccccagct ccaagccccc cagggctgct cccccacccc gtcccggacc gaagagaaca cgggggtcga ggttcggggg gtcccgacga
9601 catggaggag caacagtggg gagaaggtgg ctccccacca gcccccttcc tgtggtcaag gtacctcctc gttgtcaccc ctcttccacc gaggggtggt cgggggaagg acaccagttc
9661 gttggggtgc cgtggtctct gagccaatcc tgctggaatc aaggtgcacc ccctctccct caaccccacg gcaccagaga ctcggttagg acgaccttag ttccacgtgg gggagaggga
9721 gtgggcaggt gggagtgccc ggggtaactg ggcaccaggt ccatatcttg ctgctgctgg cacccgtcca ccctcacggg ccccattgac ccgtggtcca ggtatagaac gacgacgacc
9781 ggccctcagt ttccccattt gagggtatga agtgtttgca tgccactggt cacgactgtg ccgggagtca aaggggtaaa ctcccatact tcacaaacgt acggtgacca gtgctgacac
9841 ttaggggcgg agacctcatg tctgcggccc cagggagagc gtggagctgg ggtgtgagtc aatccccgcc tctggagtac agacgccggg gtccctctcg cacctcgacc ccacactcag
9901 cttaccagcc tggctggggc tgccagggga ggtctgcacc ccctccccac ttggcgcagc gaatggtcgg accgaccccg acggtcccct ccagacgtgg gggaggggtg aaccgcgtcg
9961 tgcgacctgg gggagtcctg aagccggctg gaccctgcac cctgggggag tcctgaagcc acgctggacc ccctcaggac ttcggccgac ctgggacgtg ggaccccctc aggacttcgg
10021 ggctggaccc tgcaccctgg gggagtcctg aagccggctg gaccctgcac cctgggggag ccgacctggg acgtgggacc ccctcaggac ttcggccgac ctgggacgtg ggaccccctc
10081 tcctgaagcc ggctggaccc tgcaccctgg gggagtcctg aagccggctg gaccctgcac aggacttcgg ccgacctggg acgtgggacc ccctcaggac ttcggccgac ctgggacgtg
10141 cctgggggag tcctgaagcc ggctggaccc tgcaccctgg gggagtcctg aagccggctg ggaccccctc aggacttcgg ccgacctggg acgtgggacc ccctcaggac ttcggccgac
10201 gaccctgcac cctgggggag tcctgaagcc ggctggaccc tgcaccctgg gggagtcctg ctgggacgtg ggaccccctc aggacttcgg ccgacctggg acgtgggacc ccctcaggac
10261 aagccggctg gaccctgcac cctgggggag tcctgaagcc ggctggacce tgcaccctgg ttcggccgac ctgggacgtg ggaccccctc aggacttcgg ccgacctggg acgtgggacc

10321 gggagtcctg aagctggctg gaccctgcac cctggggtag ctccagcacg cacagggcct ccctcaggac ttcgaccgac ctgggacgtg ggaccccatc gaggtcgtgc gtgtcccgga
10381 ccgatcagct gaggggctgg tgggcactcg tggcaccgag gtcagtccag gcgcaggact ggctagtcga ctccccgacc acccgtgagc accgtggctc cagtcaggtc cgcgtcctga
10441 ttgagggggc ccagatggag agctggaact ggcctcggcc tgaaccacaa cccccaacga aactcccccg ggtctacctc tcgaccttga ccggagccgg acttggtgtt gggggttgct
10501 tcagagatat gggggctttc agacacaggg tgggcccaat gtgaaccaaa accacctcct agtctctata cccccgaaag tctgtgtccc acccgggtta cacttggttt tggtggagga
10561 ccaggaagtt acagcagggt gcaggttctg tttttgaccc caggacggga aggatcaagg ggtccttcaa tgtcgtccca cgtccaagac aaaaactggg gtcctgccct tcctagttcc
10621 gttccctttt gacatttttg tattctgatt ataattttgc tgggaaccag ggaatgaggt caagggaaaa ctgtaaaaac ataagactaa tattaaaacg acccttggtc ccttactcca
10681 cactgccaaa acaacaggac cttggggctt gttggggcgt ctaagctgtg ggatcctcag gtgacggttt tgttgtcctg gaaccccgaa caaccccgca gattcgacac cctaggagtc
10741 atgggcccct ggtgggtgtt ccgttggcct cagagccccc agcacaggtg cagggctttc tacccgggga ccacccacaa ggcaaccgga gtctcggggg tcgtgtccac gtcccgaaag
10801 gggggatgga ggggagccag ggcctgtggc tgagggcacc cgagcctccc gggccacagc ccccctacct cccctcggtc ccggacaccg actcccgtgg gctcggaggg cccggtgtcg
10861 acagcaccca gcgctgtctt caaagccggc aggccctagt ttggggctct gctggcttgg tgtcgtgggt cgcgacagaa gtttcggccg tccgggatca aaccccgaga cgaccgaacc
10921 gtctgcccac tgaaccttgg gtggccctgc cagcaccacc cgagggaagg gtgtggggca cagacgggtg acttggaacc caccgggacg gtcgtggtgg gctcccttcc cacaccccgt
10981 ggaacgcccc agcagcctgg gaggccgtga caggaactgg catttgctgg aggaggtgtg ccttgcgggg tcgtcggacc ctccggcact gtccttgacc gtaaacgacc tcctccacac
11041 ttgatctttg cttctgacaa cagccccagg gtaacggaaa tgtcttcctg cccaccatga aactagaaac gaagactgtt gtcggggtcc cattgccttt acagaaggac gggtggtact
11101 caggaaggct cctgggtacc ccttctcagt acctctccac cttcttggtg ccagataaaa gtccttccga ggacccatgg ggaagagtca tggagaggtg gaagaaccac ggtctatttt
11161 gggtcccagc cagctataat gttctcaggt cctgggaggt cagacccagg ggcccacccc cccagggtcg gtcgatatta caagagtcca ggaccctcca gtctgggtcc ccgggtgggg
11221 acctaggaag ccctaagacc ccccaaggag ccgacctagg cctggtttgg gcacctcccg tggatccttc gggattctgg ggggttcctc ggctggatcc ggaccaaacc cgtggagggc
11281 gatggaggga agggaaccca ggggtggaac ggccctgctg cccgctagga ggtgctgccg ctacctccct tcccttgggt ccccaccttg ccgggacgac gggcgatcct ccacgacggc
11341 gcctccagca ccaggctgag gcctgcttcc gtcattttgc ctctggggcc gagggcctgt cggaggtcgt ggtccgactc cggacgaagg cagtaaaacg gagaccccgg ctcccggaca
11401 gagtgaccac tgagccgccc ctggctgtgg atgcgtgggt taaggacggg gtgggtgtgt ctcactggtg actcggcggg gaccgacacc tacgcaccca attcctgccc cacccacaca
11461 atctggaaac aaaagggcag gaccccggca actggtggct tgggcaagtc aggtgacttt tagacctttg ttttcccgtc ctggggccgt tgaccaccga acccgttcag tccactgaaa
11521 ttttcccacc tggcagggac tgcagggctc ctgggcaggc ctagcacaca gcttggcacc aaaagggtgg accgtccctg acgtcccgag gacccgtccg gatcgtgtgt cgaaccgtgg
11581 cgcgggcatc aagccagctt tctcttcctg gcacacagcc tggcacccgc gggcatccag gcgcccgtag ttcggtcgaa agagaaggac cgtgtgtcgg accgtgggcg cccgtaggtc
11641 ccagctttct cttcctggct atgtttctgc cacagagaga cagggccaag cccagctctc ggtcgaaaga gaaggaccga tacaaagacg gtgtctctct gtcccggttc gggtcgagag
11701 ctccaagcac aggctctgct cacggaggac ctggggcaca ctggggcaga gtgcacacac gaggttcgtg tccgagacga gtgcctcctg gaccccgtgt gaccccgtct cacgtgtgtg
11761 cgggtgtgtg gggttctggt gtttgtgtgc gtgtacattt gcacgtgtgt atggtgtgtg gcccacacac cccaagacca caaacacacg cacatgtaaa cgtgcacaca taccacacac
11821 cgtgtggtgt gtgtgtctgt gtggtgtgta tgtgtgtgta cgtgtgcatg tgtatatttg gcacaccaca cacacagaca caccacacat acacacacat gcacacgtac acatataaac
11881 tgtgtggtgt gtagtgtctt gcggggagta tggtgtgggt aaggggtgta tacatctctg acacaccaca catcacagaa cgcccctcat accacaccca ttccccacat atgtagagac
11941 tgtgtacata tgtgtggtgt gatgcatgta tatgtgtggt gtgtgcatgc acgtggtgtg acacatgtat acacaccaca ctacgtacat atacacacca cacacgtacg tgcaccacac
12001 tacatgcagt caggcacaca tctgtatgca cgtacacgtg catgtgtggt gtgtgtacac atgtacgtca gtccgtgtgt agacatacgt gcatgtgcac gtacacacca cacacatgtg $\rightarrow$
12061 gtctgcatgt gtgtgtgtgc ctgtgttgtg tgcctctgca ggggcccgct gcagctgtcc cagacgtaca cacacacacg gacacaacac acggagacgt ccccgggcga cgtcgacagg
12121 tgggtgcctg gctgagcagg tgtgctgccc gcactgctga gcccgggatc cctccctccc acccacggac cgactcgtcc acacgacggg cgtgacgact cgggccctag ggagggaggg
12181 tggctgttgg cagctgggcc cagtgcagct gtgctggcct gaccctaacc caccccctgc accgacaacc gtcgacccgg gtcacgtcga cacgaccgga ctgggattgg gtgggggacg
12241 agcagggaca gggccccagc tccttgtccc aggccagctg gaggggagct tcatgggtgg tcgtccctgt cccggggtcg aggaacaggg tccggtcgac ctcccctcga agtacccacc $\rightarrow E X O N 2$ M
12301 tcagccccag gggttgggcc tgagaccgtc accaagaccc cttccctcca cagGACATGC agtcggggtc cccaacccgg actctggcag tggttctggg gaagggaggt gtcCTGTACG

12361 TGGGCCTGCG CCCCCCACTG CTCGCCCTGG TGGGGCTGCT CTCCCTCGGG TGCGgtgagt ACCCGGACGC GGGGGGTGAC GAGCGGGACC ACCCCGACGA GAGGGAGCCC ACGCcactca EXON $2 \leftarrow$
12421 tctgtgttcc acggagggga ctccacgcgt gacattccaa tcagggagaa cagggcctgc agacacaagg tgcctcccct gaggtgcgca ctgtaaggtt agtccctctt gtcccggacg

12481 cctcactttc ccatagacag aagttcccgg gaggctgggg aaggaccttg ccactggggg ggagtgaaag ggtatctgtc ttcaagggcc ctccgacccc ttcctggaac ggtgaccccc
12541 taggtcgggg actcaggcct cgtctccgca cgtgccagga ggctgtagct gtgctgtttc atccagcccc tgagtccgga gcagaggcgt gcacggtcct ccgacatcga cacgacaaag
12601 ccgagggccc ctctttgcac ctgggggtgg tggcgaatgc ggggtctcct cctggaggct ggctcccggg gagaaacgtg gacccccacc accgcttacg ccccagagga ggacctccga
12661 gcctccctgt ctttttgggc ctcctctgac cttcctgctc tactccccct tgggtgtggg cggagggaca gaaaaacccg gaggagactg gaaggacgag atgaggggga acccacaccc
$\rightarrow E X O N \quad 3 \quad V$
12721 cagggcggcc cagagcaccc actcaccagc cggcctcgtc cctcagTCCT CTCTCAGGAG gtcccgccgg gtctcgtggg tgagtggtcg gccggagcag ggagtcAGGA GAGAGTCCTC

12781 TGCACGAAGT TCAAGGTCAG CAGCTGCCGG GAATGCATCG AGTCGGGGCC CGGCTGCACC ACGTGCTTCA AGTTCCAGTC GTCGACGGCC CTTACGTAGC TCAGCCCCGG GCCGACGTGG W C Q K L
12841 TGGTGCCAGA AGCTGgtaag tgcctcctgg acccctcccc acctgcccag ctcctgggtg ACCACGGTCT TCGACcattc acggaggacc tggggagggg tggacgggtc gaggacccac EXON $3 \leftarrow$
12901 agggcccttt cctgcccctg aagccagtgg cacccagagg accaacaact ccattttctt tcccgggaaa ggacggggac ttcggtcacc gtgggtctcc tggttgttga ggtaaaagaa $\leftarrow$
12961 tttgaggggg tccccagacc ctggcctcac cctcccctca aggggcccct gaggttctga aaactccccc aggggtctgg gaccggagtg ggaggggagt tccccgggga ctccaagact
13021 tgtgtggtct gtttggaaac acacacagaa cacacacttg tattttttct cagccatttc acacaccaga caaacctttg tgtgtgtctt gtgtgtgaac ataaaaaaga gtcggtaaag
13081 cgagtaggtt gtagacatgg agccctttgt ctcttagtac ttcaagggag gtctttctta gctcatccaa catctgtacc tcgggaaaca gagaatcatg aagttccctc cagaaagaat
13141 agaacaaggt tttcacttac aaaactacgt aagcaagtgc agaaacacgc gcgcttctgg tcttgttcca aaagtgaatg ttttgatgca ttcgttcacg tctttgtgcg cgcgaagacc
13201 ccccgcatcg ccagcacgct ttcctgcaga gtcagcaggt ttggggccag tgacgggacc ggggcgtagc ggtcgtgcga aaggacgtct cagtcgtcca aaccccggtc actgccctgg
13261 caccaaacac agcccttatc agaatccagg ggctttgcag ggacaaggtg gccaagagga gtggtttgtg tcgggaatag tcttaggtcc ccgaaacgtc cctgttccac cggttctcct
13321 aatgtacttt gatgtaaatg tattttgata catgtatatc tatatttgat aaattcatgt ttacatgaaa ctacatttac ataaaactat gtacatatag atataaacta tttaagtaca
13381 atcaaattca ggaaattcag cattggtatc tcattatcat caaacctaca cactacgcat tagtttaagt cctttaagtc gtaaccatag agtaatagta gtttggatgt gtgatgcgta
13441 attttattgg ttgtcccaat gatgtcttcg tgagcatatt ttcctggtgg attttgtcac taaaataacc aacagggtta ctacagaagc actcgtataa aaggaccacc taaaacagtg
13501 ccaggctcac cctgtgctca gctgtctcag gctcagaatc tcctcactca ggctcaccct ggtccgagtg ggacacgagt cgacagagtc cgagtcttag aggagtgagt ccgagtggga
13561 gcactcagcc atctcaggcc ccaaatctcc taacccaggt tcacacactg cgctcagccg cgtgagtcgg tagagtccgg ggtttagagg attgggtcca agtgtgtgac gcgagtcggc
13621 tctaacgctc tgaatctcct cacccaggct cgccctacgc tcagccgtct ccttgggctt agattgcgag acttagagga gtgggtccga gcgggatgcg agtcggcaga ggaacccgaa
13681 cttgaatctg gacaagtcct cagctttctt tgtctttcct gtccttgggc gtttctgagg gaacttagac ctgttcagga gtcgaaagaa acagaaagga caggaacccg caaagactcc
13741 aagtcaggag agcagctttg tagcatgacc ctcgggggtc ttgtctggag tttcctctgc ttcagtcctc tcgtcgaaac atcgtactgg gagcccccag aacagacctc aaaggagacg
13801 ctggactcag gtgtgctttg gcaggagccc ctcagtgacg gtgtcctaag tgcccatgag gacctgagtc cacacgaaac cgtcctcggg gagtcactgc cacaggattc acgggtactc
13861 aggcactgct gtggccccgt cccagtcttg atgaggtcag cgtggacggc tctgctgaga tccgtgacga caccggggca gggtcagaac tactccagtc gcacctgccg agacgactct
13921 cggcatcccc ccggcctccc catcccgctg catgatcccc tcgcagccac tgtgggctgg gccgtagggg ggccggaggg gtagggcgac gtactagggg agcgtcggtg acacccgacc
13981 tgaggttctg ccctcatcaa atcgcacccg caggctcagc atccgcggag gtccccgcct actccaagac gggagtagtt tagcgtgggc gtccgagtcg taggcgcctc caggggcgga
14041 gaagccgcca tcactgcgat ggctgcagat tggtgccgtg ctggtccccg cctgccactc cttcggcggt agtgacgcta ccgacgtcta accacggcac gaccaggggc ggacggtgag
14101 acggctgcca acctaaggaa gggtctctcc ttctctgcag cctctctctc atcagcgtga tgccgacggt tggattcctt cccagagagg aagagacgtc ggagagagag tagtcgcact
14161 cccacagttt tattttgttc agtagtgcat gatctgtggc tgtcctttct tccccgagtt gggtgtcaaa ataaaacaag tcatcacgta ctagacaccg acaggaaaga aggggctcaa
14221 gctcaaatcc gctgggtctg gctgcgtgtc ctgtccatgc atccccgttc taccgttctg cgagtttagg cgacccagac cgacgcacag gacaggtacg taggggcaag atggcaagac
14281 agcactttct tgctttctgg agcaacatga tgctgcaggc ccatctccat gtcccttgcc tcgtgaaaga acgaaagacc tcgttgtact acgacgtccg ggtagaggta cagggaacgg
14341 ccagcctgga atcagatggt tttctgacgg ctcgggttcc tttccctgag agtctgggct ggtcggacct tagtctacca aaagactgcc gagcccaagg aaagggactc tcagacccga
14401 ccagcaccgt tgtagccgga caagccgcag acagaattcc tcagacactc ggttaaagaa ggtcgtggca acatcggcct gttcggcgtc tgtcttaagg agtctgtgag ccaatttctt
14461 ggaagggctt tatttggcca ggagtgtcag cagacttgcg tcttaagagc cgagctcccc ccttcccgaa ataaaccggt cctcacagtc gtctgaacgc agaattctcg gctcgagggg
14521 gaaaaagaaa ttcctagccc ttttaagggc ttacaactct aaggggtcca catgaaaggg
ctttttcttt aaggatcggg aaaattcccg aatgttgaga ttccccaggt gtactttccc 14581 ttgtgataga tagagcaagt gtggggaact tgactggggg ttacatgcat cagctgacag aacactatct atctcgttca caccccttga actgaccccc aatgtacgta gtcgactgtc 14641 aacagaaagt tttgcaatgc tttttcatac aatgtctgga atttacagat aacaccagta ttgtctttca aaacgttacg aaaaagtatg ttacagacct taaatgtcta ttgtggtcat
14701 gttttggtca ggggttaata ataataataa taataatatt attagtttaa ccaccagggc caaaaccagt ccccaattat tattattatt attattataa taatcaaatt ggtggtcccg
14761 cgggtggtgg cgccaaggtc atctggctat ttatcttact tctgtttctt tccaactttt gcccaccacc gcggttccag tagaccgata aatagaatga agacaaagaa aggttgaaaa
14821 tgctttctct cttttctcct gtcttataaa ctagagaagg ggagggggtg gagaagaagg acgaaagaga gaaaagagga cagaatattt gatctcttcc cctcccccac ctcttcttcc
14881 ggagggcagc aggagaagtg gtgtctcatt ccttaccgtc agcagcttcc tcatttctca cctcccgtcg tcctcttcac cacagagtaa ggaatggcag tcgtcgaagg agtaaagagt
14941 attttgcgtc ccacagggag cggctttggg tcctccagcc cacgccaccg cggcgagcag taaaacgcag ggtgtccctc gccgaaaccc aggaggtcgg gtgcggtggc gccgctcgtc
15001 tccccggagg agaagccggg tggctcccgt ctgtcctcgg actgagcagg gtggccaggc aggggcctcc tcttcggccc accgagggca gacaggagcc tgactcgtcc caccggtccg
15061 cccaggtgga aactaccttg gtcaggctgg gttttttccc tcccctttct ggaggttttc gggtccacct ttgatggaac cagtccgacc caaaaaaggg aggggaaaga cctccaaaag
15121 agctcctctg aaataaacac ctttcagtga gatttgtttg tttctgtgtg gattacattt tcgaggagac tttatttgtg gaaagtcact ctaaacaaac aaagacacac ctaatgtaaa
15181 aggctttctt ttctctcttg ctgggatact tgctgatgta cttttatttc tggaggactt tccgaaagaa aagagagaac gaccctatga acgactacat gaaaataaag acctcctgaa
15241 ctttaaaaaa ttctgtggtt ccaaaagtca gagccctaca aagaggcacc ctggggactg gaaatttttt aagacaccaa ggttttcagt ctcgggatgt ttctccgtgg gacccctgac
15301 tcgccctccc ctcccacctg cccccctgcc tgcgggactc accagcetct gggggccggg agcgggaggg gagggtggac ggggggacgg acgccctgag tggtcggaga cccccggccc
15361 cttatcctcc tgagcttccc tctacgaaag caagtgcaca aacacgcgct tctggccccg gaataggagg actcgaaggg agatgctttc gttcacgtgt ttgtgcgcga agaccggggc
15421 catcaccagc acccttccct gtagaggcag caggctcggg gccagtgacg ggacccacca gtagtggtcg tgggaaggga catctccgtc gtccgagccc cggtcactgc cctgggtggt
15481 aacacagccc tcatcagaat ccaggggatt tgcagggaca aagtggctga gaggaaagcg ttgtgtcggg agtagtctta ggtcccctaa acgtccctgt ttcaccgact ctcctttcgc
15541 ccctgggccc aggcctccca ggctcccctc cagcttttgt tcttctaaac tttgaggggc gggacccggg tccggagggt ccgaggggag gtcgaaaaca agaagatttg aaactccccg
15601 ttggcccaga accctgccca tgttctgggt ttcttggggg cctagggact gccacgttcc aaccgggtct tgggacgggt acaagaccca aagaaccccc ggatccctga cggtgcaagg
15661 tttgggacac agcaactccc ttggacaaag gttctgctga tgaagggctt gaagcccctg aaaccctgtg tcgttgaggg aacctgtttc caagacgact acttcccgaa cttcggggac
15721 ggccagctgc tagaggacgc tgaggcctct cctcttgggt cagcctcggg ctggagagta ccggtcgacg atctcctgcg actccggaga ggagaaccca gtcggagccc gacctctcat
15781 ggtcctgagt cctttctgcc gctagaggca gaggaggagg cgataccctg ccccccaggg ccaggactca ggaaagacgg cgatctccgt ctcctcctcc gctatgggac ggggggtccc
15841 gtctgctgtg gccgggagcc ttcagggagg ggctgtctgg gaggctgaag gagctccctg cagacgacac cggccetcgg aagtccctcc ccgacagacc ctccgacttc ctcgagggac $\rightarrow$

15901 tctctgtcca cagggacttc tgtcccaaat tccaggcgca gaggaggagg ggacagggtg agagacaggt gtccctgaag acagggttta aggtccgcgt ctcctcctcc cctgtcccac
15961 ggtcctcagt cctccatggc cagtcagcct cctgcaggtc ccgcagtgtg gccccctcct ccaggagtca ggaggtaccg gtcagtcgga ggacgtccag ggcgtcacac cgggggagga $\rightarrow$ EXON 4 N $\mathrm{F} \quad \mathrm{T} \quad \mathrm{G} \quad \mathrm{P} \quad \mathrm{G} \quad \mathrm{D} \quad \mathrm{P} \quad \mathrm{D} \quad \mathrm{S} \quad \mathrm{I} \quad \mathrm{R}$
16021 gacccctgac tccccctccc cagAACTTCA CAGGGCCGGG GGATCCTGAC TCCATTCGCT ctggggactg agggggaggg gtcTTGAAGT GTCCCGGCCC CCTAGGACTG AGGTAAGCGA

16081 GCGACACCCG GCCACAGCTG CTCATGAGGG GCTGTGCGGC TGACGACATC ATGGACCCCA CGCTGTGGGC CGGTGTCGAC GAGTACTCCC CGACACGCCG ACTGCTGTAG TACCTGGGGT

16141 CAAGCCTCGC TGAAACCCAG GAAGACCACA ATGGGGGCCA GAAGCAGCTG TCCCCACAAA GTTCGGAGCG ACTTTGGGTC CTTCTGGTGT TACCCCCGGT CTTCGTCGAC AGGGGTGTTT $\begin{array}{llllllll}\mathrm{K} & \mathrm{V} & \mathrm{T} & \mathrm{L} & \mathrm{Y} & \mathrm{L} & \mathrm{R} & \mathrm{P}\end{array}$
16201 AAGTGACGCT TTACCTGCGA CCAGgtaggc ttggcctcgg tggtggtgcc aggcaccgtc TTCACTGCGA AATGGACGCT GGTCcatccg aaccggagcc accaccacgg tccgtggcag EXON $4 \leftarrow$
16261 tgtgctggtt attggctctg gctgggagtc cggccctgct ctggggtgtc agggctgccc acacgaccaa taaccgagac cgaccctcag gccgggacga gaccccacag tcccgacggg $\leftarrow$
16321 agaaggcatg tgtccgggtc ggacgggtgt gtgggccagt gaagccccaa ggctgctgga tcttccgtac acaggcccag cctgcccaca cacccggtca cttcggggtt ccgacgacct
16381 aagctgagct cctcagtgcc cttccgctct gtccccaggt ggagctgccc cagcccttac ttcgactcga ggagtcacgg gaaggcgaga caggggtcca cctcgacggg gtcgggaatg
16441 catcggtgcc cagtgtctgc actctctcgc agcccgacgt tctgcacctt ttttttttaa gtagccacgg gtcacagacg tgagagagcg tcgggctgca agacgtggaa aaaaaaatt
16501 agggagacca aaccatttgc aaaacactat tccattacac actcagaagg ccagggtact tccctctggt ttggtaaacg ttttgtgata aggtaatgtg tgagtcttcc ggtcccatga
16561 cttagaagag taaaggctct gatcagtctt gcagccaaga gacctaagta actgtttaaa
gaatcttctc atttccgaga ctagtcagaa cgtcggttct ctggattcat tgacaaattt 16621 ccagtgttgc ccaaactgac ttgaccagag cacctgtttc gacttgacca gagcacctgt ggtcacaacg ggtttgactg aactggtctc gtggacaaag ctgaactggt ctcgtggaca 16681 ttcagaacac agtcccctgt cagtgctttg agaaacgccc tccttatcgg tggctcacgc aagtcttgtg tcaggggaca gtcacgaaac tctttgcggg aggaatagcc accgagtgcg
16741 ctgtaatccc agcactttgg gaggccgagg cgggtggatc acttgaggtc gggggttcga gacattaggg tcgtgaaacc ctccggctcc gcccacctag tgaactccag cccccaagct
16801 gaccagcctg gccaacattc tcctgctgcg gcaggagaat cacttgaacc tgggaggcag ctggtcggac cggttgtaag aggacgacgc cgtcctctta gtgaacttgg accctccgtc 16861 aggttgcagt gagccaagat tgtgccactg cactttagcc tgggcgacag ggcgagattc tccaacgtca ctcggttcta acacggtgac gtgaaatcgg acccgctgtc ccgctctaag
16921 tgtctcaaaa tgaaaagaaa ggaagggaag agaaggggag ggaagggagg gagggaaggg acagagtttt acttttcttt ccttcccttc tcttcccctc ccttccctcc ctcccttccc
16981 agggaaggga gggaagggga agggagggaa gaaacgaagc tccctccttg ctccaggaag tcccttccct cccttcccct tccctccctt ctttgcttcg agggaggaac gaggtccttc
17041 gcaggtgtgg tggtgcagca gatgggagaa ggaggcctgg tgtggccacc gacctgtgac cgtccacacc accacgtcgt ctaccctctt cctccggacc acaccggtgg ctggacactg
17101 ctcccaggcc acagtgaggc cagggaagga aggggaagcc atctcccagg gagagggtgc gagggtccgg tgtcactccg gtcccttcct tccccttcgg tagagggtcc ctctcccacg
17161 agacggggcc agggtgactt tagaggaacc agctggtgtc ggacaggctg gggctcagag tctgccccgg tcccactgaa atctccttgg tcgaccacag cctgtccgac cccgagtctc
17221 ctctgccctc aggggaagcc caggtcagcg gccactgagt tagccttggg gcagagtggc gagacgggag tccccttcgg gtccagtcgc cggtgactca atcggaaccc cgtctcaccg
17281 ccagggtgca ggcggccagg tgacagccag cagcagccgg gagggaggca gctgggagac ggtcccacgt ccgccggtcc actgtcggtc gtcgtcggcc ctccctccgt cgaccctctg
17341 agtgagggat ggggccgagc cattggggag gtcctgcagg agcagggccc aggtgtccct tcactcccta ccccggctcg gtaacccctc caggacgtcc tcgtcccggg tccacaggga
17401 gggggcagca ggggacaaaa ccccaccctg gtggtagggt caggcctgcc cctcgctgcc cccccgtcgt cccctgtttt ggggtgggac caccatccca gtccggacgg ggagcgacgg
17461 ccgtgagcag gcacaagccc cttcgtgctg gaacctccac ctctgacctc aaactcgagg ggcactcgtc cgtgttcggg gaagcacgac cttggaggtg gagactggag tttgagctcc
17521 gcctggccct gtggcccctg gaggcagctc tgatgtattt tccccacgcc ccacttcaca cggaccggga caccggggac ctccgtcgag actacataaa aggggtgcgg ggtgaagtgt
17581 gtgaggaaac tgaggctgac agaggtgttg cctgtggcca gagtttgtgc agctctgtgg cactcctttg actccgactg tctccacaac ggacaccggt ctcaaacacg tcgagacacc
17641 tgctggaggt gtcaggaagg atcgtcctgg agggtggggg tgcgcctgag tgggggcatc acgacctcca cagtccttcc tagcaggacc tcccaccccc acgcggactc accccegtag
17701 cttgtgccgt ggggaagggt gggggagtgg ctgcccacgg cctggcatgc atccccactg gaacacggca ccccttccca ccccctcacc gacgggtgcc ggaccgtacg taggggtgac
17761 tgcagggggc aaatgctgtg gcctctggga gggcacaggt gagctcgggt ggatgcaggt acgtcccccg tttacgacac cggagaccct cccgtgtcca ctcgagccca cctacgtcca
17821 gagggtaggc aaactcgggt gagcaccagg gagagatctt aacagcaccc cagggccaca ctcccatccg tttgagccca ctcgtggtcc ctctctagaa ttgtcgtggg gtcccggtgt
17881 ggctcccaga gcagggcggc aggatctggg aagggactgt cctctcgcgg ggtcctggcc ccgagggtct cgtcccgccg tcctagaccc ttccctgaca ggagagcgcc ccaggaccgg
17941 ctacccccgc agcctgctgg cccttcgctg agcctctact ctggaaggga agggacgggg gatgggggcg tcggacgacc gggaagcgac tcggagatga gaccttccct tccctgcccc
18001 ccccttggtc agccaccaaa ggcctatgtc tgtgcccttt gactgcagca tggccacatt ggggaaccag tcggtggttt ccggatacag acacgggaaa ctgacgtcgt accggtgtaa
18061 gctcgtgggg acatggtctt ggctgcctcc tccatgtcgc tgcagacgcc gcacggcccc cgagcacccc tgtaccagaa ccgacggagg aggtacagcg acgtctgcgg cgtgccgggg
18121 caaagccaaa cacatttact gtctggccet tcatagaaag tgattggggt ccttggccca gtttcggttt gtgtaaatga cagaccggga agtatctttc actaacccca ggaaccgggt
18181 gaggccaagt gaagcctctg agggctgcaa ggcacagcag ggcgagggcc agggccagga ctccggttca cttcggagac tcccgacgtt cogtgtcgtc cogctcccgg tccoggtcct
18241 gtgccgggtg tgaatgtggg tgagtgtgtg tgtgtgtgag cgcatgtgtg tgtggtgtgt cacggcccac acttacaccc actcacacac acacacactc gcgtacacac acaccacaca
18301 gtgtatgcat ggtgtgtgtg gtatgtatgt gtttagcatt gtgtggggtg tgtgtatatg cacatacgta ccacacacac catacataca caaatcgtaa cacaccccac acacatatac
18361 tgtggcacgt gtgtgtatgg gtgtgaatgc atgtgtgtgg catgtgtata tgtgtgtggc acaccgtgca cacacatacc cacacttacg tacacacacc gtacacatat acacacaccg
18421 atgcatgtgt atgtgagtgc atgtgtgtgg tgtatgtgcg tgagtgcatg tgtgtgtgtt tacgtacaca tacactcacg tacacacacc acatacacgc actcacgtac acacacacaa
18481 tggtgtgtgt atgtgtgtgc gagtgtgtgt gcgtgcatgt ggtgtgtttg tgtatatgaa accacacaca tacacacacg ctcacacaca cgcacgtaca ccacacaaac acatatactt
18541 caagtacata tagtatgtgt gctttgtgtc tggtgttgtg tatgagtgtg aggtgtgtgt gttcatgtat atcatacaca cgaaacacag accacaacac atactcacac tccacacaca
18601 gtatgcgtgc atgtgtggtg gttggtgttt gtggtgtgtg tgcatatttg tggtgtgttt catacgcacg tacacaccac caaccacaaa caccacacac acgtataaac accacacaaa
18661 gctgtgtgtg ggtgccatgt gcgtgtatat gtgggtgcgt gtgtgctacg tgtgactgtg cgacacacac ccacggtaca cgcacatata cacccacgca cacacgatgc acactgacac
18721 tgtgcatgag ttatatttgt gtgtgagtat tgtgtgtgtg catatatgtg ggtgtgtatg acacgtactc aatataaaca cacactcata acacacacac gtatatacac ccacacatac
18781 ccatgtgtga ctgtgtgtgc atgtggtgta tttatgtgta ggagtgcgtg tgtgtgtgtc ggtacacact gacacacacg tacaccacat aatacacat cctcacgcac acacacacag
18841 tgtgaggagg ggcacatacc ctttctccca gggtgtggca gcaggcaggt tgggggtgga
acactcctcc ccgtgtatgg gaaagagggt cccacaccgt cgtccgtcca acccccacct 18901 gtgggctctg cacttgtcct gccgcccagg ggggctgaag gctggagggt ggggcagagg cacccgagac gtgaacagga cggcgggtcc ccccgacttc cgacctccca ccccgtctcc 18961 ggacaggctt cgtcaccctc cctccctgct cattcattaa catgatgtgt gtttatggag cctgtccgaa gcagtgggag ggagggacga gtaagtaatt gtactacaca caaatacctc
19021 agctgacttg taccaggcac ggttctggga gccagggaca caggctggtt gcggtgacag tcgactgaac atggtccgtg ccaagaccct cggtccctgt gtccgaccaa cgccactgtc
19081 gtaaacagac agctacacaa ccgcacgtgt gactgagacg cgggcgagcg gaagtccagg catttgtctg tcgatgtgtt ggcgtgcaca ctgactctgc gcccgctcgc cttcaggtcc
19141 gctgggatgg agaaaggggc ctcctgggga cggcccogcgt gggccttgct gggaagtgcc cgaccctacc tctttccccg gaggacccct gccgggcgca cccggaacga cccttcacgg
19201 aggttcgtgt gatgcctgtc tatgtggctg ctgatgaggg cgcccgtggg actgcaggca tccaagcaca ctacggacag atacaccgac gactactccc gcgggcaccc tgacgtccgt
19261 tggggccggg ttaggtgtga gcccaggtga gccgctgcca ctctcaatga aaggcgtccc accccggccc aatccacact cgggtccact cggcgacggt gagagttact ttccgcaggg
19321 caatgtcgct ggcgtcaact ctaagcaccc caggcgggtt tcagtgtcac tgcgaggcac gttacagcga ccgcagttga gattcgtggg gtccgcccaa agtcacagtg acgctccgtg
19381 atgcctgcgt ccgggagtca tgtccttcct ctgaaactcc agcctgcctg gatgtcctgt tacggacgca ggccctcagt acaggaagga gactttgagg tcggacggac ctacaggaca $\rightarrow$
19441 gtctggtcct cagctctcct gccaggcggg gagaggggag gtcaggagga tttgggggct cagaccagga gtcgagagga cggtccgccc ctctcccctc cagtcctcct aaacccccga
19501 ggaggagagg gtgccagtgc acagcccgct cggggccaca cgcccccaga caccctgtgt cctcctctcc cacggtcacg tgtcgggcga gccccggtgt gcgggggtct gtgggacaca $\rightarrow$ EXON 5 G Q A A A F
19561 gagatgtgag gtgtggctcc ttttgttctg tccccaccgg cagGCCAGGC AGCAGCGTTC ctctacactc cacaccgagg aaaacaagac aggggtggcc gtcCGGTCCG TCGTCGCAAG $\begin{array}{lllllllllllllllllllll}\mathrm{N} & \mathrm{V} & \mathrm{T} & \mathrm{F} & \mathrm{R} & \mathrm{R} & \mathrm{A} & \mathrm{K} & \mathrm{G} & \mathrm{Y} & \mathrm{P} & \mathrm{I} & \mathrm{D} & \mathrm{L} & \mathrm{I} & \mathrm{Y} & \mathrm{L} & \mathrm{M} & \mathrm{D} & \mathrm{L}\end{array}$
19621 AACGTGACCT TCCGGCGGGC CAAGGGCTAC CCCATCGACC TGTACTATCT GATGGACCTC TTGCACTGGA AGGCCGCCCG GTTCCCGATG GGGTAGCTGG ACATGATAGA CTACCTGGAG $\begin{array}{llllllllllllllllllll}S & Y & S & M & L & D & D & L & R & N & V & K & K & L & G & G & D & L & L & R\end{array}$
19681 TCCTACTCCA TGCTTGATGA CCTCAGGAAT GTCAAGAAGC TAGGTGGCGA CCTGCTCCGG AGGATGAGGT ACGAACTACT GGAGTCCTTA CAGTTCTTCG ATCCACCGCT GGACGAGGCC $\begin{array}{lllllllllll}\text { A } & \mathrm{L} & \mathrm{N} & \mathrm{E} & \mathrm{I} & \mathrm{T} & \mathrm{E} & \mathrm{S} & \mathrm{G} & \mathrm{R} & \mathrm{I}\end{array}$
19741 GCCCTCAACG AGATCACCGA GTCCGGCCGC ATTGgtgagg cccaggcact gcaggacaaa CGGGAGTTGC TCTAGTGGCT CAGGCCGGCG TAACcactcc gggtccgtga cgtcctgttt EXON $5 \leftarrow$
19801 acccagtcct ttcccagacc ctggccactc ctgcctgtgg cccacacaac tcctcgcact tgggtcagga aagggtctgg gaccggtgag gacggacacc gggtgtgttg aggagcgtga

19861 caggtcccca gggtagatct gggggcccca gatgcaggtc cccccagggc atccacaggg gtccaggggt cccatctaga cccccggggt ctacgtccag gggggtcccg taggtgtccc
19921 ctgcagtcgg cctgcctccc tgccaggggc cacacccagt ccctggctcc ccaacacttg gacgtcagcc ggacggaggg acggtccccg gtgtgggtca gggaccgagg ggttgtgaac
19981 tagacattgt ccagccctcc acttcccctt ccatcttcta cttcctttat ttatttcttg atctgtaaca ggtcgggagg tgaaggggaa ggtagaagat gaaggaaata aataaagaac
20041 gtatgtttaa aagacaaaat tctgacagtg ccaattatta aacattaaag accacaacac catacaaatt ttctgtttta agactgtcac ggttaataat ttgtaatttc tggtgttgtg
20101 atatagacgg cacagcgacg gcagcactga agaaaccaca cagcaagctc tgctgggcca tatatctgcc gtgtcgctgc cgtcgtgact tctttggtgt gtcgttcgag acgacccggt
20161 agggctccgc tctacaccag cactgccttg ccacagagca ggtctccagc acacgtgtgt tcccgaggcg agatgtggtc gtgacggaac ggtgtctcgt ccagaggtcg tgtgcacaca
20221 gtcagtcaca gcgcatgtct tccccgccgg ccgtggcttc tcacatcggc actctgccac cagtcagtgt cgcgtacaga aggggcggcc ggcaccgaag agtgtagccg tgagacggtg
20281 cttccctaca tgcggggatg cacacatgca cgataacaca cacacacaca agcaaatgca gaagggatgt acgcccctac gtgtgtacgt gctattgtgt gtgtgtgtgt tcgtttacgt
20341 tgctcacaca aatgcacgct cacacacaaa tgcacgctca tgcacgtgaa tgcacgctca acgagtgtgt ttacgtgcga gtgtgtgttt acgtgcgagt acgtgcactt acgtgcgagt
20401 cacacaaatg catgttcaca catgcaatgc acactcacac acaaatgcat gctcaaacgc gtgtgtttac gtacaagtgt gtacgttacg tgtgagtgtg tgtttacgta cgagtttgcg
20461 aaatgcacgc tcacacatgc gaatgcacac tcacacatga atgtatcctc acatgcacaa tttacgtgcg agtgtgtacg cttacgtgtg agtgtgtact tacataggag tgtacgtgtt
20521 atggatgctc acacacgcga atgcatgctc acacacaaaa gcacacccaa actcaaatgc tacctacgag tgtgtgcgct tacgtacgag tgtgtgtttt cgtgtgggtt tgagtttacg
20581 acacacacgt gaatgcatgc tcacacacgt gaatgtatgc tcacacacat gctcaaacac tgtgtgtgca cttacgtacg agtgtgtgca cttacatacg agtgtgtgta cgagtttgtg
20641 acaaatgcat gctctcacat gtgaatgtat gctcacacat gaatgtatgc tcacacgcaa tgtttacgta cgagagtgta cacttacata cgagtgtgta cttacatacg agtgtgcgtt
20701 atgcatgctc acacacgcga atgtatgctc acatgcacaa atgcacgctc acacacgtga tacgtacgag tgtgtgcgct tacatacgag tgtacgtgtt tacgtgcgag tgtgtgcact
20761 atgcacgctc acacacaaat gcatgctcac acacatgaat gcatgctcac acttgcatgt tacgtgcgag tgtgtgttta cgtacgagtg tgtgtactta cgtacgagtg tgaacgtaca
20821 gcacacacat gcacacacag tcatacacac gtgcaagacc acatgtgcac acagacatgc cgtgtgtgta cgtgtgtgtc agtatgtgtg cacgttctgg tgtacacgtg tgtctgtacg
20881 tcacatgcac actcatgatc acccatgtat gtcttcactc acatgtataa atcgtacata agtgtacgtg tgagtactag tgggtacata cagaagtgag tgtacatatt tagcatgtat

20941 aacacacatg cccacatatg tacctggtca cctacgtgaa ctaggtgtgt ccatacctga ttgtgtgtac gggtgtatac atggaccagt ggatgcactt gatccacaca ggtatggact
21001 acatgctcac atgcatacag gaatacgtgc acacacagtt gtgctacgca tgtacacact tgtacgagtg tacgtatgtc cttatgcacg tgtgtgtcaa cacgatgcgt acatgtgtga
21061 catgtttaca catgcagatg gctgcacaca cagatacaca tacgcacctg ttcacatgcc gtacaaatgt gtacgtctac cgacgtgtgt gtctatgtgt atgcgtggac aagtgtacgg
21121 cacatgtata cctggtcaca cacatacatg caggaacaca catgcataca cttatgctca gtgtacatat ggaccagtgt gtgtatgtac gtccttgtgt gtacgtatgt gaatacgagt
21181 tccatgtaca ttgcacactc ccaggttcac acacacgccc acatacgttc acatgtgtgc aggtacatgt aacgtgtgag ggtccaagtg tgtgtgcggg tgtatgcaag tgtacacacg
21241 ttgcacaacg cacatgctca cacacacact ccacatgtgc catcttgctt tctccacctc aacgtgttgc gtgtacgagt gtgtgtgtga ggtgtacacg gtagaacgaa agaggtggag $\rightarrow$
21301 cctgctagga gagaggaaac aggcttgccc tcgtgcccct ttgaagctgg ccctcccacc ggacgatcct ctctcctttg tccgaacggg agcacgggga aacttcgacc gggagggtgg $\rightarrow$ EXON 6 G F G S F
21361 tgggcccagc ctgccacctc cccagcccct ccatgtgccc tgcagGCTTC GGGTCCTTCG acccgggtcg gacggtggag gggtcgggga ggtacacggg acgtcCGAAG CCCAGGAAGC $\begin{array}{lllllllllllllllllllll}V & D & K & T & V & L & P & F & N & V & T & H & P & D & K & L & R & N & P & C\end{array}$
21421 TGGACAAGAC CGTGCTGCCG TTCGTGAACA CGCACCCTGA TAAGCTGCGA AACCCATGCC ACCTGTTCTG GCACGACGGC AAGCACTTGT GCGTGGGACT ATTCGACGCT TTGGGTACGG $\begin{array}{llllllllllllllllllll}\text { P } & \mathrm{N} & \mathrm{K} & \mathrm{E} & \mathrm{K} & \mathrm{E} & \mathrm{C} & \mathrm{Q} & \mathrm{P} & \mathrm{P} & \mathrm{F} & \mathrm{A} & \mathrm{F} & \mathrm{R} & \mathrm{H} & \mathrm{V} & \mathrm{L} & \mathrm{K} & \mathrm{L} & \mathrm{T}\end{array}$
21481 CCAACAAGGA GAAAGAGTGC CAGCCCCCGT TTGCCTTCAG GCACGTGCTG AAGCTGACCA GGTTGTTCCT CTTTCTCACG GTCGGGGGCA AACGGAAGTC CGTGCACGAC TTCGACTGGT $\begin{array}{lllllllllllllllllllll}N & N & S & N & Q & F & Q & T & E & V & G & K & \text { Q } & L & I & S & G & N & L & D\end{array}$
21541 ACAACTCCAA CCAGTTTCAG ACCGAGGTCG GGAAGCAGCT GATTTCCGGA AACCTGGATG TGTTGAGGTT GGTCAAAGTC TGGCTCCAGC CCTTCGTCGA CTAAAGGCCT TTGGACCTAC

21601 CACCCGAGGG TGGGCTGGAC GCCATGATGC AGGTCGCCGC CTGCCCGgtg aggccgctgc GTGGGCTCCC ACCCGACCTG CGGTACTACG TCCAGCGGCG GACGGGCcac tccggcgacg EXON $6 \leftarrow$
21661 ccctgcttgg gtccccacgt tcccccagtg tggggggctc tgtcaggcgc tgtgtggggg gggacgaacc caggggtgca agggggtcac accccccgag acagtccgcg acacaccccc
21721 caggggggta ccctggccgg caggtcgtct gagggactcc agcctgggcc tggggagagc gtccccccat gggaccggcc gtccagcaga ctccctgagg tcggacccgg acccctctcg
21781 agctcctcac cctgctgctg cccatgcagg aaagggcagg aggaagggag ggtgaccacg tcgaggagtg ggacgacgac gggtacgtcc tttcccgtcc tccttccctc ccactggtgc
21841 agggctcgaa aatcatgcgc tgatttcatt gtttgctgtg agttaaattt gtgaccctgg tcccgagctt ttagtacgcg actaaagtaa caaacgacac tcaatttaaa cactgggacc

21901 gcaaacctta tgaagcctaa acattttgcg agcggccaag ggtggaacag ggccgctagg cgtttggaat acttcggatt tgtaaaacgc tcgccggttc ccaccttgtc ccggcgatcc

21961 cagcaggtcc cgtggttcct ggggtggaac caccctcctc cctccaggcc gggtgggcgc gtcgtccagg gcaccaagga ccccaccttg gtgggaggag ggaggtccgg cccacccgcg
22021 tggctgccca gtgtgtgtcg ctctccacgt cccgcccacc cagtttgtga ctgtcagcct accgacgggt cacacacagc gagaggtgca gggcgggtgg gtcaaacact gacagtcgga
22081 aacagagcag ccacagtccc tgggccatgt tgaggccaca tgcactgtcc tcccctcaca ttgtctcgtc ggtgtcaggg acccggtaca actccggtgt acgtgacagg aggggagtgt
22141 acccacctta gctctcctgc ttccctgcag tcccagcgca gacacactgc tcacgagtgt tgggtggaat cgagaggacg aagggacgtc agggtcgcgt ctgtgtgacg agtgctcaca
22201 gaggtcacag cctaggacac gcacggcccc cacatgcgcc tgccaggggc tgtcaatggt ctccagtgtc ggatcctgtg cgtgccgggg gtgtacgcgg acggtccccg acagttacca
22261 ggactcaagg cctgctgggc ccctcaggct cgttttcctc catgttgcac tcgttctgcg cctgagttcc ggacgacccg gggagtccga gcaaaaggag gtacaacgtg agcaagacgc
22321 tgtggcccac acgtctgcga gggtttttcc ccttggggat ggtttcctga gggtttgttc acaccgggtg tgcagacgct cccaaaaagg ggaaccccta ccaaaggact cccaaacaag
22381 ccagctctga agtgcaggga ccgcagccgt gtctctctcc ttgcctttcc acacgtgcct ggtcgagact tcacgtccct ggcgtcggca cagagagagg aacggaaagg tgtgcacgga
22441 ccactccctg cgcctcacat gtgcacgggg tgtttgaggg cogttgtctc ctcctgggcg ggtgagggac gcggagtgta cacgtgcccc acaaactccc ggcaacagag gaggacccgc
22501 tggcgtctgg gggaacacag ctgggtccct gagacccagg ggaaaggggg acccacacca accgcagacc cccttgtgtc gacccaggga ctctgggtcc cctttccccc tgggtgtggt $\rightarrow$
22561 gcctccgggc caccccacca gctctgcatc gtctcctctg ctccagaggg ccgggtctgg cggaggcccg gtggggtggt cgagacgtag cagaggagac gaggtctccc ggcccagacc

$$
\rightarrow \text { EXON } 7 \quad \text { E } \quad \mathrm{E} \quad \mathrm{I} \quad \mathrm{G} \quad \mathrm{~W} \quad \mathrm{R}
$$

22621 gaaactgagg caggtaaccc tgccccccac gccttcttct cagGAGGAAA TCGGCTGGCG ctttgactcc gtccattggg acggggggtg cggaagaaga gtcCTCCTTT AGCCGACCGC $\begin{array}{llllllllllllllllllll}\mathrm{N} & \mathrm{V} & \mathrm{T} & \mathrm{R} & \mathrm{L} & \mathrm{L} & \mathrm{V} & \mathrm{F} & \mathrm{A} & \mathrm{T} & \mathrm{D} & \mathrm{D} & \mathrm{G} & \mathrm{F} & \mathrm{H} & \mathrm{F} & \mathrm{A} & \mathrm{G} & \mathrm{D} & \mathrm{G}\end{array}$
22681 CAACGTCACG CGGCTGCTGG TGTTTGCCAC TGATGACGGC TTCCATTTCG CGGGCGACGG GTTGCAGTGC GCCGACGACC ACAAACGGTG ACTACTGCCG AAGGTAAAGC GCCCGCTGCC $\begin{array}{llllllllllllllllllll}\text { K } & \text { L } & G & A & I & L & T & P & N & D & G & R & C & H & L & E & D & N & L & Y\end{array}$
22741 GAAGCTGGGC GCCATCCTGA CCCCCAACGA CGGCCGCTGT CACCTGGAGG ACAACTTGTA CTTCGACCCG CGGTAGGACT GGGGGTTGCT GCCGGCGACA GTGGACCTCC TGTTGAACAT
$\begin{array}{llllll}K & R & S & N & E\end{array}$
22801 CAAGAGGAGC AACGAATTCg taagtcccca ccccaggcac ccaggcaccg cctggcagga GTTCTCCTCG TTGCTTAAGc attcaggggt ggggtccgtg ggtccgtggc ggaccgtcct EXON $7 \leftarrow$
22861 caccactgac ggaggagaca agggtggggt ctccacctga cagagcctcc ttctgaccca gtggtgactg cctcctctgt tcccacccca gaggtggact gtctcggagg aagactgggt

22921 aggagcagat tcctgaggaa acttctggaa gccccaagtg gcagcgtggg gtctccccgg tcctcgtcta aggactcctt tgaagacctt cggggttcac cgtcgcaccc cagaggggcc
22981 actggcctca ggccagggga gggtagggtt ggtgggggca gccaggctga ggcccggcct tgaccggagt ccggtcccct cccatcccaa ccacccccgt cggtccgact cogggccgga
23041 cgcccttgtg ggacacccag gctctgttgg tctccagccc cactgccccc ctacctgggc gcgggaacac cctgtgggtc cgagacaacc agaggtcggg gtgacggggg gatggacccg
23101 tgacagctgc tgtggaggta tagtaaccgc ccccaggcta cggctgcacc ctgccgtccc actgtcgacg acacctccat atcattggcg ggggtccgat gccgacgtgg gacggcaggg
23161 cgcctctggc cagggtccca tccccaagca tccgcctcct ccccctcccg gcctccactg gcggagaccg gtcccagggt aggggttcgt aggcggagga gggggagggc cggaggtgac
23221 tacgttccct gctgcccctg agtccgcctc ctccagtgtg gcccctccct gcccccattg atgcaaggga cgacggggac tcaggcggag gaggtcacac cggggaggga cgggggtaac
23281 cctgagctgg gtgagcggct gcggggacca tgaggaacat gcagggaggg acagaggcga ggactcgacc cactcgccga cgcccctggt actccttgta cgtccctccc tgtctccgct
23341 gaaggagccc aggatgcacg ggttaggatg agcctctctg cggaggcatc tcaatggctc cttcctcggg tcctacgtgc ccaatcctac tcggagagac gcctccgtag agttaccgag
23401 agaggggcca ggacttctgg ctgggatcag ccgtgggccg agaggcaacc actggtcaca tctccccggt cctgaagacc gaccctagtc ggcacccggc tctccgttgg tgaccagtgt
23461 aggggcttgt cctcgctggc cgtagtgcga cgctcttgca gcctgacgtt gtaggctctg tccccgaaca ggagcgaccg gcatcacgct gcgagaacgt cggactgcaa catccgagac
23521 ggggccgcaa aggactttag agatacaaga ctcaggtcct ccgccgggag ccacagacgg ccccggcgtt tcctgaaatc tctatgttct gagtccagga ggcggccctc ggtgtctgcc
23581 gagggacggc cctcgggtcc cggaactcgg gtagggagag ccgcacctga cactcatggc ctccctgccg ggagcccagg gccttgagcc catccctctc ggcgtggact gtgagtaccg
23641 ctctaccgaa actgagtgtc cctcagtgcg aaagcaggtc caccgtgtgg ggaaggctgg gagatggctt tgactcacag ggagtcacgc tttcgtccag gtggcacacc ccttccgacc
23701 gattctgccc ccgtggacat ccccagtccc acggtgagac gcctcaaatg ctgctcatgc ctaagacggg ggcacctgta ggggtcaggg tgccactctg cggagtttac gacgagtacg
23761 ctggactctg aaagcccgag tctatgcaca cattgcccag agggcgtggc agctctctgc gacctgagac tttcgggctc agatacgtgt gtaacgggtc tcccgcaccg tcgagagacg
23821 cctgcactcc tgcgtggcag cctctgcctc tccagccttc cccagagagg gttcgaagca ggacgtgagg acgcaccgtc ggagacggag aggtcggaag gggtctctcc caagcttcgt $\rightarrow$
23881 cgggcagggc tgacgctgag cggggcagac aggggcgggt acctggaagc cttgtcctgg gccegtcceg actgcgactc gccccgtctg tccccgccca tggaccttcg gaacaggacc $\rightarrow$ EXON $8 \quad D \quad Y \quad P \quad S \quad V \quad G \quad Q \quad L$
23941 actcggggcc aactgagcag gacctcctct ctccagGACT ACCCATCGGT GGGCCAGCTG tgagccccgg ttgactcgtc ctggaggaga gaggtcCTGA TGGGTAGCCA CCCGGTCGAC

24001 GCGCACAAGC TGGCTGAAAA CAACATCCAG CCCATCTTCG CGGTGACCAG TAGGATGGTG CGCGTGTTCG ACCGACTTTT GTTGTAGGTC GGGTAGAAGC GCCACTGGTC ATCCTACCAC K T Y E
24061 AAGACCTACG AGgtgagtgc tgttgggtcc cgagcatcca ttgggtgggg gcatgtttca TTCTGGATGC TCcactcacg acaacccagg gctcgtaggt aacccacccc cgtacaaagt EXON $8 \leftarrow$
24121 gccacaggcc agacccagcg ccactgcagg tctgagcgcc acctctgcat gctgggcacg cggtgtccgg tctgggtcgc ggtgacgtcc agactcgcgg tggagacgta cgacccgtgc

24181 tgagcaacca gccggcccag gtcctggatc cccctaggtg aggagcccga ggcttcccgt actcgttggt cggccgggtc caggacctag ggggatccac tcctcgggct ccgaagggca $\leftarrow$
24241 cggggacagg acgcctgcgt gtcgatagga ggaggaattc ctccccgtcc ctgaaactca gcccetgtcc tgcggacgca cagctatcct cctccttaag gaggggcagg gactttgagt

24301 gctgtgttaa cagcatgtct tggcattggt gatatttctg ttttttcctg gaatacaaca cgacacaatt gtcgtacaga accgtaacca ctataaagac aaaaaggac cttatgttgt
24361 caattcttta ttttatggaa acacacttat tttatgctca catacatttg aacatcaaca gttaagaaat aaaatacctt tgtgtgaata aaatacgagt gtatgtaaac ttgtagttgt
24421 ctctccgtga agataaaatt acacaaaaac gttcacaagc ttggagtgcg gagtgggatg gagaggcact tctattttaa tgtgtttttg caagtgttcg aacctcacgc ctcaccctac
24481 aatttcacga aggaagcaca gctgataacc agcctctgct cctagaaaaa gaggccccca ttaaagtgct tccttcgtgt cgactattgg tcggagacga ggatcttttt ctccgggggt
24541 gtgaaacacg ttcatcctgc atttccgctg ctagtgccat gtgttcagcc tgagcaggtc cactttgtgc aagtaggacg taaaggcgac gatcacggta cacaagtcgg actcgtccag
24601 agtgcgtcaa gacgctcagg tgctgttggg gttcccaaga cctttctcag gatcagtgat tcacgcagtt ctgcgagtcc acgacaaccc caagggttct ggaaagagtc ctagtcacta
24661 ttgccggaag gacttgcaga gctcaggaac cagtaaactc acagttacag tttattgcag aacggccttc ctgaacgtct cgagtccttg gtcatttgag tgtcaatgtc aaataacgtc
24721 agccgaccet cctgggtgcc agccgctgcc cgaggtggga ggatacagac aaaagtcagc
tcggctggga ggacccacgg tcggcgacgg gctccaccct cctatgtctg ttttcagtcg 24781 aaaggtcaaa ggtgcacagg agagaaacca ggcaccagct cccagtgtcc cctccctgca tttccagttt ccacgtgtcc tctctttggt ccgtggtcga gggtcacagg ggagggacgt
24841 ggggtgtggg cggcgctcca tcctcccagc gatgatgtgt ggcctcatgc acagaacatg ccccacaccc gccgcgaggt aggagggtcg ctactacaca ccggagtacg tgtcttgtac
24901 gtccccaagg aggctcttga gcctcggtgt ctactggggg tcagtcacag cggcgcggct caggggttcc tccgagaact cggagccaca gatgaccccc agtcagtgtc gccgcgccga
24961 gacaccccat ggctgacctc agtctctagc cctccagagg tcaagctgaa aactgtccag ctgtggggta ccgactggag tcagagatcg ggaggtctcc agttcgactt ttgacaggtc
25021 ggccccaccc caaatcacac tgtgggacta tccggtgtga cctgaggccc caggaacaat ccggggtggg gtttagtgtg acaccctgat aggccacact ggactccggg gtccttgtta
25081 gacaccccca tcgggtagga tgtaccaagg gctcagaggc tacctctcag gagctcagcc ctgtgggggt agcccatcct acatggttcc cgagtctccg atggagagtc ctcgagtcgg
25141 ggggtctacc actctttggc atgtgtgggg tgtggccagc caggcctgct gggtggagcc ccccagatgg tgagaaaccg tacacacccc acaccggtcg gtccggacga cccacctcgg
25201 ctgactgtgt aacagggatt cgctagaagg aaggaaacac tgcacgccct cctgaaagga gactgacaca ttgtccctaa gcgatcttcc ttcctttgtg acgtgcggga ggactttcct
25261 gcctccgccc atagcagcct gcacagtgtg tggccctgag gccacaggga cagaaggact cggaggcggg tatcgtcgga cgtgtcacac accgggactc cggtgtccct gtcttcctga
25321 gagctggagg aaggaggtgg ttgtcattgt actttacaca gccacgtgcc ccgttcttca ctcgacctcc ttcctccacc aacagtaaca tgaaatgtgt cggtgcacgg ggcaagaagt
25381 gggtggagct ggaggaacga gatagttgtc attgtacttt acacagccac gtgccccgtt cccacctcga cctccttgct ctatcaacag taacatgaaa tgtgtcggtg cacggggcaa
25441 cttcagggtg gacagggccc tgggtggctc acctcacacc cagctctgtt acttaccaca gaagtcccac ctgtcccggg acccaccgag tggagtgtgg gtcgagacaa tgaatggtgt
25501 ggacgcccca ggccettcct taccccgagc cctcagtttc ctcatctgtg aaaggggtag cctgcggggt ccgggaagga atggggctcg ggagtcaaag gagtagacac tttccccatc
25561 tggtgcccgg cgggcagtgc gtcgtgggta ctgagctcac ccactcaggg aattcacaga accacgggcc gcccgtcacg cagcacccat gactcgagtg ggtgagtccc ttaagtgtct
25621 gcagccgccc taaagggcce gattatcctc atccggacag tgaccgacct gctggccatt cgtcggcggg atttcccggg ctaataggag taggcctgtc actggctgga cgaccggtaa
25681 ccagggccat caccgccatc ttcttcaaaa tggcacgcag gaaaccaggg ccccctcttc ggtcccggta gtggcggtag aagaagtttt accgtgcgtc ctttggtccc gggggagaag
25741 cccctgcctg gggctctgcc cagctgggca gcctcacagc ccctgccctg ctgcttaggc ggggacggac cccgagacgg gtcgacccgt cggagtgtcg gggacgggac gacgaatccg
25801 tgaggagaca tactcggggt ttggggaagg aaagtgcccg ggagggccgc tgcttttgta actcctctgt atgagcccca aaccccttcc tttcacgggc cctcccggcg acgaaaacat
25861 aagtgaacat ttgctgaact acagggatcc attgtcaggc ccgtatccca gctacaagct ttcacttgta aacgacttga tgtccctagg taacagtccg ggcatagggt cgatgttcga
25921 caacctcgtg gtgtgttggg gccgacagcg agaccagcgg ttcctgccat aggttccccg gttggagcac cacacaaccc cggctgtcgc tctggtcgcc aaggacggta tccaaggggc
25981 ggctctggag tgggagcagg accctgcggg agacacggga tgggggcagg tgtctggggg ccgagacctc accctcgtcc tgggacgccc tctgtgccet acccccgtcc acagaccccc
26041 accctccgag tgggcaccca gagctggtca ggctgccggg acagtgggtc agcgcagggc tgggaggctc acccgtgggt ctcgaccagt ccgacggccc tgtcacccag tcgcgtcccg
26101 agggctgggg ccggacatcc acagagacca cctgctggtc aggatgtcac agccctgtgt tcccgacccc ggcctgtagg tgtctctggt ggacgaccag tcctacagtg tcgggacaca
26161 cccatctgta ggtggccatg acggtcccct cctcaccagc agtagcaggt cccaatcctt gggtagacat ccaccggtac tgccagggga ggagtggtcg tcatcgtcca gggttaggaa
26221 tgccatttgt catggcagcc agactccagg cgatggtgtg gcaccagggc caggcactgg acggtaaaca gtaccgtcgg tctgaggtcc gctaccacac cgtggtcccg gtccgtgacc
26281 aggcaggtag agtgcacccc atgcctactg ctgagccttg gggcacacaa gccttcctca tccgtccatc tcacgtgggg tacggatgac gactcggaac cccgtgtgtt cggaaggagt
26341 gcctttcctg gtcaccttcg ggggtggtca aggcaacagg cagcgagtgt ggacacacgt cggaaaggac cagtggaagc ccccaccagt tccgttgtcc gtcgctcaca cctgtgtgca
26401 gccacgcccc ctccgagtgt gtttctgcac caagcgcaca gcaggctcag caggacagga cggtgcgggg gaggctcaca caaagacgtg gttcgcgtgt cgtccgagtc gtcctgtcct
26461 tgtgggaggg cccagctctc ccaagacagg gatggcccca gactgggacg gatggaagtg acaccctccc gggtcgagag ggttctgtcc ctaccggggt ctgaccctgc ctaccttcac
26521 ggcggtgggc ggagggcgca gaggagggtg gtggccaggc tgtgaccggc cccaggagct ccgccacccg cctcccgcgt ctcctcccac caccggtccg acactggccg gggtcctcga
26581 ccatggaagg gactgggccc atgcagggca gtcatgggag ggtttgggcg gggaggtgac ggtaccttcc ctgacccggg tacgtcccgt cagtaccctc ccaaacccgc ccctccactg
26641 ccgaccatct gctccgtgtt cctggaaggt gaaatgatgc cgcaatcttc catgaaagaa ggctggtaga cgaggcacaa ggaccttcca ctttactacg gcgttagaag gtactttctt
26701 gctgagcatg tgtgagcatg ggggaggctg gctgcgcccc tgtgcatctg gccggccaga cgactcgtac acactcgtac cccctccgac cgacgcgggg acacgtagac cggccggtct
26761 accgatgtca ccgggctgtt taacacccat ggaaactcag tagagtgcca agagccgatg tggctacagt ggcccgacaa attgtgggta cctttgagtc atctcacggt tctcggctac
26821 gcttcctggt gattttactg tgatccaccc tcgtggggct acttccgcct accatctgtg cgaaggacca ctaaaatgac actaggtggg agcaccccga tgaaggcgga tggtagacac
26881 tggtggtagg aattccctac ctgtgtggtg gaaattctgg ccaacactgg ccaccatgca accaccatcc ttaagggatg gacacaccac ctttaagacc ggttgtgacc ggtggtacgt
26941 cgtcctccca acccagtgat attgcgagct tgaaatccac caggcaggat cactcacacc gcaggagggt tgggtcacta taacgctcga actttaggtg gtccgtccta gtgagtgtgg
27001 acaggcagga gtactcacac cacaggcggg agcagtcaca ccgcaggcag gatcactcac
tgtccgtcct catgagtgtg gtgtccgccc tcgtcagtgt ggcgtccgtc ctagtgagtg 27061 actgcaggcg ggatcactca caccgcaggc gggagcactc acaccgcagg cgggatcact tgacgtccgc cctagtgagt gtggcgtccg ccctcgtgag tgtggcgtcc gccctagtga 27121 cacaccgcag gcaggaggac tcacacgggc aggcctggac aatgctgcag ctcagggctt gtgtggcgtc cgtcctcctg agtgtgcccg tccggacctg ttacgacgtc gagtcccgaa
27181 tgctttcttt tcttatcctc atgatgcaac agtaattttt tttttatttt ttatttattt acgaaagaaa agaataggag tactacgttg tcattaaaaa aaaaataaaa aataaataaa
27241 tattttattt tatttatttt tttatttcat ttcatttcat ttcatttcat ttcattttat ataaaataaa ataaataaaa aaataaagta aagtaaagta aagtaaagta aagtaaaata
27301 tttattgaga cagagttttg ctcttgttgc ccaggctgga gtgccatggt gtgatctcag aaataactct gtctcaaaac gagaacaacg ggtccgacct cacggtacca cactagagtc
27361 ctcattgcga cctccgcctc ctgggttctc ctgcctcagt ctcctgagta gctgggctta gagtaacgct ggaggcggag gacccaagag gacggagtca gaggactcat cgacccgaat
27421 caggtgccca ccaccacgcc cagctaattt ttgtattttt agtagagacg gggtttcacc gtccacgggt ggtggtgcgg gtcgattaaa aacataaaaa tcatctctgc cccaaagtgg
27481 atgttggcca ggctggtctc aaactcctga cctcagtgat ccacccactt tggcctccca tacaaccggt ccgaccagag tttgaggact ggagtcacta ggtgggtgaa accggagggt
27541 aagtgctggg atcacaggcg tgagccatca tgcctggcct gggttttttt tgttttgttt ttcacgaccc tagtgtccgc actcggtagt acggaccgga cccaaaaaaa acaaaacaaa
27601 tgtttttttg agacaaggtc ttgctctgtc gcccaggctg gagtgcagtg gtgcgatctc acaaaaaaac tctgttccag aacgagacag cgggtccgac ctcacgtcac cacgctagag
27661 agctcactgc aacctccacc tcctgggttt aagtgatcct cctgcctcag cctcccgagt tcgagtgacg ttggaggtgg aggacccaaa ttcactagga ggacggagtc ggagggctca
27721 atctggtact acaggcatgt gacgccacac cactctaata tttgtacttt ttgtagagat tagaccatga tgtccgtaca ctgcggtgtg gtgagattat aaacatgaaa aacatctcta
27781 ggggtcttct tatgttgccc aggctggtcc tgaactcctg ggctcaggca atcttcccac ccccagaaga atacaacggg tccgaccagg acttgaggac ccgagtccgt tagaagggtg
27841 atcggcctcc cgagtactgg tattccaggc gtggcccctg ggcagctcag ggctttttac tagccggagg gctcatgacc ataaggtccg caccggggac ccgtcgagtc ccgaaaaatg
27901 ccccaagaac ctcttaccac acatcactga ccacacactc tggggaccaa gtgcaaaact ggggttcttg gagaatggtg tgtagtgact ggtgtgtgag acccctggtt cacgttttga $\rightarrow$
27961 gcgttgagga ccgtcccagc cattttgcag gtgcagaagc ccctggggtg agcccagtgc cgcaactcct ggcagggtcg gtaaaacgtc cacgtcttcg gggaccccac tcgggtcacg
$\rightarrow$ EXON 9 K
28021 gtggagatgc cgtggcacag gtcctagccg tgccatgtct ggtcttcttg cccaacagAA cacctctacg gcaccgtgtc caggatcggc acggtacaga ccagaagaac gggttgtcTT $\begin{array}{llllllllllllllllllll}\mathrm{L} & \mathrm{T} & \mathrm{E} & \mathrm{I} & \mathrm{I} & \mathrm{P} & \mathrm{K} & \mathrm{S} & \mathrm{A} & \mathrm{V} & \mathrm{G} & \mathrm{E} & \mathrm{L} & \mathrm{S} & \mathrm{E} & \mathrm{D} & \mathrm{S} & \mathrm{S} & \mathrm{N} & \mathrm{V}\end{array}$
28081 ACTCACCGAG ATCATCCCCA AGTCAGCCGT GGGGGAGCTG TCTGAGGACT CCAGCAATGT TGAGTGGCTC TAGTAGGGGT TCAGTCGGCA CCCCCTCGAC AGACTCCTGA GGTCGTTACA
$\begin{array}{lllllllll}\mathrm{V} & \mathrm{H} & \mathrm{L} & \mathrm{I} & \mathrm{K} & \mathrm{N} & \mathrm{A} & \mathrm{Y} & \mathrm{N}\end{array}$
28141 GGTCCAACTC ATTAAGAATG CTTACAATgt gagtctcccg caccactcag ctgggaccca CCAGGTTGAG TAATTCTTAC GAATGTTAca ctcagagggc gtggtgagtc gaccctgggt EXON $9 \leftarrow$
28201 tggggtggcc agcaatctcc ccactgcagg tcagctcctg ctgcaggtca gcaggagggt accccaccgg tcgttagagg ggtgacgtcc agtcgaggac gacgtccagt cgtcctccca

28261 ggtctgggag gaaagccect gccccgtggg ccagctgcag cctcccgacc tgccctcagg ccagaccctc ctttcgggga cggggcaccc ggtcgacgtc ggagggctgg acgggagtcc
28321 cctccaggac accctggccc ggcacttcca cggggagctc tgggtcctct ctcgttgatg ggaggtcctg tgggaccggg ccgtgaaggt gcccctcgag acccaggaga gagcaactac
28381 ttgtgggcga tgctccctga aacctgagct ccggactctt cattccagac tcttctctgc aacacccgct acgagggact ttggactcga ggcctgagaa gtaaggtctg agaagagacg
28441 gggtctctat ggagaaggga ggtggacagc tctggtggca aacagtgcag ctggggcagg cccagagata cctcttccct ccacctgtcg agaccaccgt ttgtcacgtc gaccccgtcc
28501 aggaggccac aggaaacccc cacagtggac ttgggccetg ggaagacacc ctggggggtg tcctccggtg tcctttgggg gtgtcacctg aacccgggac ccttctgtgg gaccccccac
28561 gaattcctct tgctctgttc acctgtcctg ggcaagctcc agagaagccc ttaacccctg cttaaggaga acgagacaag tggacaggac ccgttcgagg tctcttcggg aattggggac
28621 aatttctccc tggtgcacgg cccatcgagc tgcaccctgg gctgctcccc tggggagccc ttaaagaggg accacgtgcc gggtagctcg acgtgggacc cgacgagggg acccctcggg
28681 atggccctgg tgggcatagg taggcatagg tgcgcactag gtgccattca gggccccagg taccgggacc acccgtatcc atccgtatcc acgcgtgatc cacggtaagt cccggggtcc
28741 caccagaggt cccgactgct ctggaaatct acccccaggt gagagatgtt ctcagaattc gtggtctcca gggctgacga gacctttaga tgggggtcca ctctctacaa gagtcttaag
28801 tggcctggat ccttctaggc caaaggtcct gctggcttga gggagctcag cacgttagcg accggaccta ggaagatccg gtttccagga cgaccgaact ccctcgagtc gtgcaatcgc
28861 cctccgacaa acgttgggtc attgaattaa actgaaggcc cgtggaggag gtgccatttg ggaggctgtt tgcaacccag taacttaatt tgacttccgg gcacctcctc cacggtaaac
28921 cttctcgtag cctcccgtga gccccaaacc cccagagcag accctcctgg gtgccagcca gaagagcatc ggagggcact cggggtttgg gggtctcgtc tgggaggacc cacggtcggt
28981 ctgcccgagg tgtttctgtc cttacacaga ggatgcaggt gacgccactg ccctggccat gacgggctcc acaaagacag gaatgtgtct cctacgtcca ctgcggtgac gggaccggta
29041 tcacccgatg cccctatgaa cctgtccctc tgcctcctct caatttccgt ctgtctctca
agtgggctac ggggatactt ggacagggag acggaggaga gttaaaggca gacagagagt 29101 gtctttgtct ctctttctct gtctctctct gtctgtctct ctctgtctct gccccatctc cagaaacaga gagaaagaga cagagagaga cagacagaga gagacagaga cggggtagag 29161 tgtctgtctc tgtctctgcc tctctctgtt tctctccatc tctctgtctc tggctctggc acagacagag acagagacgg agagagacaa agagaggtag agagacagag accgagaccg
29221 tctgcccatc tctgtctgtc tctgcctctc tctgtttctc cccatctctc ttatcgcttg agacgggtag agacagacag agacggagag agacaaagag gggtagagag aatagcgaac
29281 aacaatttat gcagctcctg ccctgtgcta ggtcctgggg actcaaggct gaggagcagc ttgttaaata cgtcgaggac gggacacgat ccaggacccc tgagttccga ctcctcgtcg
29341 agaagctccc acccctgagg ctcacaccct ggctccatgc tagtccctgt ggtcaggctg tcttcgaggg tggggactcc gagtgtggga ccgaggtacg atcagggaca ccagtccgac
29401 tctcgagggc aggactgtgg ccctggcagc cctgccatcc tcagcacatc ctgctgtgct agagctcccg tcctgacacc gggaccgtcg ggacggtagg agtcgtgtag gacgacacga $\rightarrow$
29461 ctcctctgtg tcattagctg acactgggga cttgcactga gaggagaagt ggagacgggt gaggagacac agtaatcgac tgtgacccct gaacgtgact ctcctcttca cctctgccca
29521 ctaaaccgct gggccaggca gggccaggac aacagaaaca ctcgccccca agagtaaccg gatttggcga cccggtccgt cccggtcctg ttgtctttgt gagcgggggt tctcattggc

## F40318

 29581 ctctctgccc gcagAAACTC TCCTCCAGGG TCTTCCTGGA TCACAACGCC CTCCCCGACA gagagacggg cgtcTTTGAG AGGAGGTCCC AGAAGGACCT AGTGTTGCGG GAGGGGCTGT $\begin{array}{llllllllllllllllllll}T & L & K & V & T & Y & D & S & F & C & S & N & G & V & T & H & R & N & Q & P\end{array}$
29641 CCCTGAAAGT CACCTACGAC TCCTTCTGCA GCAATGGAGT GACGCACAGG AACCAGCCCA GGGACTTTCA GTGGATGCTG AGGAAGACGT CGTTACCTCA CTGCGTGTCC TTGGTCGGGT R G D C D G V Q I $\quad \mathrm{N} \quad \mathrm{V} \quad \mathrm{P}$
29701 GAGGTGACTG TGATGGCGTG CAGATCAATG TCCCGgtgag cctggccaca agggtcccca CTCCACTGAC ACTACCGCAC GTCTAGTTAC AGGGCcactc ggaccggtgt tcccaggggt EXON $10 \leftarrow$
29761 tcccagcttg ctgaggggct ctggggacag ccagaaggag ccagggtccc catcccagca agggtcgaac gactccccga gacccctgtc ggtcttcctc ggtcccaggg gtagggtcgt
29821 cactgagggg tcctggggac agccagaagg agccagggtc cccatcccag cactctgacg gtgactcccc aggacccctg tcggtcttcc tcggtcccag gggtagggtc gtgagactgc
29881 ggggcaggga ggggcacaca ggcatcagag gctgtctggg agcaggcagc atccactgga ccccgtccct ccccgtgtgt ccgtagtctc cgacagaccc tcgtccgtcg taggtgacct
29941 aggaccctgt ttctttcctt cacatcctct ctgccgaggc cagacgcttt ccttgtgaag tcctgggaca aagaaaggaa gtgtaggaga gacggctccg gtctgcgaaa ggaacacttc $\leftarrow$
30001 ctgatgtcag gaacgtcagt gagagggtgc cgagagggtg cttctcagtg agagggtgct gactacagtc cttgcagtca ctctcccacg gctctcccac gaagagtcac tctcccacga
30061 gggagggttc ttcacacacg ccttggcccc caagagtcct actcatcaac ctggagccca ccctcccaag aagtgtgtgc ggaaccgggg gttctcagga tgagtagttg gacctcgggt
30121 aatccctggg tcaggccact gcagactgag ggatggagtg atggctgcgc cccctggggg ttagggaccc agtccggtga cgtctgactc cctacctcac taccgacgeg ggggaccccc
30181 cagacgccgg acacacagaa caggcgcaaa cggcaaacct gggtagggtt agttgtcatg gtctgcggcc tgtgtgtctt gtccgcgttt gccgtttgga cccatcccaa tcaacagtac
30241 attttctgcg tgtttcattc tgtttttttg acagaaacat taaggcatgg gccacacagg taaaagacgc acaaagtaag acaaaaaaac tgtctttgta attccgtacc cggtgtgtcc
30301 tggttaggaa tggacactga gttccttctc taatctttca ggtttaagtt tattatccaa accaatcctt acctgtgact caaggaagag attagaaagt ccaaattcaa ataataggtt
30361 tcatttgttg tttctttttg ttttttgttt tttgtttttt tttttgctag ctgtgaaata agtaaacaac aaagaaaaac aaaaaacaaa aaacaaaaaa aaaaacgatc gacactttat
30421 atattcagca agacacttgc tttcattgca ttttcttaat atgactttgg aattccttca tataagtcgt tctgtgaacg aaagtaacgt aaaagaatta tactgaaacc ttaaggaagt
30481 aacatgtttc ttggtaaatg ttgcctgtgg attttttttt tttttttttt ttttgagatg ttgtacaaag aaccatttac aacggacacc taaaaaaaa aaaaaaaaa aaaactctac
30541 gagtttcgct ctgtcgccca ggctggagtg cactggtgcg atcttggctc actacaatcc ctcaaagcga gacagcgggt ccgacctcac gtgaccacgc tagaaccgag tgatgttagg
30601 aagtgattct cctgactcag cctcctgagt agctgggact acaggcgtgt gccaccacgc ttcactaaga ggactgagtc ggaggactca tcgaccctga tgtccgcaca cggtggtgcg
30661 tcggctaatt tttgtatttt tagtagagac agagtttcgc catgttggcc aggctggtct agccgattaa aaacataaaa atcatctctg tctcaaagcg gtacaaccgg tccgaccaga
30721 caaactcttg ccctcagatg atctacccgc ctcggcctcc cacagtgctg ggattacagg gtttgagaac gggagtctac tagatgggcg gagccggagg gtgtcacgac cctaatgtcc
30781 cgtgagccgc ggcgcccagc tgtcccatgg attctcacca gaacgtttcc ctcatgttgg gcactcggcg ccgcgggtcg acagggtacc taagagtggt cttgcaaagg gagtacaacc
30841 gtggtctgag tgggagtgtg ctcagcttct cctgatttgc tgtctgcgtg tcacatgtct caccagactc accctcacac gagtcgaaga ggactaaacg acagacgcac agtgtacaga
30901 ctgctgatgt gtggccgggg agatcctggt gccggccagt gcgggagtgc gggcagagca gacgactaca caccggcccc tctaggacca cggccggtca cgccctcacg cccgtctcgt

30961 cctccttgcc ggctgacttt ggctctcgga tctgagcatc agctcttcct gctcctggtc ggaggaacgg ccgactgaaa ccgagagcct agactcgtag tcgagaagga cgaggaccag

31021 accccttcct ccccacgctg aaccccccag gacccctgca gaggccagga ggtgagggtt tggggaagga ggggtgcgac ttggggggtc ctggggacgt ctccggtcct ccactcccaa
31081 gggagctggg ctgccaccct ggccaccgag ggtccccacc tgagcccagc actgccctgc ccctcgaccc gacggtggga ccggtggctc ccaggggtgg actcgggtcg tgacgggacg $\rightarrow$ EXON 11
$\begin{array}{llllllllllllllllll}\text { I } & T & F & Q & V & K & V & T & \mathrm{~T} & \mathrm{~T} & \mathrm{E} & \mathrm{C} & \mathrm{I} & \mathrm{Q} & \mathrm{E} & \mathrm{Q} & \mathrm{S} & \mathrm{F} \\ \mathrm{V}\end{array}$
31141 agatcacctt ccaggtgaag gtcacggcca cagagtgcat ccaggagcag tcgtttgtca tctagtggaa ggtccacttc cagtgccggt gtctcacgta ggtcctcgtc agcaaacagt I R A $\mathrm{L} \quad \mathrm{G} \quad \mathrm{F} \quad \mathrm{T} \quad \mathrm{D} \quad \mathrm{I} \quad \mathrm{V} \quad \mathrm{T} \quad \mathrm{V} \quad \mathrm{Q} \quad \mathrm{V} \quad \mathrm{L} \quad \mathrm{P} \quad \mathrm{Q} \quad \mathrm{C} \quad \mathrm{E} \quad \mathrm{C}$
31201 tccgggcgct gggcttcacg gacatagtga ccgtgcaggt Ecttccccag tgtgagtgcc aggcccgcga cccgaagtgc ctgtatcact ggcacgtcca agaaggggtc acactcacgg $\begin{array}{lllllllllllllllllllll}R & C & R & D & Q & S & R & D & R & S & L & C & H & G & K & G & F & L & E & C\end{array}$
31261 ggtgccggga ccagagcaga gaccgcagcc tctgccatgg caagggcttc ttggagtgcg ccacggccct ggtctcgtct ctggcgtcgg agacggtacc gttcccgaag aacctcacgc G I $C \quad R$
31321 gcatctgcag gtgaggcggg cgcaggtccc tgttgaacca tccccgagcc ccaccccaca cgtagacgtc cactccgccc gcgtccaggg acaacttggt aggggctcgg ggtggggtgt EXON $11 \leftarrow$
31381 gggcaggtgt cggcaggtga ggtgggtgca ggtggctcca tcccttcccc caccccacgg cccgtccaca gccgtccact ccacccacgt ccaccgaggt agggaagggg gtggggtgcc
31441 agcaggcatc ctgggggccc ccttctgcag gaggtgggga gatttgggat tggaacccag tcgtccgtag gacccccggg ggaagacgtc ctccacccct ctaaacccta accttgggtc

31501 aacccctcgg caagccctca cccgtgctcc tgagccaggg caccagcaat gccagcctgt ttggggagcc gttcgggagt gggcacgagg actcggtccc gtggtcgtta cggtcggaca
31561 actctacact ctctggccat ggaggaccca aggcagggga cttgtgggca ggaggacccc tgagatgtga gagaccggta cctcctgggt tccgtcccct gaacacccgt cctcctgggg
31621 ctcaccagag gatggaggcc catctcaggg gcacaaggag tccccgctgg gcgtgcccat gagtggtctc ctacctccgg gtagagtccc cgtgttcctc aggggcgacc cgcacgggta
31681 ggccttgagt gggtcacttt ctctgcccga ccttccttga ccctaggagc tggcttctct ccggaactca cccagtgaaa gagacgggct ggaaggaact gggatcctcg accgaagaga
31741 ctgaccttgg ggacccaagg ccatcaggct ggcagccctg gggcatgggt ggctagagag gactggaacc cctgggttcc ggtagtccga ccgtcgggac cccgtaccca ccgatctctc
31801 cctcctggga caggtgatgg cttctgggag cctgggcccc ataggccacc tcttgggcgt ggaggaccct gtccactacc gaagaccctc ggacccgggg tatccggtgg agaacccgca
31861 gcacaccccc tgccaggcgt gggcacccca cctcagggag aagtggggga caggtggacc cgtgtggggg acggtccgca cccgtggggt ggagtccctc ttcaccccct gtccacctgg
31921 cagtggggag gttcacatgg cactcagcca ggcggcctgc ccgggggctg gatggctgtc gtcacccctc caagtgtacc gtgagtcggt ccgccggacg ggcccccgac ctaccgacag
31981 cccctttcct tgcttgcctg gccagagctg ctaggctgac ctcccaggac tctgccccct ggggaaagga acgaacggac cggtctcgac gatccgactg gagggtcctg agacggggga
32041 ttgctctggt cacctcctcc cctgtcctga ccgcagtgcc aatccacgtc ctggtggcag aacgagacca gtggaggagg ggacaggact ggcgtcacgg ttaggtgcag gaccaccgtc
32101 tgcccttcct caccttcagg tccctgctgc ccccagcctc ctcccatatc tcaggacaat acgggaagga gtggaagtcc agggacgacg ggggtcggag gagggtatag agtcctgtta
32161 gatggtctct gtggtcttgg tggtctctgt tgctgtcctc gtgcctcagt ggccaccccg ctaccagaga caccagaacc accagagaca acgacaggag cacggagtca ccggtggggc
32221 cccactgtcc tgggtgtgca gccctggggc ggcctcttca cgggactgta ggggtgactc gggtgacagg acccacacgt cgggaccccg ccggagaagt gccctgacat ccccactgag
32281 agtcggaaca cttggcaggt gtcggggggt gccctgccac aggtactcag acactcggtc tcagccttgt gaaccgtcca cagcccccca cgggacggtg tccatgagtc tgtgagccag
32341 ctgcagcttc ctcctctggg gagcaggaga gcccgtggca gtgccgtctc tgaggatgcc gacgtcgaag gaggagaccc ctcgtcctct cgggcaccgt cacggcagag actcctacgg
32401 tcagggtgtg tgtggctcct gcctgccagc ggcttgatgc cagcttccac atctgtaaaa agtcccacac acaccgagga cggacggtcg ccgaactacg gtcgaaggtg tagacatttt
32461 accccttgtc cctgcattct cccacaacct ggcctgtcag tgcacgaacg caggagggag tggggaacag ggacgtaaga gggtgttgga ccggacagtc acgtgcttgc gtcctccctc
32521 ggcggggctg tggcaggagg cagtgccgag tccccactgg ccaccatgtg gggcacgcag ccgccccgac accgtcctcc gtcacggctc aggggtgacc ggtggtacac cccotgcgtc
32581 gagcagacac cagccactca cagagcaccg gtatcgcagc acaggccatg tgcatggctg ctcgtctgtg gtcggtgagt gtctcgtggc catagcgtcg tgtccggtac acgtaccgac
32641 caacacagct ccacagtgag cctgccgagc cactgcatgc gcaggccaag gccaggtgtc gttgtgtcga ggtgtcactc ggacggctcg gtgacgtacg cgtccggttc cggtccacag
32701 tgctttcccc tcgggacatc ctggggaggg aacgcggggc ccctgcacgt tttgagcaca acgaaagggg agccctgtag gacccctccc ttgcgccccg gggacgtgca aaactcgtgt $\rightarrow$
32761 tttagtccgt ggcaagtgct cagtaggcct cgaggaaagc aaaatgtgcc ataggggtgt aatcaggca ccgttcacga gtcatccgga gctcctttcg ttttacacgg tatccccaca

32821 tctccaccaa gggcaagggg cctggaggac agggcggctg tcccttgtcc cacatcactg agaggtggtt cccgttcccc ggacctcctg tcccgccgac agggaacagg gtgtagtgac $\rightarrow$ EXON $12 \quad$ C $\quad$ D $\quad$ T $\quad$ G $\quad$ Y $\quad$ I $\quad$ G $\quad$ K
32881 gggggagcct gaccttgggg cccctcctgt ccccagGTGT GACACTGGCT ACATTGGGAA ccccctcgga ctggaacccc ggggaggaca ggggtcCACA CTGTGACCGA TGTAACCCTT $\begin{array}{llllllllllllllllllll}N & C & E & C & Q & T & Q & G & R & S & S & Q & E & L & E & G & S & C & R & K\end{array}$
32941 AAACTGTGAG TGCCAGACAC AGGGCCGGAG CAGCCAGGAG CTGGAAGGAA GCTGCCGGAA

TTTGACACTC ACGGTCTGTG TCCCGGCCTC GTCGGTCCTC GACCTTCCTT CGACGGCCTT $\begin{array}{lllllllllllllllllll}D & N & N & S & I & I & C & S & G & L & G & D & C & V & C & G & Q & L & L\end{array}$
33001 GGACAACAAC TCCATCATCT GCTCAGGGCT GGGGGACTGT GTCTGCGGGC AGTGCCTGTG CCTGTTGTTG AGGTAGTAGA CGAGTCCCGA CCCCCTGACA CAGACGCCCG TCACGGACAC $\begin{array}{llllllllllllllllllll}H & T & S & D & V & F & G & K & L & I & Y & G & Q & Y & C & E & C & D & T & I\end{array}$
33061 CCACACCAGC GACGTCCCCG GCAAGCTGAT ATACGGGCAG TACTGCGAGT GTGACACCAT GGTGTGGTCG CTGCAGGGGC CGTTCGACTA TATGCCCGTC ATGACGCTCA CACTGTGGTA $\begin{array}{lllllllllllll}N & C & E & R & Y & N & G & Q & V & C & G & G & P\end{array}$
33121 CAACTGTGAG CGCTACAACG GCCAGGTCTG CGGCGGCCCG Ggtgagcccg tgggtccctg GTTGACACTC GCGATGTTGC CGGTCCAGAC GCCGCCGGGC Ccactcgggc acccagggac EXON $12 \leftarrow$
33181 ctggacaacg gccgtcctgc ggcacagact tggtgttggg gggtgcgttc tggaccattc gacctgttgc cggcaggacg ccgtgtctga accacaaccc cccacgcaag acctggtaag

33241 aacgagtgga acagaaaggg gggatcctga ggacagacag tgcaagtctc gcccagccac ttgctcacct tgtctttccc ccctaggact cctgtctgtc acgttcagag cgggtcggtg

33301 cgggcgtcag caggaacttg gatggccctg caccccgccc tgcctcagtt tctgtgtcca gcccgcagtc gtccttgaac ctaccgggac gtggggcggg acggagtcaa agacacaggt
33361 tgaggacaag gcgtccttgt ggggagggcc cacaggagga acggtcatag ccctattgag actcctgttc cgcaggaaca cccctcccgg gtgtcctcct tgccagtatc gggataactc
33421 gtcccattca cagcagggca ggggaggtcc agactggcca ggaggcggcc acgtggctgg cagggtaagt gtcgtcccgt cccctccagg tctgaccggt cctccgccgg tgcaccgacc
$\rightarrow$
33481 cgggggcaga agggcctggc tagtggcccc atgccttttg ctcaggaaca tcccccaggc gcccccgtct tcccggaccg atcaccgggg tacggaaaac gagtccttgt agggggtccg

33541 tggaggaggc aggcgggccg gagggagcaa ccgcagtggg gtttcgggcg ggtctcggtt acctcctccg tccgcccggc ctccctcgtt ggcgtcaccc caaagcccgc ccagagccaa $\rightarrow$ EXON $13 \quad G \quad R \quad G \quad L \quad C \quad E$
33601 cccaggccaa gcaggagctt agccgtgctg ccccgtcttc cagGGAGGGG GCTCTGCTTC gggtccggtt cgtcctcgaa tcggcacgac ggggcagaag gtcCCTCCCC CGAGACGAAG

33661 TGCGGGAAGT GCCGCTGCCA CCCGGGCTTT GAGGGCTCAG CGTGCCAGTG CGAGAGGACC ACGCCCTTCA CGGCGACGGT GGGCCCGAAA CTCCCGAGTC GCACGGTCAC GCTCTCCTGG $\begin{array}{llllllllllllllllllll}\mathrm{T} & \mathrm{E} & \mathrm{G} & \mathrm{C} & \mathrm{L} & \mathrm{N} & \mathrm{P} & \mathrm{R} & \mathrm{R} & \mathrm{V} & \mathrm{E} & \mathrm{C} & \mathrm{S} & \mathrm{G} & \mathrm{R} & \mathrm{G} & \mathrm{R} & \mathrm{C} & \mathrm{R} & \mathrm{C}\end{array}$
33721 ACTGAGGGCT GCCTGAACCC GCGGCGTGTT GAGTGTAGTG GTCGTGGCCG GTGCCGCTGC TGACTCCCGA CGGACTTGGG CGCCGCACAA CTCACATCAC CAGCACCGGC CACGGCGACG $\begin{array}{llllllllllllllllllll}\mathrm{N} & \mathrm{V} & \mathrm{C} & \mathrm{E} & \mathrm{C} & \mathrm{H} & \mathrm{S} & \mathrm{G} & \mathrm{Y} & \mathrm{Q} & \mathrm{L} & \mathrm{P} & \mathrm{L} & \mathrm{C} & \mathrm{Q} & \mathrm{E} & \mathrm{C} & \mathrm{P} & \mathrm{G} & \mathrm{C}\end{array}$
33781 AACGTATGCG AGTGCCATTC AGGCTACCAG CTGCCTCTGT GCCAGGAGTG CCCCGGCTGC TTGCATACGC TCACGGTAAG TCCGATGGTC GACGGAGACA CGGTCCTCAC GGGGCCGACG P S P C G K Y I
33841 CCCTCACCCT GTGGCAAGTA CATgtgagtg caggcggagc aggcagggcg ggcagggatc GGGAGTGGGA CACCGTTCAT GTAcactcac gtccgcctcg tccgtcccgc ccgtccctag EXON $13 \leftarrow$
33901 cccactcccc ggccacctac ccagcagccg ggctcactcg gggaccccag tgagcacctc gggtgagggg ccggtggatg ggtcgtcggc ccgagtgagc ccctggggtc actcgtggag
33961 tgcaccgcgg ggttctctgc gggtgcgtct ggactgaggg gccctccgcg gccgccgtgc acgtggcgcc ccaagagacg cccacgcaga cctgactccc cgggaggcgc cggcggcacg
34021 ccctgtgtcc ctcctccctc ctgtgcgcac ctgccgcggg ccctgggatc aggcttccag gggacacagg gaggagggag gacacgcgtg gacggcgccc gggaccctag tccgaaggtc
34081 ttcacccgct gcacccgcat ctgccccact tgggccccct ccttggcctg agggcgcccc aagtgggcga cgtgggcgta gacggggtga acccggggga ggaaccggac tcccgcgggg
34141 aaccccacct gccccccetg gtggacacac acccaagccc gagcccccgt tgccggagcc ttggggtgga cgggggggac cacctgtgtg tgggttcggg ctcgggggca acggcctcgg $\rightarrow$ EXON $14 \quad$ S C A E C
34201 cccgcacacc ctggcaccga tgctcacacc tggctccccg cagCTCCTGC GCCGAGTGCC gggcgtgtgg gaccgtggct acgagtgtgg accgaggggc gtcGAGGACG CGGCTCACGG L $\quad \mathrm{K} \quad \mathrm{F} \quad \mathrm{E} \quad \mathrm{K} \quad \mathrm{G} \quad \mathrm{P} \quad \mathrm{F} \quad \mathrm{G} \quad \mathrm{K} \quad \mathrm{N} \quad \mathrm{C}$
34261 TGAAGTTCGA AAAGGGCCCC TTTGGGAAGA ACTGCAGCGC GGCGTGTCCG GGCCTGCAGC ACTTCAAGCT TTTCCCGGGG AAACCCTTCT TGACGTCGCG CCGCACAGGC CCGGACGTCG $\begin{array}{lllllllllllllllllllll}\mathrm{L} & \mathrm{S} & \mathrm{N} & \mathrm{N} & \mathrm{P} & \mathrm{V} & \mathrm{K} & \mathrm{G} & \mathrm{R} & \mathrm{T} & \mathrm{C} & \mathrm{K} & \mathrm{E} & \mathrm{R} & \mathrm{D} & \mathrm{S} & \mathrm{E} & \mathrm{G} & \mathrm{C} & \mathrm{W}\end{array}$
34321 TGTCGAACAA CCCCGTGAAG GGCAGGACCT GCAAGGAGAG GGACTCAGAG GGCTGCTGGG ACAGCTTGTT GGGGCACTTC CCGTCCTGGA CGTTCCTCTC CCTGAGTCTC CCGACGACCC $\begin{array}{lllllllllllllllllll}V & A & Y & \mathrm{~L} & \mathrm{E} & \mathrm{Q} & \mathrm{Q} & \mathrm{D} & \mathrm{G} & \mathrm{M} & \mathrm{D} & \mathrm{R} & \mathrm{Y} & \mathrm{L} & \mathrm{I} & \mathrm{Y} & \mathrm{V} & \mathrm{D} & \mathrm{E}\end{array}$
34381 TGGCCTACAC GCTGGAGCAG CAGGACGGGA TGGACCGCTA CCTCATCTAT GTGGATGAGA ACCGGATGTG CGACCTCGTC GTCCTGCCCT ACCTGGCGAT GGAGTAGATA CACCTACTCT S R
34441 GCCGAGgtga ggccgctggg gtgcagcggg gaccggctcc agagctccgg ctgcggggtc CGGCTCcact ccggcgaccc cacgtcgccc ctggccgagg tctcgaggcc gacgccccag EXON $14 \leftarrow$
34501 tccgcagagg cgagggtggt gcggaaggag gatctcacct ccagtgttgt gggcccagca aggcgtctcc gctcccacca cgccttcctc ctagagtgga ggtcacaaca cccgggtcgt

34561 gggtggatga gccatgcggg atgcccttgg cattgcaagg aggccagagg ccggtcccct cccacctact cggtacgccc tacgggaacc gtaacgttcc tccggtctcc ggccagggga
34621 cacccccagc agcctcccct ttccctcccc tcctcacctg accccacatc aagctgggca gtgggggtcg tcggagggga aagggagggg aggagtggac tggggtgtag ttcgacccgt
34681 gcagacggcc tgctccctgg gggcaccact gctccacacc gggcaggccc accagcatgc cgtctgccgg acgagggacc cccgtggtga cgaggtgtgg cccgtccggg tggtcgtacg
34741 cacacggccc gggccatctc ccgctggatg cctcatgggg caaggcttgc agttgcaggt gtgtgccggg cccggtagag ggcgacctac ggagtacccc gttccgaacg tcaacgtcca
34801 ccccaggggt ggcccgcggt caggttgctg cgtgaggctc ctttgagggg attgggagat ggggtcccca ccgggcgcca gtccaacgac gcactccgag gaaactcccc taaccctcta
34861 cctggtgctc ggctctcagc cctggagggg cccagggcca ccaggcacca ccgtccccat ggaccacgag ccgagagtcg ggacctcccc gggtcccggt ggtccgtggt ggcaggggta
34921 cactcgttcg gctccagtcc agtcggaacc agtcccagcc ccacagtgac cttggtggca gtgagcaagc cgaggtcagg tcagccttgg tcagggtcgg ggtgtcactg gaaccaccgt
34981 ctgaccctgt ctcattcggc tccagtcgga accagtccca gccccacagt gaccttggtg gactgggaca gagtaagccg aggtcagcct tggtcagggt cggggtgtca ctggaaccac
35041 gcactgaccc tgtctcattc agctccagtt ggaaccaggc ccagccccac ggtgaccttg cgtgactggg acagagtaag tcgaggtcaa ccttggtccg ggtcggggtg ccactggaac
35101 gtgacactgg ccctgcttct tcaggtctca tgttccacag gctcctgaga ggcagagggg cactgtgacc gggacgaaga agtccagagt acaaggtgtc cgaggactct ccgtctcccc
35161 ctggggctgg ggggccgctg gcggactcca tcaggcatgg ccagcagccg gccaggactc gaccccgacc ccccggcgac cgcctgaggt agtccgtacc ggtcgtcggc cggtcctgag
35221 ttgcagaact gagacctggg agctcctccc tcaccaggaa gggacccgag gcaggcggcg aacgtcttga ctctggaccc tcgaggaggg agtggtcctt ccctgggctc cgtccgccgc
35281 ggggtctctg agcctgtagg tgacatgcct ggagctcaca acccacatgt gggcctgggg ccccagagac tcggacatcc actgtacgga cctcgagtgt tgggtgtaca cccggacccc
35341 ctgcagaccc ccacctgcca gggcaggaac ccctgcgggc cactttgagg gacagacttg gacgtctggg ggtggacggt cccgtccttg gggacgcccg gtgaaactcc ctgtctgaac
35401 ggtcacgtct ggtttgtgca atcctggcca agtcacttca cctttgagga caggtgaagg ccagtgcaga ccaaacacgt taggaccggt tcagtgaagt ggaaactcct gtccacttcc
35461 aaggcaggtc ccccaggaaa ggcaggatgg caggaggggg cacggggagt gagagtgcag ttccgtccag ggggtccttt ccgtcctacc gtcctccccc gtgcccctca ctctcacgtc
35521 gtcaggatgg caggaggggg cacagggagt gagggtgcag atcaggatgg caggagcaac cagtcctacc gtcctccccc gtgtccctca ctcccacgtc tagtcctacc gtcctcgttg
35581 cacagggagt gagggtgcag gtcaggatgg caggaggggt acagggagtg aggatgccag gtgtccctca ctcccacgtc cagtcctacc gtcctcccca tgtccctcac tcctacggtc
35641 accaggccag ggactgcact gaacgctagt gtctgtccca gggtcagcag ctgtgggcct tggtccggtc cctgacgtga cttgcgatca cagacagggt cccagtcgtc gacacccgga
35701 ggcccactat ggcccacagt ccccatgctc aggaggtggg cacagctgcc gtgggaacca ccgggtgata ccgggtgtca ggggtacgag tcctccaccc gtgtcgacgg cacccttggt
35761 ggaccttgcg ttgctcatcc tggcaccggc ctcacctgct gagctctggt ctctggtctt cctggaacgc aacgagtagg accgtggccg gagtggacga ctcgagacca gagaccagaa
35821 gtccttgacg accacaaatc actttcacag tatggggagt gaactccagc agggaccctg caggaactgc tggtgtttag tgaaagtgtc atacccctca cttgaggtcg tccctgggac
35881 agcctggttc tgaactctgc tgtaggagag cggggtccct cgcactgtgt cctacccccc tcggaccaag acttgagacg acatcctctc gccccaggga gcgtgacaca ggatgggggg
35941 agtattagca ctctgtcctg gctgcccggt accctggccc tgccttcttc tctgtcagca tcataatcgt gagacaggac cgacgggcca tgggaccggg acggaagaag agacagtcgt
36001 ggggtggtgg caatggccct taggtttcat ttggaaccac agggcctggc ctggaatcca ccccaccacc gttaccggga atccaaagta aaccttggtg tcccggaccg gaccttaggt
36061 gcaggggctg ggccaggttc tcactgaccc ctggccatgg tgatggatgg tgagccgggg cgtccccgac ccggtccaag agtgactggg gaccggtacc actacctacc actcggcccc $\rightarrow$
36121 ccctgggaac ctccgtctct ctaagtgctg tgttgtgggg tgacctcgga cctgtgggca gggacccttg gaggcagaga gattcacgac acaacacccc actggagcct ggacacccgt
$\rightarrow$ EXON 15 E C
36181 cacggctcag tgagatgggg ctggactgac gctcaggtgt gctgcttctt ccctagAGTG gtgccgagtc actctacccc gacctgactg cgagtccaca cgacgaagaa gggatcTCAC $\begin{array}{llllllllllllllllllll}\mathrm{V} & \mathrm{A} & \mathrm{G} & \mathrm{P} & \mathrm{N} & \mathrm{I} & \mathrm{A} & \mathrm{A} & \mathrm{I} & \mathrm{V} & \mathrm{G} & \mathrm{G} & \mathrm{T} & \mathrm{V} & \mathrm{A} & \mathrm{G} & \mathrm{I} & \mathrm{V} & \mathrm{L} & \mathrm{I}\end{array}$
36241 TGTGGCAGGC CCCAACATCG CCGCCATCGT CGGGGGCACC GTGGCAGGCA TCGTGCTGAT ACACCGTCCG GGGTTGTAGC GGCGGTAGCA GCCCCCGTGG CACCGTCCGT AGCACGACTA $\begin{array}{lllllllllllllllllll}\text { G I } & \mathrm{L} & \mathrm{L} & \mathrm{L} & \mathrm{V} & \mathrm{I} & \mathrm{W} & \mathrm{K} & \mathrm{A} & \mathrm{L} & \mathrm{I} & \mathrm{H} & \mathrm{L} & \mathrm{S} & \mathrm{D} & \mathrm{L} & \mathrm{R} & \mathrm{E} & \mathrm{Y}\end{array}$
36301 CGGCATTCTC CTGCTGGTCA TCTGGAAGGC TCTGATCCAC CTGAGCGACC TCCGGGAGTA GCCGTAAGAG GACGACCAGT AGACCTTCCG AGACTAGGTG GACTCGCTGG AGGCCCTCAT $\begin{array}{llllllllllllll}R & R & F & E & K & E & K & L & K & S & \text { Q } & W & N & N\end{array}$
36361 CAGGCGCTTT GAGAAGGAGA AGCTCAAGTC CCAGTGGAAC AATgtaagtg gccgtccttg GTCCGCGAAA CTCTTCCTCT TCGAGTTCAG GGTCACCTTG TTAcattcac cggcaggaac EXON $15 \leftarrow$
36421 ggggtcccac gcagaggggc acgtgcgtcc cacactatgc gacctcctgc tgcgggaggc ccccagggtg cgtctccccg tgcacgcagg gtgtgatacg ctggaggacg acgccctccg

## $\leftarrow$

36481 cgtggacacc cgtgtgtggc tgcaccccgg cctccccagg cagcttgtcc tctgggcgtc gcacctgtgg gcacacaccg acgtggggcc ggaggggtcc gtcgaacagg agacccgcag
36541 cgccgctgtg ggtgagcttg gctgccggtg ctgggaccct gggggggctg ccccacgtct gcggcgacac ccactcgaac cgacggccac gaccctggga cccccccgac ggggtgcaga $\rightarrow$

|  | agctcttcct tcgagaagga | caaagtgacg gtttcactgc | gcttccatgc ggggagtgtc |  | gagtccctct | ttctaagacg |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  | EXON 16 D | N P L |
| 36661 | tctgaaaacc | tcccacactc | acaagcccct | gtttgcattt | cccaccagGA | TAATCCCCTT |
|  | agacttttgg | agggtgtgag | tgttcgggga | caaacgtaaa | gggtggtcCT | ATTAGGGGAA |
|  | F K S | A T T T | V M N | P K F | A E S * |  |
| 36721 | TTCAAGAGCG | CCACCACGAC | GGTCATGAAC | CCCAAGTTTG | CTGAGAGTTA | GgAGCACTTG |
|  | AAGTTCTCGC | GGTGGTGCTG | CCAGTACTTG | GGGTTCAAAC | GACTCTCAAT | CCTCGTGAAC |
| 36781 | GTGAAGACAA | GGCCGTCAGG | ACCCACCATG | TCTGCCCCAT | CACGCGGCCG | AGACATGGCT |
|  | CACTTCTGTT | CCGGCAGTC | TGGGTGGTA | AGACGGGG | GTGC | TCTGTACCGA |
| 36841 | TGCCACAGCT | CTTGAGGATG | TCACCAATTA | ACCAGAAATC | CAGTTATTTT | GGCCCTCAA |
|  | ACGGTGTCGA | GAACTCCTAC | AGTGGTTAAT | TGGTCTITAG | GTCAATAAAA | GGCGGGAGTT |
| 36901 | AATGACAGCC | ATGGCCGGCC | GGTGCTTCT | GGGGCTCGT | CGGGGGGACA | GCTCCACTCT |
|  | TTACTGTCGG | TACCGGCCGG | CCCACGAAGA | CCCCCGAGCA | GCCCCCCTGT | CGAGGTGAGA |
| 36961 | GACTGGCACA | GTCTTTGCAT | GGAGACTTGA | GGAGGGAGGG | CTTGAGGTTG | GTGAGGTTAG |
|  | CTGACCGTGT | CAGAAACGTA | CCTCTGAACT | CCTCCCTCCC | GAACTCCAAC | CACTCCAATC |
| 37021 | GTGCGTGTTT | CCTGTGCAAG | TCAGGACATC | AGTCTGATTA | AAGGTGGTGC | CAATTTATTT |
|  | CACGCACAAA | GGACACGTTC | AGTCCTGTAG | TCAGACTAAT | TTCCACCACG | GTTAAATAAA |
| 37081 | ACATTTAAAC | TTGTCAGGGT | ATAAAATGAC | ATCCCATTAA | TTATATTGTT | ATCAATCAC |
|  | TGTAAATTTG | AACAGTCCCA | TATTTTACTG | TAGGGTAATT | AATATAACAA | TAGTTAGTG |
| 37141 | GTGTATAGAA | AAAAAATAAA | ACTTCAATAC | AGGCTGTCCA | TGGAAActgg | gcactgtgtc |
|  | CACATATCTT | TTTTTTATTT | TGAAGTTATG | TCCGACAGGT | ACCTTTgacc | cgtgacacag |
|  |  |  |  |  | $16 \leftarrow$ |  |
| 37201 | cgctgtattc | ccgactgg | caaggtggcc | ggcacacgt | gggccotgtc | tgccegctg |
|  | gcgacataag | ggggctgacc | gttccaccgg | tccgtgtgca | cccgggacag | gacgggcgac |
| 37261 | acctttgcca | cacaggcaca | cagctggctt | cagaccatgg | ggcaggttct | gttttccctt |
|  | tggaaacggt | gtgtccgtgt | gtcgaccgaa | gtctggtacc | ccgtccaaga | aaaagggaa |
|  |  | $\leftarrow$ |  |  |  |  |
| 37321 | ttccttcctc | acctgctggt | ctatacagag | cccogctctg | aggccetgtc | caaatccta |
|  | aaggaaggag | tggacgacca | gatatgtctc | ggggcgagac t | tccgggacag | ggtttaggat |
| 37381 | ggtgcatttc | cgtaaacggt | ggtgg |  |  |  |

