Discovery and validation of novel substrates for the rhomboid protease RHBDL2

Lucrezia Zarcone

Sir William Dunn School of Pathology

Lincoln college

University of Oxford

A thesis submitted to the University of Oxford

for the degree of Doctor of Philosophy

Hilary term 2023

Table of Contents

Acknowledgements 5
Abstract 7
Abbreviations 9
1. Introduction 12
1.1 Rhomboid-like superfamily 12
1.1.1 Discovery of rhomboid proteases 12
1.1.2 The rhomboid-like superfamily comprises active proteases and pseudoproteases 13
1.2 Structure of a bacterial rhomboid proteases and mechanism of substrate recognition 16
1.2.1 Structure of GlpG and defining features of rhomboid proteases 16
1.2.2 Substrates gain access to the active site through a lateral gate 18
1.2.3 Features of substrates of rhomboid proteases 19
1.3 Validated substrates for RHBDL2 21
1.3.1 Overview of RHBDL2 21
1.3.2 RHBDL2 substrates 23
1.4 VE-Cadherin 27
1.4.1 VE-Cadherin is responsible for cell-cell junctions in endothelial cells 27
1.4.2 VE-Cadherin regulation at the plasma membrane via endocytosis 29
1.4.3 VE-Cadherin regulation by proteolysis 30
1.5 Aim of the project 31
2. Materials and methods 34
2.1 Molecular biology 34
2.1.1 Polymerase chain reaction 34
2.1.2 Restriction enzyme digestion 35
2.1.3 Agarose gel electrophoresis and gel purification 36
2.1.4 Site directed mutagenesis 36
2.1.5 DNA assembly 37
2.1.6 Bacterial transformation 37
2.1.7 Plasmid preparation and Sequencing 38
2.2 Cell biology 40
2.2.1 Cell culture 40
2.2.2 DNA transfection 40
2.2.3 Lentiviral production and transduction 40
2.2.4 Monoclonal cell lines 41

2.3 Biochemical methods 42
2.3.1 Cell lysis 42
2.3.2 AP shedding assay 42
2.3.3 SDS-page and Western blotting 43
2.3.4 Immunofluorescence and confocal microscopy 44
2.3.5 Flow cytometry and fluorescent-activated cell sorting (FACS) 44
2.4 TMD library methods 45
2.4.1 Bioinformatics to generate TMD library 45
2.4.2 Bicistronic vector cloning 46
2.4.3 TMD lentiviral vector cloning 47
2.4.4 TMD pool virus production and transduction 47
2.4.5 MOI titration 48
2.4.6 TMD pool sorting 48
2.4.7 DNA extraction and PCR amplification 49
2.4.8 TMD library preparation and sequencing 50
2.4.9 Data analysis 50
3. Proof of concept and troubleshooting of a new high throughput genetic screen to identify novel substrate for RHBDL2 53
3.1 Introduction 53
3.2 Proof of concept experiments using a bicistronic vector expressing two fluorescently tagged TMDs 54
3.2.1 Schematic of the screen 54
3.2.2 Proof of concept experiments confirm that decrease in mNG/mSc ratio can be used as a readout for RHBDL2 cleavage 57
3.3 Testing the new method using a larger range of substrates 60
3.3.1 Four positive controls do not show any decrease of fluorescence when co-expressed with RHBDL2 60
3.3.2 Using surface labelling allowed for the detection of cleavage for all the tested substrate TMDs 63
3.3.3 Replacing mNeonGreen tag with FLAG-TwinStrep 66
3.4 Generation of stable cell lines expressing theTMDs 68
3.4.1 Stable cells transduced with bicistronic vector did not express both TMDs 68
3.4.2 Switching the set up of the screen to use only one TMD 69
3.4.3 Selecting cells express a higher level of RHBDL2 increases detection of the cleavage for both VE-Cadherin and Spitz TMDs 71
3.5 Generation of monoclonal cell lines for tBFP RHBDL2 and RHDBL2 SA 76
3.5.1 Characterisation of clonal cell lines 76

3.5.2 Stable expression of both TMDs and RHBDL2 lead to a less striking cleavage 78
3.5.3 Increasing expressed tBFP RHBDL2 improved cleavage while retaining specificity 81
3.6 Discussion 83
4. Identification and validation of novel substrates for RHBDL2 87
4.1 Introduction 87
4.2 Generation of the TMD library 87
4.3 MOI titration and testing of the sorting strategy 89
4.3.1 Production of the viral particles 89
4.3.2 Calculating the viral supernatant necessary for an MOI of 0.2 90
4.3.3 Testing the sorting strategy confirmed bins specificity 91
4.4 Screen of TMD pool-expressing cells to identify novel substrates 93
4.4.1 Infection and sorting of TMDs-expressing cells 93
4.4.2 Amplification of TMD sequences and sequencing library preparation 96
4.4.3 Analysis of the sequencing reads and identification of novel substrates 100
4.4.4 Preliminary validation of the newly identified substrates 104
4.5 Discussion 106
5. Characterisation of RHBDL2 cleavage of VE-Cadherin 111
5.1 Introduction 111
5.2 VE-Cadherin is cleaved by RHBDL2 111
5.2.1 RHBDL2 cleavage occurs within the transmembrane domain of VE-Cadherin 112
5.2.2 The full-length VE-Cadherin is processed by RHBDL2 114
5.3 Mapping of the cleavage site using point mutations 116
5.3.1 Mutating luminal side of the TMD partially impairs cleavage 116
5.3.2 Point mutations of TMD do not identify cleavage site 119
5.4 Discussion 121
6. Discussion and future perspectives 125
6.1 Considerations on the TMD screen 125
6.2 VE-Cadherin 131
Appendices 133
Appendix 1 133
Appendix 2 134
Appendix 3 135
Appendix 4 136
Appendix 5 161
References 165

Acknowledgements

First of all, I would like to thank Prof. Matthew Freeman for being a great supervisor and for giving me the possibility of working in his lab. I am grateful for his support through the setbacks and roadblocks during my PhD project, for his trust and for the freedom to follow my ideas.

I would also like to thank Dr. Adam Grieve for his great support and friendship before and after I started my DPhil. He has always been someone I could rely on to discuss any scientific issues and I enjoyed the time we had together inside and outside the lab.

I am also extremely grateful to all the members of the Freeman lab- Fangfang Lu, Clémence Levet, Iqbal Dulloo, Lucie Kafkova, Shaked Ashkenazi, Boyan Zhang, Owain Bryant, Xiaoli Zhou. The lunches, tea breaks, silly chats as well as the scientific discussions have made my time in the lab a memory I will remember fondly. Especially, I would like to thank Dr. Fangfang Lu for being my PhD-buddy in these last four years, and for always being available whenever I needed some advice or some encouragement. Many thanks also to Dr Clémence Lévet for keeping the lab running and for being a great desk and bench neighbour.

The work presented in this thesis would not have been possible without the help of Robert Headly and Vasiliki Tsioligka in the Dunn School sorting facility that have helped with the flow cytometry and sorting experiments. Furthermore, I would like to thank Dr. Samuel Dean for writing the code necessary to generate the library of TMDs. Many thanks to Dr. James Carrington for his work in the analysis of the sequencing data and for his help in the preparation of the sequencing library. Lastly, thank you to all the scientists in the Dunn school who, over the years, advised me on the DPhil project, especially to Dr Alberto Baena-Lopez.

I am very grateful for the support of Lincoln College and EPA Research fund in funding my DPhil.

The lab has not been only the place that allowed me to progress in my academic career, but also where I have met most of my friends here in Oxford. I am extremely grateful for Emily Graham, Emile Roberts, Violaine André, Arnau Bertran, Anna Hunt, Bertie Ansell and their friendship. Coming back home and being able to relax watching questionable TV together has honestly been the highlight of my time in Oxford. Knowing that I had a group of people with whom I could share the high and lows of the PhD life has made this journey immensely better, and it has helped in making Oxford feel like a real home.

I would also like to thank my friends back home, Roswitha, Alessia, Marzia, Stefania, Giuseppe, Giulio, Dacia for their friendship despite the distance. I am also extremely grateful for my sister, Giulia. Probably the person that has listened the most to my complaints (lab related and not) and that had the great job of reminding me that, even though there were instances where it seemed unlikely, there is a life outside the academic Oxford bubble. Grazie.

In conclusione, vorrei ringraziare i miei genitori, Rosario e Giusi, per aver ascoltato le mie lamentele di questi quattro anni e per aver cercato di capire i miei improbabili paragoni mentre cercavo di spiegare (male) su cosa stessi lavorando.

Abstract

Rhomboid proteases are a group of intramembrane proteases from the rhomboid-like superfamily. In mammals they include RHBDL1 -4, which are located along of the secretory pathway, and PARL, which is located in the mitochondria. Rhomboid proteases are highly conserved and, like other intramembrane proteases, they cleave their substrates within their transmembrane domains (TMD). However, not much is known about the biological role of rhomboid proteases, due to the very small number of validated substrates. To this end, the main project of my DPhil aimed at developing a comprehensive genetic screen that employed cell sorting to identify new potential substrates of RHBDL2. The assay was based on the use of Fluorescence-activated cell sorting (FACS) and used a decrease in the fluorescence as a readout for rhomboid cleavage. Firstly, I tested each step of the screen to ensure that this new methodology could work using control TMDs. After a long period of troubleshooting, I conducted the screen by using a newly generated lentiviral library that encodes for the TMDs of all human type I single-pass transmembrane proteins which are the most common substrates for RHBDL2. The screen was performed in the presence of either RHBDL2 or its inactive mutant, RHBDL2 SA, that acted as the negative control. The cells expressing the TMDs in both conditions were firstly labelled with a fluorescent antibody, then divided into 4 different bins according to their fluorescence. The cell populations in the bins were sorted and used for Nanopore sequencing. Analysis of the sequencing data aimed at identifying TMDs whose fluorescence decreased upon co-expression of RHBDL2. Overall, 7 TMDs were identified that showed a significant decrease of their average fluorescence. In this work, I will present the preliminary validation of 4 of them.

In parallel to this main project, I have also investigated the relationship between RHBDL2 and a newly identified substrate, VE-Cadherin. I found that both the TMD and the full-length protein were cleaved by RHBDL2. Moreover, I have also attempted at mapping the rhomboid cleavage site.

Overall, in this thesis I am going to discuss my work in the identification of novel rhomboid substrates and the validation of newly discovered ones.

Abbreviations

AA: aminoacid

ADAM: a disintegrin and metalloproteinase

AP: alkaline phosphatase

AU: average fluorescence

BafA: bafilomycin

CP: cleavage product

DMEM: Dulbecco's Modified Eagle Medium

EC: extracellular cadherin-like domain

EGF: epidermal growth factor

EGFR: epidermal growth factor receptor

ER: endoplasmic reticulum

ERAD: ER-associated degradation

EV: empty vector

FACS: fluorescence activated cell sorting

FBS: fetal bovine serum

FL: full length

FRET: Fluorescence resonance energy transfer

F-TwS: FLAG TwinStrep

HEK: Human embryonic kidney 293T

LB: Luria-Bertani

MOI: multiplicity of infection

MS: mass spectrometry

PARL: Presenilin-associated Rhomboid-like protein

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PEG: polyethylene glycol

PFA: paraformaldehyde

PNPP: p-Nitrophenyl Phosphate

PVDF: polyvinylidene difluoride

SDS-PAGE: sodium-dodecyl-sulfate acrylamide gel electrophoresis

SILAC: stable isotope labelling in cell culture

SOCE: store-operated calcium entry

SPRI: solid phase reversible immobilization

TA: tail-anchored

TBE: Tris-Borate-EDTA

TMD: transmembrane domain

Chapter 1

Introduction

1. Introduction

1.1 Rhomboid-like superfamily

1.1.1 Discovery of rhomboid proteases

Intramembrane proteases are enzymes that reside within the cellular membrane. They are made of multiple transmembrane domains (TMDs) and have lipid-embedded catalytic site. All intramembrane proteases cleave their substrates TMDs and are classified according to which residues compose the active site. To date, there are four families of intramembrane proteases: 1) metalloproteases comprises S2P and S2P-like proteases (Rawson et al., 1997). 2) Aspartyl proteases such as presenilin (active subunit of γ-secretase, De Strooper et al., 1999; Wolfe et al., 1999), signal peptide peptidases (SPPs, Weihofen et al., 2002) and SPP-like proteases (Friedmann et al., 2006). 3) Glutamyl proteases for which only one member has been identified, Rce1 ((Manolaridis et al., 2013) and 4) Serine proteases which are the rhomboid proteases (Urban et al., 2001).

Rhomboid proteases were first discovered in *Drosophila melanogaster,* where they were named after the unusual shape of the fly head caused by mutation of the rhomboid gene (Mayer & Nüsslein-Volhard, 1988). In *Drosophila* there are 5 rhomboid proteases. The first one to be identified and the most studied is Rhomboid-1, which was characterised as the master regulator of the EGFR signalling pathway in flies (Urban et al., 2001). The TGF-α homologue Spitz is synthesised as a type I single pass transmembrane protein in the endoplasmic reticulum (ER). Its traffic to the Golgi requires an interactor called Star. Once in the Golgi, the TMD of Spitz is cleaved by Rhomboid-1 leading to the formation of a soluble ectodomain that gets released outside of the cells. The domain will engage the EGF-receptor in neighbouring cells eliciting the EGF pathway (Lee et al., 2001).

1.1.2 The rhomboid-like superfamily comprises active proteases and pseudoproteases

With follow-up studies, it became apparent that rhomboid proteases were extremely conserved enzymes that were present in essentially all organisms (Lemberg & Freeman, 2007). Moreover, genetic studies identified other rhomboid-like proteins that lacked the catalytic residues of the active counterparts (Lemberg & Freeman, 2007). The discovery of these pseudoproteases promoted the family of rhomboid related protein to the rhomboid-like superfamily. Today, we know of multiple members of the superfamily that are conserved in all kingdoms of life. Although the members do not share high sequence similarity, there are some defining features that classify them as part of the superfamily. All rhomboid-like proteins have a core structure made of six transmembrane domain called rhomboid fold, however most of the mammalian rhomboid proteins have an additional TMD at either the C- or N-terminus (Freeman, 2014). Moreover, rhomboid proteases retain substantial sequence similarity in a few regions, specifically around the active site and in a domain of unknown function in loop1 (more details will be discussed in the next section).

In mammals, there are 5 rhomboid proteases and several pseudoproteases. A list of all known mammalian rhomboid-like proteins is described in Figure 1.1. The focus of my DPhil has been investigating rhomboid proteases, so for the scope of this introduction I am going to focus only on the active members of the superfamily.

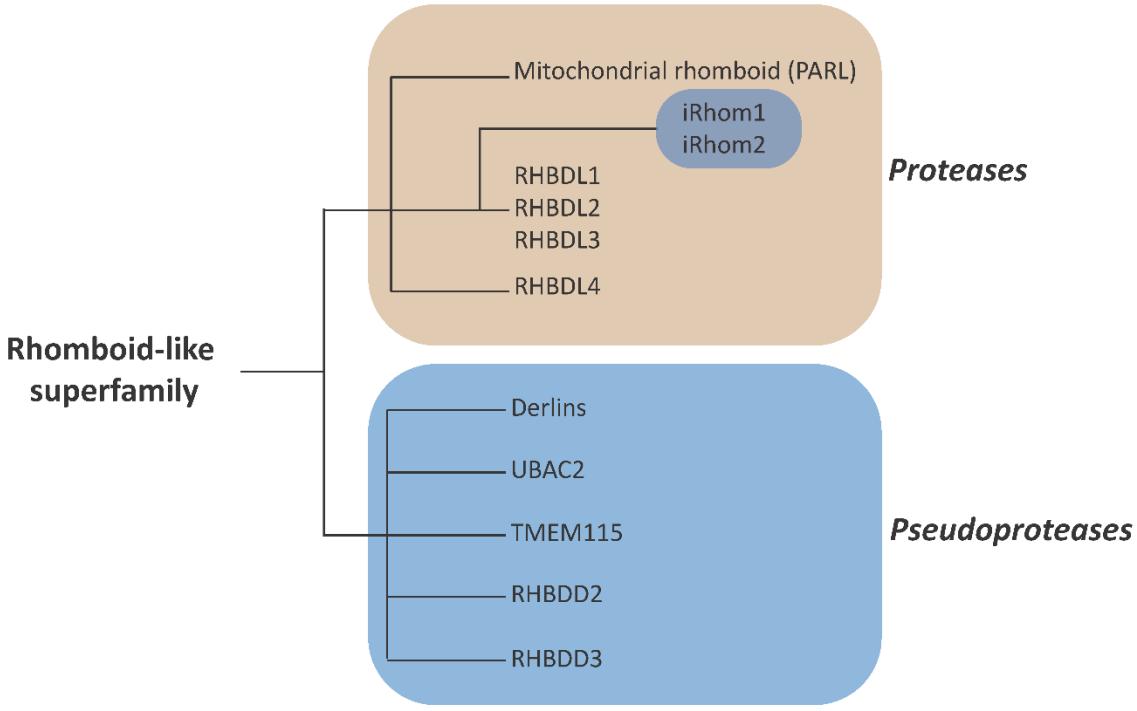


Figure 1.1 List of all the members of the rhomboid-like superfamily in mammals.

Rhomboid proteases can be further divided into two groups based on their localisation: PARL, Presenilin-associated Rhomboid-like protein, which is the only mitochondrial rhomboid proteases, and the secretase rhomboids. The latter are also called RHBDL1-4 and are localised throughout the secretory pathway (Lastun et al., 2016). The role of rhomboid proteases has been well characterised in Drosophila and apicomplexan parasites, but their biological role in mammals has only begun to be understood. The best studied is PARL, which has been correlated with mitochondrial quality control and homeostasis (Spinazzi & De Strooper, 2016). Our knowledge of secretase rhomboids has been patchier and only in recent years there has been substantial progress. Given the pivotal role of rhomboid proteases in the EGFR pathway in flies, one of the main questions was to ascertain whether a similar involvement occurred in mammals. The answer to this question is not a simple one: although the major regulator of EGF-ligands shedding in mammals are members of the ADAM family (Edwards et al., 2008), there is evidence that suggests that rhomboid proteases could be, to some extent, involved. This will be discussed in more details below.

Being proteases, the main way to elucidate what the role of rhomboid proteases is, is to identify which proteins they cleave. For this reason, most of the efforts have been towards uncovering substrates. To date, the only identified substrates are for RHBDL2 and RHBDL4. RHBDL1 and 3 are still mostly uncharacterised. Among the secretase rhomboids, RHBDL1 and RHBDL3 have the most sequence similarity, the murine proteins share ~55% similarity according to Clustal omega alignment. Moreover, according to transcriptome analysis they are expressed in neuronal cells (Cahoy et al., 2008; Jászai & Brand, 2002). Even though analysis of their sequences has identified all the hallmarks of an active site, their catalytic activity has never been experimentally confirmed (Strisovsky, 2016) so there are currently no known substrates for RHBDL1 and RHBDL3.

RHBDL4 is the only rhomboid proteases residing in the endoplasmic reticulum (ER), and it is phylogenetically divergent from RHBDL1-3. Contrary to the other mammalian rhomboid proteases, RHBDL4 has only six TMDs instead of seven (Lemberg & Freeman, 2007). Several substrates have been identified linking RHBDL4 to ER-associated degradation (ERAD, Fleig et al., 2012), APP processing (Paschkowsky et al., 2016) and protein quality control (Kühnle et al., 2019). The main peculiarity of RHBDL4 that arises from the identification of its substrates, is that it appears to be cleaving its substrates at multiple cleavage sites even far outside of the membrane. Rhomboid proteases can cleave residues that are localised at the juxtamembrane region, but the efficiency of the cleavage decreases the further the cleavage site is from the border of the lipid bilayer (Strisovsky et al., 2009). RHBDL4 seems not be affected by distance from the membrane, as cleavage of APP and other substrates can apparently occur far from the membrane. But cleaving in more distal luminal sites from the membrane is not the only peculiarity of RHBDL4, as two recent studies have highlighted that RHBDL4 can also cleave soluble proteins (Bock et al., 2022; Tang et al., 2022). By using a model substrate, Bock et al. showed that soluble luminal aggregation-prone ERAD substrates are cleaved by RHBDL4. The authors suggest that the cleavage of soluble proteins is

possible due to the interaction of RHBDL4 with the erlin complex, a putative ERAD recruitment factor (Pearce et al., 2009). However, the mechanism by which the RHBDL4-Erlin complex binds the soluble substrates and leads them to the membrane-embedded active site is yet to be characterised. In the other study that discovered RHBDL4-dependent cleavage of soluble proteins, the authors replaced the catalytic serine of RHBDL4 with an unnatural amino acid that, when illuminated, would crosslink the N-terminal fragment of RHBDL4 putative substrates. By using this approach, they identified several new substrates. Surprisingly, 10 out of the 15 ER resident proteins identified were soluble proteins. Moreover, also a small proportions of soluble nuclear proteins were also identified as substrates (Tang et al., 2022). Given that the crosslinking occurred at the active site of the protease, the identification of non-ER resident protein meant that they interacted with RHBDL4 active site. This is extremely surprisingly as there are currently no data that suggest that these nuclear proteins can be found in the ER lumen, so further investigation is needed to confirm these results.

RHBDL2 is the best characterised secretase rhomboid proteases and therefore can be considered to be the model mammalian rhomboid. A more detailed discussion of RHBDL2 role and which substrates have been identified can be found in a later section of this introduction.

1.2 Structure of a bacterial rhomboid proteases and mechanism of substrate recognition

1.2.1 Structure of GlpG and defining features of rhomboid proteases

For many years, intramembrane proteases have been mysterious enzymes as their active site is embedded in cell membranes. How is a hydrophilic reaction such as proteolysis happening in the hydrophobic environment of the lipid bilayer? Understanding how water molecules get access to the active site was for years one of the most sought-after questions. For rhomboid proteases, an

answer was possible through the resolution of the structure of a bacterial rhomboid, GlpG (Y. Wang et al., 2006). The structure showed that GlpG is made of 6 TMDs (the core structure of the rhomboid fold), and it also confirmed that the active site is made up of two catalytic residues, S201 and H254, localised in TMD4 and TMD6 respectively. This is a striking difference to soluble serine proteases that process their substrates via a catalytic triad, frequently Arg-Ser-His (Fodor et al., 2006). Moreover, the structure revealed that the catalytic dyad is below the membrane surface but open to the aqueous environment through a water-filled indentation. The indentation is made possible by the positioning of TMD4, at the top of which the catalytic serine is localised. TMD4 is shorter than the rest of the transmembrane domains, which surround it (Zhou et al., 2012). Because of this particular structure, the active site is still accessible to water molecules but shielded from the lipids of the cell membranes (Figure 1.2). Consistent with the position of the active site, rhomboid proteases mostly cleave the TMDs of their substrates in the luminal side (Lastun et al., 2016).

Unsurprisingly, the active site is one of the most conserved regions in rhomboid proteases. In fact, the catalytic serine is in most cases part of a GxSx motif (Lemberg & Freeman, 2007). Another domain that is well conserved throughout most of the rhomboid proteases is a WR motif localised in loop1, connecting TMD1 and TMD2. This loop is usually longer than the others, both in the proteases and in the inactive members of the superfamily, such as iRhoms and Derlins, where it has been shown to have a functional role (Dulloo et al., 2019; Düsterhöft et al., 2021; Rao et al., 2021). For rhomboid proteases, the Arg and Tryp are well conserved, and it has been shown that the presence of the motif is necessary for activity. In fact, mutations of one of the residues can abolish completely the catalytic activity of the protease or reduce it drastically (Lemberg et al., 2005; Urban et al., 2001). However, this motif is not part of the catalytic mechanism, and its exact function still remains elusive.

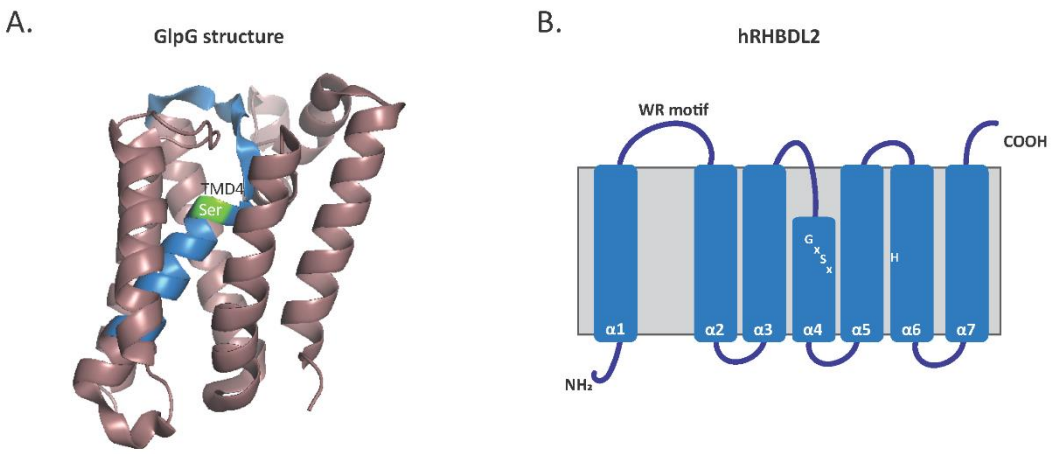


Figure 1.2 Structure of rhomboid proteases. A) Crystal structure of *E.Coli* GlpG (Y. Wang et al., 2006), TMD4 is in blue and the catalytic serine that resides at the top of the domain is in green. **B)** Schematic of hRHBDL2 highlighting the WR motif in loop1 and the two catalytic residues in TMD4 and TMD6.

1.2.2 Substrates gain access to the active site through a lateral gate

The resolution of the structure of GlpG and its homologues in other bacterial strains (Lemieux et al., 2007; Y. Wang et al., 2006) allowed for the determination of the proteolytic mechanism through the catalytic dyad. Some questions still remained regarding how the substrate gained access to the active site through the lipid bilayer. Although there is currently no structure of the protease-substrate complex, several studies have solved either the structure of the open confirmation of GlpG or the enzyme in complex with inhibitory peptides. These studies showed that the substrate enters the active site through a lateral gate made of TMD2 and TMD5 (Baker et al., 2007; Vinothkumar et al., 2010; Wu et al., 2006; Zoll et al., 2014). The helix of TMD5 is bent and, in the open state of the enzyme, it rotates and allows the access for a large polypeptide chain. Movement of helix 5 also causes the displacement of loop5, and it was suggested that this movement could allow for water molecules to get into the active site (Baker et al., 2007; Wu et al., 2006). More recently, the structure of human derlin, a rhomboid pseudoproteases, was solved. Derlins have been described to form a tetramer that acts as the retrotranslocon for membrane

proteins that are substrates for ER-associated degradation (ERAD) (Rao et al., 2021). In this study, it was shown that TMD5 of the monomeric Derlin is parallel to TMD2, and it leads to the formation of a gate that allows the access of the ERAD substrate into the channel. This is in line to what was observed in the open conformation of GlpG and would confirm that the lateral gating strategy is conserved in at least some other members of the superfamily.

1.2.3 Features of substrates of rhomboid proteases

Rhomboid proteases are very selective enzymes, and the current subset of known substrates is relatively small. This suggests that the majority of TMDs that rhomboid proteases encounter are not cleaved. So, what are the features that determine what a rhomboid protease substrate is? Over the years, a few key qualities have been identified. Notably, almost the totality of known substrates are type I transmembrane domains (i.e. luminal N-terminus and cytosol C-terminus) and, for the most part, the substrates are single pass transmembrane proteins. However, there are some exceptions, such as the of cleavage of 4th TMD (TMD4) of Orai1 (Grieve et al., 2021a). This TMD still has a type I orientation, but it is the first example of a rhomboid cleavage of a polytopic protein. This suggests that the repertoire of substrates could be larger than expected.

Most of the substrates present helix-destabilising residues along the TMD such as proline or glycine. The presence of this residues generates a region of the domain that is less prone to the formation of α-helix, and it has been shown that rhomboid proteases recognise this instability (Urban & Freeman, 2003). Moreover, inserting such residues in proteins that are usually not cleaved by rhomboid is enough to transform them into rhomboid substrates (Moin & Urban, 2012). The presence of an instable region in the α-helix of a substrate TMD can be used by rhomboid proteases to selectively interact with the TMDs. This is the case of RHBDL2 recognition of TMD4 of Orai1. It has been observed that the TMD presents a kink due to the presence of a proline (Hou et al., 2012). When the proline is mutated, the TMD straightens and binding to RHBDL2 is reduced (Grieve et al.,

2021a), this highlights that helix-destabilising residues could represent a way for rhomboid proteases to discriminate among TMDs. Due to this, the presence of helix-destabilising residues has been used in different occasions as a way to predict novel substrates (Grieve et al., 2021a; Urban & Freeman, 2003). However, not all the predicted TMDs were confirmed as rhomboid substrates when they were tested, suggesting that, although the presence of these residues is an important feature of rhomboid substrates, there may be other qualities that determines rhomboid selectivity. In fact, some validated rhomboid substrates do not have helix-destabilising residues and in the case of thrombomodulin, the transmembrane domain is not necessary for the cleavage which is in turn driven through the cytoplasmic side of thrombomodulin (Lohi et al., 2004).

Beyond helix-destabilising residues, a motif surrounding the cleavage site has been described in a number of substrates (Strisovsky et al., 2009). In that study, defined positions around the cleavage site were observed to prefer specific residues, since mutating the amino acid in these positions could affect the efficiency of the cleavage. This motif, called 'recognition motif', contained the cleavage site and it could still drive the cleavage even when re-positioned in the juxtamembrane region of the substrate TMD, although the cleavage would decrease the more distant the motif was from the membrane. These findings led to a model for substrate recognition. The model states that the interaction between substrate TMD and protease occurs through an exosite sequence present in the rhomboid protease. Once the interaction is established, the cleavage can occur through the presence of helix destabilising residues, if the recognition motif is within the membrane. The presence of these residues allows for the local unfolding of the helix that is necessary for the recognition motif to enter the active site. If the recognition motif is located outside of the membrane, there might not be helical instability, as the linker region between the motif and the TMD is flexible enough to allow the motif to access the active site.

Understanding which features of a rhomboid substrate determine recognition and processing by rhomboid proteases would help in predicting and identifying new substrates. In fact,

these characteristics have been used as guidelines in several screening studies to identify new substrates for mammalian rhomboid proteases (Grieve et al., 2021a; Lohi et al., 2004; Noy et al., 2016; Pascall & Brown, 2004). Most of these efforts have been aimed at increasing the repertoire of RHDBL2 substrates.

1.3 Validated substrates for RHBDL2

1.3.1 Overview of RHBDL2

RHBDL2 is the model rhomboid for mammalian secretase rhomboids as there are several substrates that have already been identified. However, for most of these substrates the functional role of the cleavage has not been elucidated, so the full range of the biological function of RHBDL2 cleavage still needs to be defined.

RHBDL2 is localised at the plasma membrane and is made up of 7 transmembrane domains, with the additional 7^{th} TMD fused at the N-terminus of the rhomboid fold. According to online databases, RHBDL2 is ubiquitously expressed with an increased level of expression in epithelial tissues such as keratinocytes. Not much is known about the regulation of RHBDL2, but the N-terminal tail could theoretically be used for regulation as it is the only terminus to be cytoplasmic. There are a few studies that highlight a potential regulatory role for this domain, in fact the N-terminus domain has been found to be necessary for the cleavage of thrombomodulin and C-type Lectin 14A (CLEC14a) (Lohi et al., 2004; Noy et al., 2016). Notably, there is also evidence that suggests a role for this domain in the regulation of the binding of Orai1. Although these data hint at a possible regulation of RHBDL2 via its cytosolic region, the mechanisms controlling this regulation are still unclear. There are some hints that RHBDL2 could be regulated by Ca^{2+}. In fact, Grieve et al demonstrated that the binding to active mutant forms of Orai1 was decreased when Ca^{2+} was chelated using EGTA. However, RHBDL2 catalytic activity was not affected by changes in

the calcium levels. Overall, there is still much left to understand about how RHBDL2 activity could be regulated.

Being an active enzyme, one of the main ways to uncover what is the biological role of RHBDL2 is to identify its substrates. As previously mentioned, there are already several substrates that have been identified: most of them are type I single-pass transmembrane proteins but RHBDL2 can also cleave type I TMDs within polytopic proteins (Grieve et al., 2021a) (Table 1.1). Notably, a significant proportion of RHBDL2 substrates are also cleaved by ADAM10. ADAMs are metalloproteases that are present at the plasma membrane where they are involved in ectodomain shedding of growth factors and cytokines (Edwards et al., 2008). ADAM10 cleaves its substrates in the juxtamembrane region, just outside the membrane surface, leading to the formation of a membrane-bound cleavage product that it is often processed further by gamma-secretase. The significance of the processing of the same proteins by both ADAM10 and RHBDL2 is not understood. Given how close the two cleavage sites are, a simple explanation would be that these two enzymes have redundant functions. However, there are currently no data that suggest that rhomboid cleavage is followed by further processing by other proteases. Moreover, there are proteins that are cleaved by RHBDL2 that are not ADAM10 substrates. Considering this, a perhaps more interesting explanation for this overlap in substrate repertoire is that the two cleavage events occur in response to different stimuli. A further characterisation of RHBDL2 cleavage of already identified substrates and increasing our knowledge of which proteins are cleaved by RHBDL2 will help elucidate what pathways regulate RHBDL2 activity. In the next section, I am going to discuss some of the validated substrates for RHBDL2 and the approach that has led to their discovery.

	Proposed function	**ADAM10 substrate**
Thrombomodulin(Lohi et al., 2004)	Wound healing(Cheng et al., 2011)	Yes (Menschikowski et al., 2010)

C C-type Lectin 14A (CLEC14a)(Noy et al., 2016)	Angiogenesis(Noy et al., 2016)	
EGF (Adrain et al., 2011)	Anoikis (Cheng et al., 2014)	Yes(Sahin et al., 2004)
EGFR(Adrain et al., 2011)		
Ephrin B3(Pascall & Brown, 2004)		Yes (Atapattu et al., 2014)
Orai1 (Grieve et al., 2021a)	Regulation of protein degradation	
KIRREL(Johnson et al., 2017)		
DDR1(Johnson et al., 2017)		
IL6R(Johnson et al., 2017)		
BCAM(Johnson et al., 2017)		
N-Cadherin(Johnson et al., 2017)		Yes(Reiss et al., 2005)
CLCP1(Johnson et al., 2017)		
IL11R(Koch et al., 2021)		Yes(Lokau et al., 2016)
Notch (Chen et al., 2022)		Yes (Groot & Vooijs, 2012)
VE-Cadherin (Battistini et al., 2019)		Yes (Schulz et al., 2008)

Table 1.1 List of RHBDL2 substrates

1.3.2 RHBDL2 substrates

RHBDL2 is the only mammalian secretase rhomboid able to cleave *Drosophila* Spitz (Urban et al., 2001). Spitz is the fly homologue of EGF, so this discovery raised the possibility that RHBDL2 could also regulate EGFR signalling in mammals even though, this pathway is usually regulated by

members of the ADAM family. Indeed, it was observed that RHBDL2 is able to cleave EGF in mammals generating a cleavage product that is different to the one caused by ADAM10 cleavage (Adrain et al., 2011). The ectodomain generated by the processing of RHBDL2 could interact with the EGFR and elicit the signalling pathway. However, RHDBL2 cleavage was only detectable if ADAM10 activity was inhibited, suggesting that RHBDL2 contribution to EGF processing is minor or only occurs in the absence of ADAM10. To date, the physiological role of RHBDL2 cleavage is EGF is not known but it is notable that it is increased in cancer cell lines, suggesting a potential role in the pathology of the EGFR signalling pathway (Cheng et al., 2014).

RHBDL2 cleavage of Spitz has been useful in the prediction of putative novel substrates. In fact, by testing TMDs that showed similar characteristics to Spitz TMD (such as helix-destabilising residue and similar residues in the potential cleavage site) thrombomodulin and C-type Lectin 14A (CLEC14a) were discovered as rhomboid substrates (Lohi et al., 2004; Noy et al., 2016). It has been suggested that RHBDL2 cleavage could have a role in cell migration in endothelial tissues. Both of the proteins contain EGF-like domain, and the cleavage occurs near the top of the TMDs, the canonical site for rhomboid cleavage. Surprisingly, mutating the helix-destabilising residues in thrombomodulin TMD did not abrogate the cleavage. In fact, using chimeras where thrombomodulin domains were swapped with those of TGF-α (that is usually not cleaved by RHBDL2), the authors found that the cytoplasmic domain of thrombomodulin was driving the cleavage. Moving this domain to proteins that are usually not cleaved by RHBDL2 was enough to make them rhomboid substrates. Moreover, the cytoplasmic domain of RHBDL2 was also necessary for thrombomodulin cleavage, suggesting that the two domains somehow interact. The exact mechanism of this interaction is not clear, and this finding is made even more interesting considering the similarity between the TMD of thrombomodulin and Spitz. It suggests that there are multiple recognition mechanisms that rhomboid can use to interact with their substrates. Further study will help understanding in which circumstances each mechanism is used.

Using Spitz TMD as a guideline to predict TMDs that could potentially be substrates, also led to the identification of IL11RA and EphrinB as RHBDL2 substrates (Koch et al., 2021; Pascall & Brown, 2004). While the significance of EphrinB cleavage is not clear, Koch et al. identified that RHBDL2 processing of IL11RA leads to the release of a soluble IL11RA that is able to elicit IL11 trans-signalling. In both cases, the cleavage occurs a few residues before the TMD, and it is driven by the presence of helix-destabilising residues within the transmembrane domain. It is important to highlight that all the substrates that have been described so far are also processed by ADAM10, suggesting a potential degree of redundancy between the two enzymes.

One of the best characterised substrate so far, that is exclusively cleaved by RHBDL2, is Orai1 (Grieve et al., 2021a). Orai1 was identified in a more high-throughput candidate screen that tested a large number of TMDs. The TMDs to test were obtained through a bioinformatic analysis that used Spitz transmembrane domain as a model to predict which other TMDs could have a similar secondary structure. Out of the hundreds of TMDs that were analysed, only a handful were identified as substrates, confirming the very selective nature of rhomboid proteases. Orai1 is an unusual substrate since it is a polytopic protein made up of 4 TMDs. Moreover, it assembles into a hexamer at the plasma membrane to form a calcium channel (Prakriya, 2009). RHBDL2 cleaves the 4^{th} TMD of Orai1 (TMD4) which is at the exterior of the hexamer, making it the most accessible to the rhomboid cleavage (Hou et al., 2012). At the moment, the stoichiometry between RHBDL2 and Orai1 is unknown. Strikingly, TMD4 presents a proline (P245) that causes the formation of a hinge in the TMD. When the channel is in an open conformation, the hinge undergoes conformational changes that lead to the straightening of the TMD. It has been observed that RHBDL2 recognises the open conformation of Orai1 and binds selectively to the active channel due to the presence of the proline residue. In fact, mutations of the residue decrease the binding of the protease (Grieve et al., 2021a).

Orai1 is part of store-operated calcium entry (SOCE), an important pathway that gets activated when ER calcium stores are depleted. Through the activation of the pathway via STIM (an ER resident protein that senses decreases in calcium levels in the ER) and STIM interaction with the cytoplasmic domain of Orai1, the channel opens up when the stores are depleted. This leads to the entry of calcium into the cytosol and the refill of calcium stores in the ER (Hogan & Rao, 2015). Since calcium signalling regulates important processes such as gene expression and immune responses, the SOCE pathway is tightly regulated. Moreover, mutation in Orai1 channel that lead to its constitutive activation lead to a group of diseases called channelopathies (Feske, 2010). Grieve et al. showed that RHBDL2 cleaves Orai1 in the event of an unstimulated opening of the channel, thus preventing inappropriate changes to the intracellular levels of calcium.

The substrates identified with candidate-based approaches confirm that these methods are successful in predicting substrates. Moreover, understanding more which features of the TMD affect rhomboid selective interaction will help making the prediction more stringent. Nonetheless, these methods have some shortcomings. Firstly, since they are based on the current knowledge about rhomboid proteases, they could lead to missing some important new substrates that have unusual characteristics. The discovery that RHBDL2 can cleave polytopic proteins demonstrates that the substrate repertoire could be bigger than what is currently known. Secondly, assaying one putative substrate at a time is time consuming and not suitable for the testing of a large number of TMDs. So, there is a need in the field to develop a more high throughput method that could systematically lead to the discovery of new substrates. The first attempt at such a method was done by analysing via stable isotope labelling in cell culture (SILAC) followed by mass spectrometry (MS) the extracellular proteome of cells over-expressing RHBDL2 (Johnson et al., 2017; Ong et al., 2002). The authors identified several type I single pass proteins that were cleaved by RHBDL2 and not by other intramembrane proteases present at the cell surface, such as ADAMs. By looking at the relative mRNA expression of the newly identified substrates and RHBDL2, it was proposed that RHBDL2 could have a role in epithelial homeostasis. Although there is evidence that shows a potential

function of RHBDL2 in wound healing (Cheng et al., 2011), this functional role needs to be further investigated.

1.4 VE-Cadherin

1.4.1 VE-Cadherin is responsible for cell-cell junctions in endothelial cells

VE- Cadherin is part of the cadherin family and it is responsible for maintaining adherens junctions in endothelial tissues (Bravi et al., 2014). The cadherin family is a large group of single-pass transmembrane proteins. Depending on domain structure and genomic organisation, they can be divided into 5 groups and the cadherins most frequently found in adhesions are part of the first two groups (Nollet et al., 2000; Vincent et al., 2004). VE-Cadherin is a part of the type II subgroup. Similar to all cadherins, VE-Cadherin has structure that is composed of three domains (Figure 1.3). All cadherins are translated as precursor proteins that contain a signal peptide and propeptide that immediately precede the extracellular domain, these sequences are removed during protein maturation (Ozawa & Kemler, 1990). The extracellular region is by far the largest of the protein and is made up of 5 extracellular cadherin-like domains (EC). The 5 EC repeats form a rigid rod-like structure that is stabilised by calcium ions that bind sequences in between each of the EC repeats. The rod-like structure is responsible for the homophilic interaction necessary for the formation of cell adhesion. Through the extracellular domain, VE-Cadherins can associate with another copy to form *cis*- or *trans*- interactions. Structural analysis of other type II cadherins has elucidated that the mechanism by which the dimers form involves that EC1 domain and conserved Tryp residues. In fact, the Tryp residues in one EC1 domain anchors the strand of the partner EC1 stabilising the interaction (Patel et al., 2006).

The transmembrane domain of cadherins has a type I topology and precedes the cytosolic domain. The latter is a signalling hub where several interactors of VE-Cadherin regulate its activity. One of

the classical partners of cadherins are cytoplasmic catenins (such as β-catenin, plakoglobin and p120). Catenins link VE-Cadherin to the actin cytoskeleton and mediate interaction with several molecules to regulate downstream pathways (Bravi et al., 2014). Deletion in the cytosolic domain of VE-Cadherin that abrogate binding to β-catenin and plakoglobin causes similar *in vivo* effects to the complete knock-out of VE-cadherin gene, highlighting the importance of this region in maintaining endothelial homeostasis (Carmeliet et al., 1999).

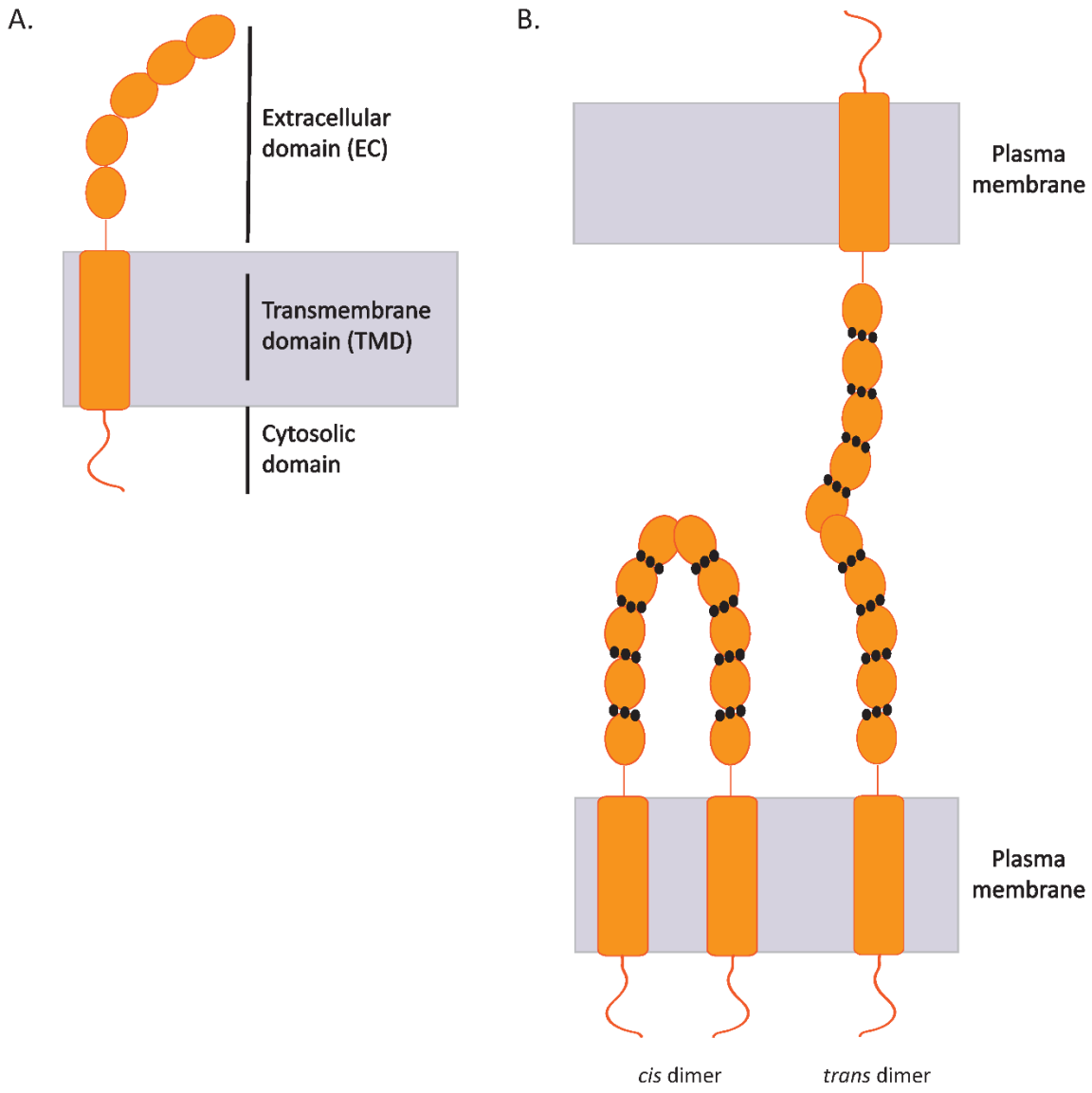


Figure 1.3 Schematic of VE-Cadherin. A) Schematic of the full-length VE-Cadherin highlighting the three domains. EC= extracellular cadherin-like domain. **B)** Schematic of *cis* and *trans* dimers formed by two

monomers of VE-Cadherin. The dimers are formed due to the rod-like structure of the EC domain that is stabilised by calcium (black dots). The two monomers of VE-Cadherin interact through the first EC domain.

1.4.2 VE-Cadherin regulation at the plasma membrane via endocytosis

Cell junctions are dynamic structures that have to respond quickly to stimuli (Kowalczyk & Nanes, 2012). A fast way of modulating cell adhesion is by regulating the levels of cadherins at the cell surface. In general, the pool of VE-Cadherin available at the plasma membrane is regulated by a balance of endocytosis and recycling. VE-Cadherin is endocytosed through clathrin-dependent mechanism that are initiated by cytoplasmic interactors. In fact, it is well established that p120 is a master regulator of VE-Cadherin trafficking. p120 interacts with the cytoplasmic juxtamembrane region of VE-Cadherin and, in contrast with other catenins, it does not link VE-Cadherin to the cytoskeleton. Instead, the binding of p120 masks a well conserved endocytic motif (DEE) in the cytoplasmic tail of VE-Cadherin that prevents the interaction with components of the clathrin-endocytic pathway (Chiasson et al., 2009; Nanes et al., 2012). Upon p120 dissociation, VE-Cadherin is quickly trafficked into clathrin rich vesicles. Interestingly, it is currently not clear what the driving force of p120 dissociation from VE-Cadherin tail is. The presence of VE-Cadherin at cell junctions inhibits cell proliferation (Caveda et al., 1996), so the decrease in its levels at the cell surface lead to cell growth that is essential in processes such as angiogenesis. It is not surprising then, that VEGF, the main growth factor regulating formation of new endothelial vessels, also regulates levels of VE-Cadherin at the plasma membrane (Gavard & Gutkind, 2006). In fact, VEGF causes phosphorylation of the cytoplasmic tail of VE-Cadherin and elicits clathrin-dependent endocytosis through β-catenin.

The fate of endocytosed VE-Cadherin is not only lysosomal degradation, as a proportion of it gets recycled back to the plasma membrane. The factors involved in this process are not completely identified but it has been suggested that Rab11 plays a role in redirecting VE-Cadherin back to the cell surface (Yan et al., 2016).

1.4.3 VE-Cadherin regulation by proteolysis

In addition to endocytosis, VE-Cadherin levels at the plasma membrane are regulated through proteolysis via intramembrane and soluble proteases. The VE-Cadherin cytoplasmic tail is cleaved by the cytosolic protease calpain upon entering endocytic vesicles. The cleavage removes the catenin-binding domain and it leads to VE-Cadherin degradation rather than its recycling (Su & Kowalczyk, 2017). Removal of VE-Cadherin from the plasma membrane leads to less adherent cells, however the exact trigger that stimulates VE-Cadherin cleavage by calpain is not completely understood.

VE-Cadherin transmembrane domain also undergoes processing mainly by ADAM metalloproteases (Dreymueller et al., 2012). ADAM proteases cleave their substrates just outside the luminal side of the membrane, so their activity leads to the formation of a soluble ectodomain, and a membrane-bound cleavage product made up of the transmembrane and cytosolic domain. The endothelial shedding of the extracellular domain of VE-Cadherin causes an increase in intercellular gaps and permeability (Schulz et al., 2008). The authors showed that ADAM10 cleavage is regulated by calcium influx and contributes to thrombin-dependent decrease of cell adhesion. The remaining membrane-bound fragment gets further processed by gamma-secretase. Similar processing also occurs for E-Cadherin where the intracellular fragment generated by gamma-secretase activity can migrate to the nucleus and acts as a transcription factor (Ferber et al., 2008). However, it is still unknown whether also VE-Cadherin product could have a similar role.

It is important to highlight that, at the moment, the mechanisms that regulate the decision of whether to endocytose and recycle VE-Cadherin or to proteolytic process it are not completely understood. But further investigation will be needed to understand in which context these two processes are required.

1.5 Aim of the project

The majority of rhomboid substrates have been identified using candidate-based approaches. As mentioned before, these approaches have proven to be useful, but they are not well suited to test a large number of candidates. Attempts at developing more systematic analysis using proteomic techniques have helped in testing a larger cohort of potential substrates (Johnson et al., 2017), but they are still dependent on the expression of those proteins in the particular cell type that is being assessed. Considering this, the main goal of my project has been to generate a novel method for the identification of rhomboid substrates. The goal was to test, in an unbiased way, all the potential substrates in the human genome among type I single pass transmembrane proteins, around 1146 proteins. To do this, I used the stripped-back TMD structure used by Grieve et al., where the TMD sequence of putative substrates was flanked by short stretches at the luminal and cytosolic side, to generate a library containing the TMD sequences of all type I single pass proteins. Concomitantly to the start of my DPhil, Koren et al. published a paper where they used a bicistronic vector system to test the stability of the human proteome (Koren et al., 2018). We decided to adapt this method for our purposing using a bicistronic vector coding for two fluorescently tagged TMDs and assaying the cleavage of one of those TMDs. The readout for cleavage would be the decrease in the fluorescence ratio.

Since a reduction in fluorescence had never been used to test rhomboid cleavage, and considering that a screen for such a large number of putative substrates had not been attempted before, I had to test each step of the method to make sure that it could work. Using positive and negative control TMDs, I did a thorough testing and troubleshooting of all the steps before doing the screen with the pool of TMDs of the library. The troubleshooting period ended up taking longer that it was first planned, and the overall strategy of the screen changed considerably. However, at the end of this the screen was more efficient than it was originally.

After having successfully did three replicates of the screen and careful analysed of the data, I obtained 7 new hits that had not been identified before as RHBDL2 substrates. Due to the time constraints of the PhD, I was only able to do preliminary validation of the results and I will discuss it in later chapters.

In the last few months of my DPhil, I started a side project following up a previous result that was obtained in the lab that showed RHBDL2 cleavage of VE-Cadherin. I validate RHBDL2 cleavage of both TMD and full-length VE-Cadherin and attempted the characterisation of the cleavage site. This project aimed at elucidating the cleavage of a newly identified substrate, but also to act as a guideline for how to begin the study of new substrates that would be useful for a quicker validation of the hits from the screen.

Overall, the main focus of my DPhil has been to advance our understanding of rhomboid proteases and to potentially uncover new functions. To do this, I set out to find new substrates by developing a new method that allows for a high-throughput and systematic screening. Moreover, I have started to investigate the relationship between RHBDL2 and VE-Cadherin. My overall aim is that my DPhil project will provide useful tools for the study of rhomboid protease and will help deepen our knowledge of the roles of these proteases.

Chapter 2

Materials and methods

2. Materials and methods

2.1 Molecular biology

2.1.1 Polymerase chain reaction

DNA fragments were amplified using polymerase chain reaction (PCR). The reaction was performed using Phusion polymerase (Thermo scientific) using the protocol described in Table 2.1. The reaction was done in a 96-well thermocycler (Applied Biosystems) with the program shown in Table 2.2. Oligo primers were ordered via Merck or Thermo-fisher. A list of the main primers used can be found in Table 2.3. The list of the primers used for the construction of the library can be found in table 2.14. To generate pcDNA3.1 HA hRHBDL2 and p5FRT tBFP RHBDL2, hRHBDL2 sequence was ordered from IDT, PCR amplified and inserted into pcDNA3.1 vector or pcDNA5/FRT between BamHI and XhoI sites by InFusion cloning (protocols described below). The inactivating S187A mutation was generated by site directed mutagenesis using the protocol described in table 2.5 and 2.6.

Table 2.1 | Phusion PCR protocol

PCR mix (50µl mix)	Volume (µl)
5x Phusion reaction buffer (Thermo scientific)	10µl
10mM dNTP mix	1µl
10 µM forward primer	2.5µl
10 µM reverse primer	2.5µl
200-300ng/µl template DNA	1µl
Phusion DNA polymerase	1µl
ddH_2O	Up to 50µl

Table 2.2 | Phusion PCR program

Step & Temperature	Time	Cycles
Initial denaturation (98°C)	1 min	1
Denaturation (98°C)	20s	30/35 cycles
Annealing (55°C)	30s	
Extension (72°C)	1min/Kb	

Final extension	10 min
Hold (4°C)	/

Table 2.3 | List of primers

Primer name	Primer sequence
p3.1Age1_HA hRHBDL2 F	TACCGGTCTCGGATCCATGTACCCATACGATGTTCCAGATTACGCTGCT GCTGTTCATGATCTGGAG
p3.1AgeI_HA hRHBDL2 R	GCCCTCTAGACTCGAGTCAGTTTGCTGGAGATAGGAAAATG
p3.1AgeI_hRHBDL2 F	TACCGGTCTCGGATCCATGGCTGCTGTTCATGATCTGGA
p3.1AgeI_hRHBDL2 R	GCCCTCTAGACTCGAGTCAGTTTGCTGGAGATAGGAAAATG
pEGFP_AP VE-Cadherin TMD F	TGCTGGTGCTGTCGACGATATGGCCGCCCAGGTGGGCGTGAGCATC CAGGCAGTGGTAGCCATCTTACTCTGCATCCTCACCATCACAG
pEGFP_AP VE-Cadherin TMD R	TCTAGAGTCGCGGCCGCTCACCGCCGCCGCAGGAAGATGAGCAGGG TGATCACTGTGATGGTGAGGATGCAGAG
pEGFP_AP VE-Cadherin TMD mut1 F	TGCTGGTGCTGTCGACGATATGGCCGCCCAGGTGGGCGTGAGCCTC AACTTAGTGGTAGCCATCTTACTCTGCATCCTCACCATCACAG
pEGFP_AP VE-Cadherin TMD mut2 F	TGCTGGTGCTGTCGACGATATGGCCGCCCAGGTGGGCGTGAGCATC CAGGCACTCCTACTGCTCTTACTCTGCATCCTCACCATCACAG
pEGFP_AP VE-Cadherin TMD mut3 F	TGCTGGTGCTGTCGACGATATGGCCGCCCAGGTGGGCGTGAGCATC CAGGCAGTGGTAGCCATCTTACTCCTCCTACTCACCATCACA
pEGFP_AP VE-Cadherin TMD mut3 R	TCTAGAGTCGCGGCCGCTCACCGCCGCCGCAGGAAGATGAGCAGGG TGATCACTGTGATGGTGAGTAGGAGGAG
pEGFP_AP VE-Cadherin TMD mut5 F	TGCTGGTGCTGTCGACGATATGGCCGCCCAGGCGGCGGCCGCAATC CAGGCAGTGGTAGCCATCTTACTCTGCATCCTCACCATCACAG
p3.1AgeI_ SP 2xHA VE-Cadherin V5 F	TACCGGTCTCGGATCCATGCAGAGGCTCATGATGCTCC
p3.1AgeI_SP 2xHA VE-Cadherin V5 R	GCCCTCTAGACTCGAGCTACGTAGAATCGAGACCGAGGAGAGGGTT AGGGATAGGCTTACCATACAGCAGCTCCTCCCGGGG

2.1.2 Restriction enzyme digestion

Plasmid DNA was incubated with restriction enzymes for 1h (test digest) or 3hs (backbone digestion) as described in Table 2.4. When the digestion was done with two enzymes, the correct buffer was used according to NEB double digest finder website.

Table 2.4 | Endonuclease restriction digestion protocol

Restriction digest mix (15µl)	Volume
10x Reaction buffer (New England Biolabs)	1.5µl
Restriction enzymes (New England Biolabs)	0.75µl
Vector DNA	~500ng
Water	up until 15µl

2.1.3 Agarose gel electrophoresis and gel purification

Digested PCR product and vectors were resolved in a 1-2% agarose gel, supplemented with SYBR safe DNA Gel Stain (1:10000 dilution; Invitrogen). 1x Tris-Borate-EDTA (TBE) buffer (Sigma-Aldrich) was used. The bands of the correct size were cut using a scalpel and purified using NucleoSpin® Gel and PCR Clean-up kit (TaKara) according to the manual's instructions.

2.1.4 Site directed mutagenesis

Site directed mutagenesis was performed using Cloned Pfu DNA polymerase (Agilent). The reaction was prepared following the protocol described in Table 2.5 and the program in Table 2.6. Control conditions were prepared without Cloned Pfu DNA polymerase. The PCR product was incubated for 1h with DpnI at 37°C (adding 1µl of DpnI to 50µl of PCR reaction) before transformation. The primers used for site directed mutagenesis are listed in Table 2.7.

Table 2.5 | Site directed mutagenesis protocol

PCR mix (50µl)	Volume
10x Cloned Pfu Reaction buffer AD	5µl
10mM dNTP mix	2µl
100ng/µl forward primer	1µl
100ng/µl reverse primer	1µl
100ng/µl template DNA	1µl
Cloned Pfu DNA polymerase AD	1µl
ddH_2O	Up to 50µl

Table 2.6 | Site directed mutagenesis PCR program

Step & temperature	Time	Cycles
Initial denaturation (95°C)	30s	1
Denaturation (95°C)	30s	25-30cycles
Annealing (55°C)	60s	
Extension (68°C)	2min/kb	
Final extension (68°C)	7min	1
Hold (4°C)	/	/

Table 2.7 | List of site directed mutagenesis primers

Primer name	Primer sequence
hRHBDL2 SA F	GATATCTTGTGGGAGCTGCAGGAGGAGTCTATGCT
hRHBDL2 SA R	AGCATAGACTCCTCCTGCAGCTCCCACAAGATATC
pIRES_NG O1 linker F	GACGAGCTGTACAAGGGTGCTGGTGCTGTCGACATCACCCCG
pIRES_NG O1 linker R	CGGGGTGATGTCGACAGCACCAGCACCCTTGTACAGCTCGTC
pIRES_Spitz linker F	ACGAGCTGTACAAGGGTGCTGGTGCTGTCGACCCGCGTCC
pIRES_Spitz linker R	GGACGCGGGTCGACAGCACCAGCACCCTTGTACAGCTCGT
pIRES_LDLR linker F	GGTGGCATGGATGAACTATACAAAGGTGCTGGTGCTGTCGAC GAGGAGC
pIRES_LDLR linker R	GCTCCTCGTCGACAGCACCAGCACCTTTGTATAGTTCATCCAT GCCACC
pIRES_TACE linker F	GGTGGCATGGATGAACTATACAAAGGTGCTGGTGCTGTCGAC GGGAAGT
pIRES_TACE linker R	ACTTCCCGTCGACAGCACCAGCACCTTTGTATAGTTCATCCAT GCCACC

2.1.5 DNA assembly

Purified PCR products and digested backbone were assembled using InFusion HD Cloning Kit (Takara). According to manufacturer's instructions the PCR product included an overlap to the backbone of 15bp. When multiple fragments were assembled, the overlap region between them was increased to 20bp. The reaction was prepared using the protocol in Table 2.8 and incubated at 50°C for 15m, before transformation.

Table 2.8 | InFusion cloning PCR protocol

DNA assembly mix (10µl)	Volume
5x Mix InFusion (TaKara)	2µl
Purified insert DNA fragment	~20ng
Purified and digested backbone plasmid	~20ng
ddH_2O	up to 10µl

2.1.6 Bacterial transformation

5µl of the InFusion assembled vector and 10µl of vector generated by site directed mutagenesis were transformed in 20-30µl of 50ul Stellar™ Competent Cells (Clontech). The bacteria+DNA mix was incubated on ice for 10m, heat-shocked for 1m at 92°C and put back on ice for 2m. Afterwards, bacteria were recovered in 200-300µl of SOC medium for 30m/1h. After being centrifuged at

10000rpm for 1m, the bacteria were resuspended in 100µl of medium and plated in ampicillin (100ng/µl) or kanamycin (30ng/µl) LB (Luria-Bertani) agar plate and incubated overnight at 37°C.

2.1.7 Plasmid preparation and Sequencing

Single colonies were picked and cultured in 5ml of LB media containing either 100ng/µl of ampicillin or 30ng/µl kanamycin. After 16hs, the bacterial cultures are centrifuged at 10000rpm for 1m. Plasmid DNA was extracted from bacteria using QIAprep Spin Miniprep kit (Qiagen) according to manufacturer's instructions. DNA sequencing data of constructed plasmids were obtained using Source BioScience, Cambridge, either using company-provided primers or custom sequencing primers that are listed in Table 2.9. The list of constructed and used vectors in this study can be found in table 2.10.

Table 2.9 | Sequencing primers

Primer name	Sequence
pIRES F	CCCTTGAACCTCCTCGTTCGACC
pIRES R	TTTTATTTTATCGATATCAGCGCTTTAAATTTGCGC
pLEX_F	CACCAAAATCAACGGGACTT
pLEX_R	ATATAGACAAACGCACACCGGCCT
CMFV_PCDNA3	CAACGGGACTTTCCAAAATG
BGH Reverse	TAGAAGGCACAGTCGAGG

Table 2.10 | List of constructs

Name of the vector	Source
pcDNA3.1	Freeman lab
pcDNA3.1_AgeI	Dr Fangfang Lu
pcDNA5FRT	Dr. Michael van de Weijer
pLEX_puro	Freeman lab
pMSCV-IRES-GFP II (pMIG II)	(Holst et al, 2006)
psPAX2	Paul Metz
pMD2.G	Paul Metz
pMSCV mNG O1 TMD/mSc LDLR TMD	Generated in this study
pMSCV mNG Spitz TMD /mSc TACE TMD	Generated in this study
pMSCV mNG LDLR TMD/mSc LDLR TMD	Generated in this study
pMSCV BCAM TMD/mSc LDLR TMD	Generated in this study

pMSCV VE-Cadherin TMD/mSc LDLR TMD	Generated in this study
pMSCV IL11RA2 TMD/mSc LDLR TMD	Generated in this study
pLEX mNG Orai1 TMD/mSc LDLR TMD	Generated in this study
pLEX mNG Spitz TMD/mSc TACE TMD	Generated in this study
pLEX mNG LDLR TMD/mSc LDLR TMD	Generated in this study
pLEX mNG VE-Cadherin TMD/mSc LDLR TMD	Generated in this study
pLEX FLAG-TwinStrep Spitz TMD	Generated in this study
pLEX FLAG-Twinstrep LDLR TMD	Generated in this study
pLEX FLAG-TwinStrep VE-Cadherin TMD	Generated in this study
pLEX FLAG-TwinStrep TACE TMD	Generated in this study
pcDNA_AgeI HA hRHBDL2	Generated in this study
pcDNA_AgeI HA hRHBDL2 SA	Generated in this study
pcDNA_AgeI tBFP hRHBDL2	Generated in this study
pcDNA_AgeI tBFP hRHBDL2 SA	Generated in this study
pcDNA_AgeI hRHBDL2	Generated in this study
pcDNA_AgeI hRHBDL2 SA	Generated in this study
pcDNA5FRT tBFP hRHBDL2	Generated in this study
pcDNA5FRT tBFP hRHBDL2 SA	Generated in this study
pcDNA_AgeI 2xHA VE-Cadherin V5	Generated in this study
pEGFP-N1 mNG-Orai1 TMD	Generated in this study
pEGFP-N1 mNG Spitz TMD	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD	Generated in this study
pEGFP-N1 AP Cadherin10 TMD	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD mut1	Generated in this study
pEGFP-N1 AP Ve-Cadherin TMD mut2	Generated in this study
pEGFP-N1 AP Ve-Cadherin TMD mut3	Generated in this study
pEGFP-N1 AP VE-Cadheri TMD mut5	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD I598P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD Q599P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD A600P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD V601P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD V602P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD A603P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD I604P	Generated in this study
pEGFP-N1 AP VE-Cadherin TMD L605P	Generated in this study

2.2 Cell biology

2.2.1 Cell culture

Human embryonic kidney (HEK) 293T or LentiX293T were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS), 2mM L-Glutamine (Gibco). Cells were cultured at 37°C with 5% CO_2 in a cell culture incubator and were split twice-three times a week using TrypLE Express (Gibco).

2.2.2 DNA transfection

Non liposomal transfection reagent FuGene (Promega) was used for transient transfection in HEK293T cells, 24hs after the cells were seeded in DMEM supplemented with 10% fetal bovine serum (FBS), 2mM L-Glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (all Gibco). A ratio of 1:4 was used to transfect DNA (i.e. 4µl of FuGene for 1µg of DNA). 500ng-2µg of DNA was mixed with OptiMEM, in a separate tube an appropriate amount of FuGene was suspended with 100µl of OptiMEM for 5 minutes. Afterwards, it was added to the DNA containing tube. After 10-15 minutes the transfection mix was added dropwise to the cells.

2.2.3 Lentiviral production and transduction

1E5 LentiX293 cells were seeded in a 6 well plate in DMEM supplemented with 10% FBS, 2mM L-Glutamine. After 24hs, 500ng of pLEX vector was co-transfected with 350ng of psPAX2 and 150ng of pMD2.G to produce the viral particles. After 24hs, cells were changed into fresh media where the viral particles will be collected for 48hs. Viral supernatant was collected and filtered through 45µM filters and either used straightaway to infect target cells or aliquoted and frozen at -80°C. The day before the infection, 1E5 HEK293T were plated. On the day of the infection, 10µg/ml of Polybrene (Sigma-Aldrich) was added to the media. Between 10-40% of viral supernatant was used to infect cells in a dropwise manner. 48hs after infection, cells were trypsonised with 500µl and seeded in

selection media with 2µg/ml puromycin and selected for another 48hs. A list of the generated stable cells can be found in Table 2.11

2.2.4 Monoclonal cell lines

2E5 HEK293T Flp-In TREX were plated in a 6 well plate. After 24hs, 100µg of p5FRT constructs and 900µg pOG44 (1:9 ratio) were transfected and cells were incubated for 24hs at 37°C. The next day, cells were trypsonised with 500µl of TrypLE Express (Gibco) and moved in a 15cm plates with DMEM supplemented with 10% FBS, 2mM L-Glutamine. After 24hs, media was replaced with fresh DMEM containing 15µg/ul of blasticidin and 100µg/µl of Hygromicin B. The selection media was replaced every 3-4 days until colonies were clearly visible. Each clone was picked using a 200µl tip with a few µl of trypsin and moved into a 6 well plate and then expanded until confluency.

Table 2.11 | List of cell lines

Cell line	Description	Source
HEK293T cells		Freeman lab
LentiX293		Hassan lab
HEK293T Flp-In TREX		Carvalho lab
F-TwS VE-Cadherin/mSc LDLR	HEK293T cells transduced with pLEX F-TwS mVE-Cadherin TMD/mSc mLDLR TMD	Generated in this study
F-TwS Spitz/mSc TACE	HEK293T cells transduced with pLEX F-TwS Spitz TMD/mSc mTACE TMD	Generated in this study
F-TwS VE-Cadherin	HEK293T cells transduced with pLEX FLAG-TwinStrep mVE-Cadherin TMD	Generated in this study
F-TwS Spitz	HEK293T cells transduced with pLEX FLAG-TwinStrep Spitz TMD	Generated in this study
F-TwS Orai1	HEK293T cells transduced with pLEX FLAG-TwinStrep mOrai1 TMD4	Generated in this study
F-TwS TACE	HEK293T cells transduced with pLEX FLAG-TwinStrep mTACE TMD	Generated in this study
F-Tws LDLR	HEK293T cells transduced with pLEX FLAG-TwinStrep mLDLR TMD	Generated in this study
tBFP hRHBDL2	HEK293 Flp-In Trex transfected with p5.1 tBFP RHBDL2 (inducible)	Generated in this study

tBFP hRHBDL2 SA	HEK293 Flp-In Trex transfected with p5.1 tBFP RHBDL2 SA (inducible)	Generated in this study

2.3 Biochemical methods

2.3.1 Cell lysis

Cells were washed once with ice-cold PBS and lysed on ice with Triton X-100 lysis buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris-HCl (pH 7.5)) supplemented with EDTA-free complete protease inhibitor mix (Roche). Cells were lysed for 10 minutes and the lysates were cleared by centrifugation at 10000g for 10m in a 4°C centrifuge.

2.3.2 AP shedding assay

5E4 were seeded in Poly-L-Lysine coated plates in triplicates, 24hs before transfection. 250ng of alkaline phosphatase (AP)-conjugated or fluorescently tagged TMD (FP) and 250ng of HA RHBDL2 or HA RHBDL2 SA were transfected using FuGene HD. The day after transfection, cells were washed once with PBS and 300µl of colourless OptiMEM (Gibco, 11058-021) were added for 16-20hs. Supernatant was collected and cells were lysed in Triton X-100 lysis buffer supplemented with EDTA-free complete protease inhibitor mix (Roche). 100µl of supernatant and diluted lysates were with 100µl incubated with AP substrate p-Nitrophenyl Phosphate (PNPP, Thermo scientific) at room temperature and absorbance was read at 404nm with a plate reader (SpectraMax M3, Molecular Devices). For the fluorescently-tagged TMD, 150µl of supernatant and cleared lysates were assayed and the absorbance was read at 488nm with the same plate reader. The percentage of AP or FP released (the ratio of released AP/FP in the media to the total AP/FP in media+lysates) was calculated to minimise variability due to transfection efficiency.

2.3.3 SDS-page and Western blotting

Protein concentration in the lysates were measured using a Bradford assay and reading absorbance at 595nm using a spectrophotometer. The same concentration of protein was separated by sodium-dodecyl-sulfate acrylamide gel electrophoresis (SDS-PAGE). Proteins were separated according to their size using Novex™ 8-16% or 4-12% Tris-Glycine Mini Gels with WedgeWell™ format (Thermo Scientific) and Tris-Glycine running buffer (25 mM Tris,192 mM glycine,0.1% SDS) was used. Proteins were transferred to methanol activated polyvinylidene difluoride (PVDF) membrane (Millipore) at 100 V for 90 min using transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol). The membranes were blocked in 5% milk in PBS+0.05% Tween20 (PBST) for 1 hour. 5% milk was used to incubate with primary and secondaries antibodies and PBST was used for washing. Enhanced chemiluminescence (ECL) reagent or ECL Select western blotting detection reagent (GE Healthcare) was used to detect antibodies. Membranes were exposed with X-ray films (Thermo scientific) or super sensitive films (Amersham Hyperfilm ECL, GE Healthcare). The films were developed in an SRX-101A medical film processor. A list of the antibodies used in this study can be found in Table 2.12.

Table 2.12 | Antibodies used for western blotting

Antibody	Dilution	Manufacturer
Anti-HA, monoclonal antibody	1:2500	Cell signaling (#3724)
Anti-V5, monoclonal antibody	1:2500	Thermo Fisher (SY30-01)
Anti-mNeonGreen, monoclonal antibody	1:2500	ProteinTech (32F6)
Anti-RFP, monoclonal antibody	1:2500	Chromotek (6g6-100)
Anti-HA HRP, rat monoclonal	1:2500	Roche (12013819001)
Anti-beta actin, mouse monoclonal	1:2500	Santa cruz (sc-47778)
Anti- beta actin HRP, mouse monoclonal	1:5000	Sigma-Aldrich (A3854)
Anti RHBDL2, rabbit monoclonal	1:2500	Protein Tech (12467-1-AP)
Anti Tubulin, mouse monoclonal	1:2500	Abcam (Ab7291)
Anti-rabbit HRP, goat polyclonal	1:5000	CST (7074)
Anti-mouse HRP, horse polyclonal	1:5000	CST (7076)

2.3.4 Immunofluorescence and confocal microscopy

4E5 HEK293T were seeded in Poly-L-Lysine coated coverslips a day before transfection. 48hs after transfection, cells were washed once with PBS and and fixed in 4% paraformaldehyde diluted in PBS (PFA) for 20m at room temperature. Washed them once with PBS and quenched with 50mM NH_4Cl for 5m. Cells were permeabilised with 0.2% Tx-100 in PBS for 20m. Cells were washed with PBS and blocked with 3% BSA for 1h. Cells were incubated with primary antibodies for 1h, washed with PBS three times and incubated with Alexa-fluor secondary antibodies for 1 hour. After three washes in PBS and once in water, coverslips were mounted using mounting medium containing DAPI (ProLong Gold; ThermoFisher Scientific). Images were taken using an inverted Olympus FV1000 confocal microscopy, A 60x/1.40 Oil UPlanSApo objective (Olympus) was used, and the images were analysed in ImageJ. A list of antibodies using for confocal microscopy can be found in Table 2.13.

Table 2.13 | List of antibodies using for confocal microscopy

Antibody	Dilution	Manufacturer
Anti-V5, rabbit monoclonal	1:100	Thermo Fisher (SY30-01)
Anti-actin phalloidin 568	1:200	Life technologies (A12380)
Anti-actin phalloidin 488	1:250	Life Technologies (A12379)
Anti-BAP31, mouse monoclonal	1:500	Enzo Life Science (ALX-804-601-C100)
Anti-rabbit Alexa fluor 568, donkey polyclonal	1:400	Invitrogen (A10042)
Anti-rabbit Alexa Fluor 647, goat monoclonal	1:200	Invitrogen (A27040)
Anti-mouse Alexa fluor 488, donkey polyclonal	1:200	Invitrogen (A-21202)

2.3.5 Flow cytometry and fluorescent-activated cell sorting (FACS)

To minimise endocytosis, all steps were performed on ice. Cells were resuspended by gently doing up and down in the media and collected. For fluorescently-tagged TMD, 3E6 of each condition were pelleted at 1000rpm for 4m and resuspended in FACS buffer (PBS+ 1% FBS) and stored in ice. For surface labelling experiments, 3E6 cells were collected, spun down at 1000rpm for 4 minutes. Cells were stained with 50µl of FLAG-DyLight 488 conjugated antibody (1:500, Sigma) in FACS buffer.

Isotype control were stained using Mouse IgG2a Isotype Control DyLight™ 488 (Invitrogen). Cells were washed twice with ice-cold FACS buffer and resuspended before analysis. Dead cells were stained with DAPI or DRAQ7 (Abcam). Cells were analysed using Cytoflex LX and FlowJo software. All antibodies used can be found in Table 2.14. For cell sorting, the same protocol was used but 5E6 cells were labelled and sorted using BD FACSAria III (Beckman). Sorted cells were collected in DMEM supplemented with 10% fetal bovine serum (FBS), 2mM L-Glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (all Gibco).

Table 2.14 | List of antibodies used for flow cytometry

Antibody	Dilution	Manufacturer
anti-FLAG M2, mouse monoclonal	1:500	Sigma (F1804)
Anti-DYKDDDDK ANTIBODY DYLIGHT 488, mouse monoclonal	1:500	Thermo scientific (**200-341-383)**
Anti-mouse 488, donkey monoclonal	1:1000	Life Technologies (A21202)
Anti-mouse 647, donkey monoclonal	1:1000	Life Technologies (A31571)

2.4 TMD library methods

2.4.1 Bioinformatics to generate TMD library

The TMD sequences of type I single pass transmembrane proteins were downloaded through Uniprot, in total 1146 sequences. The sequences were selected by the Uniprot provided list for reviewed type I single pass proteins. By using a Python script written by Dr. Samuel Dean, the protein sequences were downloaded and the corresponding 7aa were added to the luminal side and 3aa to the cytoplasmic side. Then, the FASTA sequences were turned into DNA strings keeping the GC content as close as possible to 50% for minimising any issues with cloning. Moreover, the

choice of bases for the DNA sequences was done keeping into consideration the human codon usage. SpeI and AgeI restriction sites were avoided to facilitate downstream cloning.

2.4.2 Bicistronic vector cloning

For the generation of the bicistronic vector, the sequences of Orai TMD4, LDLR TMD, VE-Cadherin TMD, Spitz TMD and TACE TMD and HB-EGF signal peptide (SP) were PCR amplified from pcDNA3.1_TMDscreen (Grieve et al., 2020) with Phusion polymerase (Thermo scientific). The PCR products were analysed by gel electrophoresis and bands of the correct size were purified using a gel and PCR clean-up kit (Macherey-Nagel). The purified PCR products were then fused to mNeonGreen or mScarlet to obtain SP-mNeonGreen-TMD and SP-mScarlet-LDLR,TACE TMD. These two final products were subsequently cloned between EcoRI-XhoI and NcoI-AgeI of pMSCV-IRES-GFP II (pMIG II) (Addgene #52107) respectively, using an In-Fusion HD Cloning kit (Takara). The PCR protocol and the program have been described in Table 2.1 and 2.2. To generate FLAG-TwinStrep (F-TwS) TMDs, the F-TwS tag was ordered from IDT, PCR amplified and inserted into the bicistronic vector replacing mNeonGreen. To generate the lentiviral bicistronic vectors, the whole insert mNG-TMD-IRES-mSc TMD was amplified by Phusion PCR and sub-cloned into pLEX within the EcoRI and NotI sites. The primers used to clone the bicistronic vectors are listed in Table 2.15.

Table 2.15 | List of primers for bicistronic vectors

Primer name	Primer sequence
HB_EGF_SP-mNG F	CTAGGCGCCGGAATTCATGAAGCTGCTGCCGTCG
HB_EGF_SP-mNG R	CCCGTCGACCTTGTACAGCTCGTCCATGC
HB_EGF_SP-mSc F	CCCGTCGACCTTGTACAGCTCGTCCATGC
HB_EGF_SP-mSc R	CCCGTCGACTTTGTATAGTTCATCCATGCC
mNG-Orai1 TMD4 R	ATTGATCCCGCTCGAGTCACAGGGAACGGTAGAAGTGAACG
mNG-Spitz TMD R	ATTGATCCCGCTCGAGTCACTCGAAGCGCAAATAGAAGGCC
mNG-LDLR TMD R	ATTGATCCCGCTCGAGTCACCAGTTCCTCCACAGC
mNG-VE-Cadherin TMD F	TGCTGGTGCTGTCGACGCAGCCCAGGCGGGTGTCAG
mNG-VE-Cadherin TMD R	ATTGATCCCGCTCGAGTCACCGCCTCCGCAGGATGATC
mSc-LDLR TMD R	TTCTAGAGGTACCGGTTCACCAGTTCCTCCACAGCAGG
mSc-TACE TMD R	TTCTAGAGGTACCGGTTTAATCCACACAGTGGACAAGAATGCTG
mSc-Orai1 TMD4 R	TTCTAGAGGTACCGGTTCACAGGGAACGGTAGAAGTGAACG

mSc-Spitz TMD R	TTCTAGAGGTACCGGTTCACTCGAAGCGCAAATAGAAGGCC
pLEX_2TMD F	CTCTACTAGAGGATCCCTTCTCTAGGCGCCGGAATTC
pLEX_mSc LDLR R	GAGAGGGGCGACCGGTTCACCAGTTCCTCCACAGCAG

2.4.3 TMD lentiviral vector cloning

To generate pLEX_FLAG TwinStrep (F-TwS) TMDs, the insert HB-EGF-SP F-TwS TMD was cloned from the bicistronic vector and sub-cloned into lentiviral vector pLEX within the BamHI and AgeI using InFusion cloning. To simplify the construction of the library, the TMD sequence could be cut out from the backbone using SpeI-AgeI. The primers used to sub-clone the control TMDs into the pLEX backbone are listed in table 2.16. The cloning of the library of TMDs into pLEX lentiviral vector was done by TwistBioscience. The company was also responsible for the quality control of the library after cloning.

Table 2.16 | List of primers for lentiviral vector cloning

Primer name	Primer sequence
pLEX_ single TMD F	CTCTACTAGAGGATCCATGAAGCTGCTGCCGTCGG
pLEX_ Orai1 TMD4 R	GAGAGGGGCGACCGGTTCACAGGGAACGGTAGAAGTG
pLEX_Spitz TMD R	GAGAGGGGCGACCGGTTCACTCGAAGCGCAAATAGAAG
pLEX VE-Cadherin TMD R	GAGAGGGGCGACCGGTTCACCGCCTCCGCAGGATG
pLEX_LDLR TMD R	GAGAGGGGCGACCGGTTCACCAGTTCCTCCACAGCAG
pLEX_TACE TMD R	GAGAGGGGCGACCGGTTTAATCCACACAGTGGACAAGAATG

2.4.4 TMD pool virus production and transduction

2E6 LentiX were seeded 24hs after transfection in 3 15cm plates with 15ml of DMEM supplemented with 10% FBS and L-Glutamine. On the day of transfection, 2µg of TMD pool library or mCherry vector were co-transfected with 1,4µg of psPAX2 and 600µg of pMD2.G. FuGene (Promega) was used as a transfection reagent using a ratio of 1:4 (4µl of FuGene for 1µg of DNA). 24hs after transfection, media was replaced with 15ml fresh media without any antibiotic where the virus was collected for 48hs. Viral supernatant was filtered through 45µM filters and aliquoted to be stored at -80°C. Around 6hs before transduction, 4E6 HEK293T cells were seeded in 15cm plates with 15ml of DMEM without any antibiotic. Fresh vials of viral supernatant were thawed on ice and 10µg/ml

of polybrene (Sigma) were added to the cells. 200µl of virus were used to infect cells to achieve an MOI of 0.15 (to maintain a fold coverage of ~500). Two days after transduction, cells were trypsonised and seeded in new plates with selection media with 2µg/ml of Puromycin (Sigma-Aldrich). Selection lasted for 4 days or until all uninfected cells of the negative control were dead. Afterwards, cells were expanded until confluency in media without any antibiotic.

2.4.5 MOI titration

3.5E6 HEK293T cells were seeded in two sets of 15cm plates on the day of infection. After 6hs, 10µg/ml of polybrene were added and cells were infected with either 0,50,70,100,200µl of viral supernatant and one plate per set was infected with 500µl of mCherry-producing virus as positive control. After 48hs, one set of 15cm plates was trypsonised and seeded in selection media with 2µg/ml of puromycin, the other set was washed once with PBS and fresh media without selection was added. 4 days after or when all the cells in the negative control plate were all dead, all cells were trypsonised with 5ml of TrypLe express and 100µl of cell suspension was mixed with 100µl of Trypan blue. Dead and alive cells were counted using a manual haemocytometer. The percentage of alive cells was calculated by dividing the number of alive cells in the selected plate divided by the number of cells in the corresponding unselected plate.

2.4.6 TMD pool sorting

5E6 of TMD pool-expressing cells were seeded in 15cm plate with full DMEM. 24hs later, 8µg of tBFP RHBDL2 or tBFP RHBDL2 SA were transfected using FuGene HD. After 48hs, cells were gently resuspended in the media and the cell number was counted. At least 7E6 cells were collected and spun down at 1000rpm for 4m in a cold centrifuge and kept on ice throughout the rest of the protocol to decrease endocytosis. Cell pellets were resuspended in 50µl of FACS buffer+FLAG DyLight 488 antibody as described in 2.3.5. Cells were washed twice with ice-cold FACS buffer before

resuspending them in 500µl of FACS buffer. 2µl of DRAQ7 viability dye (Sigma) were added right before sorting. Sorting was done using a FACSBDAria III. tBFP+ cells were first selected, and then the cell population was divided into 4 bins according to the green fluorescence. Each bin was sorted, and cells were collected in FACS buffer. Cells were then spun down, and cell pellets were frozen before being further processed.

2.4.7 DNA extraction and PCR amplification

Genomic DNA was extracted using Quick-DNA Microprep Kit according to manufacturer's instructions. DNA concentration was calculated using Qubit™ dsDNA Quantification Assay Kits and a Qubit fluorometer. TMD sequences were amplified by PCR with Q5 High-fidelity polymerase (NEB) using primers that annealed to common upstream and downstream region of the lentiviral backbone. The sequences of the primers, the Q5 PCR mix and protocol are listed in table 2.17, 2.18 and 2.19. The amplified TMD sequences were separated from remaining template and PCR mix and also from longer unspecific PCR products using solid phase reversible immobilization (SPRI) beads. Before and after the SPRI selection, the size of the PCR product was checked using a TapeStation.

Table 2.17 | List of sequencing primers

Primer name	Primer sequence
F3	CACCCGCAGTTCGAAAAAGGT
R1	GGAGAGGGGCGACCGGT

Table 2.18 | Q5 PCR protocol

PCR mix (50µl)	Volume
5x Q5 Buffer	10µl
10mM dNTPs	1µl
10µM Forward primer	1,25µl
10µM Reverse Primer	1,25µl
Template DNA	30ng
Q5 High fidelity DNA polymerase	0,5µl
5x Q5 GC Enhancer	10µl
ddH_20	Up to 50µl

Table 2.19 | Q5 PCR protocol

Step&Temperature	Time	Cycle
Initial denaturation (98°C)	30s	1
Denaturation (98°C)	10s	25 cycles
Annealing (60°C)	30s	
Extension (72°C)	15s	
Final extension (72°C)	2m	1
Hold (4°C)	/	

2.4.8 TMD library preparation and sequencing

The preparation for the nanopore sequencing library was done using Native Barcoding Kit 24 V14 and Ligation Sequencing Kit V14 from Oxford Nanopore technologies following manual's instructions. The 24 adapters were used to barcode each of the 8 bins (4 bins for tBFP RHBDL2 and 4 bins for tBFP RHBDL2 SA) for three independent replicates. The sequencing run was performed with a flow cell R10 version in a MinION Mk1B over 72hs collecting a total of ~9 million reads.

2.4.9 Data analysis

The sequencing reads were base called using a high accuracy model. ~7.7E6 reads were then demultiplexed and assigned to a specific barcoded. After assigning the read to a barcode, both barcodes and adapters got trimmed. ~7E6 reads were further processed by identifying and removing the sequencing primers. The processed reads were compared to an indexed list of the TMDs present in the library using Minimap2. Each read in the bins was assigned to one TMD using the nanopore sequences alignment mode of Salmon, this produced a read count for each TMD. The read counts were analysed using EdgeR that removed TMDs which had a low read count, calculated a normalisation factor using TMM method and normalised the read count to the library size. The average fluorescence of each TMD was calculated to identify the ones that showed a decreasing fluorescence. First, each bin got assigned an arbitrary value of 1-8, the read count of a TMD in each bin got multiplied by this value and then divided by the sum of reads for that TMD in the bins. This created a value of average fluorescence (AU) for each TMD when either tBFP RHBDL2 or SA was

expressed. One-tailed t-test was performed to identify the TMDs that shown a reduction of the AU >0.005. The full list of commands and programs used to analyse the sequencing data can be found in Appendix 5.

Chapter 3

Proof of concept and troubleshooting of a new high throughput genetic screen to identify novel substrate for RHBDL2

3. Proof of concept and troubleshooting of a new high throughput genetic screen to identify novel substrate for RHBDL2

3.1 Introduction

In the years since the discovery of rhomboid proteases, their role has been elucidated in many organisms (Freeman, 2014; Lastun et al., 2016). A fundamental step in understanding which pathways are regulated by rhomboid protease has been identifying which are their substrates and validating the cleavage event. The majority of rhomboid substrates, to date, has been identified using candidate-based approaches. Potential substrates were selected based on their similarities with previously known substrates or because they showed typical characteristics of rhomboid substrates such as helix-destabilising residues within their TMD. These methods have been successful and led to the identification of a number of substrates for both RHBDL2 and RHBDL4 (Fleig et al., 2012; Grieve et al., 2021a; Lohi et al., 2004).

However, the main disadvantage of using candidate-based approaches is that they become less effective as the number of potential substrates increases, making them ill-suited for high throughput screening. In fact, each potential substrate has to be selected and then individually tested, making the process of screening thousands of TMDs time-consuming. Developing a new method that would allow the systematic discovery of rhomboid substrates will help speed up the discovery process and potentially help expand the current repertoire of substrates.

For this reason, the main focus of my DPhil has been the development of an assay that would allow a high throughput screening of a large library of potential substrates for rhomboid proteases. This new method is based on flow cytometry, and it uses the reduction of fluorescence as a readout for the cleavage. Since I was developing a completely new method, it was essential to test each step of the screen to assess whether it was working properly (Figure 3.1). In the following chapter I will describe the set-up of the screen and experiments that confirmed the efficacy of the screen. Moreover, I will present the extensive troubleshooting that was necessary. As it will be

discussed, there were a few setbacks and changes to the overall set up of the screen that took a longer time to solve than it was originally planned. However, they were necessary to generate a sensitive and efficient screen.

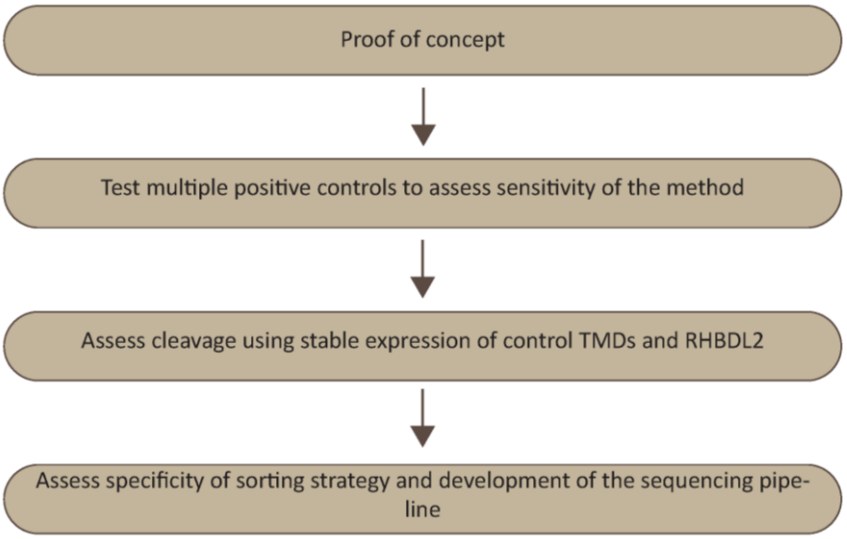


Figure 3.1 Workflow of experiments needed to set up the screen

3.2 Proof of concept experiments using a bicistronic vector expressing two fluorescently tagged TMDs

3.2.1 Schematic of the screen

The majority of identified substrates of rhomboid proteases are single pass proteins with a type I topology, i.e. luminal N-terminus and cytosolic C-terminus. For this reason, the screen will focus on determine which one of these proteins are rhomboid substrates. Specifically, I will focus on RHBDL2 (R2) which is the rhomboid protease present at the plasma membrane (PM) and the best characterised among mammalian secretase rhomboids (Grieve et al., 2021a; Johnson et al., 2017; Urban et al., 2001). Having a list of validated substrates for RHBDL2 meant that some of these

TMD could be used as positive and negative controls in the preliminary steps of the screen to test its efficacy.

Initially, the screen was based on the use of a bicistronic vector that coded for two different TMDs (Figure 3.2A). The translation of two genes from the same vector was allowed due to the presence of an internal ribosome entry site (IRES). IRES are RNA element with a secondary or tertiary structure, that permits the translation of a gene in a CAP-independent manner by recruiting the 40S ribosome subunit (Martínez-Salas, 1999). Using a bicistronic vector allows for a tighter control of the two genes expression since they are both expressed from the same promoter. Both TMDs were N-terminally tagged with a fluorescent tag and since they had a type I topology it meant that the tag was luminal or extracellular. Moreover, the signal peptide of HB-EGF (SP) was added to ensure insertion of the tag and TMD through the cellular membrane in the right topology. One of these TMDs would be part of a library of putative substrates and was tagged with an mNeonGreen (mNG) tag, while the second TMD was tagged with mScarlet (mSc) and acted as an internal control. Each of the TMDs was expressed with RHBDL2 (Figure 3.2B) and the rational of the screen was: R2 would cleave the mNG-tagged TMD causing the release or degradation of the fluorescent protein, leading to a decrease in green fluorescence. Meanwhile, the mSc-tagged TMD would not be cleaved so the red fluorescence would remain unchanged. The change in mNG/mSc ratio could be assessed via fluorescent-activated cell sorting (FACS) (Figure 3.2C) and cells that showed the decrease in fluorescence ratio would be sorted. The sorted cells could then be analysed via next generation sequencing (NGS) using a pair of primers (Figure 3.2D) that would specifically anneal to the upstream and downstream region of mNG-tagged TMD. By sequencing which TMDs were present in the population of sorted cells, a list of putative substrates would be generated and later validated.

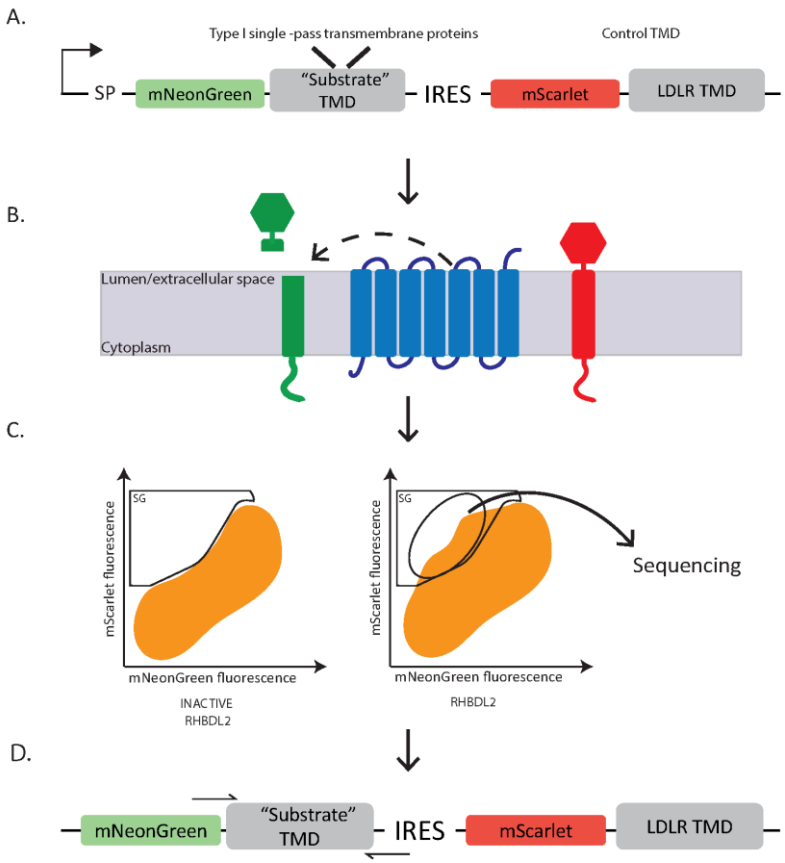


Figure 3.2 Schematic of the setup of the screen using a bicistronic vector. A) A bicistronic vector containing an IRES will code for two type I TMDs, both of them fluorescently tagged. The mNeonGreen (mNG) tagged TMD will be part of a library including the TMD of all type I single pass transmembrane proteins in the human genome. The mSc TMD is LDLR TMD which would act as an internal negative control. To ensure correct insertion there is a signal peptide sequence upstream of the two TMDs. **B)** The vector will be coexpressed with either RHBDL2 or inactive RHBDL2. The cleavage of the green TMD will lead to a decrease of the mNG/mSc ratio that can be assessed via flow cytometry. **C)** The cells whose fluorescence ratio has decreased will be selected by using a gate (SG) that, in the absence of active RHBDL2, contains almost no cells. **D)** Cells that enter the SG after expression of RHBDL2 will be sorted and sequenced using a pair of primers that anneal to the common regions upstream and downstream of the mNG-tagged TMD.

3.2.2 Proof of concept experiments confirm that decrease in mNG/mSc ratio can be used as a readout for RHBDL2 cleavage

The first preliminary experiments were aimed at assessing whether the decrease in fluorescence could be used as a readout for cleavage. To test this, I used the 4th TMD of mouse Orai1, and the TMD of mouse LDLR as positive and negative control, respectively (Grieve et al., 2021a) (Figure 3.3A). Each TMD sequence was flanked by a short stretch of amino acid on the luminal and cytosolic side. The cytosolic region was only 3aa long: this would be enough to ensure the TMD stability but too short to provide any sort of cytosolic regulation. The luminal region was 7aa long to take into consideration the proteolytic mechanism of rhomboid proteases (Vinothkumar et al., 2010). The active site of rhomboid is open to the lumen and the cleavage event can occur at the juxtamembrane region of the luminal side. By having a slightly longer luminal region, we ensured that any substrate that gets cleaved in the juxtamembrane region would not be missed. Moreover, this would also account for potential errors in setting the TMD border.

A bicistronic vector coding for mNeonGreen Orai1 TMD4 (mNG O1) and mSc LDLR TMD (mSc LDLR) were transiently transfected into HEK293T cells with either 3xHA RHBDL2 (RHBDL2) or a catalytically version of the protease where the catalytic serine S187 was mutated into an alanine (RHBDL2 SA). The inactive form of RHBDL2 was used as a negative control as it is still able to bind a substrate TMD but cannot cleave it.

The transfected cells were first analysed by immunoblot and showed good expression of both mNG O1 and mSC LDLR (Figure 3.3B). In the presence of RHBDL2 there was a significant decrease of mNG O1 levels *(green arrow),* while mSc LDLR levels did not show any appreciable change. Moreover, the decrease of mNG O1 expression did not occur when co-expressed with RHBDL2 SA or in the absence of rhomboid. This was consistent with the idea that, upon cleavage, mNG gets released into the media. However, the immunoblot detected other unexpected bands. A smaller band, around 25kDa, *(orange arrow)* was consistent with the size of mNG alone. Since this

band was also present in the absence of RHBDL2, it is most likely generated by an unspecific clipping of the tag by proteases localised throughout the secretory pathway and an accumulation of mNG inside the cells. A band of a higher molecular weight (*black arrow*, around 37kDa) was also detected. How this band was generated was never fully understood, but the main hypothesis was that it could arise from incomplete trimming of the signal peptide. Nonetheless, this band was also decreasing when co-expressed with RHBDL2 suggesting that the longer luminal side was not impairing the cleavage.

The transfected cells were further analysed via flow cytometry to assess the change in mNG/mSc ratio upon RHBDL2 expression (Figure 3.3C). In order to that, a gate (sorting gate, SG) was drawn in the control population of cells transfected with RHBDL2 SA. The gate was drawn at the outermost border of the cell population, in such a way that in the absence of active RHBDL2, there were almost no cells present in the gate. Upon co-expression of the active protease, mNG O1 was cleaved causing a decrease in the fluorescence ratio. This resulted in a shift of the cell population into the SG where the percentage of cells present increased from the 0.31% of the control population to 11.9% of cells when RHBDL2 was expressed. By looking at the median fluorescence intensity (MFI), it appeared that the decrease in fluorescence ratio was due to a specific reduction of the mNG MFI while the mSc MFI was unchanged between the two conditions. To further confirm that the decrease of the ratio was due to the cleavage of a substrate TMD, I repeated the experiment with the two conditions (co-transfecting either RHBDL2 or RHBDL2 SA) but using a bicistronic vector that coded for mNG LDLR and mSc LDLR, so that both green and red TMDs would be of a non-substrate (Figure 3.3D). Given that LDLR is not a rhomboid substrate, expression of RHBDL2 should not lead to an increase of cell in SG. As expected, the percentage of cells present in SG was consistent between the two conditions.

In these first experiments I established a new approach for screening that allows for the discrimination of rhomboid substrates based on the loss of substrate fluorescence. I have tested

the sensitivity of this method to rhomboid proteolytic activity, showing that the catalytic activity is necessary to decrease green fluorescence. Secondly, the loss of mNG occurs only when it is fused to a substrate TMD. Overall, these results provide encouraging evidence that sorting cells based on their fluorescence is a sensitive method to identify potential substrates for rhomboid proteases.

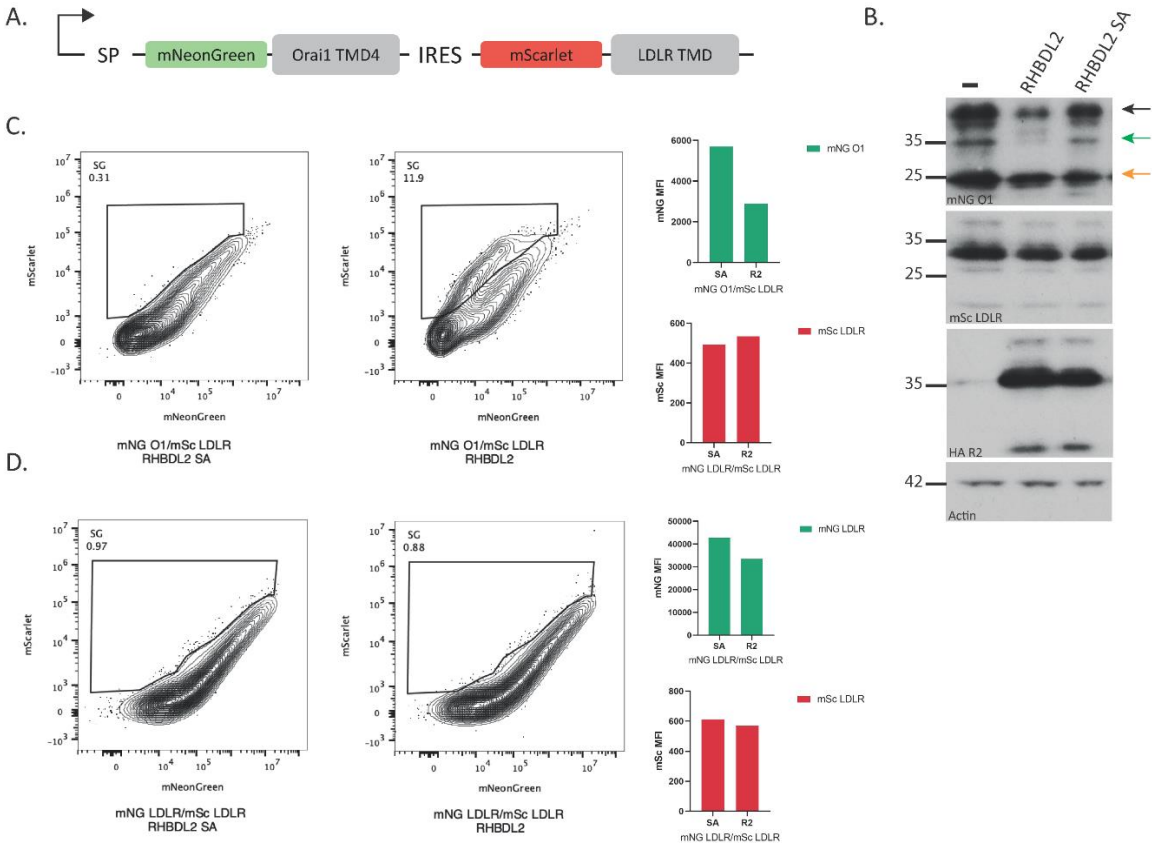


Figure 3.3 Proof of concept experiment show that decrease in mNG/mSc ratio can be used as readout for cleavage. A) Schematic of bicistronic vector coding for the 4th TMD of Orai1 and LDLR. **B)** Western blot analysis of the bicistronic vector showed reduction of full-length form of mNG O1 (green arrow) when co-expressed with RHBDL2. A 25kDa band was detected (orange arrow) consistent with mNG alone, and a bigger band (black arrow) suggests that there are higher molecular weight species potentially derived from defects in signal peptide processing. **C)** Flow cytometry analysis of mNG O1/mSc LDLR showing a shift of the cell into SG and a reduction of the green fluorescence upon co-expression of RHBDL2. Bar charts (also in Figure 3.3D) show the values of the median fluorescence intensity (MFI) for mNG and mSc. MFI defines the average fluorescence intensity of the cell population. **D)** Flow cytometry analysis of vector coding for mNG LDLR and mSc LDLR showed no decrease of fluorescence ratio in the presence of active RHBDL2.

3.3 Testing the new method using a larger range of substrates

3.3.1 Four positive controls do not show any decrease of fluorescence when co-expressed with RHBDL2

Given the positive result of the proof-of-concept experiments to test the rationale of the screen, it was important to assess the sensitivity of the screen and its ability to screen a more complex pool of substrates. To this end, I tested the cleavage of 4 validated substrates: mIL11RA2, mVE-Cadherin, mBCAM, dSpitz (Grieve et al., 2021a; Johnson et al., 2017; Urban et al., 2001). Employing the same reasoning used for the proof-of-concept experiments, the TMD of these proteins flanked by a short luminal and cytoplasmic region, was cloned into the bicistronic vector with the N-terminal mNG tag. mSc LDLR was used as internal negative control in all cases apart from vector coding mNG Spitz. In this case, the negative control was the mSc-tagged TMD of mADAM17 (Grieve et al., 2021a). The newly generated vectors were co-expressed either with RHBDL2 SA or RHBDL2, and the cleavage was assessed via flow cytometry (Figure 3.4A, left). Unexpectedly, there was no significant shift of the cell population into SG and the percentage of cells in the gate remained unchanged upon co-expression of RHBDL2, for all of the 4 tested TMDs. The mNG MFI also looked unaffected (Figure 3.4A, right) and showed only a slight increase when RHBDL2 was transfected. Given that IL11RA2, VE-Cadherin, BCAM and Spitz had already been validated as RHBDL2 substrates in previous studies, I hypothesised that the lack of cleavage was more likely due to a technical problem rather than an experimental one. To confirm that, I assessed the cleavage by two different methodologies. Firstly, I used a fluorometric assay to measure the release of mNG into the media (Figure 3.4B). Upon RHBDL2 co-expression there was a significant release on mNG into the media that was not occurring with expression of RHBDL2 SA or in the absence of any rhomboid (EV), meaning that the accumulation of mNG was due to a specific cleavage of the TMDs by RHBDL2.

Secondly, I repeated the experiments using the same conditions but assessing the cleavage using Western Blotting (Figure 3.4C). When co-transfected with RHBDL2, there was a significant decrease of the full length TMDs (*black arrow*). A longer exposure also allowed to detect the cleavage product (*green arrow*) that was not generated in the absence of RHBDL2 or with RHBDL2 SA. This result proved undeniably that the TMDs of IL11RA2, VE-Cadherin, BCAM, Spitz were being cleaved by RHBDL2. Interestingly, a third band of around 25kDa was detected (*orange arrow*). As previously observed with mNG O1 (orange arrow, Figure 3.3B), the size of the band was consistent with mNG alone that accumulates within the cells and does not undergo degradation. The presence of the fluorescent tag inside the cells would create a very strong background that could impair the readout of the cleavage by flow cytometry. If this was the case, even though the green MFI was decreasing due to the release of mNG in the media, the reduction would not visible due to the pool of mNG present in the cell. Notably, although this population of mNG was present when using all the TMDs that had been tested, it did not impair the readout of the cleavage for mNG O1. The reason for this difference was not investigated, but I decided to focus on finding a solution to visualise the cleavage for the other positive controls in the interest of setting up a more sensitive method.

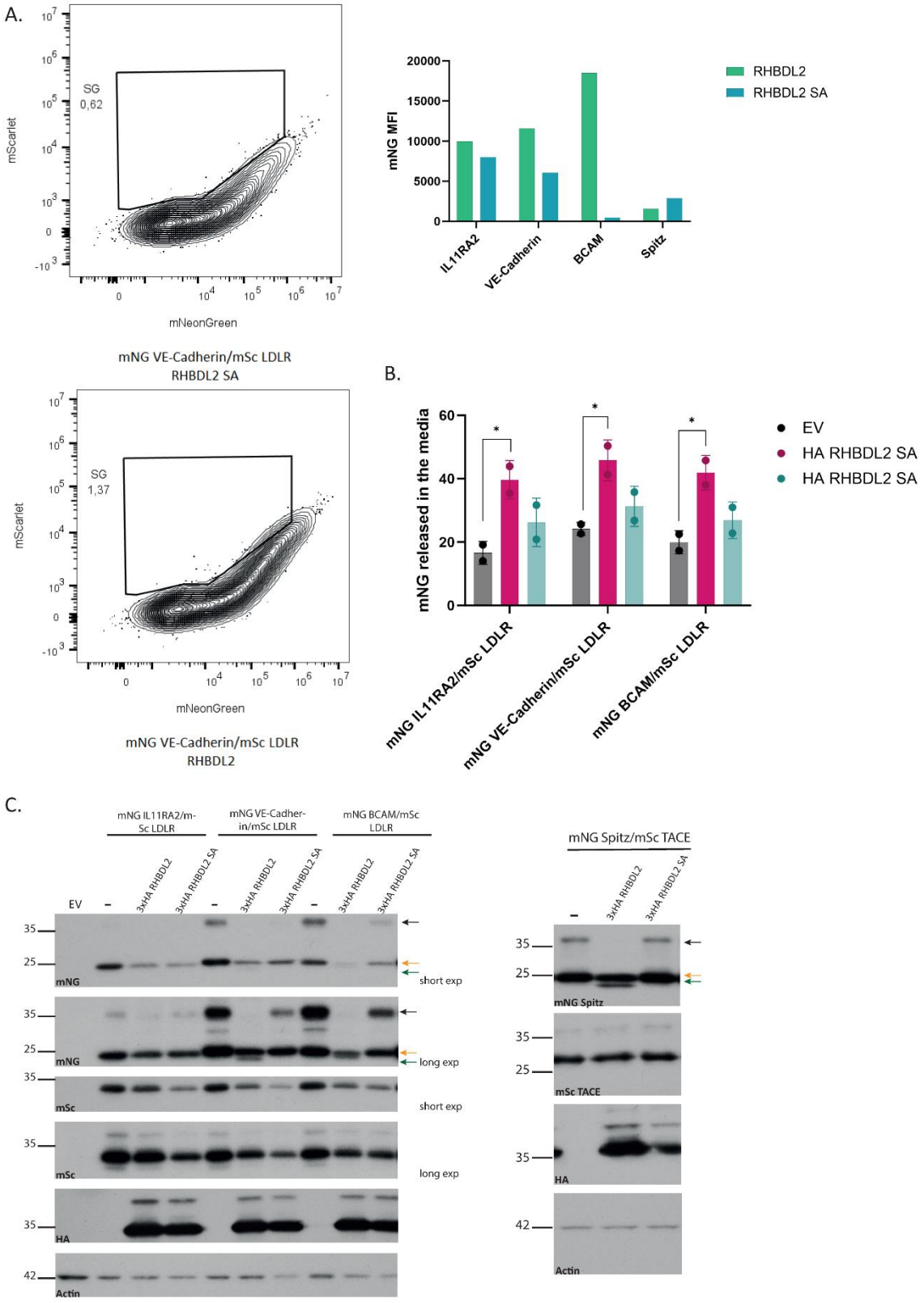


Figure 3.4 Cleavage of four positive control TMDs was not detectable. A) *Left:* contour plot of mNG VE-Cadherin/mSc LDLR shows no shift of the cells in the SG. *Right:* bar chart of mNG median fluorescence of

mNG IL11RA2, VE-Cadherin, BCAM, Spitz/mSc LDLR showing no reduction when comparing co-expression of RHBDL2 or RHBDL2 SA. **B)** Fluorometric assay where the quantity of mNG released in the media was calculated and it shows a significant shedding of mNG upon expression of RHBDL2. Error bars represent median fluorescence of two independent replicates. Statistical tests were performed using unpaired t-test. *= p value <0.01. **C)** Western blotting of positive controls TMDs and mSc LDLR shows cleavage and reduction of full-length upon RHBDL2 co-expression (green and black arrow). It also shows accumulation of mNG within cells (orange arrow).

3.3.2 Using surface labelling allowed for the detection of cleavage for all the tested substrate TMDs

The mNG TMDs were an artificial construct that did not contain the localisation information that can be usually found in full-length proteins, so I expected the tagged TMDs to be present all along the secretory pathway. Moreover, RHBDL2 is localised at the plasma membrane (PM) (Lastun et al., 2016), so it is safe to assume that most of the cleavage events happened in that compartment. In line with this, the previous results showed that mNG was released into the media upon cleavage. Keeping all of this in mind, a way of overcoming the strong fluorescent background inside the cells, was to focus on the pool of TMDs present at the PM by labelling them using a fluorescent antibody (Figure 3.5A). To this end, I repeated the flow cytometry experiment using mNG O1/mSc LDLR expressed with either RHBDL2 or RHBDL2 SA. The cleavage was assessed by measuring the MFI of the total cell population (Figure 3.5B) or by focussing on the pool at the PM using an antibody recognising mNeonGreen that had a far-red fluorescence (AF 647, Figure 3.5C). As expected, the percentage of cells in SG increased when considering the MFI of the whole cell population (from 0.43% to 5.4%). However, focussing only on the cleavage of the TMDs at the PM generated a much bigger shift, from 0.77% to 13.3%. This suggested that the fluorescent background was indeed reducing the amount of cleavage we were able to assess. Moreover, using surface labelling did not impair the specificity of the cleavage, as using a bicistronic vector coding for mNG LDLR and mSc LDLR did not cause any shift of the cells into SG (Figure 3.5D).

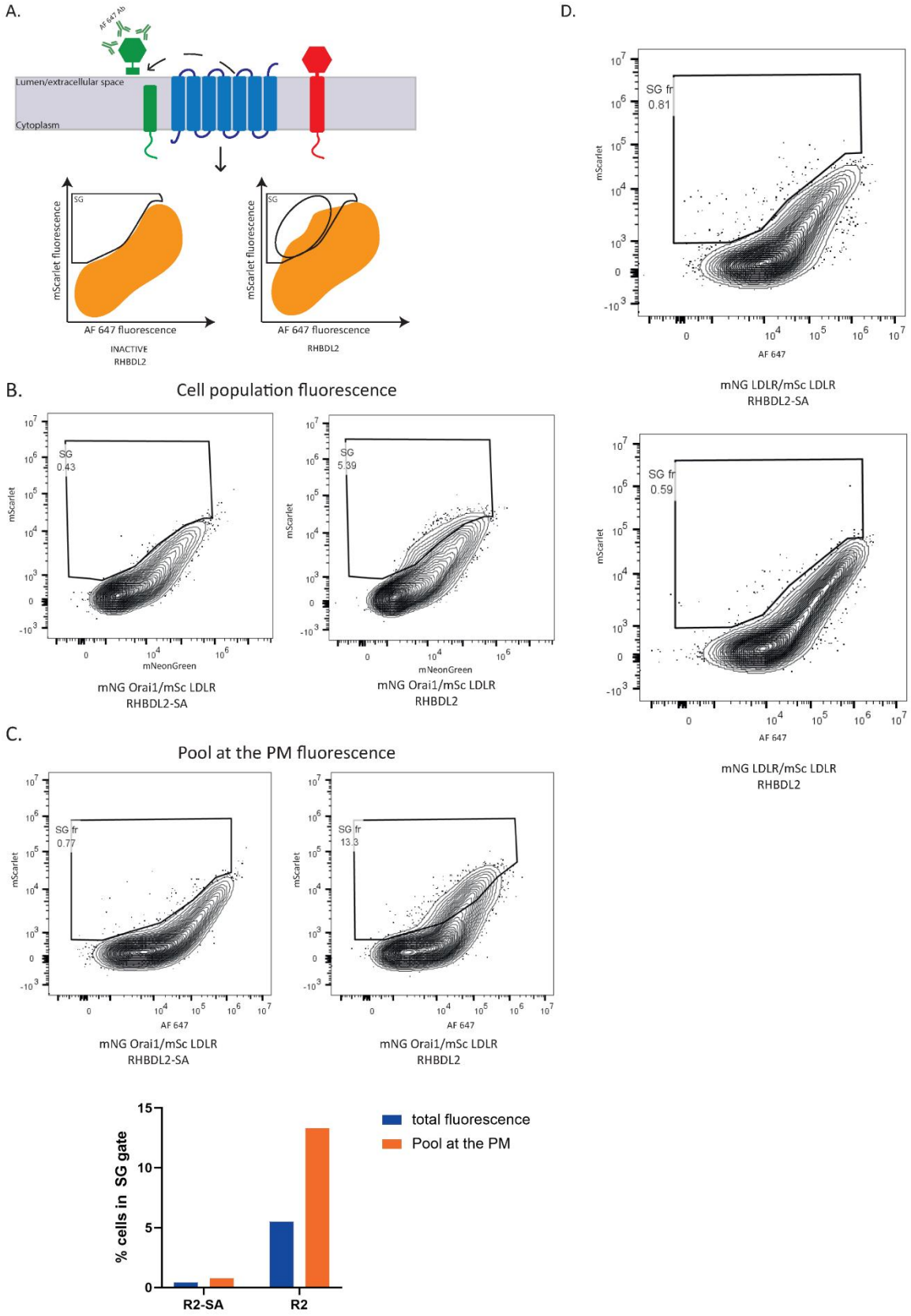


Figure 3.5 Surface labelling allows for the better detection of the cleavage. A) New setup of the screen using surface labelling. The pool of mNG-tagged TMDs is labelled using a fluorescent antibody against the

tag. The reduction in far red/mSc ratio will be the readout of cleavage and will cause the cells to move into the SG. **B)** Assessment of mNG O1/mSc LDLR cleavage based on the decrease of the total mNG fluorescence of the cell population. **C)** The pool of mNG-O1 at the plasma membrane was stained with an antibody against mNG and a with a secondary antibody conjugated with an Alexa-fluor 647. The decrease in fluorescence of the labelled pool upon R2 co-expression was assessed via flow cytometry. **D)** Same staining was used to label pool of mNG LDLR TMDs and showed no reduction of the fluorescence when RHBDL2 was expressed.

Then, I tested the 4 positive control TMDs that did not show any cleavage when looking at the total fluorescence of the cell population. Notably, using surface labelling and focussing on the pool of TMDs present at the PM revealed a significant shift of the cells into the SG for all the TMDs tested (Figure 3.6A-D).

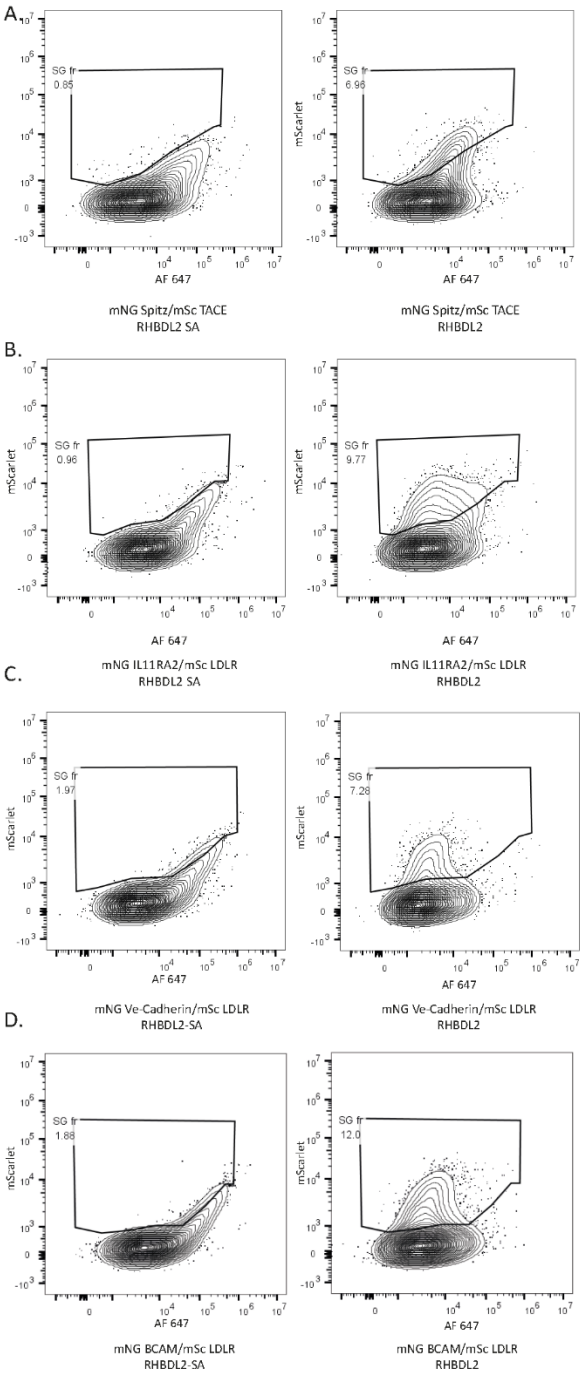


Figure 3.6 Surface labelling of the other positive control TMDs. A-D) Surface labelling of mNG Spitz, IL11RA2, VE-Cadherin, BCAM allows for the cleavage to be detected.

3.3.3 Replacing mNeonGreen tag with FLAG-TwinStrep

Given the positive results obtained with surface labelling, I decided to keep using it for future experiments. For this reason, the presence of mNeonGreen was not essential anymore, so it

was decided to switch the fluorescent tag with a smaller, non-fluorescent FLAG-TwinStrep tag (F-TwS, Figure 3.7A). Using F-TwS instead of of mNG will completely eliminate the fluorescent background within the cell. Moreover, it will make the vector smaller and easier to pack into lentiviral particles. All the control TMDs were sub-cloned to add the F-TwS tag and the cleavage of the new constructs was tested using flow cytometry to make sure that the new tag was not affecting the proteolytic event (Figure 3.7B). As expected, upon co-expression of RHBDL2 there was a significant shift of the cells into SG, the increase in the percentage of cells was not occurring when using the inactive mutant RHBDL2 SA or when using TMD that was not a substrate (F-TwS LDLR/mSc LDLR).

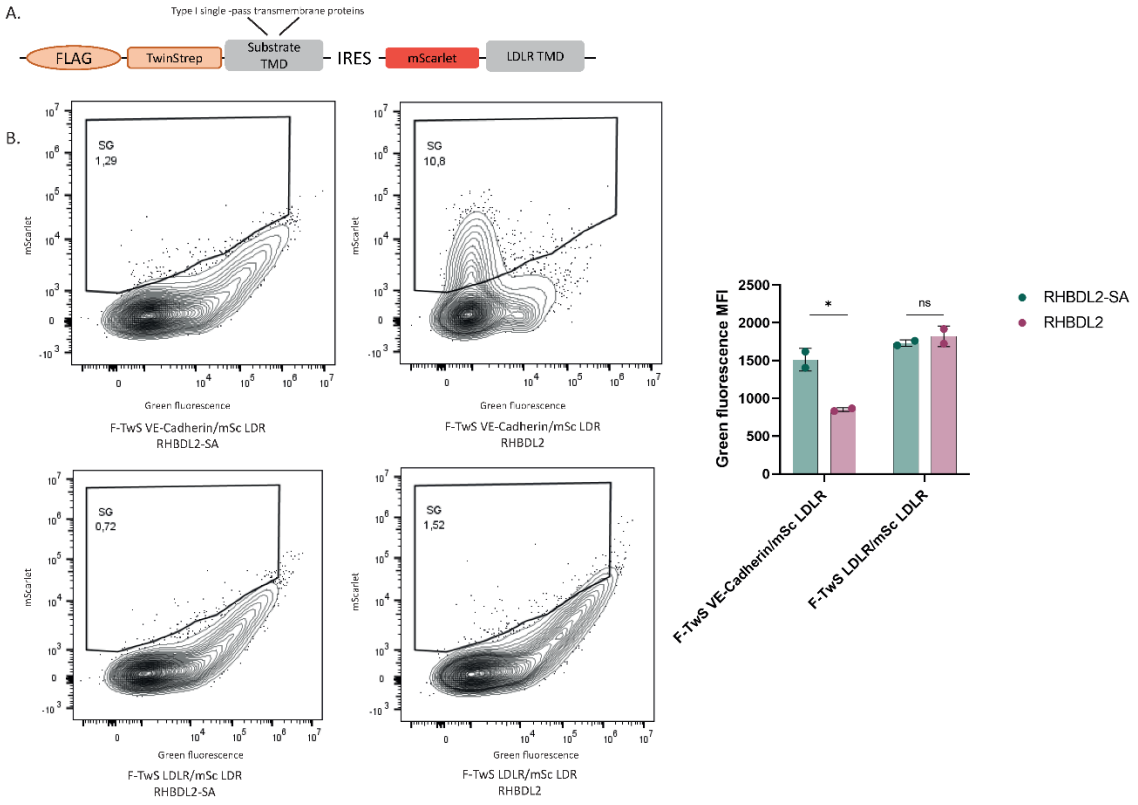


Figure 3.7 Switching mNG for a F-TwS tag. A) Schematic of the new vector where FLAG-TwinStrep tag has replaced mNG. **B)** Surface labelling of F-TwS VE-Cadherin/mSc LDLR or F-TwS LDLR/mSc LDLR using DyLight 488-conjugated antibody against FLAG tag showed significant cleavage of F-TwS VE-Cadherin but not of F-TwS LDLR.

3.4 Generation of stable cell lines expressing the TMDs

3.4.1 Stable cells transduced with bicistronic vector did not express both TMDs

The previous preliminary experiments helped to prove firstly, that the rationale behind the screen worked and a reduction in fluorescence could be used to assess the cleavage of substrates. Secondly, they showed that the new method of screening was specific and able to discriminate between a substrate TMD and a not substrate TMD. Once these essential questions were answered, I set out to fine-tune the details. First of all, the library that was going to be used for the screen was a lentiviral library, so it was important to test if the cleavage was still visible when the TMD was stably integrated into the genome instead of being transiently transfected. For this reason, the construct containing the 'substrate' TMD and mSc LDLR was sub-cloned into a lentiviral backbone. Moreover, in order to increase the efficiency of the screen, it was important to select which cells expressed RHBDL2. For this, I used a fluorescently tagged version of R2 in all subsequent experiments. This would allow for two rounds of selection of the cells during the screen: first only the cells that expressing R2 would be selected, and among this group, the population that showed a decrease of the green fluorescence would be sorted and further analysed. To have a stringent selection of RHBDL2-expressing cells, it was important to choose a fluorescent protein which had an emission spectrum that did not overlap with the green and mSc spectra. To this end, I cloned a tBFP tag at the N-terminus of the human form of RHBDL2 (tBFP RHBDL2). A preliminary experiment showed a good expression of tBFP RHBDL2 (Figure 3.8A), and when co-expressed with the new lentiviral backbone TMDs caused a significant cleavage (Figure 3.8B).

Then, I proceeded to generate cell lines that expressed stably either F-TwS VE-Cadherin or F-TwS ADAM17 as positive and negative TMD control, respectively. The internal negative control was mSc LDLR for both. Surprisingly, when these cell lines were tested using flow cytometry it was evident that each cell was expressing either F-TwS VE-Cadherin, ADAM17 (green+ population, F-TwS ADAM17 plots in Appendix 1) or mSc LDLR (red+ population) (Figure 3.8C). It was not possible

to select a group of cells that had both green and red fluorescence, making drawing the SG and assess the cleavage extremely difficult. Considering that transient expression of the same construct used to generate the cell lines did not cause the same problem (Figure 3.8B), the absence of a double positive population could be explained with a defect in the insertion of the F-TwS TMD-IRES-mSc LDLR gene into the genome. Lentiviral viruses integrate into the genome randomly, so I repeated the transduction process in case the insertion had happened in a heterochromatic domain of the genome that could preclude expression. Unfortunately, two separate transductions did not improve the result and there were two different cell populations that expressed only half of the transduced gene (data not shown). As a last attempt at solving the problem, I sub-cloned the TMDs into a retroviral vector and attempted to make stable cell lines. Unfortunately, this also did not work as there was only expression of mSc LDLR and extremely low expression of F-TwS TMDs (data not shown)

3.4.2 Switching the set up of the screen to use only one TMD

I decided to abandon the bicistronic vector and a try a simpler approach with using only one TMD (Figure 3.8D). This slightly changed the way the screen worked. In fact, the library of TMDs will be tagged with F-TwS tag. The library will be transduced into cells in such a way that each cell would express only one TMD. By surface labelling, the pool of TMDs at the plasma membrane would be labelled using a green-fluorescent antibody. Upon co-expression of tBFP RHBDL2, the cleavage of any substrate TMD would cause the release of the tag to the extracellular media and a decrease of the green fluorescence, while non-substrate TMDs would not be cleaved and consequently their fluorescence would not change. The reduction of green fluorescence would be assessed via flow cytometry, dividing the cell population into several bins according to the values of fluorescence. The cleavage would cause the cell expressing a substrate TMD to move from a bin with a higher fluorescence to one with a lower value. By sorting and sequencing the cells present in each bin in

the presence of RHBDL2 or RHBDL2 SA, we would be able to generate a list of TMDs that have shifted to a lower bin and later validate them to confirm their cleavage.

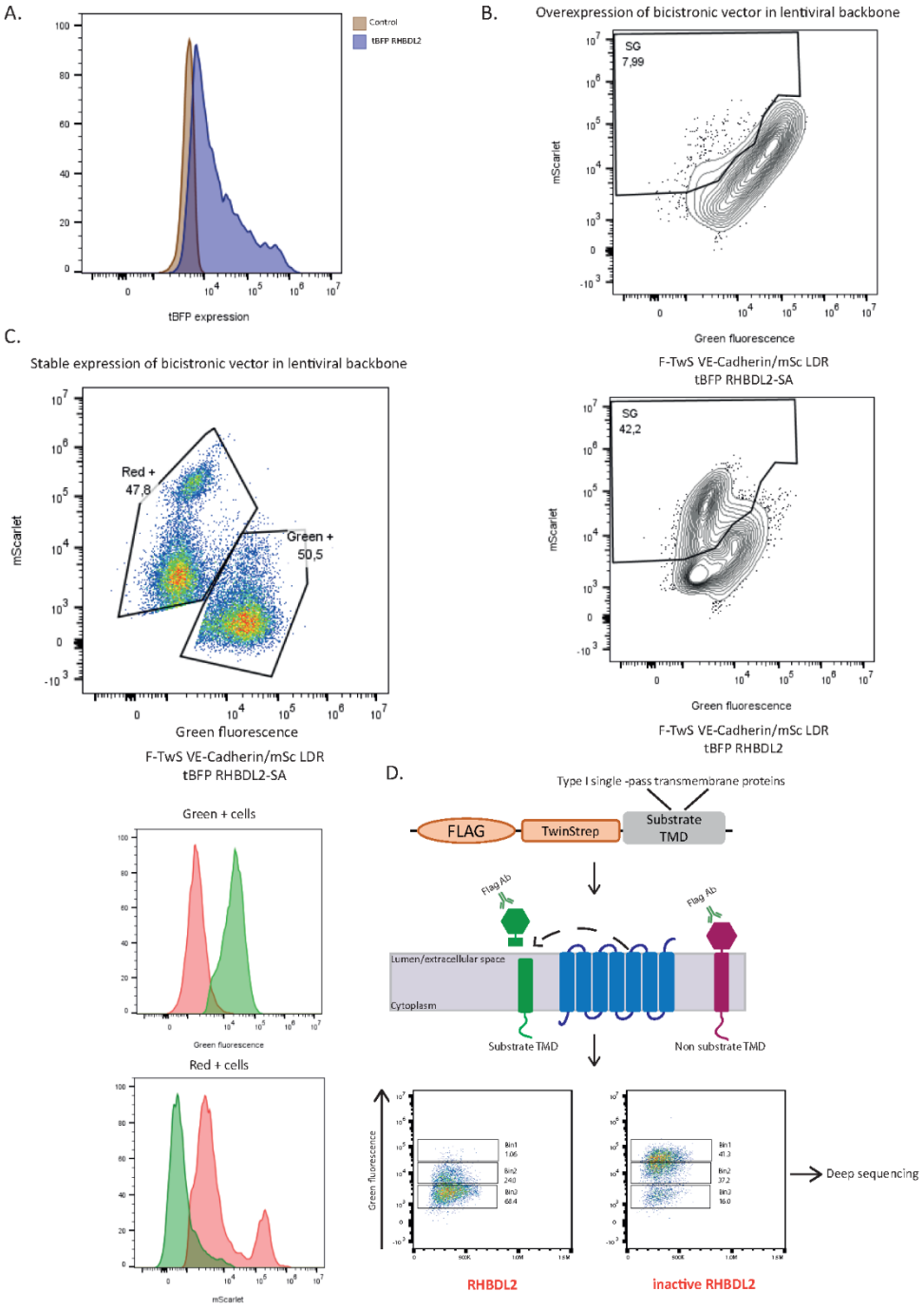


Figure 3.8 Stable expression of the bicistronic vector lead to two distinct populations. A) Histograms of tBFP RHBDL2 expression compared to untrasfected cells (control). **B)** Cells were transfected with the newly generated lentiviral bicistronic vector and the cleavage of F-TwS VE-Cadherin/mSc LDLR was assessed via flow cytometry. **C)** *Top:* Pseudocolour graph showing that in stable cells constitutively expressing F-TwS VE-

Cadherin/mSc LDLR there were two distinct population that were either green+ (expressing F-TwS Ve-Cadherin) or red + (expressing F-TwS LDLR). *Bottom:* comparison of the histograms showed that there was no overlap between green and red+ populations **D)** New setup of the screen using only one TMD. The pool of TMDs at the plasma membrane will be labelled using a fluorescent antibody against FLAG tag. If the TMD is cleaved there will be a decrease of the green fluorescence, cells that show this reduction can be selected by dividing the cell population in several bins according to the intensity of the fluorescence. When the cleavage occurs, the cell will shift from a higher bin to a lower one. By sorting, sequencing and comparing the population of each bin between the two conditions putative new substrates are going to be identified.

3.4.3 Selecting cells express a higher level of RHBDL2 increases detection of the cleavage for both VE-Cadherin and Spitz TMDs

This new strategy was tested using cell lines stably expressing F-TwS VE-Cadherin where either tBFP RHBDL2 or tBFP RHBDL2 SA were transiently transfected (Figure 3.9A). Cells expressing the protease were first selected using the tBFP gate. Afterwards, these cells were divided into 4 bins (Bin1-4) with a decreasing value of fluorescence. Upon co-expression of active R2, there was a significant shift of the cell population into lower bins. This was even more evident when comparing the histogram of the total MFI: when co-expressed with tBFP RHBDL2, a new cell population with weaker green fluorescence was generated. This was a considerable step forward in the screen sensitivity. In fact, it was the first evidence that cleavage of TMDs that are stably integrated in the genome (with a much lower expression compared to being transiently expressed) could be detected.

In a similar manner as it was done with over-expressed TMDs, the same experiment was repeated with stable cells expressing F-TwS Spitz with the same two conditions (co-transfecting tBFP RHBDL2 or tBFP RHBDL2 SA). Surprisingly, after selecting cells using the tBFP gate, there was no evident decrease in the green fluorescence (Figure 3.9B). After further analysis, it became apparent

that the population that showed a reduction and shifted down to lower bins (meaning the cells that were in bin2-3 compared to the population still present in bin1 upon R2 expression), had a higher expression of tBFP RHBDL2 (Figure 3.9C). I decided to restrict the tBFP gate and select only the cells that had a higher expression on tBFP RHBDL2 (high tBFP gate, Figure 3.9D).

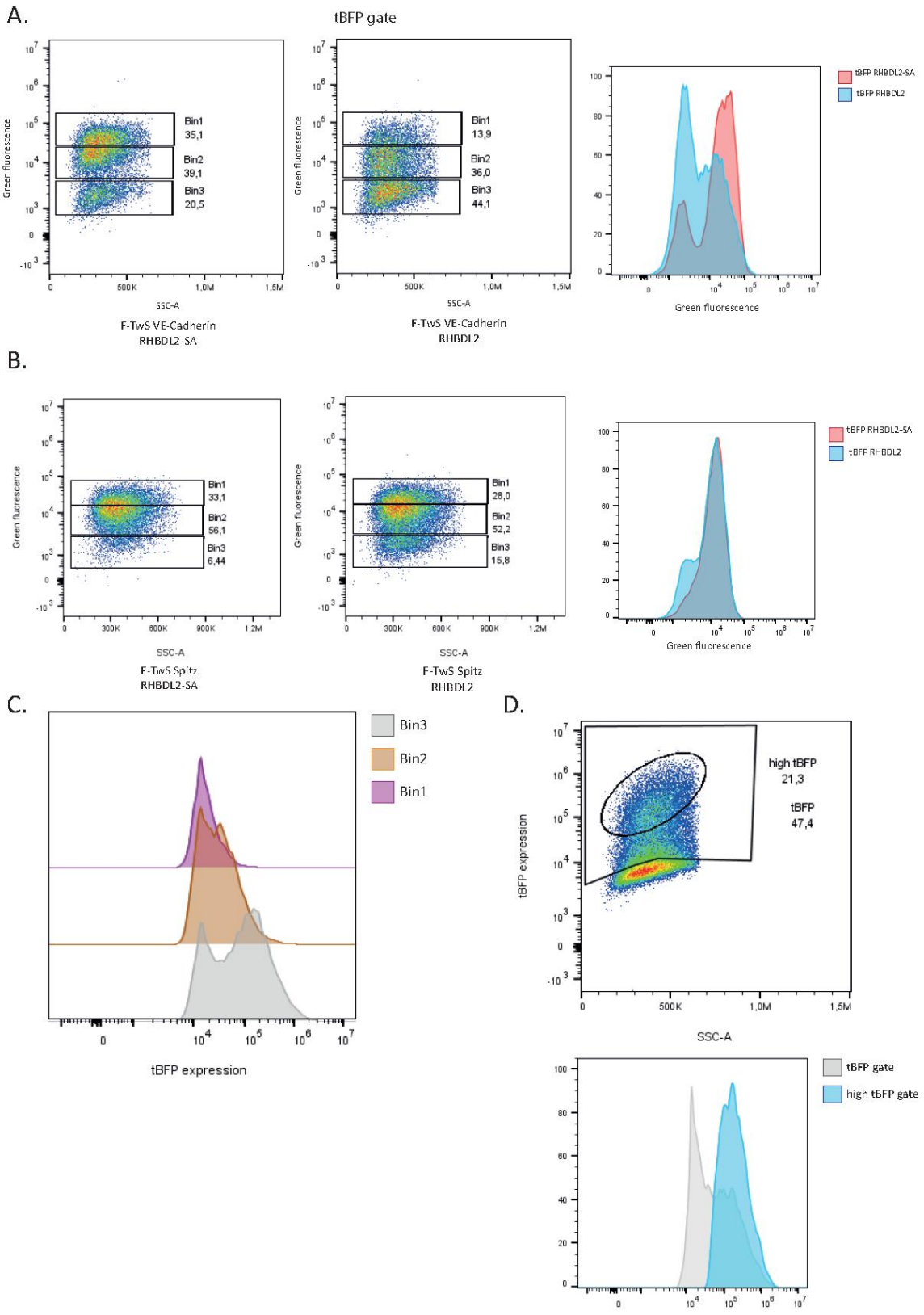


Figure 3.9 Cleavage of stably integrated TMDs was visible for F-TwS VE Cadherin but not F-TwS Spitz A)

Stable cells expressing F-TwS VE-Cadherin TMD were transfected with tBFP RHBDL2 or RHBDL2 SA and pool

of TMDs at the PM was stained using FLAG-DyLight488 antibody. Cells that were tBFP positive were selected (tBFP gate) and within those, the cell population was divided into 3 bins according to the green fluorescence. Upon expression of tBFP RHBDL2 there was a shift of cells to lower bins and a decrease of the green fluorescence. **B)** F-TwS Spitz stable cells were analysed using the same approach, but the cleavage was not detectable. **C)** Histograms showing tBFP RHBDL2 expression in the bins. Bin2-3 show a higher expression compared to bin1. **D)** Selection of cells with a higher expression of tBFP RHBDL2 using a high tBFP gate.

When F-TwS Spitz cells were selected using the high tBFP gate, a new population where the green fluorescence was decreased was generated (Figure 3.10A). Moreover, also for F-TwS VE-Cadherin the reduction in fluorescence was now more striking, confirming that having a higher expression of the protease improved the cleavage (Figure 3.10B). A potential explanation as to why it was necessary to have a higher expression on RHBDL2 in order to detect the cleavage, would be that tBFP-tagged version of the protease is expressed at a low level to begin with. To test this, I compared the expression of tBFP RHBDL2 with HA RHBDL2 via Western Blot. Notably, HA RHBDL2 was significantly more expressed than tBFP RHBDL2 (Figure 3.10C). Given that the only difference between the two version of RHBDL2 was the tag, this difference in expression could be due to the size of the tBFP tag that could reduce the stability of the protein. Fortunately, the reduced level of expression did not affect the proteolytic activity as shown by the cleavage of both VE-Cadherin and Spitz TMD. Moreover, the ability to discriminate between substrate and non-substrate TMD was also not affected. In fact, cells stably expressing F-TwS LDLR did not show any cleavage neither in the tBFP gate nor in the high tBFP one (Figure 3.10D).

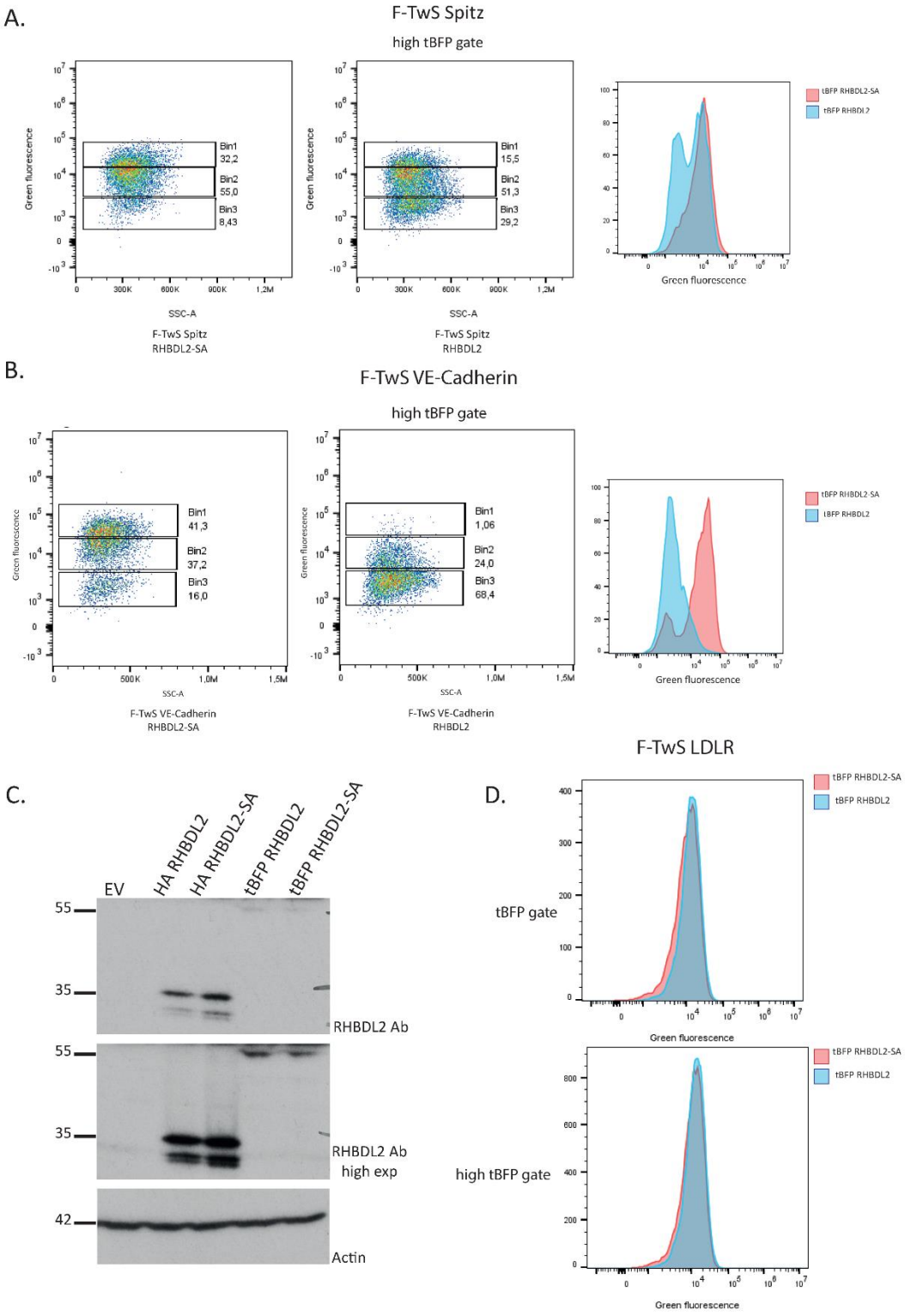


Figure 3.10 Selecting cell with a high expression of tBFP RHBDL2 improved cleavage detection A) In F-TwS Spitz stable cells the cleavage was detectable when analysing cells in the high tBFP gate. **B)** Selecting tBFP RHBDL2 high expressing cells improved the detection of the cleavage also for stable cells expressing F-TwS VE-Cadherin. **C)** Western blot confirming that tBFP RHBDL2 expression (~55kDa) was lower than

another tagged construct of RHBDL2, i.e. HA RHBDL2 (~35kDa) **D)** Flow cytometry analysis of F-TwS LDLR transfected with tBFP RHBDL2 showing that the ability to discriminate between substrate and non-substrate was still maintained.

3.5 Generation of monoclonal cell lines for tBFP RHBDL2 and RHDBL2 SA

3.5.1 Characterisation of clonal cell lines

A way of overcoming the need to select for cells expressing high levels of tBFP RHBDL2 would be to stably express it. To this end, I generated monoclonal cell lines using Flp-In HEK293T cells that expressed either tBFP RHBDL2 or RHBDL2 SA after Doxycycline treatment (Dox, 100ng/ul concentration). Six clones per cell line were selected. Expression and localisation were tested via Western blot, flow cytometry and immunofluorescence (Figure 3.11A-C). Clone #1 for tBFP RHBDL2 and #8 for tBFP RHBDL2 SA were chosen and used in further experiments.

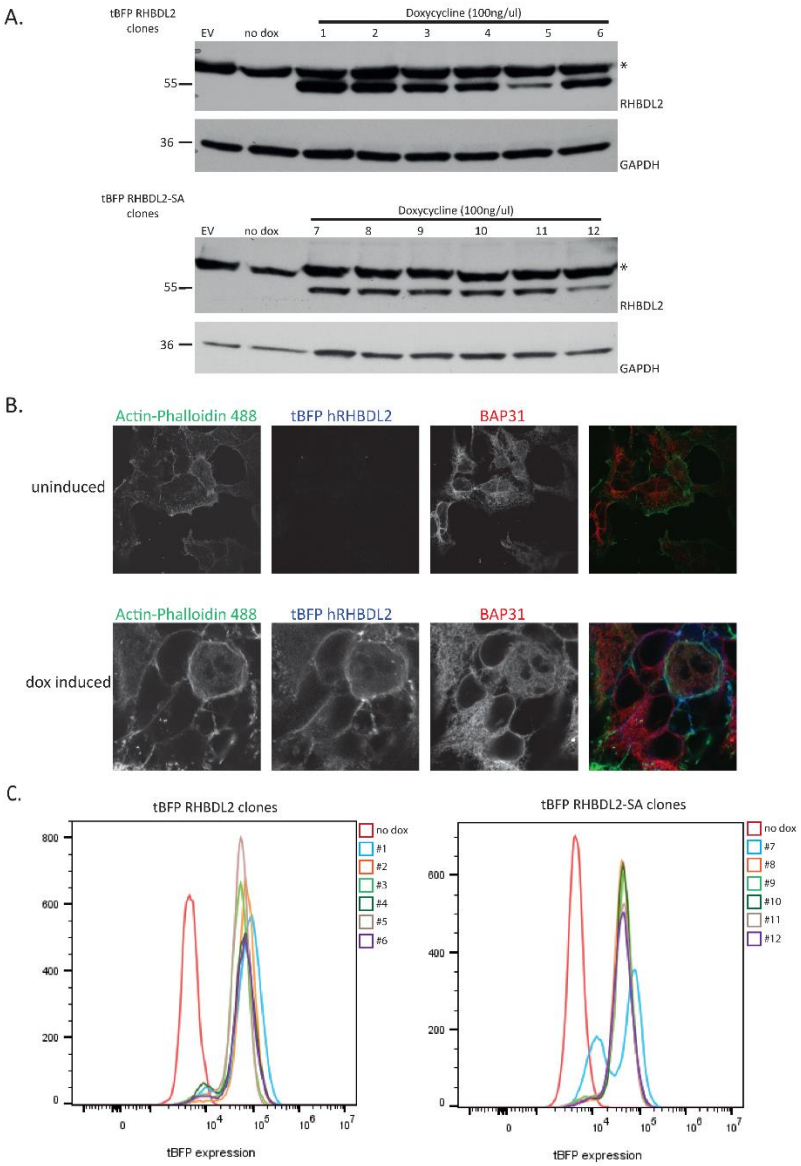


Figure 3.11 Characterisation of tBFP RHBDL2 and RHBDL2 SA monoclonal cell lines A) Six clones were amplified and tBFP RHBDL2 and RHBDL2 SA were induced with 100ng/ul Doxycycline for 16hs. Expression was tested using Western blot. **B)** Immunofluorescence of stable cells induced with 100ug/ul of Doxycycline. Cells were stained for Actin-phalloidin to visualise plasma membrane (green), BAP31 (red) and tBFP RHBDL2 (blue). **C)** Cells induced with Doxycycline were analysed via flow cytometry assessing levels of blue fluorescence.

3.5.2 Stable expression of both TMDs and RHBDL2 lead to a less striking cleavage

Once the monoclonal cell lines were obtained, I proceeded to transduce them so that they expressed stably either F-TwS VE-Cadherin, Spitz or ADAM17. Upon doxycycline treatment to induce RHBDL2 expression (Figure 3.12A), the cleavage of the TMDs was assessed via flow cytometry. Cells expressing F-TwS VE-Cadherin showed a significant decrease of their green fluorescence, compared to cells expressing F-TwS ADAM17 (Figure 3.12B-D). However, the reduction was less striking than when using overexpression of the protease. Interestingly, there was a small population that did not show any green fluorescence (lower red peak in tBFP RHBDL2 SA, Figure 3.12B) implying that the TMD was not present at the plasma membrane. The presence of this green negative population could explain why the extent of the cleavage is less significant than previous experiments where RHBDL2 was transiently expressed. The presence of this population was puzzling, as it did not occur in previous stable cell lines.

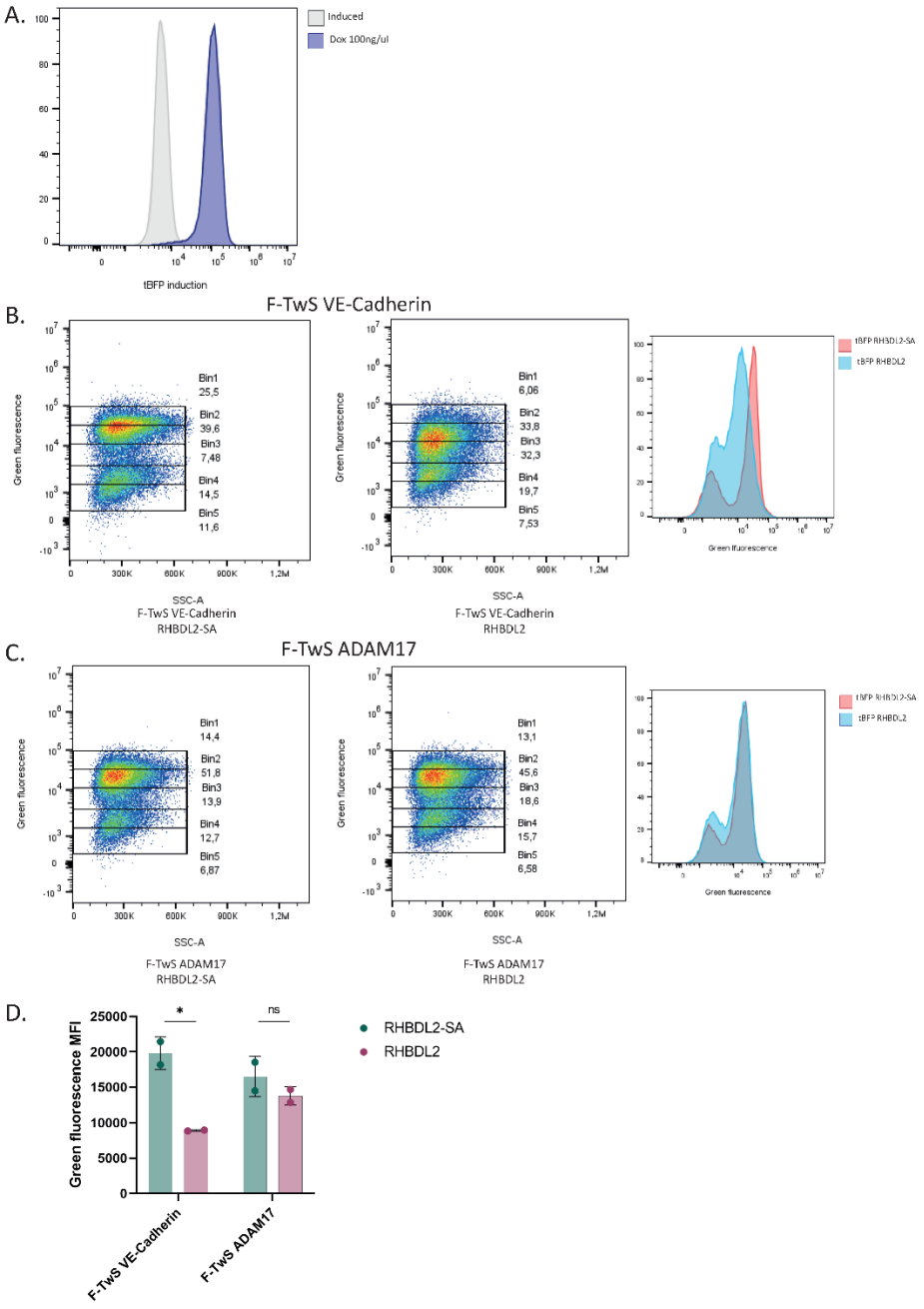


Figure 3.12 Stable expression of both RHBDL2 and TMDs led to less striking cleavage A) Histograms showing induction of tBFP RHBDL2 after treatment with 100ug/ul of Doxycycline for 16hs. **B-C)** Cells expressing both tBFP RHBDL2 or RHBDL2 SA and F-TwS VE-Cadherin were induced for 16hs with Doxycycline and stained with FLAG DyLight488 to label TMDs at the cell surface. Afterwards they were analysed via flow cytometry to assess the cleavage. **D)** Median fluorescence intensity of B and C showing a significant decrease of the green fluorescence for F-TwS VE-Cadherin but not F-TwS ADAM17. *= p value <0.01.

Importantly, when the cleavage was assessed in cells expressing F-TwS Spitz, there was no visible shift of the cell population (Figure 3.13A). This occurred even if all the cells seemed to express the TMD at the plasma membrane. It was speculated that the protocol used to transduce the cell could have overloaded the system, so that the level of substrate was massively exceeding the level of the protease. If this was the case, the cleavage could be occurring but because of the excess of substrate it was not detectable. To test this hypothesis, I repeated the transduction process but using either 5, 10 or 20% of the viral supernatant for both F-TwS Spitz and F-TwS LDLR as negative control (Figure 3.13B). Notably, there was a slight shift at lower doses of virus, but it was still not as significant as previous experiments using transiently expressed RHBDL2 (Figure 3.10A). Interestingly, increasing the dose of viral supernatant did indeed generate a population of cell that did not express the TMD at the plasma membrane (as seen for cells expressing F-TwS VE-Cadherin), suggesting that the amount of virus used to transduce cells is critical to correctly localise the TMDs to the cell surface.

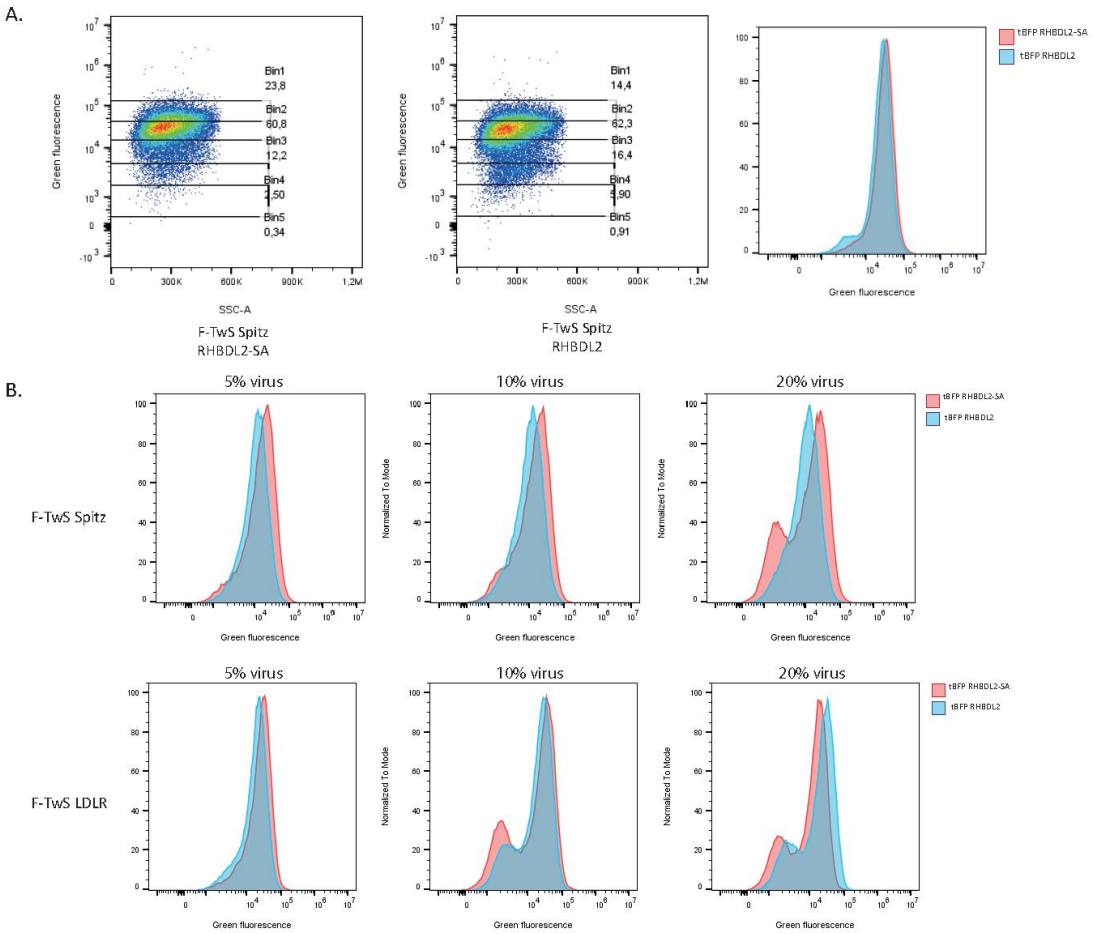


Figure 3.13 F-TwS Spitz cleavage was not detectable using stable cells A) Cells expressing F-TwS Spitz were treated with 100ng/µl of Doxycyline to induce tBFP RHBDL2 or SA and cleavage was assessed via flow cytometry after labelling the pool of Spitz TMD at the cell surface. **B)** Clone #1 and #8 were transduced with 5,10,20% of viral supernatant containing either F-TwS Spitz (*upper part*) or F-TwS LDLR *(lower part)* sequence. After expansion, these cells were induced with Doxycycline and the cleavage was assessed via flow cytometry.

3.5.3 Increasing expressed tBFP RHBDL2 improved cleavage while retaining specificity

In the end, it was decided that inducible expression of RHDBL2 was not efficient enough to screen for substrates, so I tested whether increasing the quantity of transfected protease could overcome the need to select high expressing cells. I transfected either 2, 4 ,8, 16ug of tBFP RHBDL2 and tested the cleavage of Spitz or ADAM17 TMDs either in the tBFP gate or high tBFP gate (Figure 3.14A-B). Flow cytometry analysis showed that when 4ug of RHBDL2 were transfected, there was

almost no difference in the cleavage of Spitz between the tBFP gate and high tBFP gate (Figure 3.14A). Most importantly, increasing the amount of R2 did not impair the selectivity of the protease as ADAM17 TMD was not cleaved (Figure 3.14B, pseudocolour graphs in Appendix 2).

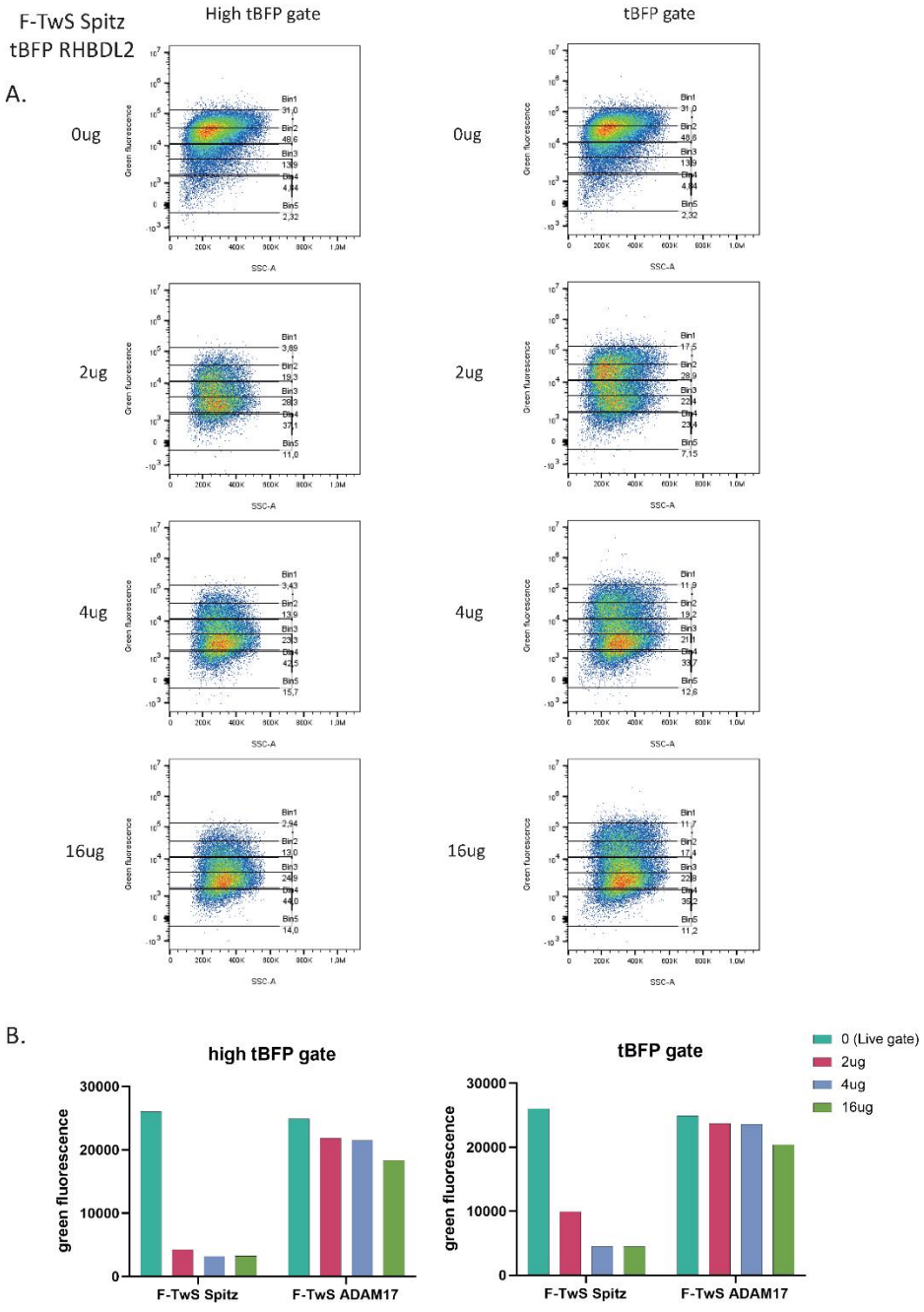


Figure 3.14 Transfecting increasing amount of tBFP RHBDL2 allows for the cleavage to be visible in the tBFP gate A) F-TwS Spitz cells were transfected with 0,2,4,16ug of tBFP RHBDL2 and cleavage was assessed in either high tBFP gate or tBFP gate. **B)** Bar chart from A and from same experiment done with F-TwS

ADAM17 cells (pseudocolour graphs in Appendix 2). MFI decreased in a similar manner in both the high tBFP and tBFP gate when transfecting 4ug of tBFP RHBDL2. Importantly F-TwS ADAM17 cells did not show any significant decrease in fluorescence.

3.6 Discussion

In this chapter, I have described the experiments and the essential troubleshooting that was necessary to set up a new genetic screen to uncover new substrates for rhomboid proteases. The screen is based on the use of flow cytometry to assess a specific decrease of the fluorescence of a substrate TMD caused by the rhomboid cleavage. In previous studies, western blotting and mass spectrometry were the main techniques used to screen for potential substrates in a more systematic way (Grieve et al., 2021a; Johnson et al., 2017). For this reason, it was essential to thoroughly test if flow cytometry was a suitable technique to use for this purpose.

Initially, I started testing an approach that used a bicistronic vector coding for two fluorescently-tagged TMDs: a 'green' TMD was the potential substrate, while mSc LDLR represented the internal control whose fluorescence would not change upon RHBDL2 expression. The use of bicistronic vectors is a common and elegant way to have a tighter control over the relative expression of two proteins. Since both TMDs were expressed from the same promoter, the green/red ratio could be used more consistently to evaluate RHBDL2 cleavage. Preliminary experiments confirmed that it was possible to use a decrease in the fluorescence ratio as a readout for cleavage, as the reduction was specific to the presence of an active rhomboid and of a substrate TMD. However, over time using the bicistronic vector became more of a problem than a brilliant solution. In fact, whenever the TMDs were stably integrated into the genome, there was a defect in the expression of the two genes so that only one of the TMDs was translated. This resulted in two distinct populations that expressed either the green TMD or the mSc one. The reason behind this defect in expression was never fully understood, but it was speculated that it could be due to the presence of two IRES close to each other. In fact, other than the one present between the two TMD

genes, the lentiviral plasmid backbone contained an additional IRES needed for the expression of the antibiotic resistance cassette. Since the three genes (the two TMDs and the antibiotic resistance gene) were in frame with each other, this would lead to the formation of a transcript that contained two IRES. Given the secondary structure of these RNA sequences, the presence of two of them in less than 1kB could have impaired the correct translation of the mRNA. However, it is important to highlight that this hypothesis has never been tested and remains a speculation.

Since the focus of my project was to develop an efficient screen, it was decided to abandon the bicistronic vector and to remodel the set-up of the screen to use only one TMD. The move to a simpler system was pivotal in increasing the method's sensitivity. In fact, by using cells expressing F-TwS VE-Cadherin, the cleavage was now clearly detectable. Moreover, the ability to discriminate between substrate and non-substrate TMD was not lost, proved by the lack of decrease in fluorescence when F-TwS LDLR stable cells were used. Another change that was implemented to increase the screen efficiency was to introduce tBFP RHDBL2. Having a fluorescent tag in R2 allowed for the selection of cells that definitely expressed the protease and discarding the cells that did not express it. In this way, it was possible to decrease the background noise generated by cells whose fluorescence would not change, not because of lack of cleavage but because of the absence of the protease.

However, it became apparent that tBFP RHBDL2 was expressed at a much lower level than HA RHBDL2. This was probably due to the size of tBFP and its effect on the stability of the protein. In an attempt to overcome the problem, I made monoclonal stable cells expressing tBFP RHBDL2 under the control of a doxycycline inducible expression system. Afterwards, these cell lines were transduced to constitutively express the TMDs. Having a stable expression of both RHBDL2 and TMDs was the most beneficial set-up as it would have allowed a better control of the expression of both substrate and protease. Nonetheless, the cleavage was less significant compared to transiently transfect RHBDL2. This could be due to two different explanations: firstly, the expression level of

tBFP RHBDL2 would be lower when stably expressed compared to being transiently co-transfected. This would cause a lower cleavage. Secondly, a few of the cell lines presented a population of cells that did not show the TMDs at the cell surface. This only happened when transducing Flp-In monoclonal cells and not when using Hek293T. But having a pool of TMDs that does not reach the plasma membrane could lower the detectable cleavage. In the end, it was decided to move forward with a simpler approach and to use overexpressed tBFP RHBDL2. This allowed to detect the cleavage of all the control TMDs and to maintain the specificity as non-substrate TMDs were not cleaved.

Overall, in this chapter I described the experiments necessary to set up a new genetic screen to uncover rhomboid substrates. The essential troubleshooting to make the screen as effective as possible was a lot more extensive than planned, as each of the steps had to tested and reviewed. Ultimately, the overarching theme that comes across via these experiments is that most of the time, a simpler way is the best approach.

Chapter 4

Identification and validation of novel substrates for RHBDL2

4. Identification and validation of novel substrates for RHBDL2

4.1 Introduction

After the long troubleshooting period to identify and solve all the potential issues that could arise in the screen, I set out to test the method using a library containing the TMDs of all single-pass transmembrane proteins. In this chapter, I will describe the different steps of the screen: from the generation to the library, the virus production and lentiviral infection, to the sorting of the TMD pool and the analysis of the sequencing data. A schematic of the several steps of the screen is described in Figure 4.1A-F. Given that the screening of this library had not been performed before, there was no clear protocol regarding how to perform the experiment or how to analyse the data that were going to be produced. This is in stark contrast to what occurs when commercial libraries are used. This meant that the protocol was being written as the screen was being done. Because of this, it is likely that there are some improvements that could be implemented if the screen gets repeated that would help to give a better output. Nonetheless, the screen has provided a list of putative new substrates for RHBDL2, all of which were not identified in previous studies.

Due to the time constraints of the DPhil, the validation of these hits is not complete, but I will present preliminary experiments to confirm the results of the screen.

4.2 Generation of the TMD library

Since there was not an available library containing the sequences that were needed for the screen, I generated a new one containing the TMDs of all human type I single pass transmembrane proteins. To do this, I selected the reviewed type I single pass transmembrane proteins on the online database Uniprot and downloaded only the transmembrane region. This provided a list of 1146 proteins, according to the number of reviewed proteins in 2020. To transform the protein sequences to DNA strings, I established a collaboration with Dr. Samuel Dean (University of Warwick). Dr Dean

used a script that added a luminal and cytosolic flanking region to the TMD sequence. As mentioned in the previous chapter, the 7aa luminal stretch compensated for potential errors in setting the N-terminal border of the TMD and for the possibility of the rhomboid cleavage occurring in the juxtamembrane region. While the shorter 3aa cytosolic region allowed for the insertion of the TMD in the membrane, without being long enough to cause cytoplasmic regulation. The full TMD sequence was then transformed into DNA strings (Figure 4.1A). The conversion was done keeping into consideration the human codon usage and the GC content of the DNA sequences, generating sequences that had a GC content below 50%. This would be helpful in facilitating the cloning of the sequences into the lentiviral backbone. Lastly, in converting the protein sequences into DNA strings, the sequences forming two restriction enzymes sites were avoided as the enzymes were going to be used to insert the TMDs into the backbone.

Then, the TMD sequences were cloned into a pLEX backbone, a non-inducible lentiviral vector based on HIV-1 (Kappes & Wu, 2001). The cloning was done by TwistBioscience and the DNA strings were inserted in the vector using SpeI and AgeI restriction sites (Figure 4.1B). The rationale of having this restriction sites after the F-TwS tag was to avoid unnecessary cloning of large sequences. In fact, by flanking the TMD region with two restriction sites it was possible to cut out only the TMD and to replace it with the other candidates of the library. The newly-generated library was tested by TwistBioscience via next-generation sequencing to check for potential PCR bias that could have enriched for one specific sequence. The sequencing confirmed that the distribution of the different TMDs within the pool was equal.

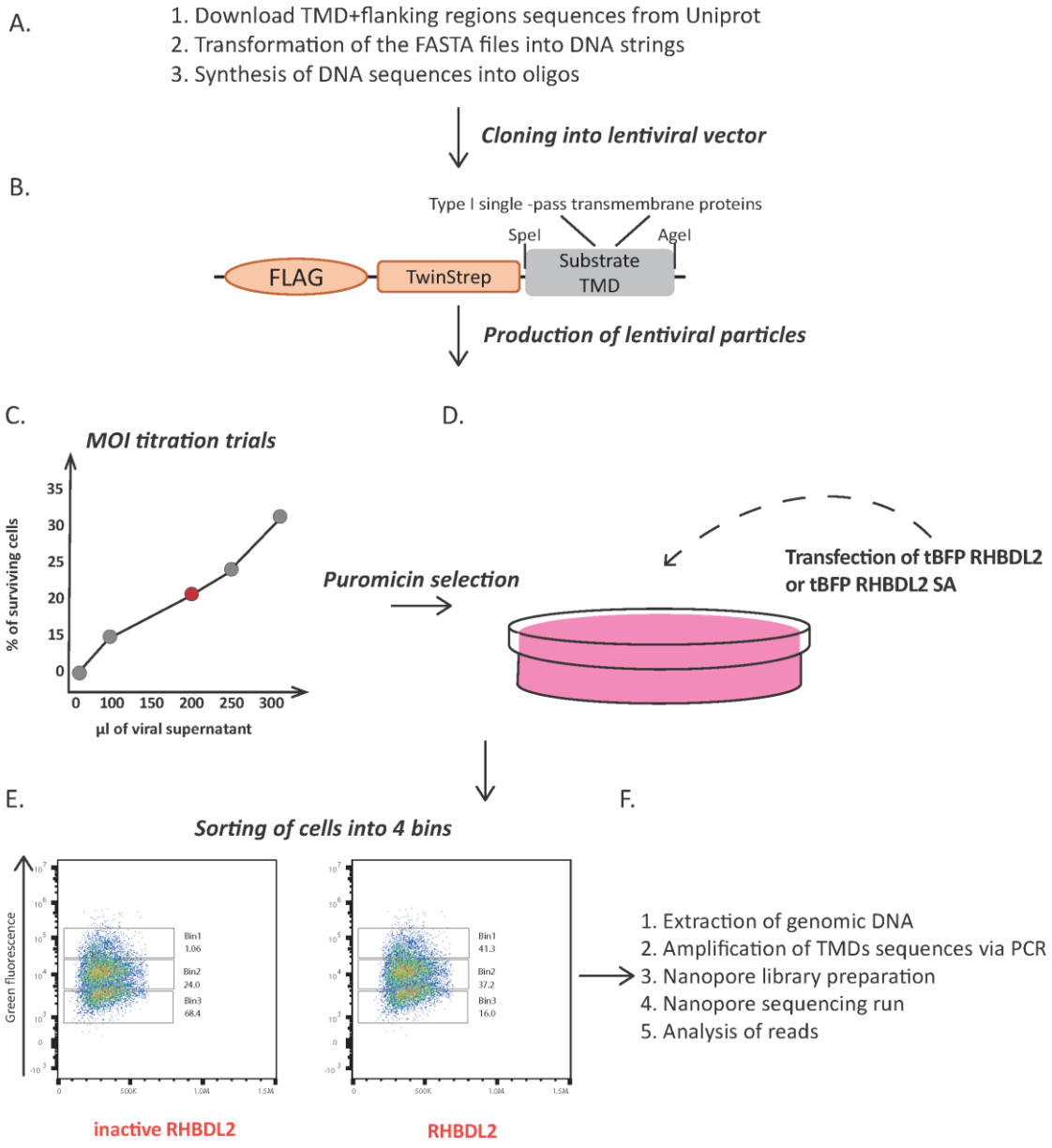


Figure 4.1 Workflow of the screen

4.3 MOI titration and testing of the sorting strategy

4.3.1 Production of the viral particles

Once the lentiviral library passed the quality control tests run by TwistBioscience, I proceeded to produce the virus that would be used to infect the target cells. The viral particles were

generated using 4µg of the TMD library, co-transfected with 1.4µg of psPAX2 and 600ng of pMD2.G as described in the Materials and methods (2.4.4). The packaging vectors coded for the group-antigens (gag), the reverse polymerase and the envelope proteins respectively. These viral proteins are necessary for the correct packaging of the TMDs into the viral particles.

The TMD library was a pooled one, meaning that the target cells were infected with the whole complexity of the library at the same time. To ensure an effective screen using such a library, there are two aspects that were important to consider. Firstly, it was essential to ensure that each cell only got infected with one viral particle, therefore expressing only one TMD. To make sure of this, cells needed to be infected at a low multiplicity of infection (MOI) that is usually between 0.2-0.3 or lower. The MOI is defined as the ratio of viral particles to the number of infected cells (Sinha et al., 2018), so infecting with an MOI of 0.2 means that only 20% of the cells would be successfully infected while the rest of the cell population would die during the selection process. The second important aspect was that each plasmid of the library needed to be represented in the cell population. The representation of each plasmid is called coverage, and it is quite common for pooled libraries (such as CRISPR libraries) to use a 200x-250x coverage, meaning that 200-250 cells express one specific item of the library. Given the small size of the TMD library, it was possible to keep a 500x coverage, so that each TMD would be expressed on average by 500 cells. In order to consider both the MOI and the coverage, the number of cells that needed to be infected was calculated as #TMDs in the library (1146)/MOI (0,2) *coverage (500x). According to this calculation a minimum of 2.9E6 needed to be infected to ensure only one integration per cell and that each of the TMDs was represented at least 500x.

4.3.2 Calculating the viral supernatant necessary for an MOI of 0.2

To calculate the correct volume of viral supernatant that was needed for achieving an MOI of 0,2 I performed a virus titration by infecting HEK293T. To simulate what was going to occur during

the screen, I used the same number of cells and plate format by infecting 3.5E6 cells (a higher number to what was needed to compensate for potential growth fluctuations) of HEK293T in a 15cm plate (Figure 4.2A). Two sets of plates were infected with increasing volume of viral supernatant between 0-250µl. One of the sets was selected with puromycin for 96hs, while the other set acted as a negative control and was not treated with the antibiotic. After selection, the cell number of each plate in the selection set was counted, and the percentage of living cells was calculated by comparing it to the cell number of the corresponding plate in the control set. By doing this, 200µl was identified as the correct volume that infected only 15% of the cells. This is lower than the usual 0,2 that is used in other common pooled libraries, but given the size of the TMD library, it could be accommodated by infecting a higher number of cells while keeping the screen manageable.

4.3.3 Testing the sorting strategy confirmed bins specificity

Before proceeding with the screen using the TMD pool-expressing cells, it was important to test the specificity of the sorting strategy that was going to be employed. In fact, prior to sorting, the cell population is divided into 4 bins according to decreasing levels of green fluorescence and the cells in each bin get sorted. The 4 bins were close to each other, so it was critical to assess the specificity of the sorting for each bin and to ensure that there was no 'cross-contamination' between them. To test this, cells expressing F-TwS ADAM17 TMD were labelled with a FLAG-fluorescent antibody and the green positive cells were divided into 4 bins (Figure 4.2B). Each population of the bins got sorted and then re-run through the instrument to assess whether it was overlapping with the bin it originally came from, i.e., if the population sorted from Bin4 was actually contained within Bin4 (Figure 4.2C). Interestingly, the pseudocolour graphs showed a general decrease of the fluorescence when the samples got re-run through the sorter. This was due to a technical issue more than an unspecificity of the sorting strategy. In fact, sorting the samples a

second time caused the bleaching of the antibody leading to a general decrease in the signal. This was exemplified by the bigger decrease for Bin1 and Bin2, which were originally the bins with the higher fluorescence. Considering this, the overall result of the test is that the sorting strategy is able to sort cells from one specific bin with only a minority of cells mis-sorted to the adjacent bin.

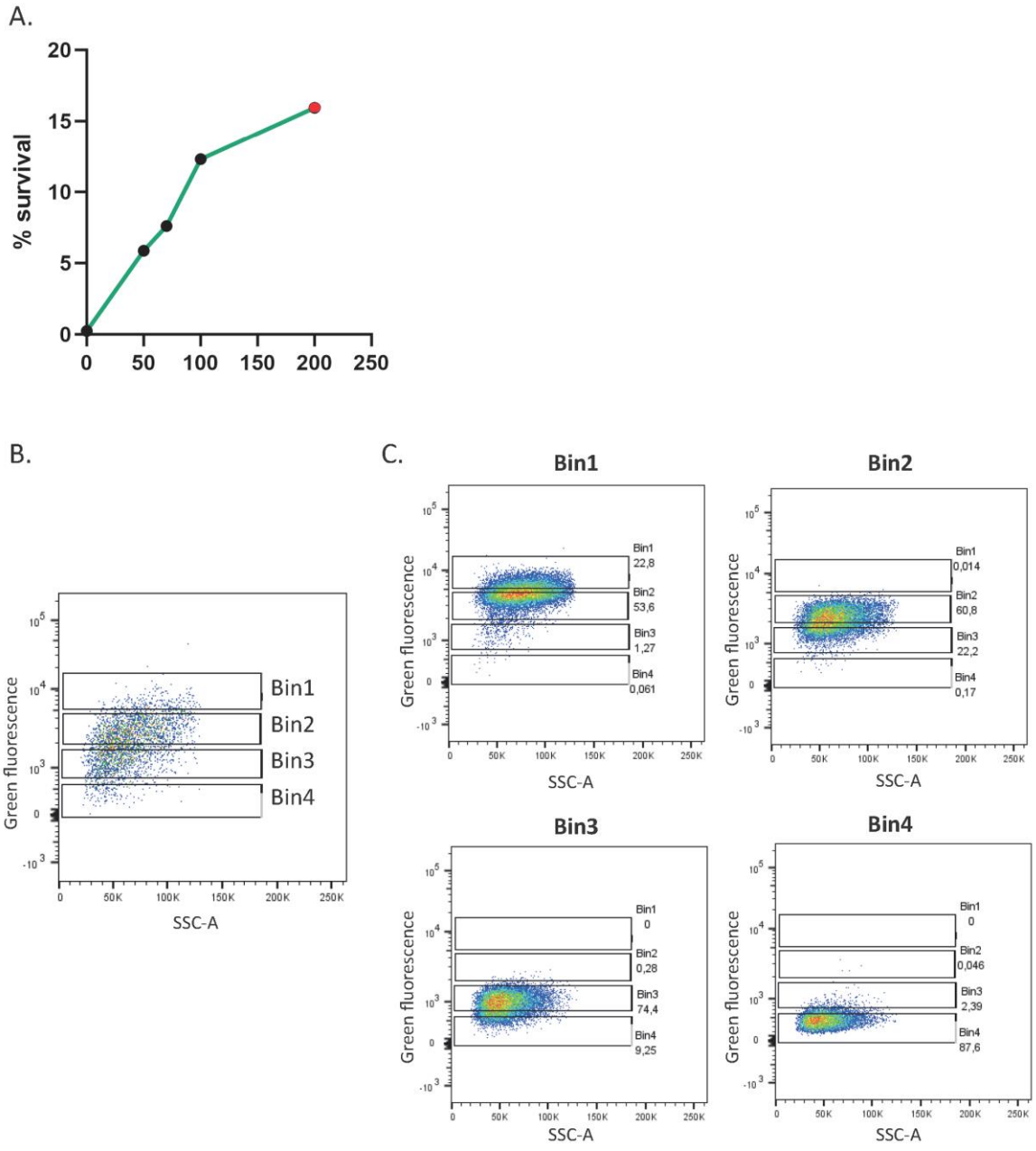


Figure 4.2 MOI titration and purity test. A) MOI was calculated by infecting two sets of 15cm plates with the same volume of viral supernatant. One set was selected with 2µg/ml of puromycin for 96hs before

counting the number of live cells. The second plate was not selected and acted as the negative control. The percentage of live cells was calculated as the number of cells in the selected set compared to the number of cells in the unselected set. **B)** Cells expressing F-TwS ADAM17 TMD were labelled with FLAG-DyLight488 antibody and analysed through FACSBDAria II. Only 1E5 cells were recorded and used to set the 4 bins according to the green fluorescence. **C)** The sorted population of Bin1-4 was sorted and re-run through the sorter to assess if the sorting was able to separate cells from adjacent bins.

4.4 Screen of TMD pool-expressing cells to identify novel substrates

4.4.1 Infection and sorting of TMDs-expressing cells

Once the virus that was necessary to obtain an MOI <,2 had been determined and the efficiency of the sorting strategy had been tested, it was finally time to proceed by using the TMD pool virus and to sort the infected cells. The target HEK293T cells were infected using the protocol described in Material and Method 2.4.4. A total of 5E6 cells were infected in triplicates to keep the coverage of the library 500x throughout the experiment and to have the three biological replicates necessary for statistical analysis. Moreover, multiple controls were used to ensure that the system worked as it should have (Figure 4.3A). First of all, a transduction control was generated by infecting the same number of cells with viral particles that produced mCherry. This viral supernatant was made using the same transfection mix used to produce the virus for the TMD pool, this control would ensure that the infection worked, and that the viral particles were viable. A second plate acted as the negative control for the selection process. These cells were not infected with the virus, but they were selected for 96hs with 2µg/ml of puromycin. Only when all the cells in this control were dead the selection process was stopped. Lastly, the third control was a set of cells that got infected with the TMD pool virus but that was not selected (positive control). The reason for this control was to make sure that the death of the cells was only caused by the selection process and not by toxicity of the infection.

After the selection process, the TMD-expressing cells were expanded and 5E6 were transfected with either tBFP RHBDL2 or the inactive tBFP RHBDL2-SA. After labelling the cells with a fluorescent antibody against the FLAG tag, a minimum of 7E6 cells were analysed through the sorter for the three replicates (one of the replicates is shown in Figure 4.3B, the other two can be found in Appendix 3). To maintain the coverage of 500x after selection, only 6E5 cells were required. The excess of cells that were initially analysed accounted for the selection that was going to be applied by the gating of the cells. The sorting bins were drawn as close as possible but not overlapping and named P4-7, where P4 was corresponding to the highest value of fluorescence and P7 to the lowest. The three replicates showed similar distribution of the cell population, however Replicate 2 showed a general decrease of the fluorescence and for this reason the bins had to be slightly moved down to include the whole cell population. However, the dimensions and respective position of the bins were not changed. Moreover, there was no evident shift in the total fluorescence upon RHBDL2 co-expression. This was different to what was seen in previous preliminary experiments using control TMDs. This observation was not surprising, as the library contains over a thousand TMDs, most of which it was hypothesized were not going to be RHBDL2 substrates. The shift in fluorescence caused by the cleavage of the minority of TMDs that are putative substrates most likely was too small to cause an evident decrease in fluorescence. Overall, an average of 1.5E6 cells (almost three times the minimum number of cells to keep the coverage) were sorted collectively among the 4 bins, for both tBFP RHBDL2 and tBFP RHBDL2 SA samples in all the replicates. This was enough to compensate for potential loss of cells in the subsequent steps of the DNA extraction and library preparation.

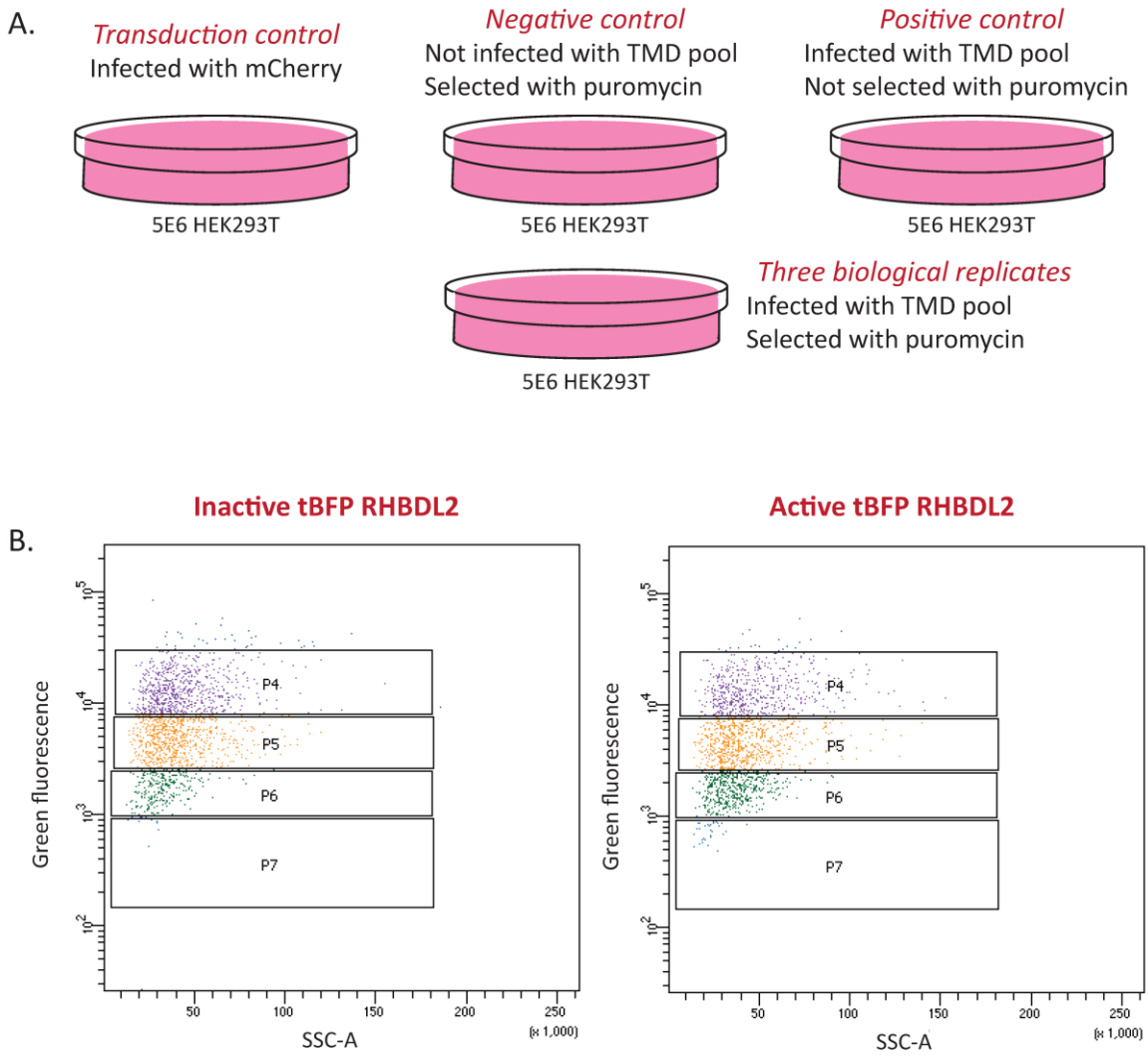


Figure 4.3 Sorting of TMD-infected cells. A) Schematic of the controls used in the screen. 5E6 HEK293T were infected with 200µl of viral supernatant producing either TMDs pool or mCherry. Cells were infected for 48hs and then selected with 2µg/ml of puromycin for 96hs (apart from the positive control) or until the cells of the negative control were completely dead. Afterwards, cells were left to recover and expanded until confluency. **B)** Graph of one of the replicates of the screen. TMD-infected cells were transfected with 8µg of either tBFP RHBDL2 or tBFP RHBDL2 SA for 48hs before being labelled and sorted. The experiment has been repeated three times, graphs of the other two replicates can be found in Appendix 3.

4.4.2 Amplification of TMD sequences and sequencing library preparation

After sorting, the genomic DNA of the cells in each bin was extracted and the TMD sequences were amplified by PCR. In order to only amplify the sequences I was interested in, I used a pair of primers that annealed to common regions upstream and downstream of the TMD sequence, so that the same pair of primers could be used for all the TMDs of the library (Figure 4.4A). The number of PCR cycles was carefully calibrated to ensure the amplification of the TMD sequences but also to avoid the creation of PCR bias that would enrich for sequences that were easier to amplify and that could affect the final TMD count. After 25 cycles, a PCR product of the correct size (~160bp) was enriched and could be visualised in an agarose gel (data not shown).

In collaboration with Dr. James Carrington (University of Oxford), the amplified TMD were separated from the rest of the PCR mix and the remaining template genomic DNA by using Solid Phase Reversible Immobilization (SPRI, Figure 4.4B) (DeAngelis et al., 1995). This technique is routinely used in sequencing library preparation, and it utilises carboxyl-coated paramagnetic beads that can bind to DNA in the presence of a crowding agent such as polyethylene glycol (PEG). The magnetic beads serve the double purpose of cleaning up the PCR product from the PCR mix, and of selecting products of the desired size. By modifying the ratio between the SPRI beads:DNA, fragments of different sizes bind to the beads with a smaller ratio leading to the binding of bigger DNA fragments. The TMD products were size-selected twice: firstly, a high ratio of beads:DNA was used. By doing this, potential longer fragments and the genomic DNA were bound to the beads that would later be discarded. In this way, fragments that were smaller than 250bp could be found in the supernatant and could be isolated by a second round of size selection. This time, a small ratio of beads:DNA was used so that small sequences would bind to the beads and leading to the enrichment of sequences that were around 160bp. After the size selection, a few of the isolated TMDs PCR products were analysed in a TapeStation for a more accurate calculation of the length of the DNA sequences (Figure 4.5). All the tested sample had a narrow peak around 160bp which is

the expected length of the amplified TMD sequences (TMD sequence+ sequencing primers), confirming that the size selection worked.

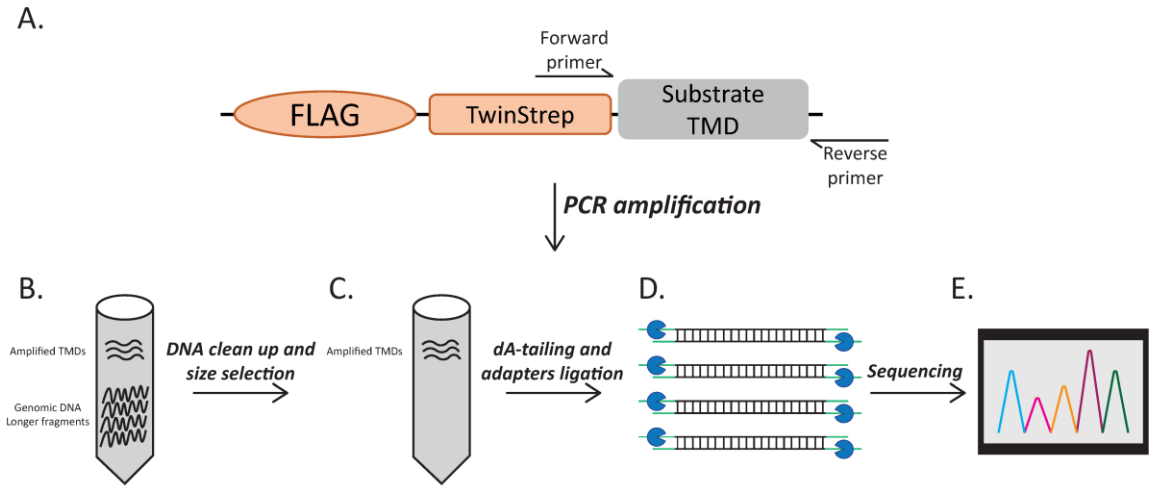


Figure 4.4 Schematic of the sequencing library preparation. A) The TMDs were amplified from the genomic DNA of the sorted cells using a pair of primers that annealed to flanking regions upstream and downstream of the TMDs. **B)** The PCR products were cleaned up and separated from genomic DNA using double side size selection with SPRI beads. **C-E)** The 24 sequencing libraries were prepared by adding A-tails to the end of each PCR fragment, linking a different barcode per library, and adding the adapters which provided the motor proteins necessary for the sequencing of the DNA strand through the nanopore.

The newly generated TMD libraries (24 in total, 8 bins for each of the three replicates) were then further processed by following the necessary steps to prepare a Nanopore sequencing library (Figure 4.4C,D,E). Contrary to other deep sequencing techniques such as Illumina sequencing, Nanopore technology does not rely on synthesis for the sequencing of a DNA strand (Y. Wang et al., 2021). In fact, the sequencing library is inserted into a flow cell that contains thousands of pores (nanopores) that are embedded in a membrane. Each pore is connected to its own electrode that generates a constant current that goes through the pore. When the DNA strand enters the pore, each base causes a disruption of the current, and the entity of the disruption is dependent on which one of the four bases has gone through the pore. The disruption is then decoded with a basecalling

algorithm that identifies the sequence of the DNA strand. The DNA strand goes in and out of the pore due to a motor protein that is attached at the end of the fragment through adapters. The motor protein forces the strand to unwind to create single-strand DNA. Then, the single strand goes into the pore one base at the time, and then detaches from the pore when the whole strand has been sequenced. To prepare the TMD library to be sequenced, Dr. Carrington and I added barcodes at the end of each TMD library. In total, we used 24 different barcodes. By doing so, the TMD library of one bin would be uniquely labelled and could be set apart from the others. This meant that, after adding the adapters providing the motor proteins, we were able to pool the 24 libraries together and run them in a single sequencing run using a MinION machine. At the end of this, a total of ~9 million reads were collected.

A.

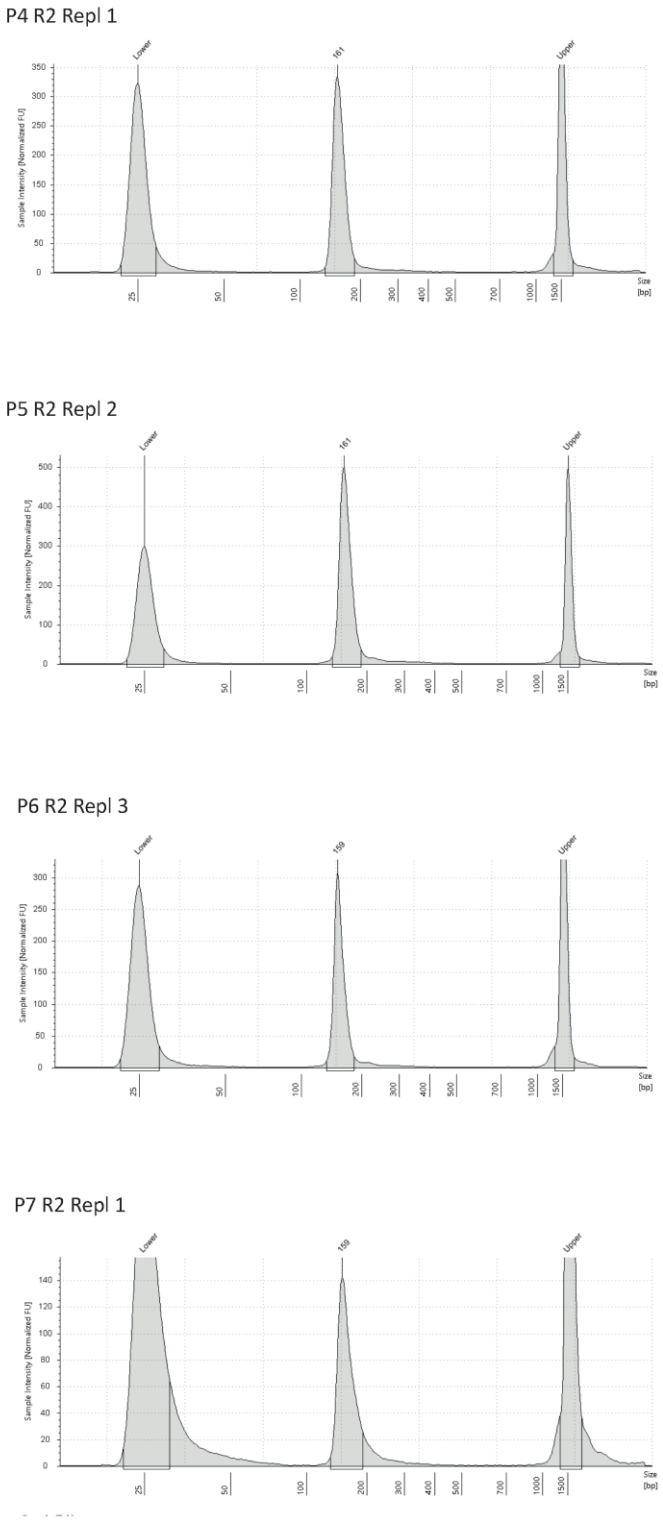


Figure 4.5 SPRI beads size selection. A) The peaks show the size of the PCR fragments after amplification from genomic DNA and SPRI size selection. A subset of samples was analysed via TapeStation and confirmed that the fragments had the expected size of ~160bp. The lower and upper peaks correspond to markers used as control during the analysis.

4.4.3 Analysis of the sequencing reads and identification of novel substrates

The data analysis was carried out by Dr. Carrington. Firstly, the sequencing reads were base called using Guppy (version 6.4.6). This allowed to correlate the voltage disruption of the nanopore to the base the caused it, leading to generation of an actual DNA sequence (Figure 4.6, grey histogram). Afterwards, the same program was used to assign each read to one barcode. Notably, the majority of the reads were uniquely assigned to one barcode as only 8% of the reads could not be assigned. Unfortunately, only a few reads were assigned to barcode20 and it was decided to not consider them. Barcode20 was used to label bin P7 in the tBFP RHBDL2 sample of the third replicate, meaning that this specific bin will be analysed as a duplicate instead of a triplicate. Once the reads had been assigned to one barcode, the adapters were trimmed (Figure 4.6, blue histogram). Moreover, a common short sequence that was amplified during the amplification of the TMDs from genomic DNA was also deleted. Satisfyingly, the processed reads had exactly the length that was expected (~100bp) as it was confirmed by the overlap between the green and pink histograms in Figure 4.6, confirming that the number of PCR cycles and the library preparation had worked as expected. At this stage, ~6 million processed reads could be taken into the next step.

The next step used Minimap2 (Li, 2021) to index the read count to the TMDs, meaning that the processed reads were assigned to each of the 1146 items of the TMDs library. To understand how many reads were assigned to one TMD, the program Salmon was used taking advantage of its nanopore specific alignment mode (Patro et al., 2017). Salmon generated a table where each column was one of the 24 barcodes and each row was a TMD, in this way the number of reads that aligned with each item of the library was calculated for the 8 bins. A quality control check was done using EdgeR (Robinson et al., 2010). The program eliminated 11 TMDs that had a read count that was too low and normalised the read count to the size of the library. A complete list of the commands used for the analysis of the sequencing data can be found in Appendix 5.

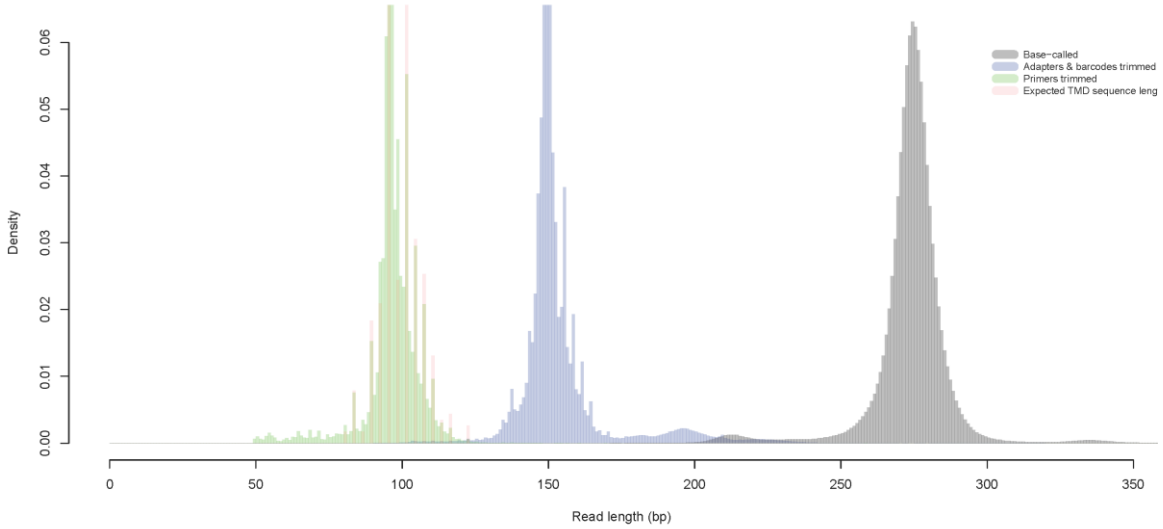


Figure 4.6 Trimming of the amplified TMD fragments. Histograms showing the size of the TMD fragments after base calling (grey peak). Once each read got assigned to one barcode, the barcode and adapters sequences got trimmed (purple peak). Afterwards, the primers sequences used to amplify the TMDs got trimmed (green peak) and each read of ~100bp got assigned to one TMD.

A number of methods were trialled to derive the potential hits from the processed read counts. Given that this type of analysis had not been performed before, there was no straightforward strategy to quantify the results. In the end, we settled on using the average fluorescence of each TMD and defining a potential hit as a TMD that showed a reduction of the average fluorescence (AU) in the condition where tBFP RHBDL2 was expressed. To calculate the average fluorescence, the three replicates were combined to find a mean read count for each TMD in each bin. Then, an arbitrary value between 1-4 and 5-8 was attributed to each bin of the tBFP RHBDL2 and tBFP RHBDL2 SA sample respectively, where 1 and 5 were assigned to P7, and 4 and 8 to P4. These values were multiplied by the number of reads of each TMD per bin and then divided by the total sum of reads for the specific TMD. In this way, each TMD had a value of fluorescence for the two conditions used in the screen (co-expressed with R2 or with R2 SA) (Figure 4.7). TMDs

that showed a decrease in the AU when RHBDL2 was co-expressed could be found in the bottom right corner of the scatter plot.

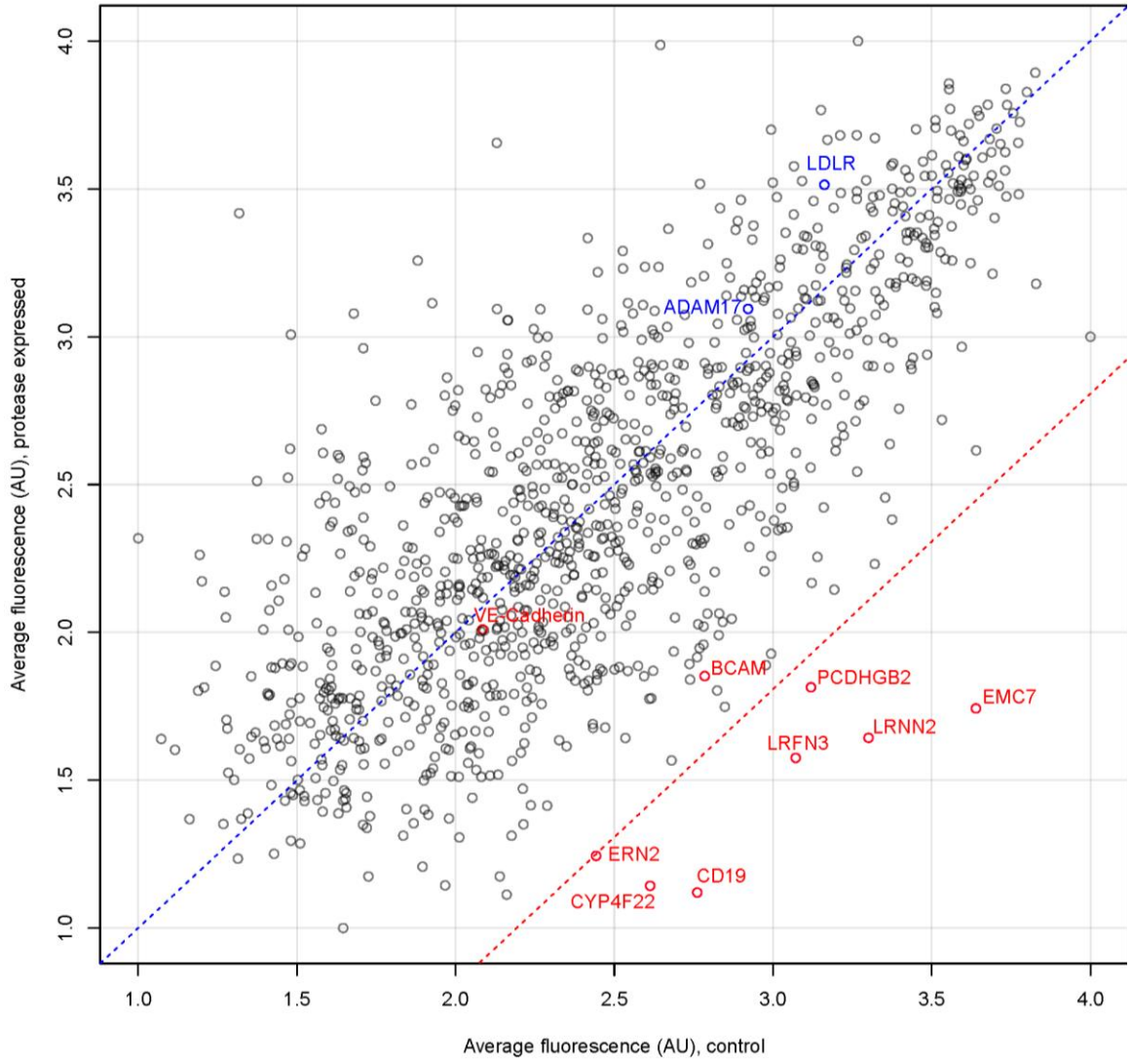


Figure 4.7 Scatter plot of TMDs average fluorescence. The plot shows the difference the average fluorescence (AU) of the 1146 TMDs between the control condition and upon co-expression of tBFP RHBDL2. The blue line identifies no change in fluorescence. The red line identifies the TMDs that showed a significant decrease (p value <0.05). The two hits highlighted in red outside of the red line are two validated substrates (VE-Cadherin, BCAM) while the two highlighted in blue are known non-substrates (LDLR, ADAM17).

Overall, the data showed some variability among the three replicates which increased the background noisiness, but nonetheless it showed a strong correlation between a decrease in fluorescence and the expression of RHBDL2. A list of the TMDs that showed a significant decrease (below red line) can be found in Table 4.1.

TMDs	p value		Name
Q9NPA0	0.004		EMC7
P15391	0.011		CD19
O75325	0.011		LRRN2
Q9BTN0	0.019		LRFN3
Q6NT55	0.021		CYP4F22
Q9Y5G2	0.036		PCDHGB2
Q76MJ5	0.049		ERN2

Table 4.1 List of TMDs that showed a significant decrease of average fluorescence (AU). Complete list can be found in Appendix 4.

Surprisingly, known substrates such as VE-Cadherin and BCAM were not found significant in this assay. This was unexpected but could be attributable to the use of a different method to identify substrates, and to the inherent variability of the assay due to the random integration of the lentiviral vectors into the genome that could affect the expression of the TMDs. This finding was, however, surprising and it highlighted how a thorough validation of the hits is necessary before drawing any conclusions. Even though the decrease in the average fluorescence of BCAM was not found to be significant, it is worth noting that it is included in the first 20 TMDs that showed a reduction. This suggests that, although the dataset needs to be improved, methodically testing the TMDs that showed a decrease could lead to the identification of real substrates.

Overall, the putative substrates that were found significant in this assay had not been identified before as RHBDL2 substrates. A few of them, such as EMC7, ERN2, CYP4F22 are ER resident proteins, while CD19, LRNN3, LRFN3, PCDHGB2 are localised at the plasma membrane. Identifying ER proteins as potential substrates could, at a first glance, appear as an artifact of the system. It is entirely possible that RHBDL2 physiologically never cleaves these proteins as they are

in the wrong compartment, however RHBDL2 is required to pass through the ER to be trafficked towards the plasma membrane. Moreover, it has been shown by both me and Dr. Adam Grieve (data not shown) that tagging RHBDL2 with a KDEL tag to restrict its localisation to the ER, does not impair the catalytic activity. If EMC7, ERN2 and CYP4F22 cleavage will be confirmed by further experiments, it could be interesting to investigate when and under which conditions their cleavage occurs.

The other 4 putative substrates are localised at the plasma membrane and do not show any correlation among them. LRNN2 and LRFN3 are both leucine-rich proteins and they are usually expressed in neuronal cells although not much is known about them. CD19 is an important surface marker for B cells, and it plays an essential role in B cell development. It is involved in the activation and modulation of the B-cell receptor signalling (K. Wang et al., 2012). Lastly, PCDHGB2 is part of the protocadherin family, a subgroup of the larger cadherin superfamily (Morishita & Yagi, 2007). These proteins are mainly expressed in neuronal cells and are thought to be responsible for the incredible cell diversity that is present in the brain. Although these TMDs showed a significant decrease in their AU, further validation experiments are needed to confirm if they are really RHBDL2 substrates.

4.4.4 Preliminary validation of the newly identified substrates

Due to the time constraints of the PhD, I could only do a preliminary validation of 4 of the 7 TMDs that were identified. Further investigations of these 4 and the remaining 3 TMDs is currently ongoing. The first step of validation was to use the same artificial constructs that were used in the screen and test again whether the TMDs were confirmed to be cleaved. To do this, I cloned the TMD sequences and the additional flanking regions of ERN2, EMC7, CYP4F22 and CD19 and fused an N-terminal alkaline phosphatase (AP) tag. Given that the TMDs have a type I topology, the AP tag would be exposed to the extracellular side, so RHBDL2 cleavage will lead to a release of AP into the

media. The media can be collected and by adding AP substrate p-Nitrophenyl Phosphate (PNPP), the amount of AP released in the media can be quantified as a readout for the cleavage using a colorimetric reaction. I tested the cleavage by expressing the AP-tagged TMDs either in the absence of rhomboid or co-transfecting either HA RHBDL2 or HA RHBDL2 SA (Figure 4.8). The TMD of VE-Cadherin was used as a positive control and, unsurprisingly, showed significant release of AP upon co-expression of active RHBDL2. Among the 4 new putative substrates that were tested, only EMC7 showed a significant cleavage. On one hand, this result further confirmed that the noisiness of the data made the identification of real substrates more difficult. However, it also highlighted how the methodology and data analysis were able to enrich for potential substrates, despite the need for improvements.

The cleavage of EMC7 TMD is only the first step of validation that needs to be further investigated by repeating the cleavage assay. Moreover, before claiming that EMC7 is a RHBDL2 substrate, the full-length protein will need to be tested and its cleavage confirmed.

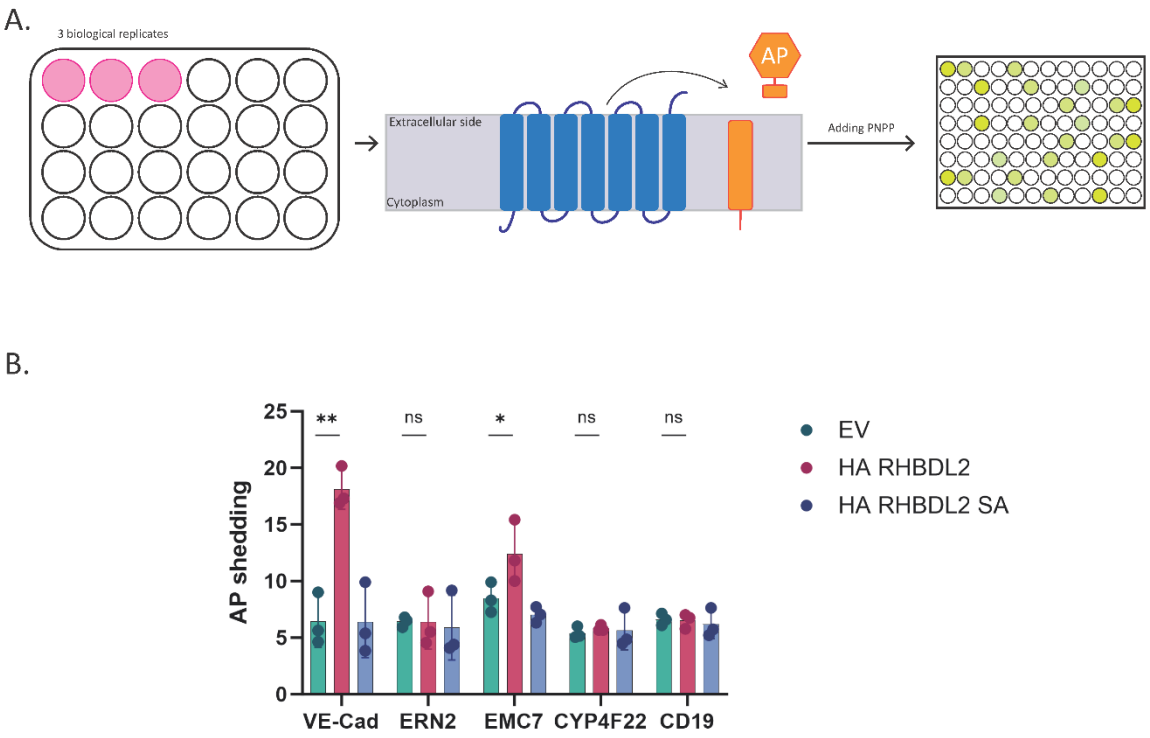


Figure 4.8 AP assay of putative substrates confirmed EMC7 cleavage. A) Schematic of AP shedding assay. Each condition is plated on poly-L-lysine coated plates in three biological replicates. Upon expression of RHBDL2, the AP-tagged TMDs are cleaved and the AP tag is shed in the extracellular media. Media is collected for 24hs before adding PNPP and read the chemiluminescence using a spectrophotometer. **B)** AP assay of 4 of the 7 hits from the screen. The assay confirmed significant cleavage for EMC7 TMD but not for the other 3 TMDs. VE-Cadherin TMD was used as positive control. The graph combines the results of three independent experiments and the error bars represent standard deviations. The statistical test that was used was unpaired student t-test comparing HA RHBDL2 samples vs HA RHBDL2 SA ones. **=<0.01, *=<0.05.

4.5 Discussion

In this chapter, I have described a screening experiment using the TMD library to identify novel substrates for RHBDL2. The TMD library was generated in collaboration with Dr Samuel Dean and Twistbioscience, which took care of the bioinformatics for retrieving the DNA sequences and the cloning into the lentiviral vector respectively. After producing viable lentiviral particles, I infected enough cells to keep a coverage of the library of 500x while also ensuring to use an MOI<0.2, so that each cell would only express one TMD. Avoiding double integration is essential for the success of any screen when using pooled libraries. In fact, having two TMDs expressed at the surface of the same cell could have generated potential problems in determining the decrease of fluorescence, for example if only one of the two was cleaved. The infected cells were sorted, and the cell population divided into 4 bins according to the value of fluorescence.

After the TMDs were amplified from the genomic DNA extracted from the sorted cells, in collaboration with Dr. James Carrington we proceeded to the sequencing of the PCR products using Nanopore technology. Using Nanopore instead of more common sequencing services such as Illumina allowed us to have more control over the process, as it could be done inside our labs, and it required a shorter turnaround time. We labelled each bin of the three replicates with one of the

24 barcodes that were provided. By doing that, we were able to use a multiplex approach by pooling the libraries and run them together. The sequencing run technically worked perfectly, with a total of 6 million reads left after quality control and processing of the reads.

Understanding how to analyse the processed data was not straightforward as there was not a clear strategy that could be used. The analysis was also made more difficult by the noisiness of the data. In fact, there was some variability among the three replicates. This was, in part, expected given the way the cells were infected. Lentivirus integrates randomly in the genome and the expression of the TMDs would be influenced by the genomic environment in which they are. This can change between replicates and potentially leading to different levels of expression that could be part of the reason behind the noisiness of the data. After trying different methods to quantify the data, we settled on calculating the average fluorescence of each TMD. This led to the discovery of 7 TMDs that showed a significant reduction in their fluorescence when co-expressed with RHBDL2. ERN2, EMC7, CYP4F22 and CD19 were tested with preliminary experiments to confirm their cleavage. This showed that only EMC7 was really cleaved by RHBDL2. As mentioned above, EMC7 is an ER resident protein. The possibility of finding TMDs that were not in the plasma membrane had been accounted for from the planning stage of this project. In fact, the library contained all the TMDs of single pass proteins irrespective of their localisation. First of all, the cleavage of EMC7 needs to be validated through further experiments. If it is confirmed, it would be interesting to understand what its role could be. RHBDL2 is a membrane protein that needs to be trafficked from the ER through the secretory pathway, to get to the final destination of the plasma membrane. However, according to current knowledge, the catalytic activity of RHBDL2 is not inhibited by its localisation. In fact, limiting RHBDL2 to the ER using a KDEL tag did not impair its ability to cleave the TMDs of Spitz and Orai1 (data not shown). EMC7 is part of the EMC complex, a multi-protein complex that is required for the insertion of membrane proteins in the ER membrane (Pleiner et al., 2020). Although, the majority of RHBDL2 substrates to date have been studied as single proteins, there are multiple evidence that RHBDL2 can cleave proteins that are assembled in

complexes. In fact, RHBDL2 cleaves Orai1 that is known to form a hexamer at the plasma membrane (Grieve et al., 2021b). Specifically, the 4th TMD of Orai1 (which is the domain cleaved by RHBDL2) is located at the exterior of the hexamer structure, making it more accessible to be processed (Hou et al., 2012). Interestingly, a recent structure of the EMC complex has showed that EMC7 should also be localised on the outside on the luminal side of the protein complex (O'Donnell et al., 2020; Pleiner et al., 2020). If the cleavage of the full-length EMC7 is indeed confirmed, it would be interesting to investigating what the role of RHBDL2 could be in EMC7 regulation.

Surprisingly, the analysis of the sequencing data did not highlight any of the known substrates as significant hits. The TMD of BCAM is the one that is the closest to the significancy threshold. This is most likely due to the noisiness of the data, and it suggests that there are some hits that are probably being missed with the current set of data. However, it is worth noting that all the previously validated substrates have been identified using very different approaches. This makes the comparison of how significant each hit is between different studies more difficult. Nonetheless, it is a puzzling finding and it is a big reminder that a screen is only the first step. Thorough validation of the results is needed to confirm if this novel method is actually working and if the analysis of the data can be improved. An indication that the analysis has somewhat worked in enriching for TMDs that show a decrease in the fluorescence is given by the preliminary confirmation of EMC7, but also by LDLR and ADAM17 (TMDs that are known not to be cleaved by RHBDL2) being in top part of the graph. This suggests that methodically testing the TMDs in the bottom part will still be useful in identifying putative substrates. Moreover, decreasing the noisiness of the data by having more replicates and potentially increasing the number of reads will help improving the dataset and give more confidence to the analysis.

Overall, in this chapter I have presented the final stages of the development of a novel assay. I have tested it by screening a large lentiviral library. The analysis of the sequencing data has

identified some interesting hits which are currently being tested to confirm if they are real substrates.

Chapter 5

Characterisation of RHBDL2 cleavage of VE-Cadherin

5. Characterisation of RHBDL2 cleavage of VE-Cadherin

5.1 Introduction

During the last months on my DPhil, I have started investigating the cleavage of VE-Cadherin by RHBDL2. There were several reasons to start this side project. Firstly, VE-Cadherin has been recently identified as a RHBDL2 substrate by two independent studies (Battistini et al., 2019; Grieve et al., 2021a) and the cleavage has not been fully investigated. Secondly, during the preliminary experiments for the setting up of the screen, cleavage of VE-Cadherin TMD was often more significant compared to the cleavage of other positive control TMDs such as Orai1 and Spitz (Figure 3.10A-B). This further encouraged me to better understand the relationship between RHBDL2 and VE-Cadherin. Lastly, studying the cleavage of VE-Cadherin would help set up a list of essential experiments necessary to investigate new substrates. This knowledge would prove extremely useful for a swifter validation of any potential hits derived from the screen.

In this chapter, I am going to present preliminary experiments that confirm the cleavage of VE-Cadherin TMD and of the full-length protein. Moreover, I used point mutants of the transmembrane domain to map the cleavage site. These results represent only a first step in the characterisation of the relationship between VE-Cadherin and RHBDL2 and they will have to be further explored with future experiments.

5.2 VE-Cadherin is cleaved by RHBDL2

5.2.1 RHBDL2 cleavage occurs within the transmembrane domain of VE-Cadherin

As mentioned before, VE-Cadherin had already been identified as a RHBDL2 substrate. At first, I confirmed that the TMD of VE-Cadherin was cleaved by using an alkaline phosphatase (AP) reporter system (Figure 5.1A). Alkaline phosphatase was fused at the N-terminus of the TMD sequence of mouse VE-Cadherin and mouse Cadherin10 that was used as a negative control (Grieve et al., 2021a). Each TMD sequence was flanked by short stretches at the luminal and cytoplasmic side. Similar to the rationale used for designing the TMDs of the library, a 7aa long luminal domain was added to accommodate for potential inaccuracy in setting the N-terminal border of the TMD. While the 3aa long cytoplasmic domain was to ensure TMD stability, but to avoid any cytosolic regulation.

The AP-tagged TMDs of VE-Cadherin and Cadherin10 were transfected in HEK293T with either an empty vector (EV), HA RHBDL2 or RHBDL2 SA (Figure 5.1B). Upon co-expression with active RHBDL2, there was a significant release of AP after cleavage of VE-Cadherin TMD. The release did not occur when expressing Cadherin10 TMD or upon co-expression of the inactive RHBDL2 SA. I further confirmed this result, by repeating the same experiment but using stable cells that expressed the F-TwS VE-Cadherin TMD construct used for the screen (Figure 5.1C). I assessed the cleavage with flow cytometry and, when RHBDL2 was expressed, there was a striking reduction of the green fluorescence and almost the totality of the cell population shifted to lower bins, confirming RHBDL2 cleavage.

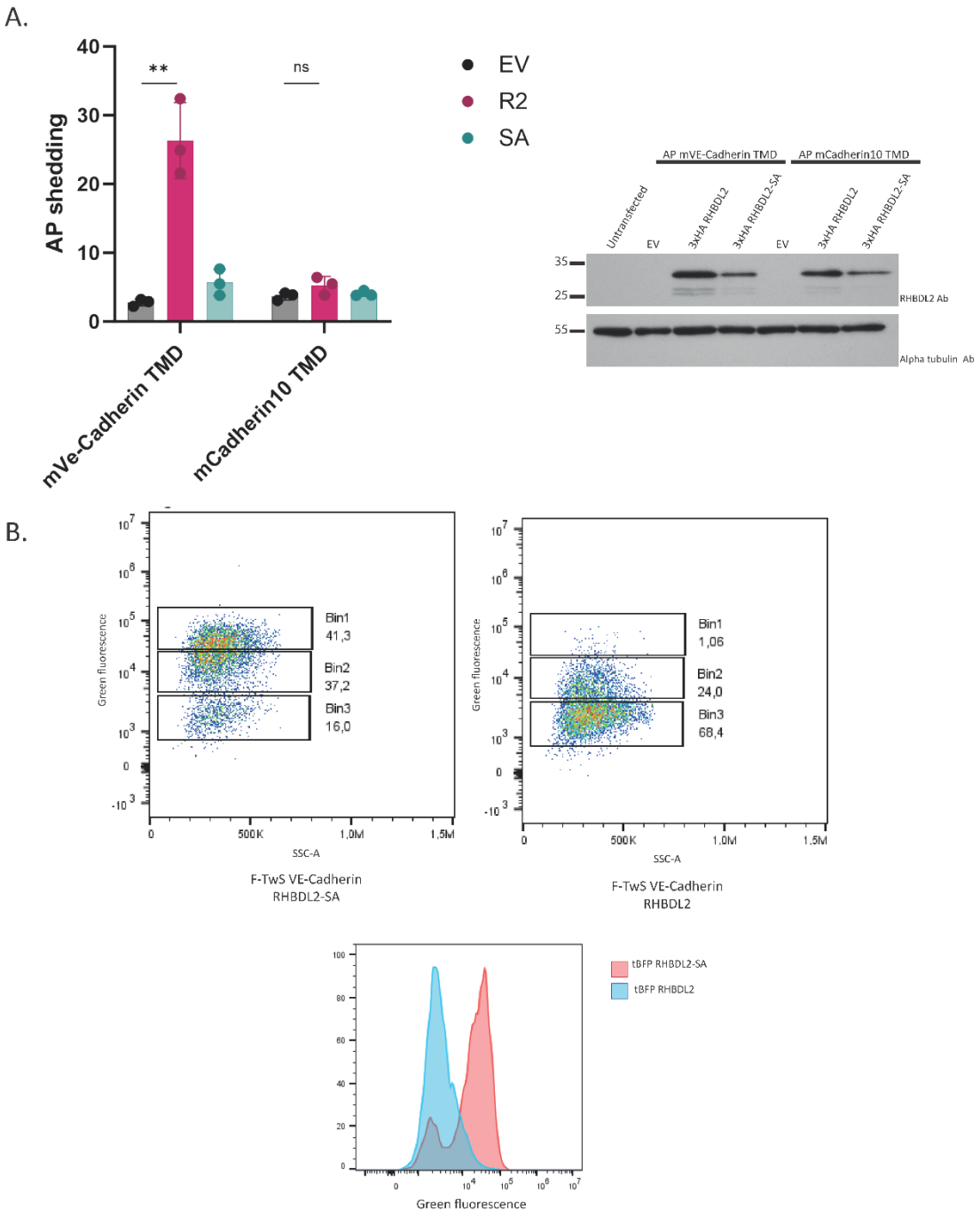


Figure 5.1 TMD of hVE-Cadherin is cleaved by RHBDL2. A) *Left:* AP shedding assay of mVE-Cadherin and Cadherin10 AP-tagged TMDs, co-transfected with empty vector (EV), HA RHBDL2 and HA RHBDL2 SA. Shed AP was collected in OptiMEM for 24hs. Unpaired t-test was performed using Prism GraphPad software; error bars represent the standard deviation, ns= not significant, ** = p value<0.01. *Right:* Western blot showing expression of HA RHBDL2 and HA RHBDL2 SA. Alpha tubulin was used as loading control. **B)** Flow cytometry analysis of cells stably expression F-TwS VE-Cadherin where either tBFP RHBDL2 or tBFP RHBDL2

SA was overexpressed for 24hs. Cells were labelled with a FLAG-DyLight 488 antibody and the green fluorescence was assessed using a flow cytometer.

5.2.2 The full-length VE-Cadherin is processed by RHBDL2

Next, I aimed at confirming that the human full-length VE-Cadherin was cleaved (Figure 5.2A). To do this, I cloned a version of hVE-Cadherin that was tagged with 2xHA tag at the N-terminus and a V5 tag at the C-terminus. Having tags at both termini will allow the detection of the two cleavage products. Moreover, the 2xHA tag was positioned after the signal peptide (SP) and propeptide sequence (Nollet et al., 2000; Posthaus et al., 1998) to ensure correct insertion into the membrane and complete maturation of VE-Cadherin. 2xHA VE-Cadherin V5 was transiently transfected in HEK293T cells with either empty vector (EV), RHBDL2 or RHBDL2 SA. Upon co-expression of RHBDL2, it was possible to detect a cleavage product of 25-30kDa using an antibody against the V5 tag. This product was consistent with a rhomboid cleavage occurring on the luminal half of the TMD (Figure 5.2B). The corresponding N-terminal product was never detected in the cell lysates using antibodies against the HA tag, most probably because it gets shed into the extracellular media after cleavage. In order to assess this, the conditioned media will need to be tested. The proportion of the C-terminal cleavage product (CP) was relatively small compared to the level of the full-length protein (FL), this could suggest that the product undergoes further processing via protein degradation pathways. Considering that the cleavage occurs at the luminal side of the TMD, the cleavage product is, at least in part, still embedded in the plasma membrane. Because of this, I investigated the fate of this C-terminal cleavage product by repeating the experiment with the same conditions but inhibiting lysosomal protein degradation with bafilomycin (BafA, Figure 5.2C). Notably, stopping protein degradation rescued the cleavage product, confirming that it is indeed degraded via the lysosomes. However, the effect was mild and did not affect the level of the full-length protein. There are multiple potential explanations to elucidate this result. The simpler explanation is that RHBDL2 processing of VE-Cadherin is not a major one, at

least in an unstimulated condition. It would be interesting to test if there are any stimulus that could increase RHBDL2 processing. Secondly, the partial rescue of the cleavage product could mean that multiple degradation pathways, such as the proteasome, are involved in its processing. This can be assessed by using the appropriate inhibitors. Lastly, the result could also suggest that the largest proportion of the C-terminal product is not degraded at all. The cytoplasmic domain of VE-Cadherin is an important signalling hub (Bravi et al., 2014) and for other members of the cadherin family, intramembrane proteolysis can lead to the formation of cleavage products that have new functions (Maretzky et al., 2005) . It is tempting to speculate that the cleavage product is not degraded as it could have a new biological role. However, further experiments will be needed to investigate this.

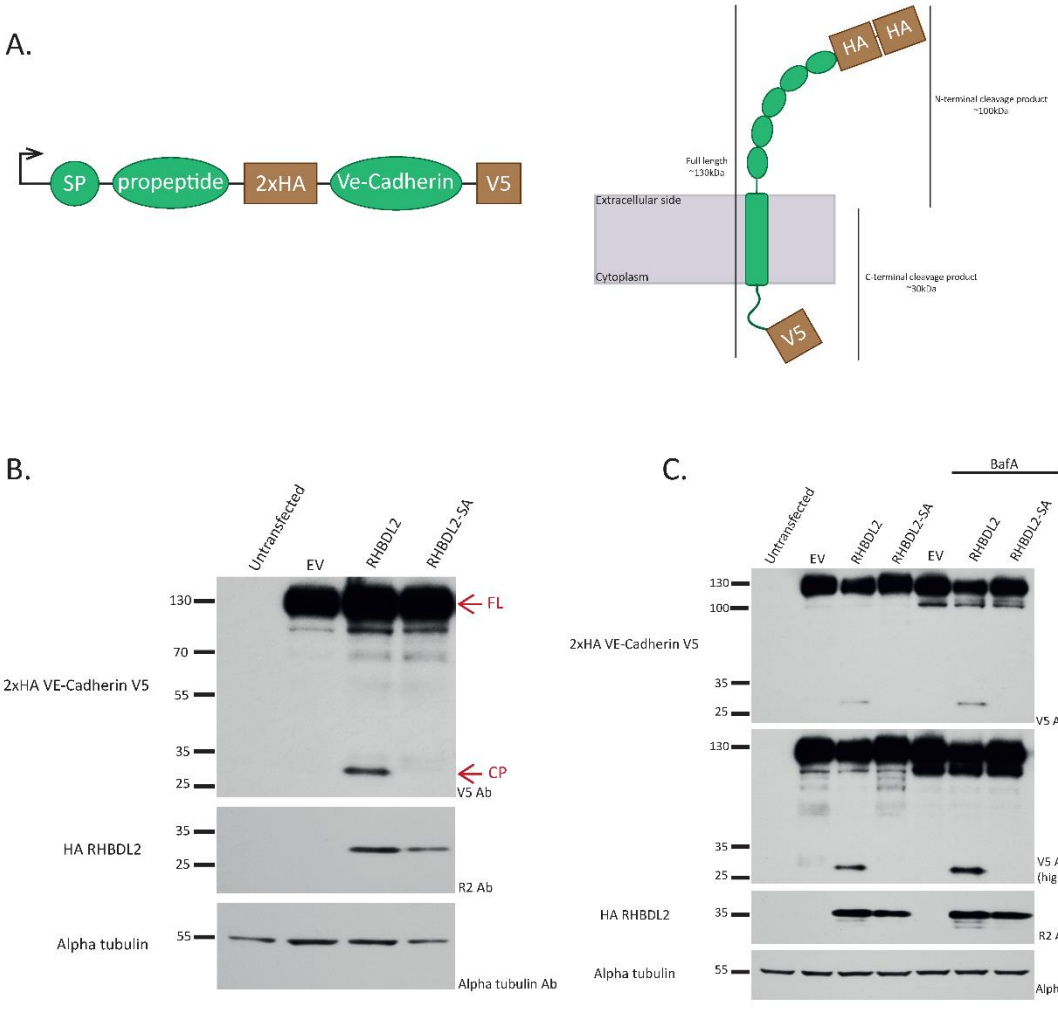


Figure 5.2 VE-Cadherin full length is cleaved by RHBDL2. A) Schematic of 2xHA VE-Cadherin V5 construct and of the N- and C-terminus cleavage products. **B)** Western blot analysis of 2xHA VE-Cadherin co-transfected with RHBDL2 or RHBDL2 SA. Upon RHBDL2 expression, the C-terminus cleavage product (CP) can be detected around 30kDa. CP: cleavage product, FL: full length. **C)** Western blot of cells treated with Bafilomycin (BafA) for 16hs to stop lysosomal degradation. The inhibition only mildly rescued the cleavage product degradation. Alpha tubulin was used a loading control in B and C.

5.3 Mapping of the cleavage site using point mutations

5.3.1 Mutating luminal side of the TMD partially impairs cleavage

After confirming that VE-Cadherin was cleaved by RHBDL2, I set out to map the cleavage site. To do this, I used the alkaline phosphatase reporter system, and I fused AP to the N-terminus of the TMD of hVE-Cadherin. I generated three point mutants where 3 or 4 amino acids were mutated at the same time (mut1-3, Figure 5.3A). In mutating the TMD, I focussed on the luminal half since RHBDL2 usually cleaves at the juxtamembrane region of the luminal side. Moreover, each amino acid that was not already a Leucine (L) was mutated into one. The choice to mutate into leucine was made for two main reasons: firstly, leucines are hydrophobic amino acids so that the overall negative charge of the TMD could be maintained. Secondly, it has been described that poly leucine peptides form alpha helices that could be inserted into the membrane (Chipot & Pohorille, 1998), so replacing multiple amino acids to leucine and generating stretches of poly leucines should not affect the insertion of the mutated TMD into the membrane. The only exception was Q599 that was mutated into an asparagine (N) to keep the polarity. According to the proposed structure of Alphafold, this residue should be at the very beginning of the transmembrane domain, so Q599 could interact with the heads of the phospholipids as the TMD enters the bilayer.

Mut1,2,3 were co-transfected with either empty vector, HA RHBDL2 or HA RHBDL2 SA (Figure 5.3B). Surprisingly, none of the mutants completely abolished the cleavage. However, mut1

and mut2 caused a reduction of the cleavage compared to the controls, although only the decrease caused by mut1 was statistically significant (Figure 5.3C). Mut3 did not show any statistical difference compared to WT. The mutations of mut1 and mut2 affect the most luminal part of the TMD, so their effect on the cleavage is consistent with previous evidence that RHBDL2 cleavage occurs at the juxtamembrane region of the luminal side of TMDs (Lastun et al., 2016). Given that mut1 and mut2 are not overlapping, it was speculated that the reason the cleavage was not completely inhibited was that the cleavage site was located between the two mutants. To test this, I combined the two mutations and replaced the first 7aa of the TMD with leucine (aside from Q599N) generating mut1+2 (Figure 5.3D). Surprisingly, the double mutant only caused a moderate additional decrease of the cleavage compared to the single mutants, reducing the cleavage of 42% against the 30% reduction caused by mut1 and mut2.

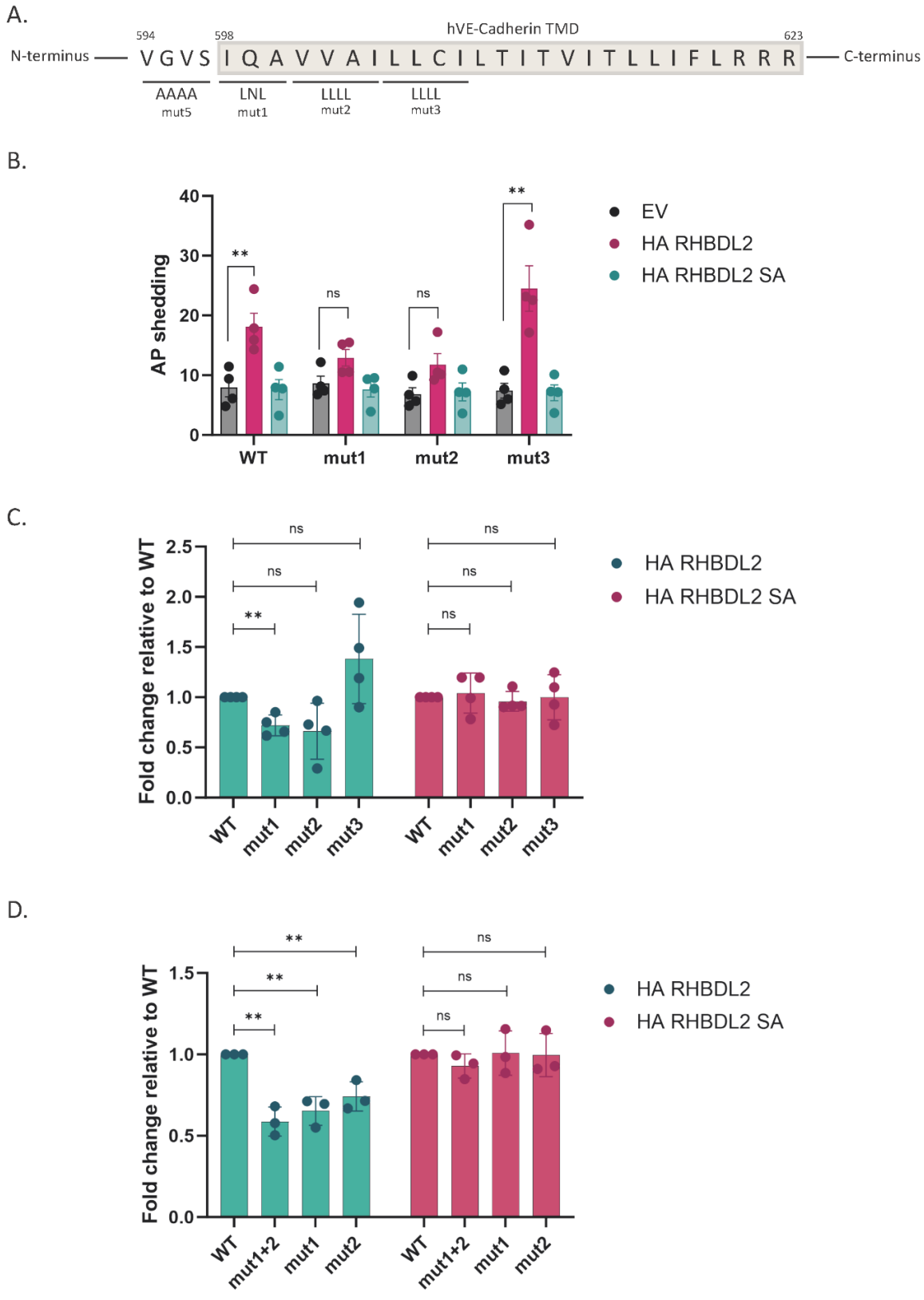


Figure 5.3 Using point mutants did not identified the cleavage site. A) Schematic of hVE-Cadherin TMD and point mutants used in the experiments. **B)** AP shedding assay of WT TMD or mut1, 2, 3 TMD

transfected with empty vector (EV), HA RHBDL2 or HA RHBDL SA. Media were collected for 24hs before analysis. **C)** Fold change of the cleavage of mut1, 2, 3 compared to WT TMD. **D)** Fold change of the cleavage of mut1+2, mut1 and mut2 compared to WT TMD. Unpaired t-test was performed using Prism GraphPad software; error bars represent the standard deviation; each dot represents an independent experiment; ns=not significant, ** = p value<0.01

5.3.2 Point mutations of TMD do not identify cleavage site

It has been shown that rhomboid proteases are sensitive to which amino acid is present at the cleavage site, preferring amino acids with a smaller side chain (Strisovsky et al., 2009). Given this evidence, it could be that mutating into leucine was not disrupting the cleavage site enough, so I mutated each of the amino acids of mut1 and mut2 into phenylalanine which has a larger side chain and has been proved to inhibit cleavage when it is part of the cleavage site (Strisovsky et al., 2009) (Figure 5.4A). Unfortunately, none of the mutants that were assessed abolished it, as none of them significantly reduced the cleavage compared to the WT TMD (Figure 5.4B). The exception was A600F: this mutant was shed in all three conditions equally, even in the absence of RHBDL2 or when the inactive protease was expressed. This could be due to several reasons; the mutation could make the TMD more unstable leading to unspecific clipping of the AP tag. Moreover, VE-Cadherin is also cleaved by the metalloprotease ADAM10 (Schulz et al., 2008), so it could be possible that A600F mutation increases the processing by this enzyme. However, further experiments are needed to identify the cause of A600F shedding.

Nevertheless, these experiments did not succeed in mapping the cleavage site. Yet, the stripped back nature of the AP construct assures that the cleavage must happen within the TMD sequence or in the juxtamembrane region. To evaluate whether this domain was involved in the cleavage, 4 of the 7amino acids directly upstream of the TMD were mutated to Alanine (from V594 to S597) (Figure 5.4C). However, the cleavage was still unaffected by these mutations. In conclusion, point mutants of the luminal half of the TMD or the juxtamembrane region did not

abolish RHBDL2 processing. This could be due to the mutations not being disruptive enough, especially for the juxtamembrane region where the residues where only mutated to alanine which is permitted in a rhomboid cleavage site (Strisovsky et al., 2009). Replacing these residues with phenylalanine will help elucidate whether the cleavage site is present in this region.

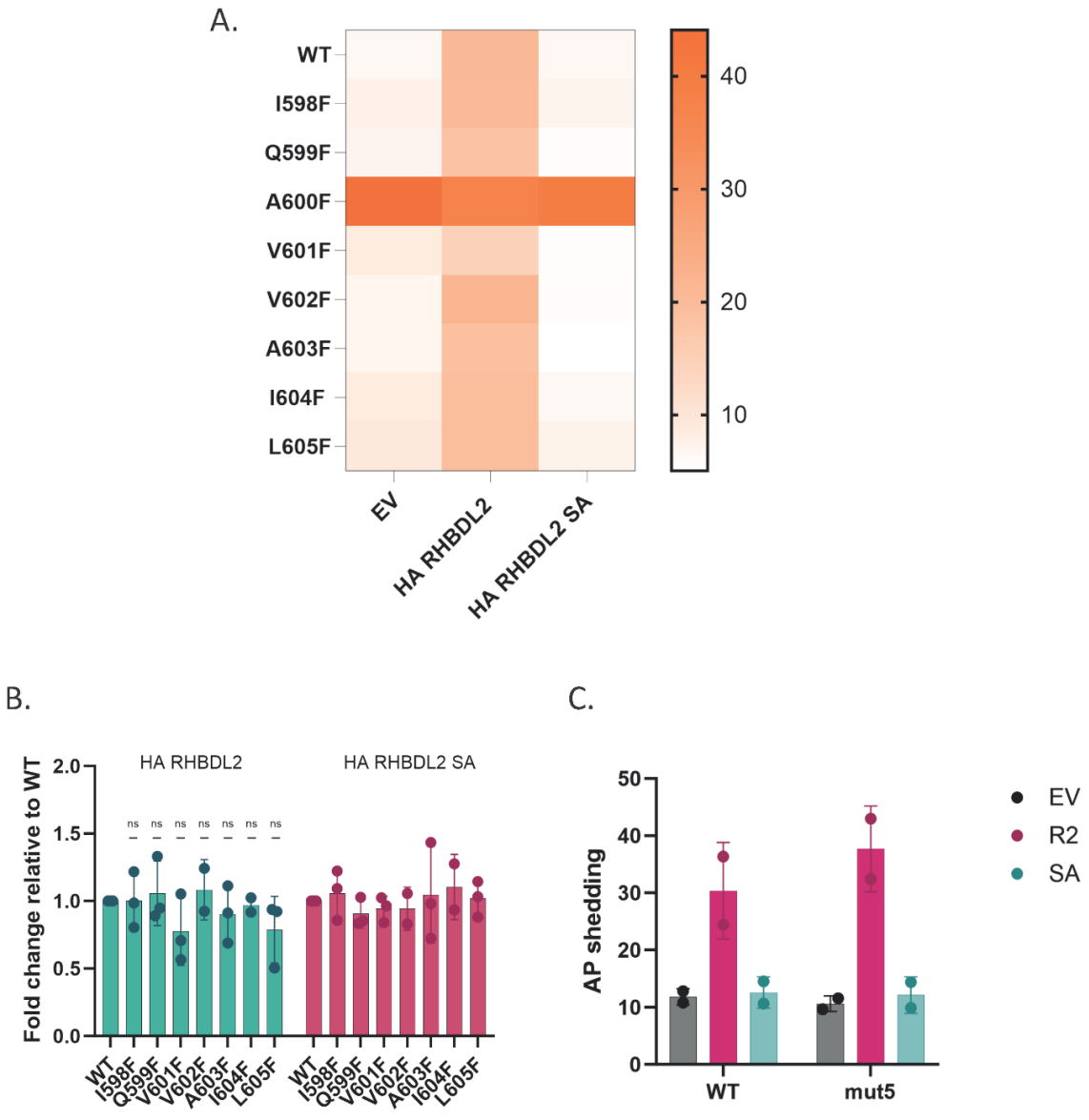


Figure 5.4 Phenylalanine mutants of the TMD did not identify the cleavage site. A) Heatmap showing the shed AP of the phenylalanine mutants of VE-Cadherin TMD. **B)** Fold change of the cleavage of phenylalanine mutants compared to WT TMD. **C)** AP shedding assay of mut5 compared to WT TMD

Unpaired t-test was performed using Prism GraphPad software; error bars represent the standard deviation; each dot represent an independent experiment, ns= not significant.

5.4 Discussion

In this chapter, I have presented the preliminary work that started to elucidate the cleavage of VE-Cadherin by RHBDL2. Firstly, I have confirmed that both the TMD and full-length protein of VE-Cadherin are cleaved only in the presence of active RHBDL2 and not when the inactive enzyme is expressed.

Afterwards, I set out to map the cleavage site using point mutations of the luminal side of the TMD. This region, and the juxtamembrane domain closest to it, is highly conserved in mammals and, to a lesser extent, in more distant organisms. This could suggests a potential functional role on this domain. To identify which of the conserved residues were involved in the cleavage I mutated group of 3-4 aa into leucine and assessed the cleavage. Two of the three mutants showed a decrease of the cleavage, but they did not completely abolish it. Moreover, when the two mutants were combined there was no additional reduction. A potential explanation for this could be that mutating the residues into leucine is not enough to disrupt the cleavage site. In fact, Strisovsky et al. (Strisovsky et al., 2009) demonstrated the presence of a sequence dependent motif that determines the cleavage site. By saturation mutagenesis they identified which amino acids were permitted at the site. Most amino acids with a smaller side chain were allowed, while amino acids with larger side chains were not tolerated. For this reason, I decided to mutate the first seven aa of the TMD into phenylalanine and assess whether the cleavage would be disrupted. Surprisingly, none of the mutants abolished the cleavage apart from A600F. This mutant was shed in all the tested conditions, making it difficult to conclude if the shedding was RHBDL2 dependent. It could be that the mutation of the alanine would cause an unspecific clipping of the AP tag leading to its accumulation in the media. It would be interesting to assess if ADAM10 has a role in the shedding.

In fact, ADAM10 putative cleavage site should be included in the AP-tagged construct so it could be tested whether the unregulated shedding of A600F is due to an increase of the processing by ADAM10.

Given that the AP-tagged TMD only contained the transmembrane domain sequence and the short sequences flanking it, the fact that the point mutants of the TMD did not abolish the cleavage was puzzling. On one hand, it confirmed that the cleavage site was localised within the sequence that was tested. On the other hand, it opened the possibility that the cleavage was not occurring within the TMD. The active site of rhomboid proteases is open to the lumen (Y. Wang et al., 2006), and there has been evidence showing that the cleavage could still occur if the site was in the juxtamembrane region (Strisovsky et al., 2009). To assess if this region was involved in the cleavage, the 4aa directly upstream from the TMD were mutated into alanine but unfortunately, the cleavage was again unaffected. However, alanine is the most common amino acid that forms a rhomboid cleavage site. It would be interesting to test again the role of this region in the cleavage by replacing into phenylalanine, which is, on the contrary, poorly tolerated when part of the cleavage site.

These results represent the first step in understanding how RHBDL2 may be involved in regulating VE-Cadherin activity. Moving forward, it would be interesting to investigate how this regulation might occur in a more physiological setting, for example using endothelial cells. It is well established that the level of VE-Cadherin at the cell surface is tightly regulated and is essential to maintain the integrity of the endothelial barrier (Cadwell et al., 2016). The main way the dynamics of VE-Cadherin are controlled is by removing it from the membrane. In fact, all members of the cadherins family can undergo endocytosis through clathrin-dependent and independent mechanisms (Kowalczyk & Nanes, 2012). Once in vesicles, cadherins can either get recycled back to the plasma membrane or get degraded. However, there are several evidence of regulations of VE-Cadherin and other members of the family through proteolytic shedding by soluble and

intramembrane proteases (Dreymueller et al., 2012; Schulz et al., 2008; Su & Kowalczyk, 2017). Cleaving VE-Cadherin leads to a fast removal of the protein from the membrane and the generation of fragments that could have a new biological role. For example, it has been shown that the ADAM10-dependent shedding of E-Cadherin ectodomain has an effect on cell motility (Maretzky et al., 2005). Whether cleavage of VE-Cadherin could have a biological consequence is still unclear and will need further investigation.

Chapter 6

Discussion and future perspectives

6. Discussion and future perspectives

6.1 Considerations on the TMD screen

The work that I conducted during my DPhil aimed at the development of a novel assay for the identification of rhomboid substrates in a systematic way. Up until now, all the methods that have been used were candidate-based approaches where a TMD that had similar characteristics to already validated substrates was assessed for rhomboid cleavage (Lastun et al., 2016). These methodologies are certainly successful as they have led to the identification of all the currently known substrates for RHBDL2 and RHBDL4. However, candidate-based approaches are usually time consuming and in general not well suited for the systematic analysis of a large number of TMDs. For these reasons, the goal of my DPhil project was to develop a new method that can screen large libraries of potential substrates. This would help in testing the majority of potential rhomboid substrates in one go and, moreover, it could also help in clarifying rhomboid proteases' broader role within cells. In fact, although our knowledge of rhomboid proteases has increased substantially over the last ten years, there is still a gap in a general role that these proteases play inside cells. It is entirely possible that rhomboid proteases are actually involved in a multitude of pathways, but testing a large number of substrates in one time would give a more comprehensive view of which TMDs are cleaved by these proteases that could highlight a potential effect that rhomboid proteases have in specific pathways. Moreover, the screen was more unbiased than previous studies that were performed in the past, as it used all the type I TMDs expressed in the human genome independently of their expression in specific cell types. However, the exogenous expression of the TMDs is also one of the limitations of this method. In fact, for the screen to succeed, it relied on the all the TMDs to be expressed at a similar level and to have similar stability within the cell. If this was not the case, there could be an increase in the number of false negatives.

This novel assay is based on the use of flow cytometry to interrogate the cleavage event. More specifically, I proved that a decrease in fluorescence can be used as a readout for the cleavage. Since this assay had not been used to test cleavage of substrates, I had to make sure that the screen could actually work for this goal. To do this, I spent a large part of my PhD troubleshooting each step of the screen before being able to test the method using the TMD library. Moreover, I decided to focus on RHBDL2 which is the best characterised mammalian rhomboid. This provided several controls, both TMDs that had already been identified as substrates (such Orai1, Spitz and VE-Cadherin) and TMDs that were known not be cleaved by RHBDL2 (for example LDLR and ADAM17). Initially, I developed the assay around the use of a bicistronic vector that coded for two TMDs, one tagged with mNG and one with mSc. The mSc-tagged TMD was a non-substrate TMD, usually mSc LDLR, and it was therefore acting as a negative control, whose fluorescence would stay unchanged. The mNG-TMD was the putative substrate, and, upon cleavage, its fluorescence would decrease and lead to a change in the mNG/mSc ratio. After implementing a few changes to the initial setup (such as changing the mNG tag for a F-TwS one), my experiments showed that the reduction of the fluorescence ratio due to the cleavage of the F-TwS-tagged TMD caused a shift of the cells, and that this reduction could be used to isolate and sort the cells. However, even though the bicistronic vector was at first a clever system and it seemed to work in the preliminary tests, it became clear that its use was not feasible when using the library. In fact, when I tried to test it using stably integrated TMDs, I could never select cells that were expressing both TMDs. Given that the idea was to develop a functional assay, I decided to switch the set-up of the screen by abandoning the use of the bicistronic vector to adopt a new system with only one TMD being tested. By doing this, I was able to design a more efficient assay and to show that the system was able to detect the cleavage of stably integrated TMDs for both VE-Cadherin and Spitz. Moreover, expression of the TMD of LDLR did not cause any decrease in the fluorescence confirming that it was possible to discriminate between substrate TMDs and not substrates.

I then proceeded to do the screen, by first generating the lentiviral TMD library in collaboration with Dr Dean and TwistBioscience. As mentioned before, given that the TMD library was not a commercial one, a trial-and-error approach was necessary to perform the experiment. Overall, I repeated the screen three times and, in collaboration with Dr. Carrington, we carried out the sequencing analysis using Nanopore Technologies. 7 TMDs showed a significant decrease of their fluorescence upon co-expression with RHBDL2. However, the data was generally quite noisy due to a degree of variability among the replicates and the positive controls did not show any significant reduction. This does not necessarily mean that the identified hits are wrong, but it is an unusual result that caused to treat the hits with more caution. A thorough validation of the hits is currently being carried out, first by confirming the cleavage of the TMDs and if that result is recapitulated, by testing the cleavage of the full-length proteins. Although the data is noisy, it still showed a correlation between the decrease of the average fluorescence and co-expression of RHBDL2. This suggests that methodically testing the TMDs that showed the reduction will still be helpful in identifying putative substrates in an efficient way.

Preliminary experiments of 4 of the 7 identified hits (ERN2, EMC7, CYP4F22, CD19) confirmed the cleavage of EMC7 TMD. EMC7 is a single-pass transmembrane protein that has a large luminal side that forms a β-sandwich structure (O'Donnell et al., 2020). It is part of the EMC complex, a 10-subunits complex which plays an important role in the insertion of membrane proteins in the ER membrane (Chitwood et al., 2018; Guna et al., 2018). According to the recently resolved structure, EMC7 is located on the luminal side of the complex and in the exterior part of the structure (O'Donnell et al., 2020; Pleiner et al., 2020). This suggests that, theoretically, its position makes EMC7 accessible for processing by other proteins that are present in the membrane. The finding that RHBDL2 cleaves EMC7 was unexpected, and it indeed needs to be strengthened by further experiments using the full-length protein. But if the cleavage of full-length EMC7 is confirmed, it raises interesting questions. First of all, as it was mentioned in previous chapters, EMC7 is an ER-resident protein. However, the current knowledge of RHBDL2 does not suggest that

the catalytic activity is dependent on its localisation, so it is theoretically possible that the cleavage occurs in the ER. Secondly, EMC7 is part of a complex made up of 9 different subunits- does the cleavage occurs when EMC7 is already in the complex? Or does RHBDL2 recognises EMC7 when it is not interacting with the other subunits? The latter option could be supported by the evidence that two different bacterial rhomboids act as a quality control factor for orphan subunits. In fact, Liu et al revealed that the rhomboid proteases GlpG and Rhom7 in *S. Sonnei* recognise and selectively cleave orphan subunits of respiratory complexes only when they were not associated with the cognate complex (Liu et al., 2020). Overall, the finding that RHBDL2 cleaves the TMD of EMC7 is an interesting result and if validated further, it would be worth investigating what is the role of the cleavage and in which context it occurs.

Further experiments of the remaining 3 hits (LRNN2, LRFN3, PCDHGB2) have validated the cleavage of LRFN3 and PCDHGB2 TMDs. Both of these proteins are involved in adhesion in neuronal cells. Although the same rationale of EMC7 applies to LRFN3 and PCDHGB2 and the cleavage should also be confirmed with the full-length proteins, RHBDL2 processing of the two TMDs adds to the list of RHBDL2 substrates that are involved in adhesion. In fact, as mentioned in the previous chapter, it has been shown that RHBDL2 also cleaves other members of the cadherin family (Battistini et al., 2019; Grieve et al., 2021b) and that inhibition of RHBDL2 expression has an effect on cell motility. Although further experiments are needed to corroborate this data, it is intriguing to speculate an emerging theme in RHBDL2 substrate repertoire that could hint at a potential role of the rhomboid protease in the regulation of cell adhesion.

The long troubleshooting that was done over the years laid the bases that this methodology is a viable way of looking for substrates as it can discriminate between substrate and non-substrate TMDs. However, the low success rate in identifying substrates and the absence of any known substrate in the list of hits suggests that there are a few improvements that could be implemented. First of all, the main problem of the dataset we have generated is the noisiness due to a variability

among the replicates, this probably has led to both false negative and false positive. The most logical solution to decrease the background noise would be to include more replicates of the screen, and to increase the number of reads that will be analysed. In order to do this, Illumina sequencing could be used to have a higher number of reads per TMD and to increase the depth of the sequencing analysis. Reducing the noisiness could also help in simplifying the statistical analysis. As there was no clear strategy that could be used, different pipelines were trialled before settling on calculating the average fluorescence of each TMDs. However, this does not mean that this the only correct way of analysing the data, and it would be interesting to re-visit some of these strategies with an improved dataset. For example, an alternative strategy was to calculate the read count of each TMD in the bins and to assess whether it changed upon co-expression of RHBDL2. Given the noisiness of the data, this was not used as there was no clear distribution of the TMDs in one or more bins, and that is why in the end the average fluorescence was calculated. But increasing the read count could help clarify the distribution of the TMDs in the bins.

Another potential limitation of the current setup is the use of overexpressed RHBDL2. In fact, the efficiency of the transfection could vary significantly among the replicates which could have an effect the level of cleavage. This was kept in consideration and a similar number of cells that expressed RHBDL2 was sorted among all three replicates. However, overexpressing RHBDL2 potentially introduced some variability that could decrease the efficiency of the screen. An obvious way of solving this is to stably express both protease and TMDs. As I have described, I attempted this solution by using the Flp-In system to generate cells expressing either tBFP RHDBL2 or RHBDL2 SA. Unfortunately, this approach was not efficient enough to be pursued, however inducible and stable expression of RHBDL2 could still be obtained by using another viral vector. A perhaps more sophisticated way of regulating RHBDL2 expression could be to use a bicistronic vector coding for both the protease and the TMDs. By using this approach, not only RHBDL2 could also be stably integrated, but each cell that expresses the protease would also express the potential substrate TMD, avoiding the possibility of a false negative result due to cells not being transfected with

RHBDL2. Although the bicistronic vector I have used at the beginning of this project did not work when used to generate stable cells, there are multiple alternatives that are available that use different strategies to produce multiple proteins from the same mRNA. For example, by using a self-cleaving 2A peptide (Ryan et al., 1991). Overall, the method developed in this study has shown its effectiveness as it led to the identification of two new substrates of RHBDL2, however there are still limitations that need to be addressed to improve its efficiency.

My data shows, for the first time, that a decrease in fluorescence can efficiently be used as a readout for rhomboid cleavage. This can prove to be useful not only for identifying substrates, but also in other type of screening, for example when looking for compounds that inhibit the catalytic activity of rhomboid proteases. Currently, there are a few molecules that have shown an efficient inhibitory effect on rhomboid proteases, such β-lactams and peptidyl ketoamides (Tichá, Stanchev, Vinothkumar, et al., 2017; Yang et al., 2023). Efficient screening of rhomboid inhibitors is performed by using a purified rhomboid protease and a substrate TMD tagged with two Fluorescence resonance energy transfer (FRET) probes. The pair is incubated with small molecules and readings of the FRET signal are used to identify inhibiting molecules (Tichá, Stanchev, Škerle, et al., 2017). Most of the inhibitory screens have tested bacterial rhomboids, potentially because of the need to purify the protease. With some necessary changes, the assay I have developed could potentially be adapted for testing a large number of small molecules for RHDBL2 without the need for purification. In fact, by using cells that stably expressing the substrate TMD, the rhomboid protease could be transfected and incubated with the library of compounds. By assessing whether there is any change in the fluorescence of the substrate, it could be possible to identify molecules that affect RHBDL2 catalytic activity.

In conclusion, in this thesis I have described the development of a novel assay that, despite the need for some improvements, has led to an advance in the discovery of rhomboid proteases.

6.2 VE-Cadherin

In the last few months of my DPhil, I have also started to investigate RHBDL2 cleavage of VE-Cadherin. My results confirmed the published data that both the TMD and the full-length protein are cleaved by RHBDL2 (Battistini et al., 2019). However, mutating the luminal half of the TMD (either with 3aa mutated at the same time or with single mutations) and the juxtamembrane region failed at mapping the cleavage site. This was a puzzling result as the experiments were done using a stripped back version of VE-Cadherin that only contained the transmembrane region and a few amino acids at the luminal and cytosolic region. VE-Cadherin does not have any helix-destabilising residues within its transmembrane region. According to the current model for substrate recognition (Strisovsky et al., 2009), the absence of a kink in the TMD suggests that the cleavage occurs in the juxtamembrane region rather than within the TMD itself. Considering this, a potential explanation for not being able to identify the cleavage site could be that the mutations used for the juxtamembrane luminal region were too conservative and did not disrupt the cleavage site enough. In the future, it could be interesting to mutate this region further and assess whether the cleavage is abrogated. Another aspect that it would be worth investigating further is to determine which regions of VE-Cadherin regulate the cleavage. The experiments presented in this work show that the TMD alone is enough to trigger RHBDL2 processing, however this does not exclude that other regions of the protein could have a role in the regulation of the cleavage or the interaction. In fact, there are instances where the cytosolic region of the substrate can have an effect on the binding with RHBDL2. As an example, the cytosolic domain of Orai1 is not necessary for the cleavage event, but deletion of the domain and deletion in the cytoplasmic tail of RHBDL2 have been show increase the interaction between the two proteins (Grieve et al., 2021a). In the case of thrombomodulin, the role of the cytoplasmic domain is even more important as it is the main driver of RHBDL2 cleavage (Lohi et al., 2004). To interrogate this aspect in studying VE-Cadherin, it would be interesting to swap VE-Cadherin domains with the corresponding region of a

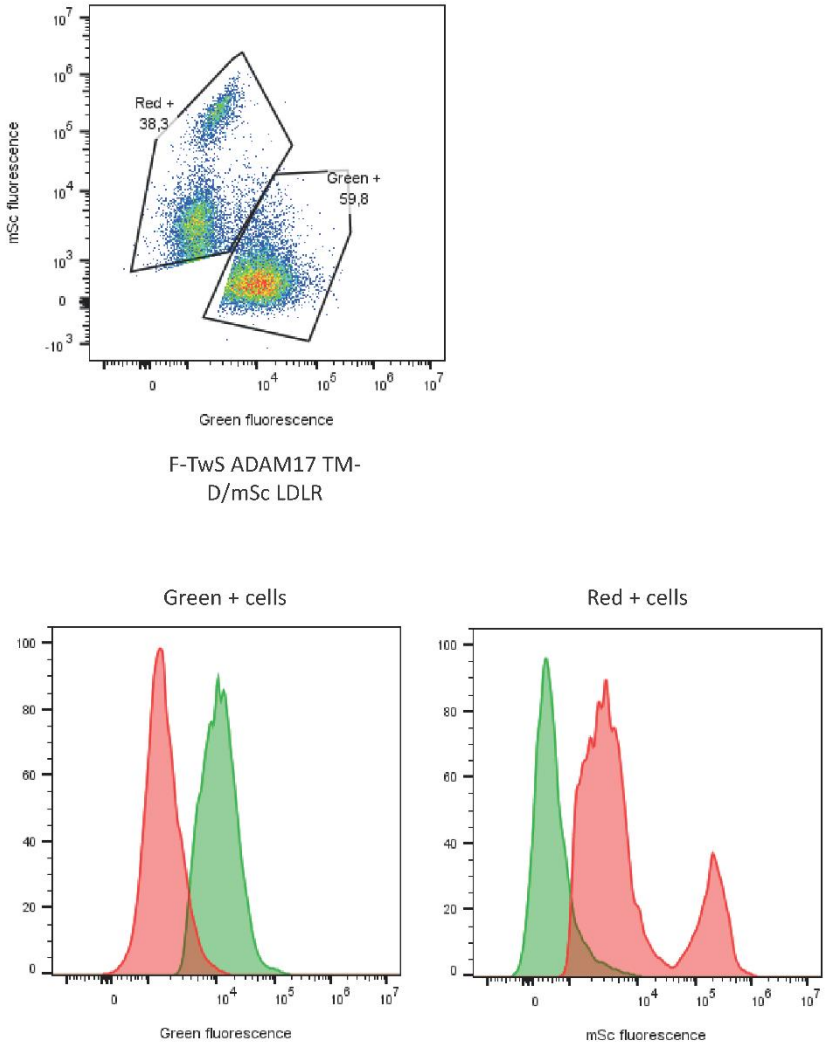
non-substrate (such as Cadherin10) to evaluate the effect of these chimeras on the cleavage. These experiments will help in clarifying the relationship between VE-Cadherin and RHBDL2.

More generally, the physiological role of RHBDL2 cleavage of VE-Cadherin is still uncharacterised. Given the importance of VE-Cadherin in the formation of new blood vessels and its role in the formation of metastasis, it is tempting to investigate RHBDL2 processing in this context. A potential role for RHBDL2 in angiogenesis had been suggested through multiple studies that have identified important proteins involved in this process, such thrombomodulin and CLEC14A, to be cleaved by RHBDL2 (Cheng et al., 2011; Noy et al., 2016). The authors of both studies have found that inhibition of RHBDL2 caused a decrease of the cleavage of thrombomodulin and CLEC14A, and it was shown that the knockdown of RHBDL2 caused a slower migration of cells in wound healing assay of keratinocytes (Cheng et al., 2011). A similar effect on cell motility was found when knocking down RHBDL2 in prostate cancer cells (Battistini et al., 2019). The authors showed that RHBDL2 cleaves E-Cadherin and causes its shedding. Moreover, overexpression of RHBDL2 caused an increase in the shedding of E-Cadherin and an increase in cell motility. Conversely, knocking down the protease slowed down cell migration. It would be interesting to investigate further these effects on cell migration to assess if the same phenomenon can be observed in endothelial cells, and if there is any cellular trigger that stimulates the cleavage. Interestingly, Battistini et al. found that TNF-α stimulated the expression of RHBDL2 in cancer and endothelial cells suggesting a potential role of the cytokine in the regulation of the protease expression level.

Overall, the work about VE-Cadherin presented in this thesis laid the foundation for the characterisation of the relationship between RHBDL2 and VE-Cadherin.

Appendices

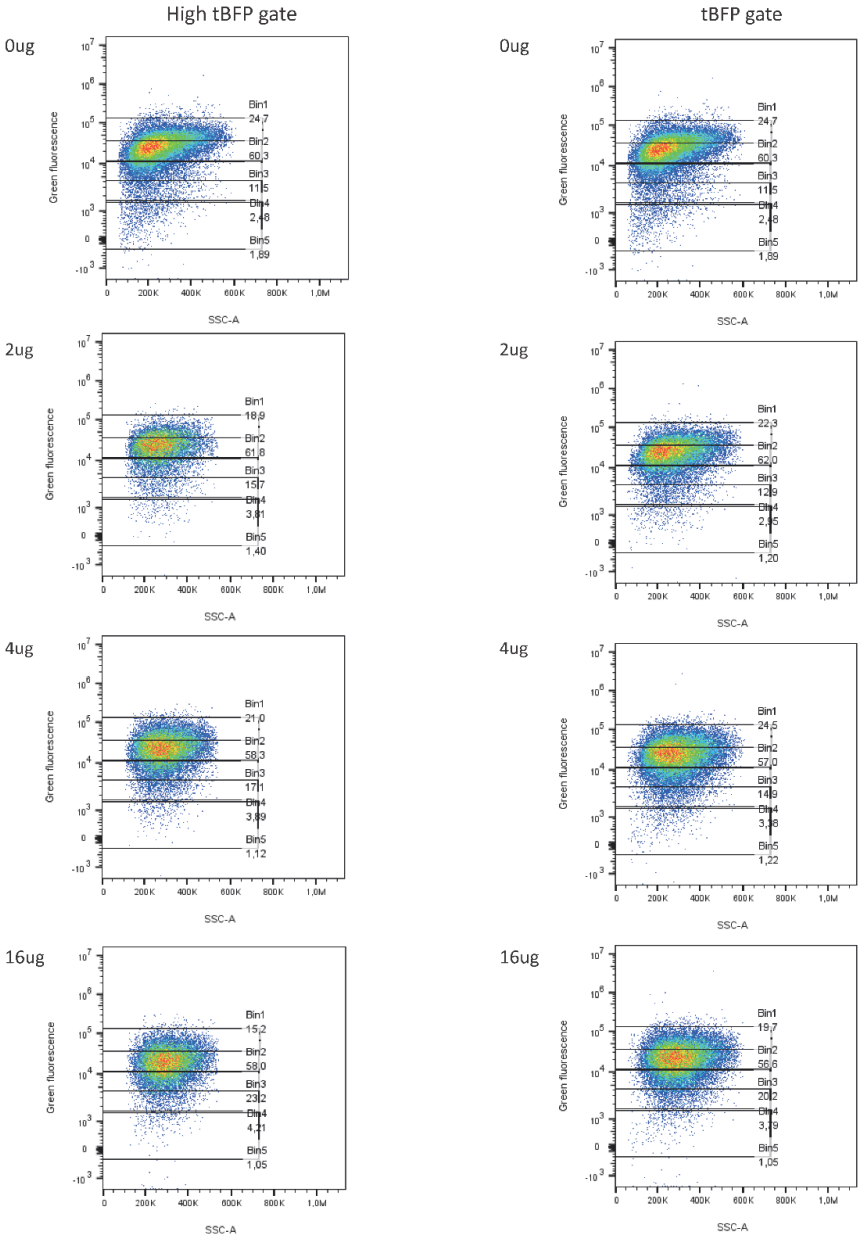
Appendix 1



Appendix 1 Stably integrated F-TwS ADAM17/mSc LDLR does not express both proteins at the same time. The pseudocolour graph of cells expressing F-TwS A17/mSc LDLR shows the formation of two distinct populations of green and red+ cells. The histograms below show how there is no overlap between the two populations meaning that the cells do not express both proteins at the same time.
Appendix refers to Figure 3.8.

Appendix 2

F-TwS ADAM17
tBFP RHBDL2

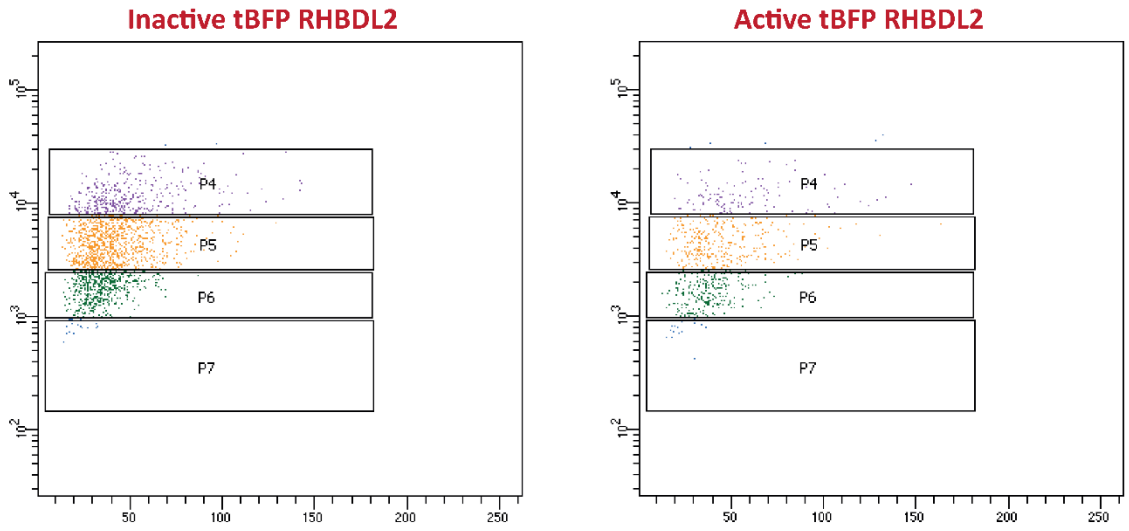


Appendix 2 Titrating RHBDL2 does not impact its specificity Pseudocolour graphs showing transfection of either 0,2,4,16µg of tBFP RHBDL2 on F-TwS ADAM17 expressing cells. Increasing the amount of RHBDL2 that is overexpressed does not have an impact on the selectivity of the protease as it does not cleave ADAM17 TMD.

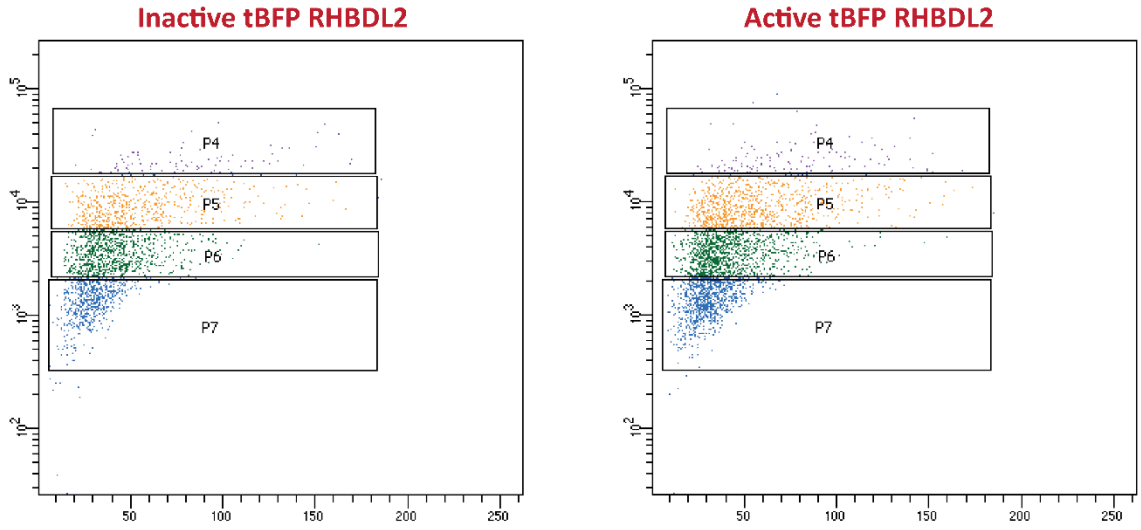
Appendix refers to Figure 3.14

Appendix 3

Replicate 1



Replicate 2



Appendix 3 Graph of the two additional replicates of the TMD screen *Appendix refers to Figure 4.3.*

Appendix 4

List of hits from TMD screen

Entry	p value	Reviewed	Entry Name
Q9NPA0	0.004	reviewed	EMC7_HUMAN
P15391	0.011	reviewed	CD19_HUMAN
O75325	0.011	reviewed	LRRN2_HUMAN
Q9BTN0	0.019	reviewed	LRFN3_HUMAN
Q6NT55	0.021	reviewed	CP4FN_HUMAN
Q9Y5G2	0.036	reviewed	PCDGE_HUMAN
Q76MJ5	0.049	reviewed	ERN2_HUMAN
P29322	0.063	reviewed	EPHA8_HUMAN
Q86XX4	0.065	reviewed	FRAS1_HUMAN
Q86VH5	0.067	reviewed	LRRT3_HUMAN
O14967	0.067	reviewed	CLGN_HUMAN
Q9Y5H9	0.071	reviewed	PCDA2_HUMAN
P59901	0.075	reviewed	LIRA4_HUMAN
Q9H1E5	0.075	reviewed	TMX4_HUMAN
P34910	0.079	reviewed	EVI2B_HUMAN
Q6UXG2	0.08	reviewed	ELAP1_HUMAN
P35968	0.085	reviewed	VGFR2_HUMAN
P50281	0.086	reviewed	MMP14_HUMAN
Q8TAB3	0.093	reviewed	PCD19_HUMAN
O15455	0.095	reviewed	TLR3_HUMAN
P50895	0.1	reviewed	BCAM_HUMAN
Q9ULL4	0.109	reviewed	PLXB3_HUMAN
Q6UXD5	0.109	reviewed	SE6L2_HUMAN
Q9NTN9	0.11	reviewed	SEM4G_HUMAN
Q8NFY4	0.112	reviewed	SEM6D_HUMAN
Q6P7N7	0.115	reviewed	TMM81_HUMAN
P06340	0.118	reviewed	DOA_HUMAN
P20701	0.118	reviewed	ITAL_HUMAN
Q8IX05	0.123	reviewed	CD302_HUMAN
Q3KNT9	0.124	reviewed	TMM95_HUMAN
P19022	0.125	reviewed	CADH2_HUMAN
Q96LA6	0.125	reviewed	FCRL1_HUMAN
Q6UWW9	0.126	reviewed	TM207_HUMAN
O95866	0.129	reviewed	G6B_HUMAN
P39656	0.13	reviewed	OST48_HUMAN
Q6EMK4	0.13	reviewed	VASN_HUMAN
O94985	0.13	reviewed	CSTN1_HUMAN
Q9H3R2	0.13	reviewed	MUC13_HUMAN
Q14118	0.133	reviewed	DAG1_HUMAN
Q5T1S8	0.135	reviewed	NCMAP_HUMAN
Q9ULB4	0.138	reviewed	CADH9_HUMAN
Q9P2S2	0.147	reviewed	NRX2A_HUMAN

O75051	0.147	reviewed	PLXA2_HUMAN
P21754	0.148	reviewed	ZP3_HUMAN
Q9Y5G7	0.15	reviewed	PCDG6_HUMAN
Q9P2J2	0.151	reviewed	TUTLA_HUMAN
Q16832	0.151	reviewed	DDR2_HUMAN
Q9H3N1	0.154	reviewed	TMX1_HUMAN
Q4KMG9	0.155	reviewed	TM52B_HUMAN
Q86WK7	0.155	reviewed	AMGO3_HUMAN
Q8WXI7	0.156	reviewed	MUC16_HUMAN
Q9BZV3	0.156	reviewed	IMPG2_HUMAN
P09619	0.157	reviewed	PGFRB_HUMAN
P21583	0.157	reviewed	SCF_HUMAN
Q8TF66	0.158	reviewed	LRC15_HUMAN
P22794	0.159	reviewed	EVI2A_HUMAN
P21860	0.162	reviewed	ERBB3_HUMAN
Q8NI17	0.167	reviewed	IL31R_HUMAN
Q9Y5H0	0.172	reviewed	PCDG3_HUMAN
Q9Y279	0.173	reviewed	VSIG4_HUMAN
Q9H665	0.173	reviewed	IGFR1_HUMAN
P24394	0.173	reviewed	IL4RA_HUMAN
Q9BUF7	0.178	reviewed	CRUM3_HUMAN
P56975	0.178	reviewed	NRG3_HUMAN
Q9Y6W8	0.179	reviewed	ICOS_HUMAN
P43121	0.187	reviewed	MUC18_HUMAN
Q6UX41	0.188	reviewed	BTNL8_HUMAN
Q6ZRH7	0.188	reviewed	CTSRG_HUMAN
Q9Y5H5	0.19	reviewed	PCDA9_HUMAN
O60939	0.19	reviewed	SCN2B_HUMAN
Q8NFT8	0.192	reviewed	DNER_HUMAN
Q96BY9	0.192	reviewed	SARAF_HUMAN
Q9BY67	0.193	reviewed	CADM1_HUMAN
P07204	0.195	reviewed	TRBM_HUMAN
B4DS77	0.196	reviewed	SHSA9_HUMAN
W5XKT8	0.196	reviewed	SACA6_HUMAN
P17948	0.2	reviewed	VGFR1_HUMAN
Q13873	0.201	reviewed	BMPR2_HUMAN
A6NFU0	0.201	reviewed	F187A_HUMAN
Q9HBJ8	0.201	reviewed	CLTRN_HUMAN
O75022	0.201	reviewed	LIRB3_HUMAN
P27930	0.201	reviewed	IL1R2_HUMAN
A6NL88	0.202	reviewed	SHSA7_HUMAN
Q9Y5E4	0.203	reviewed	PCDB5_HUMAN
Q8N386	0.204	reviewed	LRC25_HUMAN
Q16819	0.208	reviewed	MEP1A_HUMAN
A0A1B0GUA7	0.209	reviewed	TEX51_HUMAN
P20702	0.209	reviewed	ITAX_HUMAN
O60487	0.211	reviewed	MPZL2_HUMAN

Q5DID0	0.213	reviewed	UROL1_HUMAN
P53801	0.213	reviewed	PTTG_HUMAN
Q96J42	0.214	reviewed	TXD15_HUMAN
P55287	0.214	reviewed	CAD11_HUMAN
Q6UX71	0.217	reviewed	PXDC2_HUMAN
Q9Y5Z0	0.217	reviewed	BACE2_HUMAN
Q5VUB5	0.218	reviewed	F1711_HUMAN
Q96KC8	0.219	reviewed	DNJC1_HUMAN
Q9UHC6	0.221	reviewed	CNTP2_HUMAN
P54764	0.222	reviewed	EPHA4_HUMAN
Q9Y2C9	0.223	reviewed	TLR6_HUMAN
Q9BZ11	0.224	reviewed	ADA33_HUMAN
A0A1B0GTY4	0.226	reviewed	TEX50_HUMAN
P10586	0.227	reviewed	PTPRF_HUMAN
P19021	0.227	reviewed	AMD_HUMAN
Q9HCU0	0.228	reviewed	CD248_HUMAN
P07949	0.23	reviewed	RET_HUMAN
Q9Y6N8	0.23	reviewed	CAD10_HUMAN
P22607	0.23	reviewed	FGFR3_HUMAN
Q9UIB8	0.234	reviewed	SLAF5_HUMAN
Q8TDF5	0.234	reviewed	NETO1_HUMAN
P48551	0.235	reviewed	INAR2_HUMAN
Q5T292	0.235	reviewed	TM273_HUMAN
Q8NEA5	0.237	reviewed	CS018_HUMAN
Q96D42	0.238	reviewed	HAVR1_HUMAN
Q5SQ64	0.238	reviewed	LY66F_HUMAN
Q496F6	0.239	reviewed	CLM2_HUMAN
Q9Y5H6	0.241	reviewed	PCDA8_HUMAN
Q9UBG0	0.241	reviewed	MRC2_HUMAN
Q5VT99	0.241	reviewed	LRC38_HUMAN
Q96PZ7	0.242	reviewed	CSMD1_HUMAN
Q15762	0.243	reviewed	CD226_HUMAN
Q9BRK3	0.243	reviewed	MXRA8_HUMAN
Q96RD9	0.246	reviewed	FCRL5_HUMAN
P40198	0.246	reviewed	CEAM3_HUMAN
O60462	0.247	reviewed	NRP2_HUMAN
P15813	0.248	reviewed	CD1D_HUMAN
Q6UW88	0.248	reviewed	EPGN_HUMAN
P20333	0.249	reviewed	TNR1B_HUMAN
Q9BZA7	0.25	reviewed	PC11X_HUMAN
Q9NR16	0.251	reviewed	C163B_HUMAN
Q6UXG3	0.252	reviewed	CLM9_HUMAN
Q6NUS8	0.255	reviewed	UD3A1_HUMAN
Q16620	0.255	reviewed	NTRK2_HUMAN
Q9Y5H4	0.255	reviewed	PCDG1_HUMAN
Q68D85	0.255	reviewed	NR3L1_HUMAN
Q13683	0.256	reviewed	ITA7_HUMAN

Q9P2E7	0.257	reviewed	PCD10_HUMAN
Q9NYJ7	0.258	reviewed	DLL3_HUMAN
Q9Y5H1	0.259	reviewed	PCDG2_HUMAN
Q14392	0.259	reviewed	LRC32_HUMAN
P16066	0.26	reviewed	ANPRA_HUMAN
P54289	0.261	reviewed	CA2D1_HUMAN
Q9H9K5	0.261	reviewed	MER34_HUMAN
Q9NS68	0.262	reviewed	TNR19_HUMAN
O14786	0.263	reviewed	NRP1_HUMAN
Q9Y3B3	0.264	reviewed	TMED7_HUMAN
Q9Y5E6	0.265	reviewed	PCDB3_HUMAN
P49771	0.266	reviewed	FLT3L_HUMAN
Q9Y5F8	0.266	reviewed	PCDGJ_HUMAN
O60279	0.267	reviewed	SUSD5_HUMAN
Q96LC7	0.271	reviewed	SIG10_HUMAN
Q8TBQ9	0.271	reviewed	KISHA_HUMAN
P10912	0.272	reviewed	GHR_HUMAN
Q495A1	0.272	reviewed	TIGIT_HUMAN
Q9H7M9	0.273	reviewed	VISTA_HUMAN
P09848	0.274	reviewed	LPH_HUMAN
Q8NBR0	0.275	reviewed	P5I13_HUMAN
P0DPA2	0.276	reviewed	VSIG8_HUMAN
Q19T08	0.276	reviewed	ECSCR_HUMAN
Q96NY8	0.276	reviewed	NECT4_HUMAN
A8MVW5	0.28	reviewed	HECA2_HUMAN
O95727	0.281	reviewed	CRTAM_HUMAN
Q92896	0.282	reviewed	GSLG1_HUMAN
P17693	0.282	reviewed	HLAG_HUMAN
O95944	0.283	reviewed	NCTR2_HUMAN
P14151	0.283	reviewed	LYAM1_HUMAN
Q9UIK5	0.284	reviewed	TEFF2_HUMAN
Q16790	0.285	reviewed	CAH9_HUMAN
Q9NRX6	0.287	reviewed	KISHB_HUMAN
P55283	0.288	reviewed	CADH4_HUMAN
Q8IVU1	0.291	reviewed	IGDC3_HUMAN
P43630	0.291	reviewed	KI3L2_HUMAN
P08F94	0.292	reviewed	PKHD1_HUMAN
Q6ZMC9	0.293	reviewed	SIG15_HUMAN
P02724	0.293	reviewed	GLPA_HUMAN
Q8IW00	0.294	reviewed	VSTM4_HUMAN
O60245	0.294	reviewed	PCDH7_HUMAN
Q5JRM2	0.294	reviewed	CX066_HUMAN
P06213	0.295	reviewed	INSR_HUMAN
Q92542	0.295	reviewed	NICA_HUMAN
P31431	0.295	reviewed	SDC4_HUMAN
Q01151	0.296	reviewed	CD83_HUMAN
Q9Y5G8	0.296	reviewed	PCDG5_HUMAN

Q6UXL0	0.297	reviewed	I20RB_HUMAN
Q96MU8	0.297	reviewed	KREM1_HUMAN
Q9UN72	0.298	reviewed	PCDA7_HUMAN
Q9Y6N7	0.299	reviewed	ROBO1_HUMAN
Q8WW62	0.301	reviewed	TMED6_HUMAN
Q96A28	0.302	reviewed	SLAF9_HUMAN
Q8N423	0.304	reviewed	LIRB2_HUMAN
P11717	0.304	reviewed	MPRI_HUMAN
Q9Y5I0	0.304	reviewed	PCDAD_HUMAN
P05556	0.305	reviewed	ITB1_HUMAN
P78504	0.308	reviewed	JAG1_HUMAN
P10747	0.309	reviewed	CD28_HUMAN
Q6UXZ3	0.311	reviewed	CLM5_HUMAN
O75309	0.311	reviewed	CAD16_HUMAN
Q9NY72	0.312	reviewed	SCN3B_HUMAN
O14522	0.312	reviewed	PTPRT_HUMAN
O75631	0.313	reviewed	UPK3A_HUMAN
P14770	0.313	reviewed	GPIX_HUMAN
Q9NYK1	0.314	reviewed	TLR7_HUMAN
P01909	0.314	reviewed	DQA1_HUMAN
P01589	0.314	reviewed	IL2RA_HUMAN
O75197	0.314	reviewed	LRP5_HUMAN
Q12907	0.314	reviewed	LMAN2_HUMAN
A8MXK1	0.315	reviewed	VSTM5_HUMAN
Q9NZV1	0.315	reviewed	CRIM1_HUMAN
Q9BQ51	0.316	reviewed	PD1L2_HUMAN
O75054	0.317	reviewed	IGSF3_HUMAN
Q53EL9	0.318	reviewed	SEZ6_HUMAN
Q7KYR7	0.318	reviewed	BT2A1_HUMAN
Q96QE4	0.319	reviewed	LR37B_HUMAN
O75871	0.319	reviewed	CEAM4_HUMAN
Q8N114	0.321	reviewed	SHSA5_HUMAN
Q9BZZ2	0.321	reviewed	SN_HUMAN
P20023	0.322	reviewed	CR2_HUMAN
P21802	0.322	reviewed	FGFR2_HUMAN
P46531	0.326	reviewed	NOTC1_HUMAN
Q9NP99	0.33	reviewed	TREM1_HUMAN
P82279	0.33	reviewed	CRUM1_HUMAN
P58401	0.33	reviewed	NRX2B_HUMAN
Q99466	0.332	reviewed	NOTC4_HUMAN
Q13349	0.332	reviewed	ITAD_HUMAN
Q99467	0.333	reviewed	CD180_HUMAN
Q86VH4	0.333	reviewed	LRRT4_HUMAN
Q96PJ5	0.336	reviewed	FCRL4_HUMAN
Q9UGT4	0.337	reviewed	SUSD2_HUMAN
P78552	0.337	reviewed	I13R1_HUMAN
Q9HBB8	0.338	reviewed	CDHR5_HUMAN

Q86YT9	0.338	reviewed	JAML_HUMAN
Q9UPU3	0.339	reviewed	SORC3_HUMAN
P13726	0.34	reviewed	TF_HUMAN
P55286	0.341	reviewed	CADH8_HUMAN
Q8NFZ8	0.341	reviewed	CADM4_HUMAN
P09958	0.341	reviewed	FURIN_HUMAN
Q6UY09	0.341	reviewed	CEA20_HUMAN
Q9Y5F2	0.341	reviewed	PCDBB_HUMAN
Q02846	0.342	reviewed	GUC2D_HUMAN
Q8IW52	0.343	reviewed	SLIK4_HUMAN
Q8IYS2	0.343	reviewed	K2013_HUMAN
P08648	0.343	reviewed	ITA5_HUMAN
Q3MIW9	0.343	reviewed	MUCL3_HUMAN
P43629	0.344	reviewed	KI3L1_HUMAN
P11215	0.344	reviewed	ITAM_HUMAN
Q9NPH3	0.345	reviewed	IL1AP_HUMAN
Q9Y286	0.346	reviewed	SIGL7_HUMAN
P51805	0.346	reviewed	PLXA3_HUMAN
Q04771	0.346	reviewed	ACVR1_HUMAN
Q12866	0.348	reviewed	MERTK_HUMAN
Q9NR96	0.349	reviewed	TLR9_HUMAN
Q9C0C4	0.349	reviewed	SEM4C_HUMAN
P01911	0.351	reviewed	DRB1_HUMAN
P29376	0.351	reviewed	LTK_HUMAN
Q86VB7	0.352	reviewed	C163A_HUMAN
Q96LA5	0.352	reviewed	FCRL2_HUMAN
Q14954	0.352	reviewed	KI2S1_HUMAN
Q06481	0.353	reviewed	APLP2_HUMAN
P07202	0.353	reviewed	PERT_HUMAN
P37173	0.353	reviewed	TGFR2_HUMAN
Q9H159	0.354	reviewed	CAD19_HUMAN
Q902F9	0.354	reviewed	EN113_HUMAN
Q9Y6J6	0.354	reviewed	KCNE2_HUMAN
O75074	0.355	reviewed	LRP3_HUMAN
P16471	0.356	reviewed	PRLR_HUMAN
Q96F05	0.357	reviewed	CK024_HUMAN
I3L273	0.357	reviewed	GFY_HUMAN
Q8NCW0	0.358	reviewed	KREM2_HUMAN
P06729	0.359	reviewed	CD2_HUMAN
Q86WC4	0.359	reviewed	OSTM1_HUMAN
Q16820	0.36	reviewed	MEP1B_HUMAN
Q13308	0.36	reviewed	PTK7_HUMAN
Q8J025	0.361	reviewed	APCD1_HUMAN
P51841	0.361	reviewed	GUC2F_HUMAN
A6NM11	0.362	reviewed	L37A2_HUMAN
Q6MZM0	0.362	reviewed	HPHL1_HUMAN
Q6IEE7	0.364	reviewed	T132E_HUMAN

P07333	0.364	reviewed	CSF1R_HUMAN
O43184	0.364	reviewed	ADA12_HUMAN
Q92956	0.367	reviewed	TNR14_HUMAN
Q9UMZ3	0.367	reviewed	PTPRQ_HUMAN
P40126	0.368	reviewed	TYRP2_HUMAN
O00533	0.368	reviewed	NCHL1_HUMAN
Q9Y5G0	0.368	reviewed	PCDGH_HUMAN
Q9H3S1	0.37	reviewed	SEM4A_HUMAN
Q9Y5F9	0.37	reviewed	PCDGI_HUMAN
Q4V9L6	0.371	reviewed	TM119_HUMAN
P08069	0.371	reviewed	IGF1R_HUMAN
Q14626	0.371	reviewed	I11RA_HUMAN
Q14162	0.372	reviewed	SREC_HUMAN
P09693	0.372	reviewed	CD3G_HUMAN
P38484	0.372	reviewed	INGR2_HUMAN
P20273	0.374	reviewed	CD22_HUMAN
Q9NZ53	0.375	reviewed	PDXL2_HUMAN
Q6Q8B3	0.375	reviewed	MO2R2_HUMAN
Q8N3J6	0.376	reviewed	CADM2_HUMAN
P18084	0.377	reviewed	ITB5_HUMAN
Q02297	0.377	reviewed	NRG1_HUMAN
Q9H1U4	0.378	reviewed	MEGF9_HUMAN
Q4G0I0	0.38	reviewed	UQCC4_HUMAN
P04626	0.381	reviewed	ERBB2_HUMAN
P51571	0.382	reviewed	SSRD_HUMAN
P79483	0.382	reviewed	DRB3_HUMAN
Q01973	0.383	reviewed	ROR1_HUMAN
P09564	0.383	reviewed	CD7_HUMAN
Q92823	0.383	reviewed	NRCAM_HUMAN
Q9C0I4	0.384	reviewed	THS7B_HUMAN
P07359	0.384	reviewed	GP1BA_HUMAN
P29016	0.384	reviewed	CD1B_HUMAN
Q8WY21	0.384	reviewed	SORC1_HUMAN
Q9Y3P8	0.384	reviewed	SIT1_HUMAN
O60895	0.385	reviewed	RAMP2_HUMAN
P31994	0.385	reviewed	FCG2B_HUMAN
Q9Y5F7	0.386	reviewed	PCDGL_HUMAN
Q9Y5Y7	0.387	reviewed	LYVE1_HUMAN
O43493	0.388	reviewed	TGON2_HUMAN
P29320	0.388	reviewed	EPHA3_HUMAN
P12830	0.388	reviewed	CADH1_HUMAN
P43631	0.388	reviewed	KI2S2_HUMAN
O75144	0.389	reviewed	ICOSL_HUMAN
Q9BZA8	0.389	reviewed	PC11Y_HUMAN
Q14C87	0.39	reviewed	T132D_HUMAN
O14763	0.391	reviewed	TR10B_HUMAN
Q08345	0.391	reviewed	DDR1_HUMAN

P07766	0.391	reviewed	CD3E_HUMAN
Q86T13	0.391	reviewed	CLC14_HUMAN
Q86XR5	0.391	reviewed	PRIMA_HUMAN
P55808	0.392	reviewed	XG_HUMAN
P08581	0.392	reviewed	MET_HUMAN
Q9Y5H2	0.393	reviewed	PCDGB_HUMAN
P23471	0.393	reviewed	PTPRZ_HUMAN
Q01113	0.394	reviewed	IL9R_HUMAN
Q15109	0.394	reviewed	RAGE_HUMAN
Q9NS62	0.395	reviewed	THSD1_HUMAN
Q9Y5I2	0.396	reviewed	PCDAA_HUMAN
Q3MIP1	0.397	reviewed	IPIL2_HUMAN
Q9NP60	0.398	reviewed	IRPL2_HUMAN
Q5SGD2	0.399	reviewed	PPM1L_HUMAN
Q9P0V8	0.4	reviewed	SLAF8_HUMAN
E5RIL1	0.4	reviewed	UPKL2_HUMAN
Q9UKF5	0.4	reviewed	ADA29_HUMAN
Q6PI73	0.4	reviewed	LIRA6_HUMAN
P06756	0.4	reviewed	ITAV_HUMAN
Q01974	0.401	reviewed	ROR2_HUMAN
Q9NPG4	0.402	reviewed	PCD12_HUMAN
P43146	0.402	reviewed	DCC_HUMAN
Q8TCZ2	0.402	reviewed	C99L2_HUMAN
P18827	0.402	reviewed	SDC1_HUMAN
Q86SJ6	0.403	reviewed	DSG4_HUMAN
Q86VR7	0.404	reviewed	VS10L_HUMAN
P06028	0.404	reviewed	GLPB_HUMAN
Q8NEW7	0.405	reviewed	TMIE_HUMAN
Q9UQV4	0.405	reviewed	LAMP3_HUMAN
A0A087WTH5	0.405		KCE1B_HUMAN
P0DKB5	0.406	reviewed	TPBGL_HUMAN
P28067	0.406	reviewed	DMA_HUMAN
Q9HCJ2	0.406	reviewed	LRC4C_HUMAN
O43914	0.408	reviewed	TYOBP_HUMAN
Q9H756	0.408	reviewed	LRC19_HUMAN
P49257	0.408	reviewed	LMAN1_HUMAN
Q13332	0.409	reviewed	PTPRS_HUMAN
Q96GX1	0.409	reviewed	TECT2_HUMAN
Q8TDQ1	0.409	reviewed	CLM1_HUMAN
P78310	0.409	reviewed	CXAR_HUMAN
Q9NY47	0.41	reviewed	CA2D2_HUMAN
Q13705	0.411	reviewed	AVR2B_HUMAN
O94856	0.411	reviewed	NFASC_HUMAN
O14944	0.412	reviewed	EREG_HUMAN
Q9H0V9	0.412	reviewed	LMA2L_HUMAN
Q9NP84	0.413	reviewed	TNR12_HUMAN
O00526	0.413	reviewed	UPK2_HUMAN

O94991	0.415	reviewed	SLIK5_HUMAN
P57087	0.415	reviewed	JAM2_HUMAN
Q8N0Z9	0.416	reviewed	VSI10_HUMAN
Q9UF33	0.416	reviewed	EPHA6_HUMAN
A6NF34	0.416	reviewed	ANTRL_HUMAN
Q8TBF5	0.416	reviewed	PIGX_HUMAN
Q9P2V4	0.416	reviewed	LRIT1_HUMAN
O75056	0.417	reviewed	SDC3_HUMAN
P30530	0.417	reviewed	UFO_HUMAN
O75121	0.417	reviewed	MFA3L_HUMAN
Q8N109	0.418	reviewed	KI2LA_HUMAN
P09758	0.418	reviewed	TACD2_HUMAN
P32927	0.42	reviewed	IL3RB_HUMAN
Q8N126	0.421	reviewed	CADM3_HUMAN
P16070	0.421	reviewed	CD44_HUMAN
Q5VYJ5	0.422	reviewed	MALR1_HUMAN
P01133	0.422	reviewed	EGF_HUMAN
Q5TFQ8	0.423	reviewed	SIRBL_HUMAN
Q6UVK1	0.423	reviewed	CSPG4_HUMAN
P55899	0.423	reviewed	FCGRN_HUMAN
P04440	0.423	reviewed	DPB1_HUMAN
P40200	0.423	reviewed	TACT_HUMAN
Q9UKX5	0.424	reviewed	ITA11_HUMAN
P0DTE4	0.425	reviewed	UD2A1_HUMAN
O00238	0.425	reviewed	BMR1B_HUMAN
O60309	0.425	reviewed	L37A3_HUMAN
P30203	0.426	reviewed	CD6_HUMAN
P56817	0.426	reviewed	BACE1_HUMAN
Q13797	0.426	reviewed	ITA9_HUMAN
Q15262	0.426	reviewed	PTPRK_HUMAN
Q14517	0.427	reviewed	FAT1_HUMAN
P29317	0.428	reviewed	EPHA2_HUMAN
O75023	0.428	reviewed	LIRB5_HUMAN
P43308	0.429	reviewed	SSRB_HUMAN
Q92643	0.429	reviewed	GPI8_HUMAN
P78423	0.43	reviewed	X3CL1_HUMAN
Q9UJQ1	0.431	reviewed	LAMP5_HUMAN
O94898	0.431	reviewed	LRIG2_HUMAN
O95185	0.432	reviewed	UNC5C_HUMAN
P51693	0.432	reviewed	APLP1_HUMAN
Q9UQF0	0.433	reviewed	SYCY1_HUMAN
Q6UVY6	0.433	reviewed	MOXD1_HUMAN
P98155	0.434	reviewed	VLDLR_HUMAN
B0FP48	0.435	reviewed	UPK3L_HUMAN
P25445	0.436	reviewed	TNR6_HUMAN
Q9Y5H3	0.436	reviewed	PCDGA_HUMAN
P78325	0.437	reviewed	ADAM8_HUMAN

Q6UXC1	0.437	reviewed	AEGP_HUMAN
Q86UP0	0.437	reviewed	CAD24_HUMAN
B6A8C7	0.437	reviewed	TARM1_HUMAN
Q9NRJ7	0.438	reviewed	PCDBG_HUMAN
Q9UNN8	0.438	reviewed	EPCR_HUMAN
Q9Y5F0	0.439	reviewed	PCDBD_HUMAN
A8MWY0	0.44	reviewed	ELAP2_HUMAN
Q99062	0.44	reviewed	CSF3R_HUMAN
P17643	0.441	reviewed	TYRP1_HUMAN
Q86XK7	0.442	reviewed	VSIG1_HUMAN
Q9H3T2	0.442	reviewed	SEM6C_HUMAN
Q9Y5U5	0.443	reviewed	TNR18_HUMAN
P36888	0.443	reviewed	FLT3_HUMAN
Q96KG7	0.443	reviewed	MEG10_HUMAN
P13765	0.444	reviewed	DOB_HUMAN
Q03167	0.444	reviewed	TGBR3_HUMAN
P01730	0.445	reviewed	CD4_HUMAN
Q9UPZ6	0.445	reviewed	THS7A_HUMAN
P0DTF9	0.446	reviewed	PTIP2_HUMAN
Q96J84	0.446	reviewed	KIRR1_HUMAN
Q9Y5G3	0.446	reviewed	PCDGD_HUMAN
Q8N6Y1	0.446	reviewed	PCD20_HUMAN
O00548	0.447	reviewed	DLL1_HUMAN
P40189	0.447	reviewed	IL6RB_HUMAN
Q5SZK8	0.447	reviewed	FREM2_HUMAN
Q96F46	0.448	reviewed	I17RA_HUMAN
Q6PJG9	0.448	reviewed	LRFN4_HUMAN
P08922	0.449	reviewed	ROS1_HUMAN
Q9UKJ8	0.449	reviewed	ADA21_HUMAN
Q9H8J5	0.449	reviewed	MANS1_HUMAN
Q9NZR2	0.449	reviewed	LRP1B_HUMAN
A6NLX4	0.449	reviewed	TM210_HUMAN
Q8IZJ1	0.45	reviewed	UNC5B_HUMAN
Q04900	0.45	reviewed	MUC24_HUMAN
Q9UKJ1	0.45	reviewed	PILRA_HUMAN
Q5TEA6	0.451	reviewed	SE1L2_HUMAN
Q13740	0.451	reviewed	CD166_HUMAN
Q6UWI2	0.451	reviewed	PARM1_HUMAN
Q8N387	0.451	reviewed	MUC15_HUMAN
P43632	0.451	reviewed	KI2S4_HUMAN
O94933	0.452	reviewed	SLIK3_HUMAN
P31995	0.452	reviewed	FCG2C_HUMAN
Q6GPH6	0.452	reviewed	IPIL1_HUMAN
P34741	0.453	reviewed	SDC2_HUMAN
Q93038	0.453	reviewed	TNR25_HUMAN
P36941	0.454	reviewed	TNR3_HUMAN
Q9NYQ8	0.454	reviewed	FAT2_HUMAN

Q7Z6M3	0.454	reviewed	MILR1_HUMAN
Q9Y5I4	0.454	reviewed	PCDC2_HUMAN
P23467	0.455	reviewed	PTPRB_HUMAN
Q9ULH4	0.455	reviewed	LRFN2_HUMAN
Q01638	0.455	reviewed	ILRL1_HUMAN
Q9H6D8	0.456	reviewed	FNDC4_HUMAN
Q08ET2	0.457	reviewed	SIG14_HUMAN
P17927	0.457	reviewed	CR1_HUMAN
Q8N743	0.457	reviewed	KI3L3_HUMAN
Q5VWK5	0.457	reviewed	IL23R_HUMAN
P52799	0.457	reviewed	EFNB2_HUMAN
Q9Y5I1	0.457	reviewed	PCDAB_HUMAN
P20645	0.458	reviewed	MPRD_HUMAN
Q86YD3	0.459	reviewed	TMM25_HUMAN
Q02763	0.459	reviewed	TIE2_HUMAN
Q8WUT4	0.459	reviewed	LRRN4_HUMAN
Q9BQT9	0.459	reviewed	CSTN3_HUMAN
Q96JQ0	0.459	reviewed	PCD16_HUMAN
P33151	0.46	reviewed	CADH5_HUMAN
P14778	0.46	reviewed	IL1R1_HUMAN
Q8IUK5	0.46	reviewed	PLDX1_HUMAN
P18564	0.46	reviewed	ITB6_HUMAN
Q9UJ99	0.461	reviewed	CAD22_HUMAN
Q14114	0.461	reviewed	LRP8_HUMAN
Q14773	0.462	reviewed	ICAM4_HUMAN
P54753	0.462	reviewed	EPHB3_HUMAN
P25189	0.463	reviewed	MYP0_HUMAN
Q5VX71	0.465	reviewed	SUSD4_HUMAN
Q17R55	0.465	reviewed	F187B_HUMAN
P01732	0.465	reviewed	CD8A_HUMAN
Q9HAT1	0.466	reviewed	LMA1L_HUMAN
P01889	0.466	reviewed	HLAB_HUMAN
P56199	0.466	reviewed	ITA1_HUMAN
Q15223	0.466	reviewed	NECT1_HUMAN
P01833	0.467	reviewed	PIGR_HUMAN
Q96NI6	0.467	reviewed	LRFN5_HUMAN
P14784	0.467	reviewed	IL2RB_HUMAN
Q96LL3	0.468	reviewed	FIMP_HUMAN
A0A1B0GTW7	0.469	reviewed	CIROP_HUMAN
P16422	0.47	reviewed	EPCAM_HUMAN
Q9BQ49	0.47	reviewed	SMIM7_HUMAN
Q7Z5N4	0.47	reviewed	SDK1_HUMAN
Q9H156	0.471	reviewed	SLIK2_HUMAN
Q96JA1	0.471	reviewed	LRIG1_HUMAN
Q9BXS4	0.471	reviewed	TMM59_HUMAN
Q8IZU9	0.472	reviewed	KIRR3_HUMAN
Q99965	0.473	reviewed	ADAM2_HUMAN

Q8N8Z6	0.473	reviewed	DCBD1_HUMAN
Q8N441	0.473	reviewed	FGRL1_HUMAN
P38570	0.473	reviewed	ITAE_HUMAN
P13224	0.473	reviewed	GP1BB_HUMAN
Q7Z4F1	0.474	reviewed	LRP10_HUMAN
Q07954	0.474	reviewed	LRP1_HUMAN
Q6NUS6	0.474	reviewed	TECT3_HUMAN
Q96BF3	0.475	reviewed	TMIG2_HUMAN
Q16827	0.475	reviewed	PTPRO_HUMAN
P32942	0.476	reviewed	ICAM3_HUMAN
P12319	0.477	reviewed	FCERA_HUMAN
Q5IJ48	0.477	reviewed	CRUM2_HUMAN
Q15116	0.477	reviewed	PDCD1_HUMAN
P04234	0.478	reviewed	CD3D_HUMAN
Q13444	0.478	reviewed	ADA15_HUMAN
Q9HDB5	0.478	reviewed	NRX3B_HUMAN
Q8TB96	0.479	reviewed	TIP_HUMAN
Q8NET5	0.479	reviewed	NFAM1_HUMAN
A0A0K2S4Q6	0.479	reviewed	CD3CH_HUMAN
Q96RL6	0.48	reviewed	SIG11_HUMAN
Q86UP6	0.48	reviewed	CUZD1_HUMAN
P32926	0.48	reviewed	DSG3_HUMAN
Q8N766	0.481	reviewed	EMC1_HUMAN
Q9H3W5	0.481	reviewed	LRRN3_HUMAN
Q14953	0.482	reviewed	KI2S5_HUMAN
Q96MS0	0.483	reviewed	ROBO3_HUMAN
P41217	0.483	reviewed	OX2G_HUMAN
P26842	0.484	reviewed	CD27_HUMAN
Q9BZ76	0.484	reviewed	CNTP3_HUMAN
Q9UM44	0.484	reviewed	HHLA2_HUMAN
Q9Y4C0	0.484	reviewed	NRX3A_HUMAN
Q9UN67	0.485	reviewed	PCDBA_HUMAN
Q9H3T3	0.485	reviewed	SEM6B_HUMAN
Q9Y5F3	0.485	reviewed	PCDB1_HUMAN
Q14574	0.486	reviewed	DSC3_HUMAN
P21709	0.486	reviewed	EPHA1_HUMAN
Q9BYH1	0.486	reviewed	SE6L1_HUMAN
Q86YD5	0.487	reviewed	LRAD3_HUMAN
O14669	0.488	reviewed	TMG2_HUMAN
Q96QU1	0.488	reviewed	PCD15_HUMAN
Q7L0X0	0.488	reviewed	TRIL_HUMAN
Q6ZMQ8	0.488	reviewed	LMTK1_HUMAN
P08514	0.488	reviewed	ITA2B_HUMAN
Q9H2U9	0.489	reviewed	ADAM7_HUMAN
Q9HBV2	0.489	reviewed	SACA1_HUMAN
O75078	0.49	reviewed	ADA11_HUMAN
A6H8M9	0.49	reviewed	CDHR4_HUMAN

O75578	0.49	reviewed	ITA10_HUMAN
P08138	0.491	reviewed	TNR16_HUMAN
Q08174	0.491	reviewed	PCDH1_HUMAN
Q6UY11	0.491	reviewed	DLK2_HUMAN
Q14952	0.491	reviewed	KI2S3_HUMAN
Q86V40	0.492	reviewed	TIKI1_HUMAN
P42081	0.493	reviewed	CD86_HUMAN
Q9UKF2	0.493	reviewed	ADA30_HUMAN
P19256	0.493	reviewed	LFA3_HUMAN
P05106	0.495	reviewed	ITB3_HUMAN
Q04912	0.496	reviewed	RON_HUMAN
P00533	0.497	reviewed	EGFR_HUMAN
Q6UXG8	0.497	reviewed	BTNL9_HUMAN
Q6ZU64	0.497	reviewed	CFA65_HUMAN
Q86WK6	0.497	reviewed	AMGO1_HUMAN
Q9BXJ7	0.497	reviewed	AMNLS_HUMAN
Q07011	0.498	reviewed	TNR9_HUMAN
O95256	0.499	reviewed	I18RA_HUMAN
Q8N131	0.5	reviewed	PORIM_HUMAN
Q96PX8	0.5	reviewed	SLIK1_HUMAN
Q9UBN6	0.5	reviewed	TR10D_HUMAN
P15941	0.5	reviewed	MUC1_HUMAN
P0DPE3	0.5	reviewed	TMDD1_HUMAN
P05107	0.5	reviewed	ITB2_HUMAN
O95754	0.501	reviewed	SEM4F_HUMAN
Q8N3G9	0.501	reviewed	TM130_HUMAN
O75976	0.501	reviewed	CBPD_HUMAN
Q8TD46	0.501	reviewed	MO2R1_HUMAN
P26010	0.502	reviewed	ITB7_HUMAN
Q6P1J6	0.502	reviewed	PLB1_HUMAN
Q5JX71	0.504	reviewed	F209A_HUMAN
Q86YC3	0.504	reviewed	LRC33_HUMAN
Q06418	0.505	reviewed	TYRO3_HUMAN
Q9Y320	0.505	reviewed	TMX2_HUMAN
Q9Y5E8	0.505	reviewed	PCDBF_HUMAN
P37023	0.506	reviewed	ACVL1_HUMAN
Q9NRM6	0.506	reviewed	I17RB_HUMAN
P78324	0.507	reviewed	SHPS1_HUMAN
O43300	0.507	reviewed	LRRT2_HUMAN
Q9UGN4	0.508	reviewed	CLM8_HUMAN
P10966	0.509	reviewed	CD8B_HUMAN
Q9H158	0.509	reviewed	PCDC1_HUMAN
Q9Y5I3	0.51	reviewed	PCDA1_HUMAN
O75581	0.512	reviewed	LRP6_HUMAN
Q99075	0.512	reviewed	HBEGF_HUMAN
P43627	0.513	reviewed	KI2L2_HUMAN
A6NFA1	0.514	reviewed	TIKI2_HUMAN

Q86UK5	0.515	reviewed	LBN_HUMAN
Q68DV7	0.516	reviewed	RNF43_HUMAN
Q2M385	0.516	reviewed	MPEG1_HUMAN
Q15375	0.517	reviewed	EPHA7_HUMAN
P22223	0.517	reviewed	CADH3_HUMAN
O00241	0.518	reviewed	SIRB1_HUMAN
Q9Y5E1	0.518	reviewed	PCDB9_HUMAN
Q7Z692	0.518	reviewed	CEA19_HUMAN
P20916	0.518	reviewed	MAG_HUMAN
Q5SSG8	0.52	reviewed	MUC21_HUMAN
Q6UXZ4	0.521	reviewed	UNC5D_HUMAN
P01903	0.522	reviewed	DRA_HUMAN
Q9BQS7	0.522	reviewed	HEPH_HUMAN
Q5T2D2	0.522	reviewed	TRML2_HUMAN
Q8TCW7	0.524	reviewed	ZPLD1_HUMAN
Q86SJ2	0.524	reviewed	AMGO2_HUMAN
Q15399	0.525	reviewed	TLR1_HUMAN
P60852	0.525	reviewed	ZP1_HUMAN
P36896	0.525	reviewed	ACV1B_HUMAN
Q6PCB8	0.525	reviewed	EMB_HUMAN
Q9BVV8	0.525	reviewed	F174C_HUMAN
A6NGZ8	0.526	reviewed	SMIM9_HUMAN
Q8IZS8	0.526	reviewed	CA2D3_HUMAN
Q96GP6	0.527	reviewed	SREC2_HUMAN
Q9H6B4	0.527	reviewed	CLMP_HUMAN
Q92537	0.527	reviewed	SUSD6_HUMAN
Q8WWQ8	0.529	reviewed	STAB2_HUMAN
P15812	0.529	reviewed	CD1E_HUMAN
Q92824	0.529	reviewed	PCSK5_HUMAN
P27037	0.53	reviewed	AVR2A_HUMAN
Q8NHK3	0.53	reviewed	KI2LB_HUMAN
Q9P244	0.531	reviewed	LRFN1_HUMAN
P13591	0.532	reviewed	NCAM1_HUMAN
A8MVZ5	0.532	reviewed	BTNLA_HUMAN
Q9Y5F1	0.532	reviewed	PCDBC_HUMAN
P98153	0.532	reviewed	IDD_HUMAN
Q902F8	0.534	reviewed	ENK8_HUMAN
Q9H0Q3	0.534	reviewed	FXYD6_HUMAN
Q8IXH8	0.534	reviewed	CAD26_HUMAN
Q6UX15	0.534	reviewed	LAYN_HUMAN
Q13641	0.535	reviewed	TPBG_HUMAN
P13762	0.535	reviewed	DRB4_HUMAN
Q8N967	0.535	reviewed	LRTM2_HUMAN
Q14242	0.536	reviewed	SELPL_HUMAN
A6NMB1	0.536	reviewed	SIG16_HUMAN
P25942	0.536	reviewed	TNR5_HUMAN
Q9HBE5	0.536	reviewed	IL21R_HUMAN

O15394	0.537	reviewed	NCAM2_HUMAN
Q9NR97	0.537	reviewed	TLR8_HUMAN
P13688	0.537	reviewed	CEAM1_HUMAN
P04439	0.537	reviewed	HLAA_HUMAN
Q24JP5	0.538	reviewed	T132A_HUMAN
Q9UMF0	0.538	reviewed	ICAM5_HUMAN
P08575	0.538	reviewed	PTPRC_HUMAN
P09603	0.539	reviewed	CSF1_HUMAN
O14917	0.539	reviewed	PCD17_HUMAN
Q9UBS9	0.54	reviewed	SUCO_HUMAN
P05538	0.541	reviewed	DQB2_HUMAN
Q9P0T7	0.541	reviewed	TMEM9_HUMAN
Q8IYR6	0.542	reviewed	TEFF1_HUMAN
O14668	0.544	reviewed	TMG1_HUMAN
Q8N3T6	0.544	reviewed	T132C_HUMAN
Q9P0K1	0.544	reviewed	ADA22_HUMAN
Q6UWB1	0.544	reviewed	I27RA_HUMAN
Q15884	0.545	reviewed	EREP1_HUMAN
O75096	0.545	reviewed	LRP4_HUMAN
Q99795	0.545	reviewed	GPA33_HUMAN
Q8TDW7	0.546	reviewed	FAT3_HUMAN
Q9UBV2	0.546	reviewed	SE1L1_HUMAN
Q6UWL2	0.546	reviewed	SUSD1_HUMAN
Q9H251	0.547	reviewed	CAD23_HUMAN
P12821	0.548	reviewed	ACE_HUMAN
P42702	0.548	reviewed	LIFR_HUMAN
Q96FE7	0.548	reviewed	P3IP1_HUMAN
Q9Y561	0.548	reviewed	LRP12_HUMAN
Q13634	0.549	reviewed	CAD18_HUMAN
Q07699	0.549	reviewed	SCN1B_HUMAN
Q6ISU1	0.55	reviewed	PTCRA_HUMAN
Q92692	0.55	reviewed	NECT2_HUMAN
Q9Y5E5	0.551	reviewed	PCDB4_HUMAN
Q9UN66	0.551	reviewed	PCDB8_HUMAN
Q8N1Y9	0.552	reviewed	YI025_HUMAN
Q14943	0.552	reviewed	KI3S1_HUMAN
P55291	0.553	reviewed	CAD15_HUMAN
A6NLU5	0.554	reviewed	VTM2B_HUMAN
Q9NZ94	0.555	reviewed	NLGN3_HUMAN
O43157	0.556	reviewed	PLXB1_HUMAN
O75882	0.556	reviewed	ATRN_HUMAN
Q9BYF1	0.557	reviewed	ACE2_HUMAN
P36897	0.557	reviewed	TGFR1_HUMAN
F2Z333	0.558	reviewed	FND10_HUMAN
P15151	0.558	reviewed	PVR_HUMAN
P35590	0.558	reviewed	TIE1_HUMAN
O60449	0.559	reviewed	LY75_HUMAN

Q86UE6	0.559	reviewed	LRRT1_HUMAN
P28906	0.56	reviewed	CD34_HUMAN
P29323	0.56	reviewed	EPHB2_HUMAN
Q86TY3	0.56	reviewed	ARMD4_HUMAN
P16581	0.561	reviewed	LYAM2_HUMAN
Q8NFZ3	0.561	reviewed	NLGNY_HUMAN
Q6ZTQ4	0.562	reviewed	CDHR3_HUMAN
Q9UKQ2	0.562	reviewed	ADA28_HUMAN
P23229	0.562	reviewed	ITA6_HUMAN
O15197	0.562	reviewed	EPHB6_HUMAN
Q6UWL6	0.562	reviewed	KIRR2_HUMAN
P35070	0.563	reviewed	BTC_HUMAN
Q9UHF4	0.563	reviewed	I20RA_HUMAN
Q9NR61	0.564	reviewed	DLL4_HUMAN
P28908	0.565	reviewed	TNR8_HUMAN
Q969N2	0.565	reviewed	PIGT_HUMAN
O75509	0.566	reviewed	TNR21_HUMAN
P78357	0.566	reviewed	CNTP1_HUMAN
P80370	0.566	reviewed	DLK1_HUMAN
P17813	0.566	reviewed	EGLN_HUMAN
Q9NPF0	0.567	reviewed	CD320_HUMAN
P49755	0.567	reviewed	TMEDA_HUMAN
Q5BVD1	0.568	reviewed	TTMP_HUMAN
Q86SU0	0.568	reviewed	ILDR1_HUMAN
Q9Y5E9	0.568	reviewed	PCDBE_HUMAN
Q96DD7	0.569	reviewed	SHSA4_HUMAN
Q95460	0.569	reviewed	HMR1_HUMAN
Q96PQ1	0.571	reviewed	SIG12_HUMAN
O42043	0.572	reviewed	ENK18_HUMAN
P36894	0.572	reviewed	BMR1A_HUMAN
P54760	0.572	reviewed	EPHB4_HUMAN
Q14DG7	0.573	reviewed	T132B_HUMAN
Q9NY15	0.573	reviewed	STAB1_HUMAN
P30511	0.574	reviewed	HLAF_HUMAN
Q8IWB1	0.576	reviewed	IPRI_HUMAN
Q04721	0.576	reviewed	NOTC2_HUMAN
O76095	0.576	reviewed	JTB_HUMAN
O00206	0.577	reviewed	TLR4_HUMAN
Q96P31	0.577	reviewed	FCRL3_HUMAN
P10721	0.577	reviewed	KIT_HUMAN
Q8TC27	0.578	reviewed	ADA32_HUMAN
P28827	0.578	reviewed	PTPRM_HUMAN
Q6ZSJ9	0.579	reviewed	SHSA6_HUMAN
O71037	0.579	reviewed	ENK19_HUMAN
Q9Y5H7	0.58	reviewed	PCDA5_HUMAN
Q9BZD7	0.58	reviewed	TMG3_HUMAN
Q9NZN1	0.581	reviewed	IRPL1_HUMAN

Q13145	0.582	reviewed	BAMBI_HUMAN
P16109	0.582	reviewed	LYAM3_HUMAN
Q9ULX7	0.583	reviewed	CAH14_HUMAN
Q9Y624	0.583	reviewed	JAM1_HUMAN
Q02413	0.583	reviewed	DSG1_HUMAN
Q6UWV2	0.583	reviewed	MPZL3_HUMAN
P05362	0.584	reviewed	ICAM1_HUMAN
Q6DN72	0.584	reviewed	FCRL6_HUMAN
Q9Y5G9	0.584	reviewed	PCDG4_HUMAN
Q5ZPR3	0.585	reviewed	CD276_HUMAN
P43489	0.586	reviewed	TNR4_HUMAN
Q71H61	0.586	reviewed	ILDR2_HUMAN
Q6UY18	0.586	reviewed	LIGO4_HUMAN
Q8IYV9	0.586	reviewed	IZUM1_HUMAN
Q16586	0.587	reviewed	SGCA_HUMAN
O00220	0.587	reviewed	TR10A_HUMAN
Q9C0A0	0.587	reviewed	CNTP4_HUMAN
Q9Y5G4	0.587	reviewed	PCDG9_HUMAN
O60894	0.589	reviewed	RAMP1_HUMAN
Q12836	0.589	reviewed	ZP4_HUMAN
Q9BYE9	0.589	reviewed	CDHR2_HUMAN
Q9H9P2	0.589	reviewed	CHODL_HUMAN
Q13591	0.59	reviewed	SEM5A_HUMAN
P28068	0.59	reviewed	DMB_HUMAN
P25092	0.59	reviewed	GUC2C_HUMAN
Q9NXS2	0.591	reviewed	QPCTL_HUMAN
P20963	0.591	reviewed	CD3Z_HUMAN
P55289	0.592	reviewed	CAD12_HUMAN
O75445	0.593	reviewed	USH2A_HUMAN
Q14789	0.593	reviewed	GOGB1_HUMAN
P11279	0.593	reviewed	LAMP1_HUMAN
B8ZZ34	0.594	reviewed	SHSA8_HUMAN
P78536	0.594	reviewed	ADA17_HUMAN
Q96PL5	0.594	reviewed	ERMAP_HUMAN
Q9UJ90	0.594	reviewed	KCNE5_HUMAN
Q9Y5G1	0.595	reviewed	PCDGF_HUMAN
Q9ULI3	0.596	reviewed	HEG1_HUMAN
Q8ND94	0.596	reviewed	LRN4L_HUMAN
P04843	0.596	reviewed	RPN1_HUMAN
Q3KPI0	0.597	reviewed	CEA21_HUMAN
Q9BZW8	0.598	reviewed	CD244_HUMAN
Q3SXP7	0.601	reviewed	SHSL1_HUMAN
P05067	0.601	reviewed	A4_HUMAN
P51511	0.601	reviewed	MMP15_HUMAN
P15509	0.602	reviewed	CSF2R_HUMAN
Q8WYK1	0.603	reviewed	CNTP5_HUMAN
P16410	0.603	reviewed	CTLA4_HUMAN

P0C6S8	0.603	reviewed	LIGO3_HUMAN
Q2I0M4	0.605	reviewed	LRC26_HUMAN
Q8NAC3	0.605	reviewed	I17RC_HUMAN
O15389	0.606	reviewed	SIGL5_HUMAN
Q6UXK5	0.607	reviewed	LRRN1_HUMAN
Q8TD84	0.607	reviewed	DSCL1_HUMAN
P33681	0.608	reviewed	CD80_HUMAN
Q9Y639	0.608	reviewed	NPTN_HUMAN
Q9BWV1	0.609	reviewed	BOC_HUMAN
Q8TEM1	0.609	reviewed	PO210_HUMAN
Q5DX21	0.609	reviewed	IGS11_HUMAN
Q6V0I7	0.61	reviewed	FAT4_HUMAN
P16150	0.61	reviewed	LEUK_HUMAN
P14616	0.61	reviewed	INSRR_HUMAN
P40967	0.61	reviewed	PMEL_HUMAN
Q01344	0.61	reviewed	IL5RA_HUMAN
P15529	0.61	reviewed	MCP_HUMAN
P23469	0.611	reviewed	PTPRE_HUMAN
Q7Z6A9	0.611	reviewed	BTLA_HUMAN
Q9NZJ5	0.612	reviewed	E2AK3_HUMAN
B6SEH8	0.612	reviewed	ERVV1_HUMAN
Q96FE5	0.612	reviewed	LIGO1_HUMAN
Q9UK23	0.612	reviewed	NAGPA_HUMAN
Q86XM0	0.612	reviewed	CTSRD_HUMAN
P31785	0.612	reviewed	IL2RG_HUMAN
P51512	0.613	reviewed	MMP16_HUMAN
Q92859	0.614	reviewed	NEO1_HUMAN
Q9HD43	0.615	reviewed	PTPRH_HUMAN
Q6P995	0.615	reviewed	F171B_HUMAN
Q30201	0.615	reviewed	HFE_HUMAN
Q5VV43	0.615	reviewed	K0319_HUMAN
Q9UKH3	0.616	reviewed	ENK9_HUMAN
Q92729	0.617	reviewed	PTPRU_HUMAN
Q14956	0.618	reviewed	GPNMB_HUMAN
O75077	0.619	reviewed	ADA23_HUMAN
Q14802	0.619	reviewed	FXYD3_HUMAN
Q9HBW1	0.62	reviewed	LRRC4_HUMAN
P48357	0.62	reviewed	LEPR_HUMAN
Q13586	0.62	reviewed	STIM1_HUMAN
O15031	0.621	reviewed	PLXB2_HUMAN
O00481	0.621	reviewed	BT3A1_HUMAN
O60896	0.621	reviewed	RAMP3_HUMAN
Q9BZD6	0.622	reviewed	TMG4_HUMAN
Q9HCM2	0.622	reviewed	PLXA4_HUMAN
P12318	0.623	reviewed	FCG2A_HUMAN
Q9UN70	0.623	reviewed	PCDGK_HUMAN
O95297	0.624	reviewed	MPZL1_HUMAN

P17181	0.624	reviewed	INAR1_HUMAN
Q3SXY7	0.625	reviewed	LRIT3_HUMAN
Q99665	0.625	reviewed	I12R2_HUMAN
Q9UPX0	0.626	reviewed	TUTLB_HUMAN
P14679	0.627	reviewed	TYRO_HUMAN
A8MVW0	0.627	reviewed	F1712_HUMAN
A6NJW9	0.627	reviewed	CD8B2_HUMAN
Q13477	0.627	reviewed	MADCA_HUMAN
P13598	0.627	reviewed	ICAM2_HUMAN
P58335	0.628	reviewed	ANTR2_HUMAN
O95196	0.628	reviewed	CSPG5_HUMAN
A2VDJ0	0.629	reviewed	T131L_HUMAN
Q9Y3Q7	0.63	reviewed	ADA18_HUMAN
P20594	0.631	reviewed	ANPRB_HUMAN
P16144	0.631	reviewed	ITB4_HUMAN
Q9Y5E3	0.632	reviewed	PCDB6_HUMAN
Q12864	0.632	reviewed	CAD17_HUMAN
Q96AP7	0.632	reviewed	ESAM_HUMAN
P15421	0.632	reviewed	GLPE_HUMAN
Q8WWG1	0.632	reviewed	NRG4_HUMAN
Q6ZN44	0.634	reviewed	UNC5A_HUMAN
O43506	0.634	reviewed	ADA20_HUMAN
Q6GTX8	0.634	reviewed	LAIR1_HUMAN
Q5JZY3	0.634	reviewed	EPHAA_HUMAN
P22897	0.634	reviewed	MRC1_HUMAN
Q8TD07	0.635	reviewed	RAE1E_HUMAN
Q9HBG7	0.635	reviewed	LY9_HUMAN
Q8WWV6	0.635	reviewed	FCAMR_HUMAN
P10321	0.635	reviewed	HLAC_HUMAN
O00168	0.635	reviewed	PLM_HUMAN
Q8NFR9	0.636	reviewed	I17RE_HUMAN
Q5JXA9	0.637	reviewed	SIRB2_HUMAN
P58550	0.637	reviewed	FXYD8_HUMAN
P55285	0.637	reviewed	CADH6_HUMAN
Q13651	0.637	reviewed	I10R1_HUMAN
Q05996	0.638	reviewed	ZP2_HUMAN
O43291	0.638	reviewed	SPIT2_HUMAN
P19438	0.639	reviewed	TNR1A_HUMAN
Q7Z3S7	0.639	reviewed	CA2D4_HUMAN
Q5JX69	0.64	reviewed	F209B_HUMAN
Q9Y5H8	0.642	reviewed	PCDA3_HUMAN
Q16288	0.643	reviewed	NTRK3_HUMAN
Q9HB29	0.644	reviewed	ILRL2_HUMAN
A6BM72	0.645	reviewed	MEG11_HUMAN
P98164	0.646	reviewed	LRP2_HUMAN
P13612	0.649	reviewed	ITA4_HUMAN
Q9Y5F6	0.649	reviewed	PCDGM_HUMAN

Q9Y3Q3	0.649	reviewed	TMED3_HUMAN
O00592	0.65	reviewed	PODXL_HUMAN
Q9NSI5	0.651	reviewed	IGSF5_HUMAN
Q96PE5	0.651	reviewed	OPALI_HUMAN
P22455	0.652	reviewed	FGFR4_HUMAN
Q96J86	0.652	reviewed	CYYR1_HUMAN
Q9ULB1	0.654	reviewed	NRX1A_HUMAN
Q16849	0.654	reviewed	PTPRN_HUMAN
P08637	0.655	reviewed	FCG3A_HUMAN
Q8NER5	0.656	reviewed	ACV1C_HUMAN
Q29980	0.656	reviewed	MICB_HUMAN
Q5VU97	0.657	reviewed	CAHD1_HUMAN
O14511	0.659	reviewed	NRG2_HUMAN
Q9UN75	0.66	reviewed	PCDAC_HUMAN
Q9Y219	0.661	reviewed	JAG2_HUMAN
Q5SY80	0.662	reviewed	CTSRE_HUMAN
P59646	0.665	reviewed	FXYD4_HUMAN
P34925	0.666	reviewed	RYK_HUMAN
Q8N149	0.669	reviewed	LIRA2_HUMAN
A0PJX4	0.671	reviewed	SHSA3_HUMAN
Q8TDY8	0.671	reviewed	IGDC4_HUMAN
Q9UM73	0.671	reviewed	ALK_HUMAN
Q99650	0.671	reviewed	OSMR_HUMAN
P78410	0.675	reviewed	BT3A2_HUMAN
Q08708	0.675	reviewed	CLM6_HUMAN
Q6UWV6	0.675	reviewed	ENPP7_HUMAN
P43626	0.676	reviewed	KI2L1_HUMAN
Q9UBK5	0.677	reviewed	HCST_HUMAN
Q92854	0.678	reviewed	SEM4D_HUMAN
Q8NHL6	0.678	reviewed	LIRB1_HUMAN
Q96PD2	0.679	reviewed	DCBD2_HUMAN
O75787	0.679	reviewed	RENR_HUMAN
A6NMS7	0.68	reviewed	L37A1_HUMAN
Q8NHJ6	0.682	reviewed	LIRB4_HUMAN
Q969Z4	0.684	reviewed	TR19L_HUMAN
Q8IYJ0	0.684	reviewed	PIANP_HUMAN
O00478	0.685	reviewed	BT3A3_HUMAN
Q13410	0.685	reviewed	BT1A1_HUMAN
Q8NFZ4	0.686	reviewed	NLGN2_HUMAN
Q9NU53	0.686	reviewed	GINM1_HUMAN
P01130	0.686	reviewed	LDLR_HUMAN
P15382	0.687	reviewed	KCNE1_HUMAN
Q9BXR5	0.688	reviewed	TLR10_HUMAN
Q16549	0.689	reviewed	PCSK7_HUMAN
Q5UCC4	0.691	reviewed	EMC10_HUMAN
Q92673	0.692	reviewed	SORL_HUMAN
Q99523	0.692	reviewed	SORT_HUMAN

Q9HCK4	0.692	reviewed	ROBO2_HUMAN
O15146	0.693	reviewed	MUSK_HUMAN
A6NHS7	0.694	reviewed	MANS4_HUMAN
Q8N2Q7	0.695	reviewed	NLGN1_HUMAN
Q9H2A7	0.696	reviewed	CXL16_HUMAN
Q8WWB7	0.696	reviewed	GLMP_HUMAN
O95976	0.697	reviewed	IGSF6_HUMAN
P16234	0.697	reviewed	PGFRA_HUMAN
Q9NZQ7	0.699	reviewed	PD1L1_HUMAN
P06126	0.7	reviewed	CD1A_HUMAN
A2RUT3	0.702	reviewed	TMM89_HUMAN
Q16671	0.703	reviewed	AMHR2_HUMAN
Q7Z7H5	0.704	reviewed	TMED4_HUMAN
Q8N0W4	0.704	reviewed	NLGNX_HUMAN
Q13018	0.704	reviewed	PLA2R_HUMAN
P13747	0.705	reviewed	HLAE_HUMAN
P61566	0.706	reviewed	ENK24_HUMAN
Q8NC54	0.706	reviewed	KCT2_HUMAN
Q9NPR2	0.706	reviewed	SEM4B_HUMAN
Q30154	0.709	reviewed	DRB5_HUMAN
Q9Y493	0.71	reviewed	ZAN_HUMAN
Q9NPY3	0.71	reviewed	C1QR1_HUMAN
Q9Y5R2	0.71	reviewed	MMP24_HUMAN
P19235	0.712	reviewed	EPOR_HUMAN
Q9UK28	0.713	reviewed	TM59L_HUMAN
Q9ULK6	0.713	reviewed	RN150_HUMAN
Q9Y3A6	0.714	reviewed	TMED5_HUMAN
Q9BX67	0.715	reviewed	JAM3_HUMAN
Q9Y336	0.717	reviewed	SIGL9_HUMAN
Q9BX59	0.717	reviewed	TPSNR_HUMAN
P98172	0.717	reviewed	EFNB1_HUMAN
P55082	0.717	reviewed	MFAP3_HUMAN
Q9P1W8	0.718	reviewed	SIRPG_HUMAN
Q14CZ8	0.718	reviewed	HECAM_HUMAN
Q58EX2	0.718	reviewed	SDK2_HUMAN
Q96KJ4	0.719	reviewed	MSLNL_HUMAN
P40238	0.72	reviewed	TPOR_HUMAN
Q99706	0.72	reviewed	KI2L4_HUMAN
P54756	0.723	reviewed	EPHA5_HUMAN
P01920	0.723	reviewed	DQB1_HUMAN
P23468	0.725	reviewed	PTPRD_HUMAN
Q9UKN1	0.726	reviewed	MUC12_HUMAN
P26006	0.726	reviewed	ITA3_HUMAN
P53708	0.726	reviewed	ITA8_HUMAN
Q8WWF5	0.727	reviewed	ZNRF4_HUMAN
Q9NQ60	0.727	reviewed	EQTN_HUMAN
Q14126	0.728	reviewed	DSG2_HUMAN

Q8IUW5	0.728	reviewed	RELL1_HUMAN
Q8IWT1	0.729	reviewed	SCN4B_HUMAN
Q9Y6Q6	0.729	reviewed	TNR11_HUMAN
Q9ULT6	0.729	reviewed	ZNRF3_HUMAN
O60603	0.73	reviewed	TLR2_HUMAN
O14931	0.731	reviewed	NCTR3_HUMAN
P20036	0.735	reviewed	DPA1_HUMAN
O95206	0.736	reviewed	PCDH8_HUMAN
P16871	0.736	reviewed	IL7RA_HUMAN
Q8WVN6	0.736	reviewed	SCTM1_HUMAN
Q8WVV5	0.737	reviewed	BT2A2_HUMAN
P14209	0.737	reviewed	CD99_HUMAN
P43628	0.737	reviewed	KI2L3_HUMAN
Q6UXM1	0.74	reviewed	LRIG3_HUMAN
Q92932	0.74	reviewed	PTPR2_HUMAN
Q6UXZ0	0.74	reviewed	TMIG1_HUMAN
Q92637	0.741	reviewed	FCGRB_HUMAN
Q14627	0.742	reviewed	I13R2_HUMAN
Q86VZ4	0.744	reviewed	LRP11_HUMAN
P43307	0.745	reviewed	SSRA_HUMAN
P16284	0.745	reviewed	PECA1_HUMAN
Q15768	0.748	reviewed	EFNB3_HUMAN
Q9UQC9	0.748	reviewed	CLCA2_HUMAN
Q12913	0.749	reviewed	PTPRJ_HUMAN
O60330	0.751	reviewed	PCDGC_HUMAN
Q15303	0.751	reviewed	ERBB4_HUMAN
P01906	0.752	reviewed	DQA2_HUMAN
Q6UWI4	0.754	reviewed	SHSA2_HUMAN
Q9H2E6	0.754	reviewed	SEM6A_HUMAN
P17301	0.755	reviewed	ITA2_HUMAN
Q9HBL6	0.757	reviewed	LRTM1_HUMAN
Q6UXE8	0.758	reviewed	BTNL3_HUMAN
P40259	0.758	reviewed	CD79B_HUMAN
P58400	0.758	reviewed	NRX1B_HUMAN
P29017	0.761	reviewed	CD1C_HUMAN
Q13261	0.761	reviewed	I15RA_HUMAN
P30273	0.762	reviewed	FCERG_HUMAN
Q9Y5E2	0.762	reviewed	PCDB7_HUMAN
Q9BZG2	0.763	reviewed	PPAT_HUMAN
Q9UN73	0.764	reviewed	PCDA6_HUMAN
Q5VZ72	0.765	reviewed	IZUM3_HUMAN
P32004	0.767	reviewed	L1CAM_HUMAN
Q8IU57	0.768	reviewed	INLR1_HUMAN
Q8N6P7	0.769	reviewed	I22R1_HUMAN
Q7Z7M0	0.769	reviewed	MEGF8_HUMAN
Q29983	0.771	reviewed	MICA_HUMAN
P42701	0.773	reviewed	I12R1_HUMAN

Q08334	0.774	reviewed	I10R2_HUMAN
P11362	0.778	reviewed	FGFR1_HUMAN
Q9UN74	0.779	reviewed	PCDA4_HUMAN
P58658	0.781	reviewed	EVA1C_HUMAN
Q13505	0.782	reviewed	MTX1_HUMAN
P35916	0.783	reviewed	VGFR3_HUMAN
Q3ZCQ3	0.785	reviewed	F174B_HUMAN
P61565	0.79	reviewed	ENK21_HUMAN
Q9ULB5	0.794	reviewed	CADH7_HUMAN
P0C7U0	0.794	reviewed	ELFN1_HUMAN
Q9P246	0.796	reviewed	STIM2_HUMAN
P27824	0.796	reviewed	CALX_HUMAN
O75019	0.797	reviewed	LIRA1_HUMAN
O60500	0.8	reviewed	NPHN_HUMAN
Q6P9G4	0.801	reviewed	TM154_HUMAN
Q86YL7	0.801	reviewed	PDPN_HUMAN
Q9H4D0	0.803	reviewed	CSTN2_HUMAN
Q9HC56	0.803	reviewed	PCDH9_HUMAN
O43570	0.806	reviewed	CAH12_HUMAN
P04629	0.81	reviewed	NTRK1_HUMAN
Q5SZI1	0.811	reviewed	LRAD2_HUMAN
Q08554	0.814	reviewed	DSC1_HUMAN
Q9NQ25	0.815	reviewed	SLAF7_HUMAN
Q3SY77	0.818	reviewed	UD3A2_HUMAN
Q9H6X2	0.818	reviewed	ANTR1_HUMAN
Q8NFM7	0.818	reviewed	I17RD_HUMAN
P24071	0.819	reviewed	FCAR_HUMAN
Q96IQ7	0.821	reviewed	VSIG2_HUMAN
Q9Y5G6	0.822	reviewed	PCDG7_HUMAN
Q9P2B2	0.823	reviewed	FPRP_HUMAN
Q13443	0.824	reviewed	ADAM9_HUMAN
Q15256	0.824	reviewed	PTPRR_HUMAN
P17342	0.825	reviewed	ANPRC_HUMAN
Q7L985	0.825	reviewed	LIGO2_HUMAN
P01135	0.826	reviewed	TGFA_HUMAN
Q96AW1	0.828	reviewed	VOPP1_HUMAN
Q96NU0	0.828	reviewed	CNT3B_HUMAN
Q8WZ59	0.829	reviewed	TM190_HUMAN
P54762	0.83	reviewed	EPHB1_HUMAN
Q6UXU6	0.832	reviewed	TMM92_HUMAN
Q9UM47	0.833	reviewed	NOTC3_HUMAN
Q8IV31	0.834	reviewed	TM139_HUMAN
P13473	0.834	reviewed	LAMP2_HUMAN
P20138	0.836	reviewed	CD33_HUMAN
P06127	0.836	reviewed	CD5_HUMAN
A6NI73	0.839	reviewed	LIRA5_HUMAN
Q6UWJ8	0.841	reviewed	C16L2_HUMAN

Q6UXV1	0.843	reviewed	IZUM2_HUMAN
P18627	0.845	reviewed	LAG3_HUMAN
Q9H5Y7	0.845	reviewed	SLIK6_HUMAN
O15533	0.846	reviewed	TPSN_HUMAN
Q9UEF7	0.846	reviewed	KLOT_HUMAN
P08887	0.848	reviewed	IL6RA_HUMAN
P12314	0.85	reviewed	FCGR1_HUMAN
Q13445	0.851	reviewed	TMED1_HUMAN
O43699	0.852	reviewed	SIGL6_HUMAN
Q5VV63	0.852	reviewed	ATRN1_HUMAN
Q96KV6	0.853	reviewed	BT2A3_HUMAN
P19320	0.855	reviewed	VCAM1_HUMAN
Q7Z7D3	0.855	reviewed	VTCN1_HUMAN
Q02487	0.856	reviewed	DSC2_HUMAN
Q69384	0.857	reviewed	ENK6_HUMAN
Q8TDQ0	0.858	reviewed	HAVR2_HUMAN
P26951	0.858	reviewed	IL3RA_HUMAN
Q8TBP5	0.862	reviewed	F174A_HUMAN
Q9ULC0	0.868	reviewed	MUCEN_HUMAN
P40197	0.87	reviewed	GPV_HUMAN
Q13478	0.872	reviewed	IL18R_HUMAN
O60602	0.873	reviewed	TLR5_HUMAN
Q9Y6X5	0.875	reviewed	ENPP4_HUMAN
A4D0T7	0.876	reviewed	SIM30_HUMAN
Q96DB9	0.877	reviewed	FXYD5_HUMAN
A2RRL7	0.881	reviewed	TM213_HUMAN
P11912	0.883	reviewed	CD79A_HUMAN
A8K4G0	0.883	reviewed	CLM7_HUMAN
Q9Y5G5	0.884	reviewed	PCDG8_HUMAN
Q9BVK6	0.887	reviewed	TMED9_HUMAN
A8MVS5	0.888	reviewed	HIDE1_HUMAN
Q14165	0.89	reviewed	MLEC_HUMAN
Q9Y6H6	0.89	reviewed	KCNE3_HUMAN
P15260	0.89	reviewed	INGR1_HUMAN
Q9NZC2	0.89	reviewed	TREM2_HUMAN
Q86YW5	0.894	reviewed	TRML1_HUMAN
P18433	0.895	reviewed	PTPRA_HUMAN
Q8NAU1	0.897	reviewed	FNDC5_HUMAN
Q8NC67	0.903	reviewed	NETO2_HUMAN
Q14703	0.906	reviewed	MBTP1_HUMAN
Q13291	0.908	reviewed	SLAF1_HUMAN
Q15363	0.91	reviewed	TMED2_HUMAN
Q9UN71	0.91	reviewed	PCDGG_HUMAN
Q96H15	0.921	reviewed	TIMD4_HUMAN
Q9H013	0.924	reviewed	ADA19_HUMAN
Q6NUJ2	0.927	reviewed	CK087_HUMAN
Q9BT76	0.93	reviewed	UPK3B_HUMAN

O75460	0.937	reviewed	ERN1_HUMAN
Q9HC73	0.942	reviewed	CRLF2_HUMAN
A6NDA9	0.943	reviewed	LRIT2_HUMAN
Q9UIW2	0.95	reviewed	PLXA1_HUMAN
O14672	0.959	reviewed	ADA10_HUMAN
Q6UWM9	0.967	reviewed	UD2A3_HUMAN
Q9UKJ0	0.969	reviewed	PILRB_HUMAN
Q9Y5E7	0.972	reviewed	PCDB2_HUMAN
O76036	0.974	reviewed	NCTR1_HUMAN
Q96DU3	0.984	reviewed	SLAF6_HUMAN
Q96PQ0	0.984	reviewed	SORC2_HUMAN
Q93033	0.999	reviewed	IGSF2_HUMAN

Appendix 5

List of commands used for bioinformatics analysis of the sequencing data

The data analysis of the sequencing analysis was carried out by Dr. Carrington who kindly provided the commands he used during the data analysis.

Basecalling:

guppy version 6.4.6

Command 1

guppy_basecaller -i raw_reads_fast5/ -s output_directory_fastq/ -c /mnt/beegfs/software/bioinformatics/release/guppy/guppy_6.4.6 data/dna_r10.4.1_e8.2_400bps_sup.cfg -r --device cuda: 0 --min_qscore 7 --compress_fastq --do_read_splitting --calib_detect

The -c argument tells guppy to use the 'super high quality' base calling model.

The -r argument tells guppy to look for all the .fast5 files contained in the -i input directory

The --min_qscore argument tells guppy to sort basecalled reads into a 'pass' directory if they have a mean qscore of 7 or more. This filters out low quality reads.

The --do_read_splitting argument tells guppy to detect sequencing adapters in the middle of the reads (sometimes reads are ligated together during library prep).

The --calib_detect argument tells guppy that we used the spike-in that comes with the kit, and to find those reads and separate them from our reads.

Demultiplexing:

guppy version 6.4.6

Command 2

guppy_barcoder -i input_fastq -s output_demultiplexed_fastq --barcode kits "SQK-NBD114-24" --detect barcodes --detect_mid_strand_barcodes --enable_trim_barcodes --compress_fastq

The -i argument describes the input and is the output from command 1.

The --barcode kits argument tells guppy which kit we used so it knows what barcode sequences to detect.

The --detect_mid_strand_barcodes argument tells guppy to look for barcode sequences in the middle of the reads, not just the ends.

The --enable_trim_barcodes argument tells guppy to remove the barcode sequences, as well as find them.

Primer detection and removal:

hammerpede (no version numbers, but was installed from github on 22.02.16, yy.mm.dd)

pychopper version 2.7.1

Command 3

hp_bootstrap.py -f screen_primers.fa -o hammerpede_output_all_barcodes 100kSample.fastq

hammerpede creates an HMM model tailored to sequences that you want to find in the particular reads you want to find them in.

-f argument as a fasta file with just two sequences, one is the forward primer and one is the reverse primer. I extended the reverse to include that constant sequence that I emailed you about previously.

The final argument, which has no flag, is a fastq file that contains a random sample of 100,000 reads from all the barcodes (this means that the particular error rates and their biases that will be unique to your run will be accounted for in the HMM model).

Command 4

pychopper -m phmm -g hammerpede_output/screen_primers.hmm -c config.txt -r pychopper_report.pdf -u pychopper_unclassified.fastq.gz -l pychopper_short.fastq.gz -w pychopper_rescued.fastq.gz -S pychopper_stats.txt all_barcodes.fastq.gz pychopper_output_all_barcodes.fastq.gz

pychopper is the software used to find and remove primer sequences. It reports all sorts of statistics, and by default outputs reads only if both primers sequences are found in the correct orientation. The HMM model from command 3 is used by pychoppper. The reads output by command 2 are input here.

-m argument tells pychopper to use HMM models, as opposed to a distance algorithm (like what aligners use).

-g is the HMM model made in command 3.

-c is a single-line text file which tells pychopper what the expected orientation of the primers are.

-u contains the reads which did not contain both primers in the correct orientation.

-l contains the reads which were too short to classify, these would not have aligned well anyway.

-w contains reads where multiple pairs of primers in the correct orientations were found and separated into new reads, again because fragments can become ligated together during library prep.

The final argument, with no flag, is the main output which contains only reads with the primers in the correct orientation, which have also been trimmed off.

Read alignment:

minimap2 version 2.24

Command 5

minimap2 -x map-ont -d minimap2_TMD_index ../reference_fasta/TMD_sequence-list.fasta.gz

Minimap2 is the standard aligner to use for nanopore reads. This first command indexes the reference genome (in our case, fasta file containing the TMD sequences.

-x map-ont tells minimap2 that it should use the nanopore error model

Command 6

minimap2 -a -x map-ont -N 10 minimap2_TMD_index.mmi primers_trimmed_all_barcodes.fastq.gz > TMD.sam

-a tells minimap2 to output alignments in the SAM format.

-x map-ont tells minimap2 that it should use the nanopore error model

-N 10 tells minimap2 to keep up to 10 secondary alignments. I used these to assess minimap2 capability of differentiating between TMD sequences (e.g. they might have shared sequences, functionally conserved), there were very few high quality secondary alignments.

The minimap2_TMD_index.mmi is the index file from command 5.

The primers_trimmed_all_barcodes.fastq.gz is the output from command 4.

The > TMD.sam is command-line syntax which means put the output from this command into a new file called TMD.sam.

Read counting:

salmon version 1.9.0

Command 7

salmon quant --libType A --alignments TMD_barcode01.bam --targets TMD_sequence-list.fasta --ont --output TMD_barcode01_salmon_quant

Salmon is typically used in gene counting applications, but can be used in other applications. Typically people use it in 'read mode', where salmon carries out a pseudo-alignment and read counting in a single step. This is not recommended for nanopore reads, as salmon has been designed for higher accuracy short reads. Instead I used it in 'alignment mode', where it uses alignments to assign reads to TMDs, but can still use its statistical models to resolve ambiguous reads, and its normalisation methods for GC content (i.e. PCR bias).

--libtype A tells salmon to automatically detect the library type. In our case it is single-end and unstranded so would be given the libtype of 'U'. Nanopore libraries are always single-end, and either strand of an amplicon could be sequenced with no meaning attached. However, pychopper reorientates the reads to be in the same orientation as the primer sequences. Salmon detected this and assigned the libtype as 'SF', which means stranded (as all the reads aligned to the forward strand of each TMD sequence), and the F means forward strand. Salmon is good at detecting the libtype so I just told it to figure it out by using 'A' as the argument. I'm glad I did because I wouldn't have immediately anticipated the affect of the reorientation carried out by pychopper.

--alignments TMD_barcode01.bam is just for barcode 1, and was run another 23 times for the other barcodes. Here, it is .bam, but that is just the binary format of the output from command 6.

--ont tells salmon that we have used nanopore reads, so a couple of adjustments are made to its typical assumptions. As nanopore reads are as long as the fragment, whereas with Illumina reads they are limited to typically 100 bp, salmon does not need to account for target sequence length in its read counting normalisations (also, the dynamic range of TMD sequence lengths is not large). Nanopore reads have a different error profile as they are longer and have a higher error rate compared to Illumina reads.

--output TMD_barcode01_salmon_quant contains the read count table, with a value for each TMD. I combined the 'read_number' column from each barcodes salmon output to form the table I used in R where I drew graphs and ran randomised simulations.

References

Adrain, C., Strisovsky, K., Zettl, M., Hu, L., Lemberg, M. K., & Freeman, M. (2011). Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO Reports*, *12*(5), 421–427. https://doi.org/10.1038/embor.2011.50

Atapattu, L., Lackmann, M., & Janes, P. W. (2014). The role of proteases in regulating Eph/ephrin signaling. *Cell Adhesion & Migration*, *8*(4), 294–307. https://doi.org/10.4161/19336918.2014.970026

Baker, R. P., Young, K., Feng, L., Shi, Y., & Urban, S. (2007). Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proceedings of the National Academy of Sciences*, *104*(20), 8257–8262. https://doi.org/10.1073/pnas.0700814104

Battistini, C., Rehman, M., Avolio, M., Arduin, A., Valdembri, D., Serini, G., & Tamagnone, L. (2019). Rhomboid-Like-2 Intramembrane Protease Mediates Metalloprotease-Independent Regulation of Cadherins. *International Journal of Molecular Sciences*, *20*(23), 5958. https://doi.org/10.3390/ijms20235958

Bock, J., Kühnle, N., Knopf, J. D., Landscheidt, N., Lee, J.-G., Ye, Y., & Lemberg, M. K. (2022). Rhomboid protease RHBDL4 promotes retrotranslocation of aggregation-prone proteins for degradation. *Cell Reports*, *40*(6). https://doi.org/10.1016/j.celrep.2022.111175

Bravi, L., Dejana, E., & Lampugnani, M. G. (2014). VE-cadherin at a glance. *Cell and Tissue Research*, *355*(3), 515–522. https://doi.org/10.1007/s00441-014-1843-7

Cadwell, C. M., Su, W., & Kowalczyk, A. P. (2016). Cadherin tales: Regulation of cadherin function by endocytic membrane trafficking. *Traffic*, *17*(12), 1262–1271. https://doi.org/10.1111/tra.12448

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A., Thompson, W. J., & Barres, B. A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource

for understanding brain development and function. *Journal of Neuroscience*, *28*(1), 264–278. Scopus. https://doi.org/10.1523/JNEUROSCI.4178-07.2008

Carmeliet, P., Lampugnani, M.-G., Moons, L., Breviario, F., Compernolle, V., Bono, F., Balconi, G., Spagnuolo, R., Oosthuyse, B., Dewerchin, M., Zanetti, A., Angellilo, A., Mattot, V., Nuyens, D., Lutgens, E., Clotman, F., Ruiter, M. C. de, Groot, A. G., Poelmann, R., … Dejana, E. (1999). Targeted Deficiency or Cytosolic Truncation of the VE-cadherin Gene in Mice Impairs VEGF-Mediated Endothelial Survival and Angiogenesis. *Cell*, *98*(2), 147–157. https://doi.org/10.1016/S0092-8674(00)81010-7

Caveda, L., Martin-Padura, I., Navarro, P., Breviario, F., Corada, M., Gulino, D., Lampugnani, M. G., & Dejana, E. (1996). Inhibition of cultured cell growth by vascular endothelial cadherin (cadherin-5/VE-cadherin). *The Journal of Clinical Investigation*, *98*(4), 886–893. https://doi.org/10.1172/JCI118870

Chen, S., Cai, K., Zheng, D., Liu, Y., Li, L., He, Z., Sun, C., & Yu, C. (2022). RHBDL2 promotes the proliferation, migration, and invasion of pancreatic cancer by stabilizing the N1ICD via the OTUD7B and activating the Notch signaling pathway. *Cell Death & Disease*, *13*(11), Article 11. https://doi.org/10.1038/s41419-022-05379-3

Cheng, T.-L., Lai, C.-H., Jiang, S.-J., Hung, J.-H., Liu, S.-K., Chang, B.-I., Shi, G.-Y., & Wu, H.-L. (2014). *RHBDL2 Is a Critical Membrane Protease for Anoikis Resistance in Human Malignant Epithelial Cells* [Research Article]. The Scientific World Journal. https://doi.org/10.1155/2014/902987

Cheng, T.-L., Wu, Y.-T., Lin, H.-Y., Hsu, F.-C., Liu, S.-K., Chang, B.-I., Chen, W.-S., Lai, C.-H., Shi, G.-Y., & Wu, H.-L. (2011). Functions of Rhomboid Family Protease RHBDL2 and Thrombomodulin in Wound Healing. *Journal of Investigative Dermatology*, *131*(12), 2486–2494. https://doi.org/10.1038/jid.2011.230

Chiasson, C. M., Wittich, K. B., Vincent, P. A., Faundez, V., & Kowalczyk, A. P. (2009). P120-Catenin Inhibits VE-Cadherin Internalization through a Rho-independent Mechanism. *Molecular Biology of the Cell*, *20*(7), 1970–1980. https://doi.org/10.1091/mbc.e08-07-0735

Chipot, C., & Pohorille, A. (1998). Folding and Translocation of the Undecamer of Poly- L -leucine across the Water−Hexane Interface. A Molecular Dynamics Study. *Journal of the American Chemical Society*, *120*(46), 11912–11924. https://doi.org/10.1021/ja980010o

Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., & Hegde, R. S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*, *175*(6), 1507-1519.e16. https://doi.org/10.1016/j.cell.2018.10.009

De Strooper, B., Annaert, W., Cupers, P., Saftig, P., Craessaerts, K., Mumm, J. S., Schroeter, E. H., Schrijvers, V., Wolfe, M. S., Ray, W. J., Goate, A., & Kopan, R. (1999). A presenilin-1-dependent γ-secretase-like protease mediates release of Notch intracellular domain. *Nature*, *398*(6727), Article 6727. https://doi.org/10.1038/19083

DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, *23*(22), 4742–4743. https://doi.org/10.1093/nar/23.22.4742

Dreymueller, D., Pruessmeyer, J., Groth, E., & Ludwig, A. (2012). The role of ADAM-mediated shedding in vascular biology. *European Journal of Cell Biology*, *91*(6), 472–485. https://doi.org/10.1016/j.ejcb.2011.09.003

Dulloo, I., Muliyil, S., & Freeman, M. (2019). The molecular, cellular and pathophysiological roles of iRhom pseudoproteases. *Open Biology*, *9*(3), 190003. https://doi.org/10.1098/rsob.190003

Düsterhöft, S., Kahveci-Türköz, S., Wozniak, J., Seifert, A., Kasparek, P., Ohm, H., Liu, S., Kopkanova, J., Lokau, J., Garbers, C., Preisinger, C., Sedlacek, R., Freeman, M., & Ludwig, A. (2021). The iRhom homology domain is indispensable for ADAM17-mediated TNFα and EGF receptor

ligand release. *Cellular and Molecular Life Sciences: CMLS*, *78*(11), 5015–5040. https://doi.org/10.1007/s00018-021-03845-3

Edwards, D. R., Handsley, M. M., & Pennington, C. J. (2008). The ADAM metalloproteinases. *Molecular Aspects of Medicine*, *29*(5), 258–289. https://doi.org/10.1016/j.mam.2008.08.001

Ferber, E. C., Kajita, M., Wadlow, A., Tobiansky, L., Niessen, C., Ariga, H., Daniel, J., & Fujita, Y. (2008). A Role for the Cleaved Cytoplasmic Domain of E-cadherin in the Nucleus *. *Journal of Biological Chemistry*, *283*(19), 12691–12700. https://doi.org/10.1074/jbc.M708887200

Feske, S. (2010). CRAC channelopathies. *Pflugers Archiv : European Journal of Physiology*, *460*(2), 417–435. https://doi.org/10.1007/s00424-009-0777-5

Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L., & Lemberg, M. K. (2012). Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins. *Molecular Cell*, *47*(4), 558–569. https://doi.org/10.1016/j.molcel.2012.06.008

Fodor, K., Harmat, V., Neutze, R., Szilágyi, L., Gráf, L., & Katona, G. (2006). Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry*, *45*(7), 2114–2121. https://doi.org/10.1021/bi0517133

Freeman, M. (2014). The Rhomboid-Like Superfamily: Molecular Mechanisms and Biological Roles. *Annual Review of Cell and Developmental Biology*, *30*(1), 235–254. https://doi.org/10.1146/annurev-cellbio-100913-012944

Friedmann, E., Hauben, E., Maylandt, K., Schleeger, S., Vreugde, S., Lichtenthaler, S. F., Kuhn, P.-H., Stauffer, D., Rovelli, G., & Martoglio, B. (2006). SPPL2a and SPPL2b promote intramembrane proteolysis of TNFα in activated dendritic cells to trigger IL-12 production. *Nature Cell Biology*, *8*(8), Article 8. https://doi.org/10.1038/ncb1440

Gavard, J., & Gutkind, J. S. (2006). VEGF controls endothelial-cell permeability by promoting the β-arrestin-dependent endocytosis of VE-cadherin. *Nature Cell Biology*, *8*(11), Article 11. https://doi.org/10.1038/ncb1486

Grieve, A. G., Yeh, Y.-C., Chang, Y.-F., Huang, H.-Y., Zarcone, L., Breuning, J., Johnson, N., Stříšovský, K., Brown, M. H., Parekh, A. B., & Freeman, M. (2021a). Conformational surveillance of Orai1 by a rhomboid intramembrane protease prevents inappropriate CRAC channel activation. *Molecular Cell*, *81*(23), 4784-4798.e7. https://doi.org/10.1016/j.molcel.2021.10.025

Grieve, A. G., Yeh, Y.-C., Chang, Y.-F., Huang, H.-Y., Zarcone, L., Breuning, J., Johnson, N., Stříšovský, K., Brown, M. H., Parekh, A. B., & Freeman, M. (2021b). Conformational surveillance of Orai1 by a rhomboid intramembrane protease prevents inappropriate CRAC channel activation. *Molecular Cell*, *81*(23), 4784-4798.e7. https://doi.org/10.1016/j.molcel.2021.10.025

Groot, A. J., & Vooijs, M. A. (2012). The Role of Adams in Notch Signaling. *Advances in Experimental Medicine and Biology*, *727*, 15–36. https://doi.org/10.1007/978-1-4614-0899-4_2

Guna, A., Volkmar, N., Christianson, J. C., & Hegde, R. S. (2018). The ER membrane protein complex is a transmembrane domain insertase. *Science*, *359*(6374), 470–473. https://doi.org/10.1126/science.aao3099

Hogan, P. G., & Rao, A. (2015). Store-operated calcium entry: Mechanisms and modulation. *Biochemical and Biophysical Research Communications*, *460*(1), 40–49. https://doi.org/10.1016/j.bbrc.2015.02.110

Hou, X., Pedi, L., Diver, M. M., & Long, S. B. (2012). Crystal Structure of the Calcium Release–Activated Calcium Channel Orai. *Science*, *338*(6112), 1308–1313. https://doi.org/10.1126/science.1228757

Jászai, J., & Brand, M. (2002). Cloning and expression of Ventrhoid, a novel vertebrate homologue of the Drosophila EGF pathway gene rhomboid. *Mechanisms of Development*, *113*(1), 73–77. https://doi.org/10.1016/S0925-4773(01)00655-4

Johnson, N., Březinová, J., Stephens, E., Burbridge, E., Freeman, M., Adrain, C., & Strisovsky, K. (2017). Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-07556-3

Kappes, J. C., & Wu, X. (2001). Safety Considerations in Vector Development. *Somatic Cell and Molecular Genetics*, *26*(1), 147–158. https://doi.org/10.1023/A:1021082815013

Koch, L., Kespohl, B., Agthe, M., Schumertl, T., Düsterhöft, S., Lemberg, M. K., Lokau, J., & Garbers, C. (2021). Interleukin-11 (IL-11) receptor cleavage by the rhomboid protease RHBDL2 induces IL-11 trans-signaling. *The FASEB Journal*, *35*(3). https://doi.org/10.1096/fj.202002087R

Koren, I., Timms, R. T., Kula, T., Xu, Q., Li, M. Z., & Elledge, S. J. (2018). The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*, *173*(7), 1622-1635.e14. https://doi.org/10.1016/j.cell.2018.04.028

Kowalczyk, A. P., & Nanes, B. A. (2012). Adherens Junction Turnover: Regulating Adhesion Through Cadherin Endocytosis, Degradation, and Recycling. In T. Harris (Ed.), *Adherens Junctions: From Molecular Mechanisms to Tissue Development and Disease* (pp. 197–222). Springer Netherlands. https://doi.org/10.1007/978-94-007-4186-7_9

Kühnle, N., Bock, J., Knopf, J. D., Landscheidt, N., Lee, J.-G., Ye, Y., & Lemberg, M. K. (2019). *Intramembrane protease RHBDL4 interacts with erlin complex to target unstable soluble proteins for degradation* [Preprint]. Cell Biology. https://doi.org/10.1101/848754

Lastun, V. L., Grieve, A. G., & Freeman, M. (2016). Substrates and physiological functions of secretase rhomboid proteases. *Seminars in Cell & Developmental Biology*, *60*, 10–18. https://doi.org/10.1016/j.semcdb.2016.07.033

Lee, J. R., Urban, S., Garvey, C. F., & Freeman, M. (2001). Regulated Intracellular Ligand Transport and Proteolysis Control EGF Signal Activation in Drosophila. *Cell*, *107*(2), 161–171. https://doi.org/10.1016/S0092-8674(01)00526-8

Lemberg, M. K., & Freeman, M. (2007). Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Research*, *17*(11), 1634–1646. https://doi.org/10.1101/gr.6425307

Lemberg, M. K., Menendez, J., Misik, A., Garcia, M., Koth, C. M., & Freeman, M. (2005). Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *The EMBO Journal*, *24*(3), 464–472. https://doi.org/10.1038/sj.emboj.7600537

Lemieux, M. J., Fischer, S. J., Cherney, M. M., Bateman, K. S., & James, M. N. G. (2007). The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. *Proceedings of the National Academy of Sciences*, *104*(3), 750–754. https://doi.org/10.1073/pnas.0609981104

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, *37*(23), 4572–4574. https://doi.org/10.1093/bioinformatics/btab705

Liu, G., Beaton, S. E., Grieve, A. G., Evans, R., Rogers, M., Strisovsky, K., Armstrong, F. A., Freeman, M., Exley, R. M., & Tang, C. M. (2020). Bacterial rhomboid proteases mediate quality control of orphan membrane proteins. *The EMBO Journal*, *n/a*(n/a), e102922. https://doi.org/10.15252/embj.2019102922

Lohi, O., Urban, S., & Freeman, M. (2004). Diverse Substrate Recognition Mechanisms for Rhomboids: Thrombomodulin Is Cleaved by Mammalian Rhomboids. *Current Biology*, *14*(3), 236–241. https://doi.org/10.1016/j.cub.2004.01.025

Lokau, J., Nitz, R., Agthe, M., Monhasery, N., Aparicio-Siegmund, S., Schumacher, N., Wolf, J., Möller-Hackbarth, K., Waetzig, G. H., Grötzinger, J., Müller-Newen, G., Rose-John, S., Scheller, J., & Garbers, C. (2016). Proteolytic Cleavage Governs Interleukin-11 Trans-signaling. *Cell Reports*, *14*(7), 1761–1773. https://doi.org/10.1016/j.celrep.2016.01.053

Manolaridis, I., Kulkarni, K., Dodd, R. B., Ogasawara, S., Zhang, Z., Bineva, G., Reilly, N. O., Hanrahan, S. J., Thompson, A. J., Cronin, N., Iwata, S., & Barford, D. (2013). Mechanism of

farnesylated CAAX protein processing by the integral membrane protease Rce1. *Nature*, *504*(7479), 10.1038/nature12754. https://doi.org/10.1038/nature12754

Maretzky, T., Reiss, K., Ludwig, A., Buchholz, J., Scholz, F., Proksch, E., de Strooper, B., Hartmann, D., & Saftig, P. (2005). ADAM10 mediates E-cadherin shedding and regulates epithelial cell-cell adhesion, migration, and β-catenin translocation. *Proceedings of the National Academy of Sciences*, *102*(26), 9182–9187. https://doi.org/10.1073/pnas.0500918102

Martínez-Salas, E. (1999). Internal ribosome entry site biology and its use in expression vectors. *Current Opinion in Biotechnology*, *10*(5), 458–464. https://doi.org/10.1016/S0958-1669(99)00010-5

Mayer, U., & Nüsslein-Volhard, C. (1988). A group of genes required for pattern formation in the ventral ectoderm of the Drosophila embryo. *Genes & Development*, *2*(11), 1496–1511. https://doi.org/10.1101/gad.2.11.1496

Menschikowski, M., Hagelgans, A., Eisenhofer, G., Tiebel, O., & Siegert, G. (2010). Reducing agents induce thrombomodulin shedding in human endothelial cells. *Thrombosis Research*, *126*(2), e88–e93. https://doi.org/10.1016/j.thromres.2010.05.006

Moin, S. M., & Urban, S. (2012). Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *eLife*, *1*, e00173. https://doi.org/10.7554/eLife.00173

Morishita, H., & Yagi, T. (2007). Protocadherin family: Diversity, structure, and function. *Current Opinion in Cell Biology*, *19*(5), 584–592. https://doi.org/10.1016/j.ceb.2007.09.006

Nanes, B. A., Chiasson-MacKenzie, C., Lowery, A. M., Ishiyama, N., Faundez, V., Ikura, M., Vincent, P. A., & Kowalczyk, A. P. (2012). P120-catenin binding masks an endocytic signal conserved in classical cadherins. *The Journal of Cell Biology*, *199*(2), 365–380. https://doi.org/10.1083/jcb.201205029

Nollet, F., Kools, P., & Van Roy, F. (2000). Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members 1 1Edited by M.

Yaniv. *Journal of Molecular Biology*, *299*(3), 551–572. https://doi.org/10.1006/jmbi.2000.3777

Noy, P. J., Swain, R. K., Khan, K., Lodhia, P., & Bicknell, R. (2016). Sprouting angiogenesis is regulated by shedding of the C-type lectin family 14, member A (CLEC14A) ectodomain, catalyzed by rhomboid-like 2 protein (RHBDL2). *The FASEB Journal*, *30*(6), 2311–2323. https://doi.org/10.1096/fj.201500122R

O'Donnell, J. P., Phillips, B. P., Yagita, Y., Juszkiewicz, S., Wagner, A., Malinverni, D., Keenan, R. J., Miller, E. A., & Hegde, R. S. (2020). The architecture of EMC reveals a path for membrane protein insertion. *eLife*, *9*, e57887. https://doi.org/10.7554/eLife.57887

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics *. *Molecular & Cellular Proteomics*, *1*(5), 376–386. https://doi.org/10.1074/mcp.M200025-MCP200

Ozawa, M., & Kemler, R. (1990). Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. *Journal of Cell Biology*, *111*(4), 1645–1650. https://doi.org/10.1083/jcb.111.4.1645

Pascall, J. C., & Brown, K. D. (2004). Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. *Biochemical and Biophysical Research Communications*, *317*(1), 244–252. https://doi.org/10.1016/j.bbrc.2004.03.039

Paschkowsky, S., Hamzé, M., Oestereich, F., & Munter, L. M. (2016). Alternative Processing of the Amyloid Precursor Protein Family by Rhomboid Protease RHBDL4. *Journal of Biological Chemistry*, *291*(42), 21903–21912. https://doi.org/10.1074/jbc.M116.753582

Patel, S. D., Ciatto, C., Chen, C. P., Bahna, F., Rajebhosale, M., Arkus, N., Schieren, I., Jessell, T. M., Honig, B., Price, S. R., & Shapiro, L. (2006). Type II Cadherin Ectodomain Structures: Implications for Classical Cadherin Specificity. *Cell*, *124*(6), 1255–1268. https://doi.org/10.1016/j.cell.2005.12.046

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), Article 4. https://doi.org/10.1038/nmeth.4197

Pearce, M. M. P., Wormer, D. B., Wilkens, S., & Wojcikiewicz, R. J. H. (2009). An Endoplasmic Reticulum (ER) Membrane Complex Composed of SPFH1 and SPFH2 Mediates the ER-associated Degradation of Inositol 1,4,5-Trisphosphate Receptors *. *Journal of Biological Chemistry*, *284*(16), 10433–10445. https://doi.org/10.1074/jbc.M809801200

Pleiner, T., Tomaleri, G. P., Januszyk, K., Inglis, A. J., Hazu, M., & Voorhees, R. M. (2020). Structural basis for membrane insertion by the human ER membrane protein complex. *Science (New York, N.Y.)*, *369*(6502), 433–436. https://doi.org/10.1126/science.abb5008

Posthaus, H., Dubois, C. M., Laprise, M.-H., Grondin, F., Suter, M. M., & Müller, E. (1998). Proprotein cleavage of E-cadherin by furin in baculovirus over-expression system: Potential role of other convertases in mammalian cells. *FEBS Letters*, *438*(3), 306–310. https://doi.org/10.1016/S0014-5793(98)01330-1

Prakriya, M. (2009). The molecular physiology of CRAC channels. *Immunological Reviews*, *231*(1), 88–98. https://doi.org/10.1111/j.1600-065X.2009.00820.x

Rao, B., Li, S., Yao, D., Wang, Q., Xia, Y., Jia, Y., Shen, Y., & Cao, Y. (2021). The cryo-EM structure of an ERAD protein channel formed by tetrameric human Derlin-1. *Science Advances*, *7*(10), eabe8591. https://doi.org/10.1126/sciadv.abe8591

Rawson, R. B., Zelenski, N. G., Nijhawan, D., Ye, J., Sakai, J., Hasan, M. T., Chang, T. Y., Brown, M. S., & Goldstein, J. L. (1997). Complementation Cloning of S2P, a Gene Encoding a Putative Metalloprotease Required for Intramembrane Cleavage of SREBPs. *Molecular Cell*, *1*(1), 47–57. https://doi.org/10.1016/S1097-2765(00)80006-4

Reiss, K., Maretzky, T., Ludwig, A., Tousseyn, T., de Strooper, B., Hartmann, D., & Saftig, P. (2005). ADAM10 cleavage of N-cadherin and regulation of cell–cell adhesion and β-catenin

nuclear signalling. *The EMBO Journal*, *24*(9), 1762–1762. https://doi.org/10.1038/sj.emboj.7600671

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ryan, M. D., King, A. M. Q., & Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology*, *72*(11), 2727–2732. https://doi.org/10.1099/0022-1317-72-11-2727

Sahin, U., Weskamp, G., Kelly, K., Zhou, H.-M., Higashiyama, S., Peschon, J., Hartmann, D., Saftig, P., & Blobel, C. P. (2004). Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands. *The Journal of Cell Biology*, *164*(5), 769–779. https://doi.org/10.1083/jcb.200307137

Schulz, B., Pruessmeyer, J., Maretzky, T., Ludwig, A., Blobel, C. P., Saftig, P., & Reiss, K. (2008). ADAM10 Regulates Endothelial Permeability and T-Cell Transmigration by Proteolysis of Vascular Endothelial Cadherin. *Circulation Research*, *102*(10), 1192–1201. https://doi.org/10.1161/CIRCRESAHA.107.169805

Sinha, S., Grewal, R. K., & Roy, S. (2018). Chapter Three—Modeling Bacteria–Phage Interactions and Its Implications for Phage Therapy. In S. Sariaslani & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 103, pp. 103–141). Academic Press. https://doi.org/10.1016/bs.aambs.2018.01.005

Spinazzi, M., & De Strooper, B. (2016). PARL: The mitochondrial rhomboid protease. *Seminars in Cell & Developmental Biology*, *60*, 19–28. https://doi.org/10.1016/j.semcdb.2016.07.034

Strisovsky, K. (2016). Rhomboid protease inhibitors: Emerging tools and future therapeutics. *Seminars in Cell & Developmental Biology*, *60*, 52–62. https://doi.org/10.1016/j.semcdb.2016.08.021

Strisovsky, K., Sharpe, H. J., & Freeman, M. (2009). Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Molecular Cell*, *36*(6), 1048–1059. https://doi.org/10.1016/j.molcel.2009.11.006

Su, W., & Kowalczyk, A. P. (2017). The VE-cadherin cytoplasmic domain undergoes proteolytic processing during endocytosis. *Molecular Biology of the Cell*, *28*(1), 76–84. https://doi.org/10.1091/mbc.e16-09-0658

Tang, S., Beattie, A. T., Kafkova, L., Petris, G., Huguenin-Dezot, N., Fiedler, M., Freeman, M., & Chin, J. W. (2022). Mechanism-based traps enable protease and hydrolase substrate discovery. *Nature*, *602*(7898), Article 7898. https://doi.org/10.1038/s41586-022-04414-9

Tichá, A., Stanchev, S., Škerle, J., Began, J., Ingr, M., Švehlová, K., Polovinkin, L., Růžička, M., Bednárová, L., Hadravová, R., Poláchová, E., Rampírová, P., Březinová, J., Kašička, V., Majer, P., & Strisovsky, K. (2017). Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases *. *Journal of Biological Chemistry*, *292*(7), 2703–2713. https://doi.org/10.1074/jbc.M116.762849

Tichá, A., Stanchev, S., Vinothkumar, K. R., Mikles, D. C., Pachl, P., Began, J., Škerle, J., Švehlová, K., Nguyen, M. T. N., Verhelst, S. H. L., Johnson, D. C., Bachovchin, D. A., Lepšík, M., Majer, P., & Strisovsky, K. (2017). General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases. *Cell Chemical Biology*, *24*(12), 1523-1536.e4. https://doi.org/10.1016/j.chembiol.2017.09.007

Urban, S., & Freeman, M. (2003). Substrate Specificity of Rhomboid Intramembrane Proteases Is Governed by Helix-Breaking Residues in the Substrate Transmembrane Domain. *Molecular Cell*, *11*(6), 1425–1434. https://doi.org/10.1016/S1097-2765(03)00181-3

Urban, S., Lee, J. R., & Freeman, M. (2001). Drosophila Rhomboid-1 Defines a Family of Putative Intramembrane Serine Proteases. *Cell*, *107*(2), 173–182. https://doi.org/10.1016/S0092-8674(01)00525-6

Vincent, P. A., Xiao, K., Buckley, K. M., & Kowalczyk, A. P. (2004). VE-cadherin: Adhesion at arm's length. *American Journal of Physiology-Cell Physiology*, *286*(5), C987–C997. https://doi.org/10.1152/ajpcell.00522.2003

Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S., & Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *The EMBO Journal*, *29*(22), 3797–3809. https://doi.org/10.1038/emboj.2010.243

Wang, K., Wei, G., & Liu, D. (2012). CD19: A biomarker for B cell development, lymphoma diagnosis and therapy. *Experimental Hematology & Oncology*, *1*(1), 36. https://doi.org/10.1186/2162-3619-1-36

Wang, Y., Zhang, Y., & Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature*, *444*(7116), 179–180. https://doi.org/10.1038/nature05255

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), Article 11. https://doi.org/10.1038/s41587-021-01108-x

Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K., & Martoglio, B. (2002). Identification of Signal Peptide Peptidase, a Presenilin-Type Aspartic Protease. *Science*, *296*(5576), 2215–2218. https://doi.org/10.1126/science.1070925

Wolfe, M. S., Xia, W., Ostaszewski, B. L., Diehl, T. S., Kimberly, W. T., & Selkoe, D. J. (1999). Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ-secretase activity. *Nature*, *398*(6727), Article 6727. https://doi.org/10.1038/19077

Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S., & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nature Structural & Molecular Biology*, *13*(12), 1084–1091. https://doi.org/10.1038/nsmb1179

Yan, Z., Wang, Z.-G., Segev, N., Hu, S., Minshall, R. D., Dull, R. O., Zhang, M., Malik, A. B., & Hu, G. (2016). Rab11a Mediates Vascular Endothelial-Cadherin Recycling and Controls

Endothelial Barrier Function. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *36*(2), 339–349. https://doi.org/10.1161/ATVBAHA.115.306549

Yang, J., Carvalho, L. A. R., Ji, S., Chen, S., Moreira, R., & Verhelst, S. H. L. (2023). 4-Oxo-β-Lactams as Novel Inhibitors for Rhomboid Proteases. *ChemBioChem*, *24*(21), e202300418. https://doi.org/10.1002/cbic.202300418

Zhou, Y., Moin, S. M., Urban, S., & Zhang, Y. (2012). An Internal Water-Retention Site in the Rhomboid Intramembrane Protease GlpG Ensures Catalytic Efficiency. *Structure*, *20*(7), 1255–1263. https://doi.org/10.1016/j.str.2012.04.022

Zoll, S., Stanchev, S., Began, J., Škerle, J., Lepšík, M., Peclinovská, L., Majer, P., & Strisovsky, K. (2014). Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures. *The EMBO Journal*, *33*(20), 2408–2421. https://doi.org/10.15252/embj.201489367