Mapping the proteomes of endoplasmic reticulum domains by proximity biotinylation



Lilli Hahn

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

This thesis is dedicated to:

All the women out there overcoming gender barriers on step at a time.

Acknowledgements

First of all I would like to thank my supervisor Prof. Pedro Carvalho for providing me with the opportunity to do a PhD in his lab. I am very grateful for the continuous enthusiasm he has for each and every project in the lab, and for the guidance and input he provided for this project. I am thankful for the opportunities to travel to conferences and courses, present my work, be part of stimulating scientific discussions and gain experience in writing and reviewing papers.

I would also like to thank all current and previous lab members - Yuichi, Yoel, Daniela, Joana, Nik, Skyla, Jake, Vitor, Alexis and Daniel for all their support and encouragement, but also for making the lab a welcoming place to work in.
An exceptional thanks goes to Mikey, for always helping out with plasmids, protocols and scientific input. A big thanks also goes to Loges, for all the time he spend discussing results with me, trouble-shooting, coming up with interesting ideas, teaching me microscopy and biochemical fractionations, which played such a big part in my project. Thank you also to Sabrina and Tiaan, for being so patient with me and passing on their proteomics experience.
Finally a huge thanks to Mike Renne, for all the interesting conversations, the support, making the lab a fun place filled with gossip, laughter and good meals. I am not sure how I would have gotten through a PhD without knowing that there are some emergency snacks in the drawer to get me through a late night.

A special shout out goes to Line and Jesus: lab life is only half the fun on your own. Thank you for the many joyful breaks from the lab, for always having an open ear, listening to my frustration over failed experiments, but also sharing so much fun and laughter, for being amazing friends.

A big thanks also to the Dunn School Green Group: I am proud of what we managed to achieve in making the Dunn School a more sustainable research institute over the past few months, hopefully leading the way for many other departments in the university to follow.

My time in Oxford would not have been the same without Teddy Hall and all the friends I made here. Whether it was casually hanging out in the MCR, going for meals, the social events or college sports from chess, tennis to badminton. Teddy truly was a home away from home, where I always felt welcome and it will be dearly missed once I leave Oxford.

In the past four years what often helped me to unwind, travel places, start a podcast, make friends and give talks was chess and my volunteering work. I am grateful for the experience it gave me, the skills I learned, and the self-confidence I gained from it, but most of all the amazing people that were part of it.

A huge thank you goes out to my friends John, Rassin and Andrea: it was a wild-ride, but we made it together. Thanks for making my time in Oxford so much fun and so filled with good memories to keep in mind, from BBQs to hikes and pub nights, to punting and formals. And to Fabi, who even though he joined late, became a fully loved member of the group.
Thank you also to all the members of the Bennett Crescent Playa Club for offering me a home during the lockdowns. I am not a person that deals well with being trapped in a house, but it was more bearable with you guys.
Also, thanks to the old members of 101 Courtland Road, in particular Pawala and Olga, because after all it was the people that made the difference between a home and a house.

I'd also like to thank my friends from home and undergrad for their enduring support and for being part of the journey, even from a distance. In particular Ladi and Manu, for always keeping me grounded, for coming to the UK when I was too caught up in work and couldn't make it to Germany. A big thank you also goes to Alice: even though we didn't end up in the same town for a PhD, it was great to have someone who was on the same journey. You are honestly such an inspiration to me.

Thank you also to Martin! For being there from day one of the journey. Thank you for supporting me in anything (and I mean really anything!) I do. Thank you for always believing in me and making me a better person every day. For always learning from me and teaching me new things. For challenging me and my ideas, making me think and re-think things. For bringing some calm moments into my life and reminding me to take it slow sometimes, and for always being an optimist and make me see the world in a kinder way.

Last but not least, I cannot thank my family enough! Thank you to my aunt Sziv, for always cheering me on and believing in me. Thank you to my mum for

always being there for me, for being the most amazing role model and for raising me to be the strong and independent woman I am today. Thank you dad, for the continuous support, acting as a constant safety net, and for teaching me that there is a solution to any problem. Thank you also to my two amazing brothers Niklas and Hannes: I wouldn't be where I am today without you. Your support, but also your honesty and critical words, and the way you challenged me over and over again, made me the person I am today. Thanks a ton to all of you for being part of this journey, through all the happiness, but also the occasional stress and frustration. I am eternally grateful to have you all in my life!

Funding and Authorship Declaration

This work received support from multiple funding sources, which are shown here in alphabetical order:

- Medical Sciences Graduate School Studentship
- Sir William Dunn School of Pathology Graduate Studentship
- Studienstiftung des Deutschen Volkes
- Oxford University COVID-19 Scholarship Extensions Fund (Extension)

I declare that this thesis has never been submitted, either in the same or a different form, to this or another University for a degree or qualification. In some instances, the work of others is presented in this thesis. In particular, I wish to acknowledge that:

- The NLS-GFP-APEX2 cell line was generated by Viorica Liebe
- The code for the data analysis in R was written by Tiaan Heunis
- The MBRL KO, RNF185 KO, and the MBRL-mNG TMUB1/2 KO cell lines were generated by Michael Van De Weijer

The work in this thesis was carried out during my time as a DPhil student and the Sir William Dunn School of Pathology, under the supervision of Prof. Pedro Carvalho.

Approx. word count: 45,000

Lilli Hahn (22/04/2022)

Abstract

The endoplasmic reticulum (ER) is the largest organelle in the cell and is involved in a wide range of functions such as biogenesis of membrane and secretory proteins, calcium homeostasis and lipid synthesis. To be able to carry out these functions, the ER depends on an elaborate architecture consisting of a network of flat cisternae-like sheets and narrow tubules. The ER also assembles the nuclear envelope (NE) which is formed by two membranes, the inner and outer nuclear membrane (INM and ONM), that separate the cytoplasm from the nucleoplasm. Despite continuous with each other, these ER domains serve different functions. Consistent with their diverse roles, ER domains also appear to harbour distinct proteins. However, the complete proteomes of the various ER domains have not been systematically analysed and annotated. The primary objective of this project was to map the protein compositions of the distinct ER domains. To this aim, an APEX2-based proximity biotinylation assay coupled to subcellular fractionation was established and a data analysis pipeline developed. The assay was capable to identify many of the well-characterised INM, ER tubule and ER sheet proteins in their anticipated domain. Additionally, a range of *bona fide* proteins, that have not previously been shown to enrich at any of these domains, were identified. Therefore, these data provide a valuable resource to enhance our understanding of ER organisation. Additionally, the developed assay and data analysis pipeline can serve as a discovery platform to probe how the protein compositions of the ER domains change in response to disease-associated mutations or upon external perturbances. As an example, the effect of *LMNA* knockout on the protein composition of the inner nuclear membrane was investigated, whereby a small number of proteins were identified to be affected by depletion of A-type lamins.

Contents

List of Figures xiii

List of Tables xv

List of Abbreviations xvi

1 Introduction 1
- 1.1 The endoplasmic reticulum . . . 2
- 1.2 The nuclear envelope . . . 6
- 1.3 The inner nuclear membrane proteome . . . 7
- 1.4 Targeting proteins to the INM . . . 10
- 1.5 Protein degradation at the INM . . . 11
- 1.6 The nuclear lamina . . . 16
- 1.7 NE breakdown and re-assembly . . . 17
- 1.8 Proximity-dependent biotinylation approaches to study organelle domains . . . 18
- 1.9 Aim of the study . . . 20

2 Materials and methods 22
- 2.1 List of materials . . . 23
  - 2.1.1 General and molecular biology reagents . . . 23
  - 2.1.2 List of commercial kits . . . 24
  - 2.1.3 List of cell lines . . . 24
  - 2.1.4 List of plasmids . . . 24
  - 2.1.5 List of primers . . . 25
  - 2.1.6 List of gRNAs . . . 26
  - 2.1.7 List of primary antibodies . . . 26
  - 2.1.8 List of secondary antibodies . . . 27
  - 2.1.9 List of solutions and buffers . . . 28
  - 2.1.10 List of antibiotics . . . 29
- 2.2 Protocols . . . 30
  - 2.2.1 Cell Culture Methods . . . 30
    - 2.2.1.1 Maintenance of human cell lines . . . 30

2.2.1.2 Cryopreservation of human cell lines 30
2.2.1.3 Generating stable cell lines using the Flp-In system 30
2.2.1.4 Generating of CRISPR/Cas9-mediated knockout cells 31
2.2.2 Molecular biology methods 32
2.2.2.1 Restriction-enzyme based cloning 32
2.2.2.2 Designing gRNAs for CRISPR/Cas9-mediated knockout 32
2.2.2.3 Polymerase chain reaction 32
2.2.2.4 Agarose gel electrophoresis 33
2.2.2.5 Purification of DNA from agarose 33
2.2.2.6 Determining the nucleic acid concentration 33
2.2.2.7 Gibson ligation 33
2.2.2.8 Transformation of competent *E. coli* 33
2.2.2.9 Small-scale preparation of plasmid DNA 34
2.2.3 Protein biochemistry methods 35
2.2.3.1 Proximity-biotinylation assay 35
2.2.3.2 Preparation of whole-cell lysate 35
2.2.3.3 SDS-polyacrylamide gel electrophoresis (PAGE) . 35
2.2.3.4 Western blotting and detection 36
2.2.3.5 Immunofluorescence microscopy after biotinylation 36
2.2.3.6 Differential membrane permeabilisation 37
2.2.3.7 Sucrose gradient fractionation 37
2.2.3.8 Immunoprecipitation 38
2.2.4 Proteomic analysis 40
2.2.4.1 On-bead trypsin digest 40
2.2.4.2 In-solution trypsin digest 40
2.2.4.3 C18 Sep-Pak purification of peptides 41
2.2.4.4 LC-MS/MS Settings 41
2.2.4.5 MaxQuant Analysis 41
2.2.4.6 Perseus Analysis 42
2.2.4.7 R analysis 43

3 Developing an assay to study the protein composition of ER domains 44
3.1 Introduction 45
3.1.1 An overview over ER domains 45
3.1.2 Studying organelles by proximity biotinylation 46
3.2 Rational 48

3.3 Results . . . 49
3.3.1 Designing and cloning APEX2-fusion constructs . . . 49
3.3.2 Choosing a model cell line . . . 53
3.3.3 Generating stable FlpIn293 APEX2-fusion cells lines . . . 55
3.3.4 Testing and optimising protein expression . . . 55
3.3.5 Investigating the role of proteasomal degradation on protein levels . . . 57
3.3.6 Validating APEX2-PDB of different bait proteins . . . 59
3.3.7 Quality control of the APEX2-PDB system . . . 60
3.3.8 Determining quenching efficiency . . . 61
3.3.9 Validating APEX2-fusion protein localisation . . . 65
3.3.10 Optimising sample preparation for LC-MS/MS . . . 72
3.3.11 Enriching for membrane proteins . . . 75
3.4 Discussion . . . 81
3.4.1 Cloning strategy and cell line generation . . . 81
3.4.2 LC-MS/MS sample preparation optimisation . . . 85
3.4.3 Coupling proximity-dependent biotinylation to fractionation 86
3.4.4 Concluding remarks . . . 88

4 Determining the protein composition of ER domains 89
4.1 Rational . . . 90
4.2 Results . . . 92
4.2.1 A proteomic data set to study ER protein domains . . . 92
4.2.2 Optimising data analysis for PDB-fraction LC-MS/MS data 93
4.2.3 Investigating the INM proteome using enrichment analysis 104
4.2.4 Designing a tool to identify putative INM proteins . . . 105
4.2.5 Investigating protein compositions of ER sheets and tubules 110
4.2.6 An overview of methods to validate INM protein localisation 114
4.2.7 Implementing a differential permeabilisation assay . . . 116
4.2.8 Validating putative INM proteins . . . 121
4.2.8.1 ASPH . . . 121
4.2.8.2 ELOVL5 . . . 122
4.2.8.3 PGRMC1 and PGRCM2 . . . 124
4.2.8.4 TMEM109 . . . 125
4.2.9 Additional interesting putative INM proteins . . . 127
4.2.9.1 CERS6 . . . 128
4.2.9.2 MLF2 . . . 129
4.2.9.3 ARL6IP6 . . . 129
4.2.9.4 HACD2 . . . 129

4.2.9.5 TER . . . 130
4.2.9.6 SMPD4 . . . 131
4.2.9.7 REEP4 . . . 133
4.2.10 Validating ER tubule proteins . . . 134
4.2.10.1 The RNF185-Membralin complex . . . 134
4.3 Discussion . . . 139
4.3.1 Optimising a workflow to investigate the protein composition of distinct ER domains . . . 139
4.3.2 Deciphering the proteomic composition of the ER . . . 142
4.3.3 Proteomic characterisation of the INM . . . 143
4.3.4 Validating putative INM, ER sheet and ER tubule proteins 148
4.3.5 Interesting observations to follow up . . . 149
4.3.6 Concluding remarks . . . 151

5 Investigating the INM protein composition in disease phenotypes 153
5.1 Rational . . . 154
5.2 Results . . . 156
5.2.1 Generating *LMNA* and *ZMPSTE24* KO cell lines . . . 156
5.2.2 Validating *LMNA* and *ZMPSTE24* KO cell lines by WB . 156
5.2.3 Validating PDB localisation in *LMNA* KO cells using IF . 157
5.2.4 Validating subcellular fractionation of nuclei from *LMNA KO* cells . . . 158
5.2.5 Proteomic analysis of *LMNA* KO cells . . . 159
5.3 Discussion . . . 165
5.3.1 Adapting our PDB assay to study laminopathies . . . 165
5.3.2 Depletion of A-type lamins does not result in global changes of the INM protein composition . . . 166

6 Concluding remarks and future perspectives 168
6.1 Developing an assay to study the ER domains *in situ* . . . 169
6.2 *Bona fide* INM, ER tubule and ER sheet proteins . . . 171
6.3 Using the PDB assay to investigate the protein composition of ER domains in various physiological conditions . . . 174
6.4 Concluding remarks . . . 176

References 178

Appendices

A Appendix 1 **195**
A.1 Proteomic data collection 195
A.2 Organelle marker proteins 196
A.3 SUN2 profile plot hits 199
A.4 Emerin profile plot hits 201

List of Figures

1.1 Topology of ER shaping proteins. 3
1.2 The inner nuclear membrane proteome 8
1.3 The diffusion-retention model 12
1.4 Selective protein degradation at the INM 14
1.5 Autophagy-dependent degradation at the INM 15

3.1 Topology of APEX2-bait proteins. 51
3.2 Cloning strategy 52
3.3 APEX2 PDB in Flp-In T-REx HEK293 & T-REx HeLa cell lines 54
3.4 Integration of the Flp-In T-REx vector 56
3.5 Titration of tetracycline concentration of APEX2-fusion protein expression 57
3.6 APEX2-fusion protein stability assay 58
3.7 Expression and activity of APEX2 fusion protein 59
3.8 Validating the APEX2 proximity dependent biotinylation system. 60
3.9 Investigating quenching efficiency 62
3.10 Investigating quenching efficiency 64
3.11 Testing the localisation of ER APEX2-bait proteins 66
3.12 Testing the localisation of INM APEX2-bait proteins. 68
3.13 Testing the localisation of soluble APEX2-bait proteins. 69
3.14 Testing the localisation of INM APEX2-bait proteins. 71
3.15 On-bead versus in-solution trpysin proteomic sample preparation . 74
3.16 Coupling proximity-dependent biotinylation to subcellular fractionation 76
3.17 Subcellular fractionation to enrich for ER and NE 77
3.18 Cellular fractionation enriches for known INM proteins 78
3.19 Coupling proximity-biotinylation to cellular fractionation 79

4.1 Bait proteins target to all domains of the ER. 93
4.2 Using PCA assays to assess different data analysis methods. . . . 96
4.3 Investigating Perseus t-test analysis of raw intensity values using ER and INM marker protein enrichment. 98

4.4 Investigating Perseus t-test analysis of LFQ intensity values using ER and INM marker protein enrichment. 99
4.5 Investigating R-based Limma analysis of raw intensity values using ER and INM marker protein enrichment. 100
4.6 Comparing statistical testing using R-based limma and Perseus-based student t-test analysis 101
4.7 Gene Ontology cellular compartment classification of the $Nup153_{1-144}$ NE and Climp63 CM enriched proteins. 102
4.8 Investigating the INM and ER protein composition by GO-enrichment analysis using the Sec61β bait. 106
4.9 Investigating the INM and ER protein composition by Limma analysis using the Sec61β bait. 107
4.10 Combining FC analysis across multiple probes to identify *bona fide* INM proteins. 109
4.11 Investigating the protein composition of ER tubules and ER sheets. 112
4.12 Investigating the protein composition of ER tubules and ER sheets using tubular and sheet marker proteins. 113
4.13 Differential membrane permeabilisation in HeLa cells 118
4.14 Optimising the duration of differential membrane permeabilisation in FlpIn293 cells 119
4.15 Optimising differential membrane permeabilisation conditions in FlpIn293 cells 120
4.16 Validating the localisation of ASPH and ELOVL5 123
4.17 Validating the localisation of PGRMC1 and PGRMC2 126
4.18 Validating the localisation of TMEM109. 127
4.19 Mammalian FA elongation cycle 131
4.20 The REEP protein family. 135
4.21 ER morphology analysis of the MBRL-RNF185 complex using Climp63. 137
4.22 ER morphology analysis of the MBRL-RNF185 complex using Rtn4a. 138

5.1 Lamin A post-translational processing 155
5.2 Validating the *LMNA* KO cell lines by WB 157
5.3 Validating PDB in the LMNA KO cell lines 158
5.4 Validating subcellular fractionation in *LMNA* KO cells. 159
5.5 Investigating global changes in the protein composition of the INM of *LMNA* KO cells. 160
5.6 Investigating global changes in the protein composition of the bulk ER of *LMNA* KO cells. 161

List of Tables

2.1 General and molecular biology reagents 23
2.2 List of commercial kits 24
2.3 List of secondary antibodies 24
2.4 List of plasmids 24
2.5 List of primers 25
2.6 List of gRNAs 26
2.7 List of primary antibodies 26
2.8 List of secondary antibodies 27
2.9 List of solutions and buffers 28
2.10 List of antibiotics 29

3.1 List of constructs 49

A.1 Proteomic data collection 195
A.2 Organelle marker proteins 196
A.3 SUN2 reference profile top 50 hits 199
A.4 Emerin reference profile top 50 hits 201

List of Abbreviations

AB . . . . . Ammonium bicarbonate

ACN . . . . . Acetonitrile

APEX . . . . . Ascorbate peroxidase

APC . . . . . Anaphase-promoting complex

AP-MS . . . . Affinity purification and mass spectrometry

BAF . . . . . Barrier-to-autointegration factor

BAR . . . . . Biotinylation by antibody recognition

BB . . . . . Blocking buffer

BiFC . . . . . Bimolecular fluorescence complementation

BP . . . . . Biotin-phenol

CB . . . . . CB-5083 p97 inhibitor

CC . . . . . Cellular compartment

Cdk . . . . . Cyclin-dependent kinase

CerS . . . . . Ceramide synthase

CM . . . . . Cytoplasmic membrane

COP . . . . . coat protein

CRISPR . . . Clustered regularly interspaced short palindromic repeats

DMEM . . . . Dulbecco's Modified Eagle Medium

DMSO . . . . . Dimethyl sulfoxide

DTT . . . . . Dithiothreitol

EDMD . . . . Emery-Dreifuss muscular dystrophy

EGTA . . . . . Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid

EL . . . . . Elution

EM . . . . . Electron microscopy

EDMD . . . . Emery-Dreifuss muscular dystrophy

ER . . . . . Endoplasmic reticulum

ERAD . . . . . ER-associated degradation

ERES . . . . . ER exit sites

ERGIC . . . . ER-Golgi intermediate compartment

EV . . . . . Empty vector control

FASP . . . . . Filter-aided sample preparation

FA . . . . . Formic acid

FC . . . . . Fold change

FKBP . . . . . 12-kDa FK506 binding protein

FlpIn293 . . . Flp-In T-REx Human embryonic kidney 293

FRB . . . . . FKBPerapamycin binding domain

FW . . . . . Forward

gRNA . . . . . Guide RNA

GFP . . . . . Green fluorescent protein

GO . . . . . Gene ontology

HB . . . . . Hypotonic buffer

HH3 . . . . . Histone H3

HGPS . . . . . Hutchinson-Gilford progeria syndrome

HRP . . . . . Horseradish peroxidase

IB . . . . . Immunoblot

IF . . . . . Immunofluorescence

IN . . . . . Input

IP . . . . . Immunoprecipitation

iPSC . . . . . Induced pluripotent stem cells

INM . . . . . Inner nuclear membrane

KO . . . . . Knockout

KOAc . . . . . Potassium acetate

LBR . . . . . Lamin B receptor

LCMD . . . . Left-censored missing data

LC-MS/MS . Liquid chromatography tandem mass spectrometry

LEM . . . . . LAP2, emerin, MAN1 domain

LFQ . . . . . Label free quantification

LIMMA . . . . Linear models for microarray data

LINK . . . . . Linker of nucleoskeleton and cytoskeleton

LMNA . . . . Lamin A

LSP . . . . . Low speed pellet

MCSs . . . . . Multiple cloning site

MEF . . . . . Mouse embryonic fibroblast

mNG . . . . . mNeonGreen

MT . . . . . Microtubule

MW . . . . . Molecular weight

NC . . . . . Nuclear content

NE . . . . . Nuclear envelope

NEBD . . . . . Nuclear envelope breakdown

NLS . . . . . Nuclear localisation signal

NPC . . . . . Nuclear pore complex

NTR . . . . . Nuclear transport receptors

NVJ . . . . . Nucleus-vacuole junction

nSMases . . . Neutral sphingomyelinases

Nup . . . . . Nucleoporin

ONM . . . . . Outer nuclear membrane

PAGE . . . . . Polyacrylamide gel electrophoresis

PBS . . . . . Phosphate-buffered saline

PCA . . . . . Principal component analysis

PCR . . . . . Polymerase chain reaction

PDB . . . . . Proximity-dependent biotinylation

PI . . . . . Protein inhibitor cocktail

PL . . . . . Proximity labelling

PMSF . . . . . Phenylmethylsulfonyl fluoride

PNS . . . . . Post nuclear supernatant

POI . . . . . Protein of interest

p.adj . . . . . Adjusted p-value
RIPA . . . . . Radioimmunoprecipitation assay buffer
RNAi . . . . . RNA interference
RT . . . . . Room temperature
RV . . . . . Reverse
SBB . . . . . Sample binding buffer
SCF Skp1/Cullin1/F-box
SD . . . . . Standard deviation
SDS . . . . . Sodium dodecyl sulfate
SDS-PAGE . . SDS-polyacrylamide gel electrophoresis
SE . . . . . Sterol ester
SILAC . . . . . Stable isotope labelling by amino acids in cell culture
SQLE . . . . . Squalene monooxygenase
SRP . . . . . Signal recognition particle
SRPR . . . . . Signal recognition particle receptor
TA . . . . . Tail-anchored
TAE . . . . . Tris-Acetate-EDTA buffer
TET . . . . . Tetracycline
TF . . . . . T-REx^{TM}-Flp-In^{TM}
TFA . . . . . Trifluoroacetic acid
TM . . . . . Transmembrane
TMH . . . . . Transmembrane helix
TMD . . . . . Transmembrane domain
TROLOX . . . 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
TX . . . . . Triton X-100
UN . . . . . Unbound
VLCFA . . . . Very long chain fatty acid
WB . . . . . Western blot
WCL . . . . . Whole cell lysate
WT . . . . . Wild type
ZMPSTE24 . Zinc metalloproteinase Ste24 homolog

There is only one thing that makes a dream impossible to achieve: the fear of failure.

— Paulo Coelho

1 Introduction

Contents

1.1 The endoplasmic reticulum 2
1.2 The nuclear envelope 6
1.3 The inner nuclear membrane proteome 7
1.4 Targeting proteins to the INM 10
1.5 Protein degradation at the INM 11
1.6 The nuclear lamina 16
1.7 NE breakdown and re-assembly 17
1.8 Proximity-dependent biotinylation approaches to study organelle domains 18
1.9 Aim of the study 20

1.1 The endoplasmic reticulum

The endoplasmic reticulum (ER) is the largest organelle in eukaryotic cells. Formed by a continuous membrane, it spans from the nuclear envelope (NE) all the way to the periphery of the cell. Its functions include the synthesis of membrane and secretory proteins, calcium homeostasis and lipid synthesis[1]. Besides, the ER forms membrane contact sites with a wide range of organelles. These sites of membrane tethering play an important role in the transfer of lipids, ions and macromolecules and act as transmission sites for signalling information and force[2–4].

To be capable to perform such a wide-range of functions, the ER depends on its elaborate architecture, consisting of tubules, sheets and the NE[1,5–8]. ER tubules stretch far into the cytoplasm. They contain few ribosomes on their surface and are characterised by high membrane curvature and three-way junctions, resulting in a highly interconnected network[1,5–8]. Under certain conditions, ER tubules can pack into dense arrays, termed ER matrixes, close to the cell periphery[9]. ER tubules also exhibit some intrinsic structure. Data from live cell high-speed ultra resolution microscopy studies suggest that ER tubules are composed of discrete nanodomains, characterised by the periodic distribution of distinct ER membrane and luminal reporter constructs[10]. One possible explanation for these observations could be the existence of ultra thin tubules, that lead to the exclusion of luminal and membrane proteins with large interluminal domains, yet more research is needed in this area[11].

Various proteins involved in ER tubule formation have been identified. Reticulons and REEP proteins are two classes of hairpin-like protein families facilitating high membrane curvature of ER tubules, as well as at the fusion sites of the INM and ONM, and the edges of ER sheets[12–15]. Overexpression of reticulons results in increased ER tubulation in cells, whereas deletion leads to the loss of ER tubules[12]. Mechanistically, the hairpin-like morphology of these proteins is thought to induce membrane curvature by wedging of the transmembrane helices (TMH) into the outer leaflet of the phospholipid bilayer[12] (figure 1.1). Novel biochemical, genetic

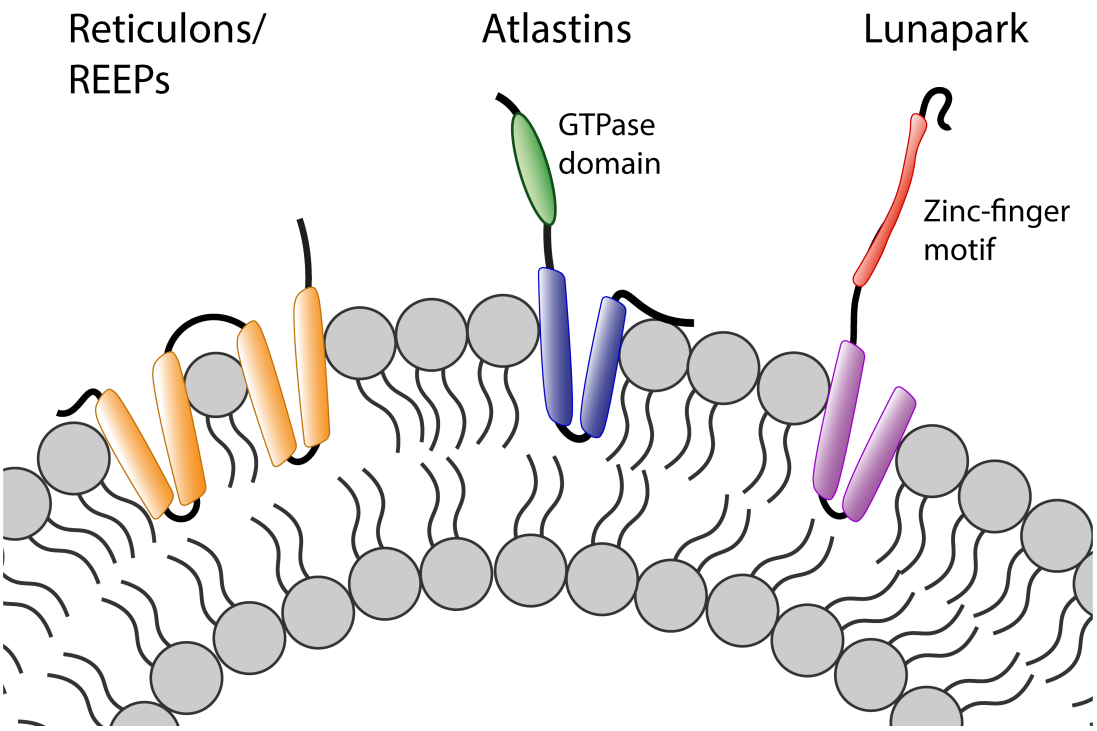


Figure 1.1: Topology of ER shaping proteins. Schematic showing the morphology of reticulon/REEP proteins (orange), atlastins (blue) and lunapark (purple). All three proteins adopt a hairpin-like structure that wedges into the outer leaflet of the phospholipid bilayer (depicted in grey), thereby inducing membrane curvature. Atlastins and lunapark are mostly localised to three-way junctions in the ER. Besides the hairpin-like TMHs, atlastins contain a N-terminal GTPase domain (green), as well as an APH, that has been shown to collaboratively with the TMD result in membrane curvature. Lunapark in contrast has a zinc-finger motif (red) localised in its N-terminal tail.

and imaging data also indicates that membrane curvature is further enhanced by amphipathic helixes (APH), as well as dimersation of REEPs and reticulons[13]. Other proteins involved in membrane shaping of the ER are atlastins, lunapark and TMCC3, which are required for the stabilisation of three-way junctions[16–20]. These proteins also have a hairpin-like structure (figure 1.1). Atlastin, similar to the mitofusin GTPase involved in homeotypic fusion in the mitochondria, also contains a N-terminal APH and was shown to not only be involved in membrane fusion, but also in the subsequent stabilisation of the tubular network[16,21].

ER sheets are generally found in closer proximity to the nucleus and can be described as flat membrane bilayers which are stacked onto each other. How ER sheets are formed is mechanistically less clear. Initially it was hypothesised that

ER sheets are formed in the absence of tubule-shaping proteins. In line with this, early modeling data predicted that all ER morphology can be formed with the help of two types of curvature-stabilising proteins: straight and negatively curved-type proteins[22]. This hypothesis was further supported by the observation that deletion of reticulons resulted in the disappearance of tubules and their conversion into ER sheets[23]. Yet, other data indicated that individual proteins may play a role in ER sheet formation, as overexpression of Climp63, known to act as a luminal ER spacer, results in strong proliferation of ER sheets[23]. Another class of proteins with ER sheet stabilising functions are ribosomes, as stripping ribosomes of ER surfaces results in the conversion of ER sheets into ER tubules[24]. One newly detected feature observed in ER sheets by live-cell stimulated emission depletion microscopy are highly dynamic nanoholes[25]. Whereas their function is not yet fully understood, one hypothesis is that they might play a role in postmitotic nuclear pore complex assembly[25,26].

Besides ER tubule and sheet forming proteins, microtubules (MTs) and the actin skeleton plays a major role in determining ER morphology. ER tubules are known to grow along MTs, and depolymerisation of these MTs results in shrinkage of ER tubules[1,6]. Various proteins responsible for the interaction of the ER and MTs have been identified, including the MT-binding domain containing proteins Climp63, p180, kinectin and REEP1. Thereby different ER proteins preferentially interact with MT populations containing distinct post-translational modifications, enabling dynamic and changeable interactions[27].

Whereas classified in distinct ER domains, the ER is a highly dynamic organelle that undergoes continuous remodeling including vesicle budding, membrane fission and fusion, formation of new junctions, as well as the transition from tubules to sheets and vice versa[28]. New state of the art high-speed super resolution microscopy experiments also indicate more subtle reformations in ER morphology, such as rearrangement, growth and shrinkage of nanodomains and nanoholes[10,27].

While advances in imaging technology greatly improved our understanding of ER morphology and the mechanisms involved in establishing and maintaining

this elaborate structure, how different ER functions are spatially organised within the ER is still less clear. Yet, there are some indications that a form of spatial organisation may exist.

One such example is protein synthesis, which requires the presence of ribosomes. Imaging data indicates that ER sheets have a higher ribosome density compared to tubules in various cells types[29]. Additionally, secretory cells show a higher ratio of ribosome-studded sheets to ER tubules[6]. One might also speculate that the increased luminal to surface area ratio of ER sheets may be beneficial and well suited for processes like protein folding. In line with this idea are observations that yeast cells display increased amounts of ER cisternae during ER stress, when high protein folding capacities are required[30]. Yet, whether protein synthesis is fully excluded from the ER tubules is still to be determined.

A large amount of proteins synthesised in the ER needs to be trafficked to other cellular localisations or secreted. Thereby protein transport is facilitated in coat-protein II (COPII) vesicles from the ER to the Golgi, upon COPII polymerisation. This vesicle formation thereby takes place in distinct ribosome-free domains termed ER exit sites (ERES), which appear as scattered punca across the ER, yet seem to preferentially localise to areas of high ER curvature[1,31–34]. Additionally, the ER-Golgi intermediate compartment (ERGIC), involved in protein trafficking, is located in close proximity to the ERES[31,32,35].

Given that sheets may play a predominant rule in protein synthesis and trafficking, there has been some speculation in the field that ER tubules may be responsible for lipid synthesis and calcium storage, yet our understanding is primordial. There is some indication of spatial organisation of calcium storage across the ER, in particular at the contact sites with other organelles. Indeed, ER-PM contact sites were shown to be involved in regulating cellular Ca^{2+} levels[36]. Additionally contact sites with organelles such as the mitochondria play a role in Ca^{2+} flux[5]. Another interesting observation supporting the speculation that ER tubules may be the predominant sites of Ca^{2+} flux and storage is that muscle cells, which required Ca^{2+} during contraction, contain large amounts of ER tubules[6].

Yet, more experimental data is required to further draw a firm conclusion of whether there is any spatial preference of Ca^{2+} storage in a distinct ER domain.

Another important function of the ER is lipid synthesis. Indeed, whereas organelles often have a distinct lipid composition, most of the lipids are synthesised in the ER. This creates the need for lipid flux between the ER and organelles such as lipid droplets, mitochondria and the plasma membrane. Interstingly, some proteins required for lipid synthesis have found to be enriched at membrane contact sites, indicating that membrane contact sites may not only play a role in lipid transfer, but also the lipid synthesis itself[1,33,37–39]. As membrane contact sites may be more predominant for ER tubules, this again leads to speculation that foci of enriched lipid synthesis exist, but it is yet unclear where the bulk of *de novo* synthesis takes place.

1.2 The nuclear envelope

A distinctive feature of eukaryotic cells is their nucleus. The NE is a highly specialised ER domain that surrounds the cells DNA and acts as a physical barrier between the cytoplasm and nucleoplasm, which results in the separation of transcription and translation. Besides acting as a physical barrier, the NE has been shown to play a role in a wide range of functions including chromatin organisation, gene regulation, DNA repair and cell signaling[40].

The NE is made up of two lipid bilayers, termed the inner and the outer nuclear membrane (INM and ONM), enclosing between them the ER lumen. Interestingly, whereas the ONM harbours similar proteins than the bulk ER, the INM contains a distinct set of proteins, many of which interact with the chromatin or the nuclear lamina, and are crucial for the versatile functions of the INM[41].

Fusions between the two membrane exist, where the INM and ONM merge with each other, thereby forming the pore membrane which encircles the nuclear pore complexes (NPC). NPCs are exceptionally large protein complexes, which are made up of multiple copies of more than 30 different proteins, collectively known as nucleoporins or Nups[42]. NPCs play an important role in preventing

free-diffusion of large proteins, thereby acting as a selective filter between the cytoplasm and the nucleoplasm for both soluble and membrane proteins.

In metazoans, the INM is lined by the nuclear lamina. The nuclear lamina is a dense protein meshwork, consisting predominantly of A-type and B-type lamins, that plays an important role in providing mechanical rigidity and stiffness to the NE, while also contributing to the maintenance of NE integrity[43].

1.3 The inner nuclear membrane proteome

The INM, and its unique protein composition, have an important and versatile role in nuclear organisation (figure 1.2). INM proteins have been shown to directly interact with chromatin, thereby tethering the DNA to the nuclear periphery. Importantly, positioning of chromatin to the nuclear periphery has been shown to influence gene expression, as well as DNA replication, repair and recombination[44,45]. Besides influencing the positioning of chromatin within the nucleus, INM proteins also play a role in regulating the position of the NE within cells, thereby impacting cell migration and cell division[46]. Furthermore, new data indicates that the INM might be metabolically active and involved in lipid synthesis[47–50].

Various proteomic analysis assays in different cell types have resulted in the identification of more than 100 INM proteins in mammalian cells, of which some are expressed in a cell-type specific manner[51–55]. Many of these INM proteins are not exclusive to the INM, but localise to both the INM and the bulk ER. Whereas no systematic data is available for metazoan cells, a genome-wide split GFP screen of all predicted transmembrane domain proteins in *S. cerevisiae* demonstrated that as many as ∼35% of all ER proteins reach the INM[56].

Some of the best characterised INM proteins in metazoans are the LEM (Lap2, emerin, Man1)-domain proteins. These proteins are non-related, but share a common LEM-domain, that is ∼45 amino acid long and folds into two α-helixes. LEM-domain proteins bind lamins and interact with the highly conserved chromatin-binding protein Barrier to Autointegration Factor (BAF)[57,58]. They

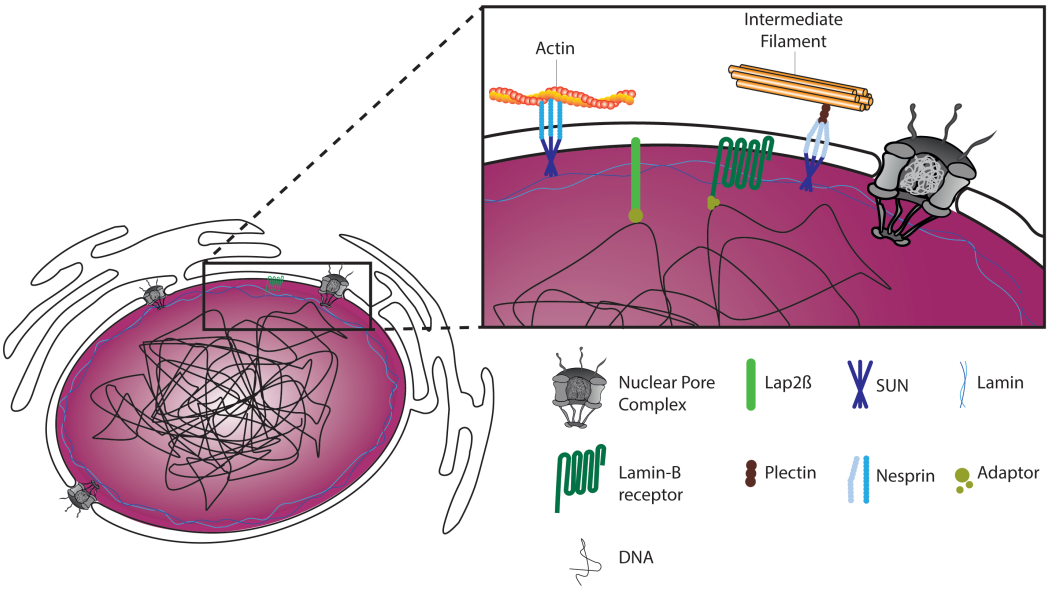


Figure 1.2: The inner nuclear membrane proteome. The NE is continuous with the ER and consists of the INM and ONM, which enclose the perinuclear space. The INM and ONM fuse at the insertion sites of the NPC. The NPCs contain a central channel in the center of the complex, as well as a narrow 10 nm wide channel at the interphase of the complex and the pore membrane. Underlying the INM is the nuclear lamina. The INM harbours a unique set of proteins, which interact both with the nuclear lamina and the chromatin. Spanning the INM and ONM are the LINC complexes, consisting of KASH-domain proteins such nesprins in the ONM and SUN proteins in the INM connecting. They are responsible for connecting the cytoskeleton, including actin, intermediate filaments and microtubules, to the nucleoskeleton.

are involved in a range of diverse functions including chromatin organisation, nuclear assembly, cell cycle control, as well as regulation of gene expression and DNA replication[58,59]. Mutations in various LEM-domain proteins have been linked to different diseases. A loss-of-function mutation of emerin was shown to result in X-linked Emery-Dreifuss Muscular dystrophy (EDMD). The clinical features involve early contractures of achillis tendons and elbows, followed by slow but progressive muscle weakness and cardiac conduction defects[60]. Additionally, mutations in the LEM-domain protein MAN1 have been linked to the Buschke-Ollendorff syndrome, as well as increased bone density. Like in humans, a tissue specific phenotype of MAN1 depletion is also seen in *Drosophila melanogaster*, with wing and reproductive tissues being affected[61,62].

Besides LEM-domain proteins, another well-characterised protein known to interact with the nuclear lamina is the lamin B receptor (LBR). This multispanning

membrane protein has a dual activity, as it is involved in cholesterol synthesis, while also tethering chromatin to the INM[63]. Mutations in LBR have also been linked to disease, more specifically the Pelger-Huët anomaly, which is characterised by abnormal nuclear morphology and chromatin disorder in blood granulocytes[62]. Another disease linked to mutations in LBR is Greenberg dysplasia, in which patients have skeletal and bone abnormalities, but also show accumulation of an abnormal sterol metabolite, indicating that this disease might be linked to the sterol reductase activity of LBR[62].

SUN (Sad1 and UNC-84) proteins constitute another class of INM proteins. Spanning the INM, they form a bridge in the lumen of the NE with the KASH-domain (Klarsicht, ANC-1, and Syne/Nesprin homology) proteins localised in the ONM. Together, the SUN and KASH-domain proteins form the linker of nucleoskeleton and cytoskeleton (LINC) complexes, which act as important transmitters of forces across the nucleus[57,64].

In humans, no mutations in SUN genes are known to be disease relevant by themselves, yet they can act as modifier genes in pre-existing diseases. Mutations in SUN genes resulted in more severe diseases in members of a family affected by disease through mutations in either the *LMNA* or *EMD* gene. Additionally, it was shown that overexpression of SUN1 in $LMNA^{-/-}$ mice resulted in severe disease, whereas loss of SUN1 alleviated pathogenic phenotypes of the A-type lamin depletion both in mice and humans, as observed in patients with X-linked Hutchinson-Gilford progeria syndrome (HGPS)[62]. Mechanistically, two hypothesis how this might happen exist: one study indicates that in absence of A-type lamins SUN1 accumulates in the Golgi resulting in cellular toxicity[65], whereas the other study suggests that SUN1 accumulation at the INM in absence of lamins leads to cell toxicity by hyper activating the DNA damage response[66]. In both cases, loss of SUN1 attenuates the damaging effect of A-type lamin depletion.

1.4 Targeting proteins to the INM

One important feature of eukaryotic cells is their compartmentalisation. mRNA transcription takes place in the nucleus, whereas protein translation of soluble proteins happens in the cytoplasm. This compartmentalisation creates the need for a nucleoplasmic/cytoplasmic transport system. This process has been well studied for soluble proteins and various excellent reviews exist discussing the process in depth[67–69]. In brief, small cargos (<40 kDa) seem to freely diffuse in and out of the nucleus. In contrast, for proteins larger than this size limit, proteins rely on an active transport pathway through the NPCs. This process is dependent on nuclear transport receptors (NTRs) of the karyopherin-β family. For the nuclear import, importins bind to a positively charged nuclear localisation signal (NLS). In contrast, exportins interact with a hydrophobic nuclear export signal (NES) within the nucleus to facilitate shuttling to the cytoplasm. Directionality of the process is dependent on the small GTPase Ran, that interacts with the NTRs and acts as a molecular switch[69]. RanGTP binds to karypherin-β with high affinity. Binding with importins thereby results in importin-cargo destabilisation. In contrast, interaction of RanGTP with exportin results in increased stability with the cargo. A Ran gradient, with RanGTP and RanGDP as the predominant form in the nucleus and the cytoplasm respectively, is achieved through the enrichment of RanGEF (guanine nucleotide exchange factor) RCC1 on chromatin, whereas RanGAP (GTPase activating protein) is predominantly found in the cytoplasm.[69]

In comparison to the transport of soluble proteins, our understanding of the translocation of membrane proteins to the INM is slightly more preliminary. Synthesis of most membrane proteins occurs in the ER regions exposed to the cytoplasm. Subsequently, INM proteins need to travel to the INM. The prevailing model of INM targeting suggests that proteins concentrate at the INM via a diffusion-retention model (figure 1.3)[70–75]. According to this model, INM proteins get synthesised in the bulk ER, from where they can freely diffuse to the ONM and INM. One factor thereby influencing targeting efficiency is the interconnection and morphology of the ER. More precisely, depletion of the GTPase atlastin,

which is required for homeotypic fusion of ER membranes, results in decreased trafficking of proteins to the INM[76]. To reach the INM, proteins need to pass by the NPC, which links the ONM and INM. Kinetic studies suggest that this is the rate-limiting step of protein targeting, as increasing the permeability of the NPC leads to faster enrichment of proteins at the INM[70]. Whereas soluble proteins utilise the central channel of the NPC to traverse from the cytoplasm to the nucleoplasm, INM proteins mostly traffic via the narrow peripheral channel. This ∼9 nm channel creates a size limit of approximately $<$ 60 kDa to the cytoplasmic/nucleoplasmic domain of INM proteins[74,77,78]. Intriguingly, some proteins seem to circumvent the size restriction of the NPC. Mechanistically, exciting new state of the art imaging studies in yeast and mammalian cells indicate that proteins with long intrinsically disordered domains make use of the central and peripheral channel simultaneously, thereby circumventing the size limit posed on proteins using the peripheral channel alone[79,80]. This dual use mechanism of the central and peripheral channel is in line with a proposed transport-factor mediated pathway, as it requires a nuclear localisation signal (NLS), as well as a long disordered linker domain to extend into the central channel[79,80]. Once proteins reach the INM, their mobility decreases significantly as they bind to chromatin or lamins[70,73,74]. In contrast, proteins that do not have interaction partners at the INM diffuse back to the bulk ER.

Besides the diffusion retention model, a small number of INM proteins have been observed to use other mechanisms. Comparable to the nuclear transport mechanism for soluble proteins, these alternative translocation mechanisms are transport factor and energy dependent, and require a NLS[77,81–83].

1.5 Protein degradation at the INM

Besides protein targeting, selective degradation of proteins plays an important role in establishing and maintaining the INM protein identity. For this, cells make use of both proteasomal and lysosomal degradation. The proteasome dependent mechanism has been termed INM-associated degradation (INMAD). Similar to

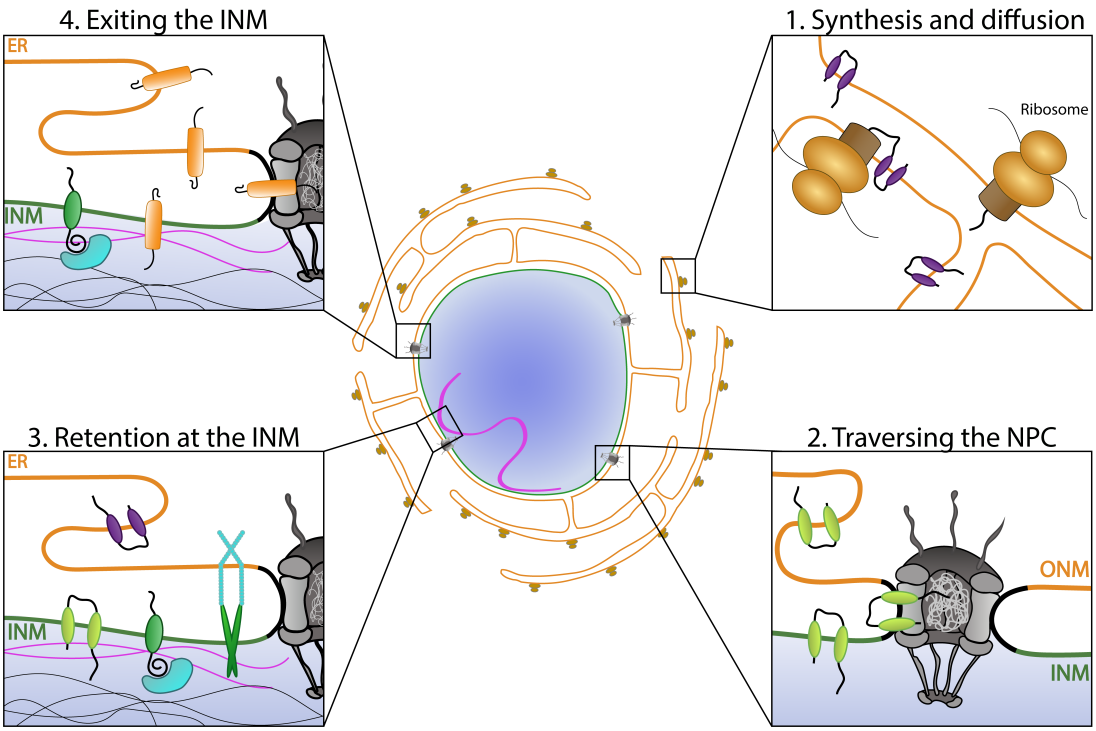


Figure 1.3: The diffusion-retention model. Transmembrane proteins get synthesised in the ER (1). From there, they freely diffuse within the ER. Proteins with cytoplasmic domains < 60 kDa can diffuse past the NPC, using its peripheral channel (2). Once proteins reach the INM, proteins that interact with the nuclear lamina, chromatin or their adaptor proteins, get retained at the INM (3). In contrast, proteins that don't have interaction partners in the nucleus, eventually diffuse back to the bulk ER (4). *Note: Figure under review at Current Opinion in Cell Biology.*

ER-associated degradation (ERAD), INMAD is a multi-step process that starts with the recognition of the substrates[84]. Next, proteins get ubiquitinated by the ubiquitin ligase complex through a cascade of reactions. Lastly, proteins get retrotranslocated into the nucleoplasm by the AAA-ATPase p97 (Cdc48 in yeast) and degraded by the proteasome.

Whereas the generic framework for INMAD degradation appears simple, our knowledge of INMAD in higher eukaryotes is still very rudimentary. Dynamic proteomic studies suggest that protein half-lifes of ER and INM proteins are comparable, yet to date very little INMAD substrates are known and no INM E3 ubiquitin ligase complex has been identified in mammalian cells[85].

One of the few INM substrates found for INMAD in mammalian cells is SUN2. Proximity-dependent biotinylation (PDB) experiments identified the soluble

Skp1/Cullin1/F-box/βTRCP (SCF^{bTRCP}) ubiquitin E3 ligase as an interaction partner and deletion of SCF^{bTRCP} results in accumulation of overexpressed SUN2[86].

Two other known INMAD substrates in mammalian cells are the disease-associated LBR mutants LBR1600* and LBR1402TΔ. Both proteins are known to be involved in Greenberg Dysplasia. Contrary to wild-type LBR, these truncated versions get turned over rapidly at the INM[63]. Whereas the exact mechanism is yet to be determined, the degradation depends both on the retrotranslocation by p97, as well as the proteasome. A similar observation was made for emerin, where a EDMD-disease variant of the protein was shown to get rapidly cleared from the INM[85]. Curiously, under ER stress by calcium depletion, a fraction of both wild type and mutant emerin trafficked out of the INM and got degraded in the lysosome, indicating the existence of a second, stress-dependent degradation pathway (figure 1.4).

Besides INMAD degradation, autophagy also exists at the INM. Lamin B, a protein of the nuclear lamina, was identified as an autophagic substrate in IMR90 cells. Degradation was dependent on the autophagy protein LC3-Atg8 and was triggered upon oncogenic insults, such as an oncogenic mutation that results in an activated RAS small GTPase[87].

Whereas some progress was made in the past decade, our understanding of protein degradation at the INM is still very primordial in higher eukaryotes. A better understanding of how protein degradation plays a role in maintaining the INM protein composition exists in yeast and drawing parallels might help unravel basic mechanism and identify promising directions of future study.

The first E3 ligase identified at the INM in yeast was Doa10. Besides the INM, Doa10 also localises to the bulk ER and is involved in the degradation of the MAT α transcription factor, as well as other membrane proteins[88–90]. Additionally, the Asi complex, consisting of the Asi1/Asi3 ubiquitin ligases and the adaptor protein Asi2, is a protein degradation branch exclusively localised to the INM (figure

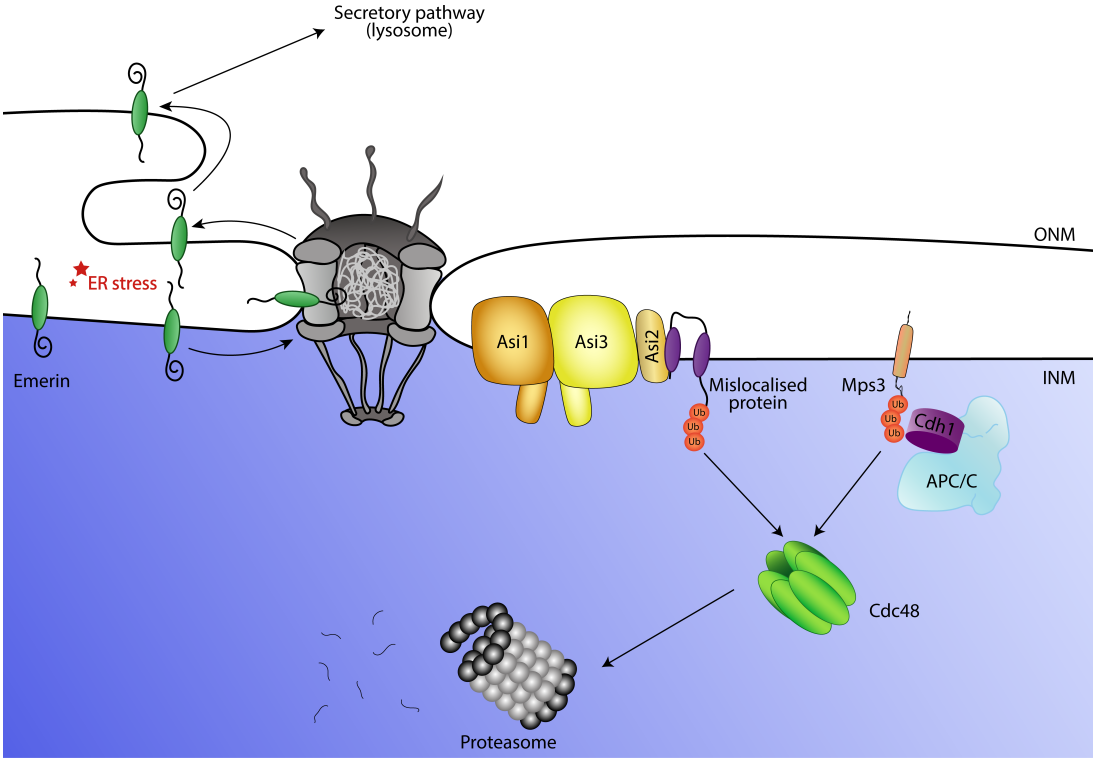


Figure 1.4: Selective protein degradation at the INM. Left: Upon ER stress, the INM protein emerin was shown to traffic to the cytoplasmic ER, from where it enters the secretory pathway for lysosomal degradation. Centre: Proteins get degraded by the INMAD pathway. Substrates are recognised by the Asi complex and get ubiquitinated. Ubiquitinated substrates then get retrotranslocated by Cdc48 and subsequently degraded by the proteasome. Right: Mps3 gets ubiquitinated by the soluble ubiquitin ligase APC/C, before subsequently getting retrotranslocated by Cdc48 and degraded by the proteasome. *Note: Figure under review at Current Opinion in Cell Biology.*

1.4)[91,92]. Mechanistically, Asi2 binds directly to the transmembrane domain (TMD) of substrates, whereas Asi1/3 are responsible for the ubiquitination (figure 1.4, centre). Thereby the Asi complex not only recognises mistargeted and mislocalised substrates, but has also been shown to identify orphan subunits, which may be a mechanism to keep subunits of protein complexes at near stoichiometric levels[93]. Besides degradation of proteins, Asi also plays a role in controlling protein distribution. In Asi mutant cells, the nucleoporin Pom33 concentrated in aberrant foci along the NE, instead of its uniform INM distribution in wild-type cells. This relocalisation was dependent on monoubiquitination of Pom33 and was not observed for any other NPC component[94]. Another ERAD branch

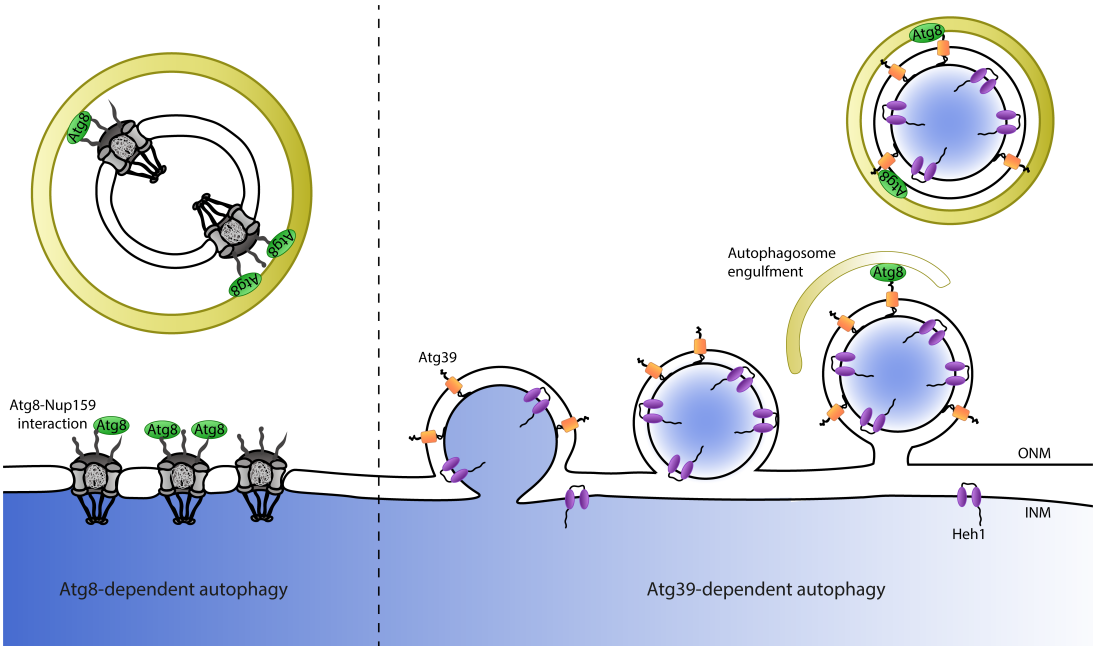


Figure 1.5: Autophagy-dependent degradation at the INM. Left: NPCs cluster and get degraded by Atg8 dependent autophagy upon starvation. Autophagic degradation of the NPCs is independent of the adaptor proteins Atg39 and Atg40. Right: Atg39, localised to the ONM, interacts through its amphipathic helix with the INM, resulting in membrane remodelling and an inside-out mechanism to create vesicles containing Heh1 for autophagosomal degradation. *Note: Figure under review at Current Opinion in Cell Biology.*

in yeast, independent of the Asi complex and Doa10, is reliant on the soluble anaphase promoting complex/cyclosome (APC/C). This branch is involved in the degradation of Mps3, which is the yeast orthologue of SUN2, that was also shown to be dependent on a soluble E3 ligase[95] (figure 1.4, right).

Besides proteasome-dependent degradation, autophagy plays a role in maintaining the INM proteome in yeast. Two main branches have been identified (figure 1.5). The first pathway was shown to be involved in the degradation of NPCs under starvation conditions and is dependent on binding of Nup159 to Atg8[96,97] (figure 1.5, left). In contrast, the second pathway is dependent on the amphipathic helix of Atg39 interacting with the INM and was required for the autophagic degradation of Heh1[98–100] (figure 1.5, right).

1.6 The nuclear lamina

Lamins are type V intermediate filaments that form the nuclear lamina underlying the nucleoplasmic side of the INM. They fall into two classes: A- and B-type lamins. In mammals, A-type lamins include lamin A and lamin C, which are splice variants expressed from the *LMNA* gene. Their expression is developmentally regulated, with gene expression appearing midway through embryonic development[101]. In contrast, lamin B1 and lamin B2 are encoded in two separate genes, *LMNB1* and *LMNB2*, and their expression is ubiquitous[43,102]. Lamins are made up of a N-terminal head domain, a central coiled-coil domain and a C-terminal tail domain[103]. Additionally, they contain a NLS sequence that ensures targeting to the nucleus. Lamins are synthesised in the cytoplasm and undergo extensive post-translational modifications, including farnesylation, phosphorylation, SUMOylation or glycosylation. Upon shuttling to the NE, single lamin proteins first dimerise and then assembly further into head-to-tail polymers. Thereby each major isoform assembles into a separate meshwork[103]. Together with their binding partners, lamins are involved in a range of functions such as signalling, cell-cycle regulation and chromatin organisation[58,59,103].

Over 70 different mutations in nuclear lamins have been linked to a wide range of skeletal and cardiac muscle diseases, such as autosomal Hutchinson-Gilford Progeria syndrom (HGPS), Limb gridle muscular dystrophy 1b and Charcot-Marie-Tooth disease[101,102]. Paradoxically, many of these diseases are tissue specific, albeit lamins being ubiquitously expressed. Reasons for this are yet to be determined, but might include factors such as tissue specific splice variances and interaction partners.

In recent years, the availability of mouse models, as well as the possibility to create patient-derived induced pluripotent stem cells (iPSCs), have led to progresses in our understanding of laminopathies.[101,104]. On an organism-wide level, $LMNA^{-/-}$ mice show no phenotype at birth, yet they later develop pathologies such as cardiomyopathies and lipodystrophies. In contrast, $LMNB1^{-/-}$ mice embryos have bone and lung abnormalities and only live until birth[101]. On a

cellular level, cells with *LMNA* mutations from humans and mice often show aberrant nuclear morphology and are more prone to nuclear rupture, as well as cell cycle defects[105–107]. Furthermore, loss of A-type lamins was demonstrated to result in mislocalisation of INM proteins and nuclei of lamin depleted cells are less contractile compared to their wild-type counterparts[108–112]. Mechanistically, depletion of lamins are thought to affect cells in two main ways. Firstly, mutations in lamins may render nuclei more prone to mechanical damage, resulting in nuclear ruptures and subsequent DNA damage. Secondly, lamins and INM proteins may interact with transcriptional interactors or influence positioning of chromatin in the cell periphery, which was shown to affect gene expression.

1.7 NE breakdown and re-assembly

Despite its important structural role in separating the nucleoplasm from the cytoplasm, the NE is highly dynamic and adaptable. Thereby the most drastic reformation happens in mitosis. During cell division, sister chromatids need to be divided into two daughter cells, which requires the attachment of the mitotic spindle. Whereas some cells undergo closed mitosis, in which the NE only locally disintegrates, in most eukaryotic cells such as mammalian cells the spindle forms in the cytoplasm[46,113,114]. To enable access of the spindle to the chromosome, the NE disassembles, resulting in the loss of cellular compartmentalisation and the unique organisation of the INM. NE breakdown (NEBD) is mostly driven by hyperphosphorylation, reliant on a whole range of kinases[114,115]. In prophase, mitotic chromosomes condense. Next, hyperphosphorylation of the NPC and other multiprotein complexes, as well as the nuclear lamina, result in their disassembly. Nuclear membrane and INM proteins get absorbed into the ER during mitosis[116]. In telophase, the two segregated masses of chromosomes then need to form two independent nuclei again. This process relies of dephosphorylation events using a range of phosphatases[117]. Interestingly, whereas the INM shows a homogeneous spread of proteins during interphase, after mitosis INM proteins and lamins are known to form distinct domains, defined as the core and non-core regions. Lamin

A, MAN1, Lap2β and emerin are enriched at the core region, close to where the MTs attach to the chromatids. In contrast, lamin B, NPCs and LBR are depleted from these regions and instead cluster at the non-core region[117,118]. How this happens mechanistically is unclear, in particular as there is no evidence supporting that these proteins are stored in different regions of the ER during mitosis[115]. Once the chromosomes are coated with membrane, the independent ER sheets need to fuse to form a continuous NE membrane. This membrane fusion was shown to involve the ESCRT-III machinery, as well as coordinated disassembly of the MTs[117,119]. Additionally to membrane fusion, evidence indicates that NPCs start assembling in the fenestrated holes[120].

1.8 Proximity-dependent biotinylation approaches to study organelle domains

Understanding the protein composition of an organelle, or distinct domains within an organelle, provides the basis for all functional and mechanistic studies. Yet, the membrane continuity between ER tubules, sheets and the NE makes it inherently difficult to study the proteome of any of these individual domains. Proximity-dependent biotinylation (PDB) assays may pose a solution to these difficulties. PDB assays rely on the basic concept that an enzyme can transfer a biotin (derivative) to proteins in its close proximity upon addition of enzyme specific substrates. Fusing enzymes to distinct proteins (referred to as "baits") enables targeting of the enzyme to different locations within cells. Importantly, the labelling process works in live cells and covalent labelling can be combined with harsh lysis conditions and wash steps in the subsequent immunoprecipitations (IPs), making PDB assays a powerful tool to study subcellular proteomes in living cells and organisms.

Various excellent reviews exist giving an in-depth analysis of different biotinylation techniques[121–124].

Briefly, there are two main classes of enzymes for PDB:

1. Biotin protein ligases

2. Peroxidases

Biotin protein ligases are based on the BirA enzyme originally found in *E. coli*, in which it only has a single known substrate for biotinylation. Interestingly, a single point mutation (R118G) makes the BirA ligase, then denoted BirA*, promiscuously active towards all substrates in a proximity-dependent manner[125]. This BirA* mutant built the basis for the BioID system, as it creates a cloud of reactive biotinyl-5'-AMP that can react with primary amines in a radius of ∼10 nM, resulting in their covalent biotinylation. The first studies genetically targeted this BioID to the nuclear lamina[126] and NPCs[127]. Since then, it has been used to investigate a wide-range of membrane-bound and membraneless organelles. With labeling times of ∼18 h, this PDB method is optimal to gain a generic understanding of a cellular compartment. Humanisation of the biotin ligase from *A. geolicus* coupled to a R40G mutation, mimicking the function of the R118G mutation in the *E. coli* BirA*, led to a 2^{nd} generation BioID, which is smaller and requires lower biotin concentrations to function[128]. Two other derivatives of the BioID generated through directed evolution are the TurboID and the miniTurboID, in which the reaction times for biotinylation can be reduced from the previous ∼18 h to just 10 min[129]. Further developments also resulted in a split-BioID and a Split-TurboID, which are separated into two parts and only result in biotinylation reactions when complementing each other[130,131].

Peroxidases are the second class of enzymes used in PDB assays. These enzymes catalyse the reduction of hydrogen peroxide. The best studied peroxidase is the horseradish peroxidase C (HRP). HRP is commonly used as a biological research tool for labelling, e.g. for western blots (WB). HRP has also been successfully used as a genetic tag in cells in certain cellular compartments for EM, however it is inactive when expressed in reducing environments such as the cytosol and attempts to remove or replace the disulfite bonds to circumvent this limitations were not successful[132,133]. Yet, directed mutagenesis of another

peroxidase, 1APX from soybeans, resulted in an engineered ascorbate peroxidase (APEX) that was used for *in vivo* labelling for EM experiments[133]. APEX creates phenoxyl-radicals by oxidising phenol derivatives such as biotin phenol. These radicals, with a short half-life of < 1 ms, covalently attach to electron-rich amino acid side chains. Three rounds of yeast display evolution resulted in a second generation APEX (termed APEX2) with enhanced cellular activity[134]. Originally used to probe the mitochondria[135] and the mitochondrial intermembrane space[136], APEX2 has now been used to study various different cellular environments. Importantly, the fast kinetics of the APEX2 system allows cellular labelling in the timescale of seconds, making APEX2 a powerful tool to gain a snapshot of a cellular environment at any given time point. This also enables to study dynamic processes and disturbances to cellular homeostasis.

1.9 Aim of the study

The ER, consisting of tubules, sheets and the NE, is the largest organelle in cells. It is involved in a wide range of functions, including synthesis of membrane and secretory proteins, calcium homeostasis and lipid synthesis. Whereas some of these functions are known take place in distinct ER domains, such as protein synthesis in ER sheets, for most functions it is unclear whether they happen uniformly throughout the ER or whether they are restricted to specific domains. Generating a map of the protein composition across the ER could help create a better understanding of ER functions and their dependency on the elaborate ER structure.

In the past two decades, various different approaches were taken by scientists to reach this goal. Subcellular fractionations coupled to proteomic analysis gave first insights into the protein composition of the rough and smooth ER and the INM[51,137]. Additionally, imaging studies were used to follow up on individual proteins[23]. The discovery of novel PDB technology also resulted in various studies investigating ER composition[53–55,138–140]. Yet, a complete overview of how proteins are distributed across the ER is still missing.

The aim of this project was to develop an assay to probe the protein composition of the different, distinct ER domains and generate a map of the protein distribution across the ER. This will result in the identification of *bona fide* proteins enriched in each of the ER domains, that will require further validation. Additionally, if functional, the assay will be a powerful tool to study whether different perturbances to cell physiology, or biological processes such as cell division, affect the protein composition within the distinct ER domains.

2

Materials and methods

Contents

2.1 List of materials 23
2.1.1 General and molecular biology reagents 23
2.1.2 List of commercial kits 24
2.1.3 List of cell lines 24
2.1.4 List of plasmids 24
2.1.5 List of primers 25
2.1.6 List of gRNAs 26
2.1.7 List of primary antibodies 26
2.1.8 List of secondary antibodies 27
2.1.9 List of solutions and buffers 28
2.1.10 List of antibiotics 29
2.2 Protocols 30
2.2.1 Cell Culture Methods 30
2.2.2 Molecular biology methods 32
2.2.3 Protein biochemistry methods 35
2.2.4 Proteomic analysis 40

2.1 List of materials

2.1.1 General and molecular biology reagents

Table 2.1: General and molecular biology reagents

Reagent	Source	Catalogue number
1 kb DNA ladder	New England Biolab	N3232L
100 kb DNA ladder	New England Biolab	N3231S
1,4-Dithiothreitol	Merck	D9779
Benzonase	Merck	E1014
Biotin	Merck	B4501
Biotin-phenol	Iris Biotech	LS-3500.1000
CB-5083	Selleckchem	S8101
cOmplete EDTA-free protease inhibitor cocktail	Roche	5056489001
Dimethyl sulfoxide	Insight Biotechnology	sc-358801
DMEM medium	Merck	D6429
FBS	Merck	F9665
Fluoromount-G (with DAPI)	Thermo Fisher Scientific	00-4959-52
Formic acid (Mass spectrometry grade)	Fisher Scientific	94318
Gel loading dye (6 ×)	New England Biolab	B7024S
L-Glutamine	Thermo Fisher Scientific	25030024
Micrococcal nuclease	New England Biolabs	M0247S
PMSF	Roche	11359061001
Precision Plus Protein Dual Color Standards	Bio-Rad	1610374
SYBR safe DNA gel stain	Thermo Fisher Scientific	S33102
TAE buffer (50 ×)	VWR	A1691.1000
Tetracycline	Merck	T7660
TransIT LT1	Mirus Bio LLC	MIR 2305
Tris/Glycine/SDS buffer (10 ×)	Bio-Rad	1610772
TrypLE express	Thermo Fisher Scientific	12604013
Trypsin (Sequencing grade modified)	Promega	V5111
OptiMEM	Thermo Fisher Scientific	31985062
(±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid	Merck	238813

2.1.2 List of commercial kits

Table 2.2: List of commercial assays

Commercial assay	Source	Catalogue number
GeneJET PCR Purification Kit	Thermo Fisher Scientific	K0702
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	K0503
Gibson Assembly Master Mix	New England Biolab	E2611L
Pierce Streptavidin Magnetic Beads	Thermo Fisher Scientific	88817
Trans-Blot Turbo Transfer Kit	Bio-Rad	1704273
Q5 High-Fidelity 2X Master Mix	New England Biolabs	M0492L
Western Lightning ECL Pro	Perkin Elmer	NEL121001EA

2.1.3 List of cell lines

Table 2.3: List of secondary antibodies

Cell line	Source	Catalogue number
Flp-In T-REx 293	Thermo Fisher Scientific	R78007
T-REx-HeLa	Thermo Fisher Scientific	R71407
HeLa	ATCC	ATCC CCL-2

2.1.4 List of plasmids

Table 2.4: List of plasmids

Name	Description	Vector	Source
pcDNA3.1	Vector expressing APEX2 in the cytosol	pcDNA3 APEX2-NES	Addgene (#49386)
pcDNA5	Vector with tetracycline inducible expression system designed for use with the Flp-In T-REx system	pcDNA5/FRT/TO	Thermo Fisher Scientific (#V652020)
RP-418	Lentiviral expression vector	RP-418	Lab stock
RP-557	Lentiviral expression vector	RP-557	Lab stock

2.1.5 List of primers

Table 2.5: List of primers

Name	Sequence (5'–3')	Description	To generate
LO-001	GTTTAAACTTAAGCTTGGTACCAC CATGGACTACAAGG	FLAG-APEX2 FW	FLAG-APEX2-POI
LO-003	GCTTGGGTTTGCTGATGCCGGATC CGGGTCAATGGACAACTACGCAG ATC	Emerin FW	FLAG-APEX2-Emerin
LO-004	GTTTAAACGGGCCCTCTAGACTCG AGCTAGAAGGGGTTGCCTTC	Emerin RV	FLAG-APEX2-Emerin
LO-008	CTTGGGTTTGCTGATGCCGGATCC GGGTCAATGACAACCTCCCAAAA GC	BAF FW	FLAG-APEX2-POI
LO-009	GTTTAAACGGGCCCTCTAGACTCG AGTCACAAGAAGGCGTCG	BAF RV	FLAG-APEX2-BAF
LO-011	CTTGGGTTTGCTGATGCCGGATCC GGGTCAATGCCGGAGTTCCTGG	Lap2β FW	FLAG-APEX2-Lap2β
LO-012	GTTTAAACGGGCCCTCTAGACTCG AGTCAGTTGGATTTTCTAGGGTC	Lap2β(*) RV	FLAG-APEX2-Lap2β(*)
LO-014	GGGCCCTCTAGACTCGAGGGGTTA CTTATCGTCGTCATCCTTGTAGT CCATGGCATCAGCAAACCCAAG	APEX2-FLAG RV	POI-APEX2-FLAG
LO-015	GCGTTTAAACTTAAGCTTGGTACC CCACCATGGCCTCG	$Nup153_{1-144}$ FW	$Nup153_{1-144}$-APEX2-FLAG
LO-016	GTTGGGTAAGACTTTCCCCTTCGA AGACCAGGGGATTCCAACATGG	$Nup153_{1-144}$ RV	$Nup153_{1-144}$-APEX2-FLAG
LO-022	CTTGGGTTTGCTGATGCCGGATCC GGGTCACAGGCATTAACTAGGGA ATCTAC	Lap2β* FW	FLAG-APEX2-Lap2β*
LO-023	CTTGGGTTTGCTGATGCCGGATCC GGGTCAATGTCCCGAAGAAGCC AGC	SUN2 FW	FLAG-APEX2-SUN2
LO-024	GGTTTAAACGGGCCCTCTAGACTC GAGGTGGGCGGGCTCCC	SUN2 RV	FLAG-APEX2-SUN2
LO-027	ACTCTAGCGTTTAAACTTAAGCTT GGTACCCCACCATGGAAGACCTG GACCAG	Rtn4a FW	Rtn4a-APEX2-FLAG
LO-033	GGGTAAGACTTTCCCCTTCGAAGA CCTTCAGCTTTGCGCTTCAATC	Rtn4 RV	Rtn4-APEX2-FLAG
LO-109	CTTGGGTTTGCTGATGCCGGATCC GGGTCAATGCCTGGTCCGACC	Sec61β FW	FLAG-APEX2-Sec61β
LO-110	GTTTAAACGGGCCCTCTAGACTCG AGCTACGaACGAGTGTACTTGC	Sec61β RV	FLAG-APEX2-Sec61β

Continuation of Table 2.5			
Name	**Sequence (5'–3')**	**Description**	**To generate**
MO-135	CTAGCGTTTAAACTTAAGCTTGGT ACCCCACCATGCCAAGTAGGAAA TTTGCC	LBR FW	LBR-APEX2-FLAG
MO-150	CAATTCCTCGCCTTTAGACATCCT TCGAAGACCGTAGATGTATGGAA ATATACGGTAGGG	LBR RV	LBR-APEX2-FLAG
End of Table			

2.1.6 List of gRNAs

Table 2.6: List of gRNAs

Name	Sequence (5'–3')	Target gene	Cloned into
LMNA_1	GCATCGACCGTGTGCGCTCGC	*LMNA*	RP-557
LMNA_2	GGCATCAAGGCCGCCTACG	*LMNA*	RP-557
LMNA_3	GCTTCAGACTCGGTGATGCGA	*LMNA*	RP-557
ZMPSTE24_1	GGCCGAGAAGCGTATCTTCG	*ZMPSTE24*	RP-418
ZMPSTE24_2	GGCATCGCTGGACGCTTTGT	*ZMPSTE24*	RP-418
ZMPSTE24_3	GATCTGTCCTAACTCCGG	*ZMPSTE24*	RP-418

2.1.7 List of primary antibodies

Table 2.7: List of primary antibodies

Antigen	Host	Source	Catalogue number	Use	Dilution
ASPH	Rabbit	Abcam	ab172475	IF	1:500
Biotin		Thermo Fisher Scientific	S911	WB	1:5000
Calnexin	Rabbit	Abcam	ab22595	IF	1:500
DCAKD	Rabbit	Human Protein Atlas	HPA060997	IF	1:500
ELOVL5	Rabbit	Invitrogen	PA583879	IF	1:500
Emerin	Rabbit	Cell Signaling	30853	WB	1:2000
FLAG	Mouse	Abcam	A8592	WB	1:2000
GAPDH	Mouse	Proteintech	60004-1	WB	1:5000
Histone H3	Rabbit	Abcam	ab1791	WB	1:10,000
Kinectin	Mouse	Santra Cruz	sc-136534	WB	1:1000
Lamin A/C	Rabbit	Abcam	ab108595	WB	1:5000

Continuation of table 2.7					
Antigen	**Host**	**Source**	**Catalogue number**	**Use**	**Dilution**
Lamin B receptor	Rabbit	Abcam	ab32535	WB	1:1000
Lap2β	Mouse	BD Transduction	611000	WB	1:2000
PGRMC1	Rabbit	ProteinTech	12990-1-AP	IF	1:500
PGRMC2	Mouse	ProteinTech	60249-1-Ig	IF	1:500
SUN1	Rabbit	Abcam	ab124770	WB	1:2000
SUN2	Rabbit	Abcam	ab124916	WB	1:1000
SUN2	Rabbit	Atlas Antibody	HPA001209	IF	1:500
SQLE	Rabbit	Protein Tech	12544-1-AP	WB	1:2000
TMEM109	Rabbit	Human Protein Atlas	HPA011785	IF	1:500
VCP	Mouse	BD Bioscience	612183	WB	1:1000
End of table					

2.1.8 List of secondary antibodies

Table 2.8: List of secondary antibodies

Antigen	**Conjugate**	**Source**	**Catalogue number**	**Use**	**Dilution**
α-rabbit IgG light chain specific	HRP	Jackson Immuno-Research	211-032-171	WB	1:10,000
α-mouse IgG light chain specific	HRP	Jackson Immuno-Research	115-035-174	WB	1:10,000
α-rabbit IgG (H+L)	Alexa Fluor 488	Thermo Fisher Scientific	A11034	IF	1:500
α-rabbit IgG (H+L)	Alexa Fluor 568	Thermo Fisher Scientific	A10042	IF	1:500
α-rabbit IgG (H+L)	Alexa Fluor 647	Thermo Fisher Scientific	A31573	IF	1:500
α-mouse IgG (H+L)	Alexa Fluor 488	Thermo Fisher Scientific	A21202	IF	1:500
α-mouse IgG (H+L)	Alexa Fluor 564	Thermo Fisher Scientific	A11004	IF	1:500
α-mouse IgG (H+L)	Alexa Fluor 647	Thermo Fisher Scientific	A31571	IF	1:500

Continuation of table 2.8					
Antigen	**Conjugate**	**Source**	**Catalogue number**	**Use**	**Dilution**
Biotin	Alexa Fluor 488	Thermo Fisher Scientific	S32354	IF	1:500
Biotin	Alexa Fluor 555	Thermo Fisher Scientific	S32355	IF	1:500
End of table					

2.1.9 List of solutions and buffers

Table 2.9: List of solutions and buffers

Solution	Component	Concentration
Elution buffer	Biotin DTT 3 × sample buffer	2 mM 100 mM 1 ×
High salt sucrose cushion buffer	Sucrose HEPES KOH (pH7.5) KOAc $Mg(OAc)_2$	500 mM 50 mM 500 mM 5 mM
Hypotonic buffer	HEPES KCl $MgCl_2$ EGTA PMSF DTT	10 mM 10 mM 1.5 mM 0.1 mM 1 mM 1 mM
PBS-T	PBS TWEEN 20	1 × 0.1% (v/v)
4% PFA	PFA PBS	4% (v/v) 1 x
PMSF 100 mM	PMSF Isopropanol	100 mM 1 ×
Quencher solution	Sodium ascorbate Sodium acide Trolox Protease inhibitor cocktail PBS	10 mM 10 mM 5 mM 1 × 1 ×
RIPA buffer	TrisHCl (pH 8) NaCl Triton X-100 Sodium deoxycholate SDS Protease inhibitor cocktail	50 mM 150 mM 1% (v/v) 0.5% 0.1% (w/v) 1 ×

Continuation of Table 2.9		
Solution	**Component**	**Concentration**
RIPA+	Sodium azide Sodium ascorbate Trolox PMSF RIPA	10 mM 10 mM 5 mM 1 mM 1 ×
RM buffer	Sucrose HEPES KOH (pH 7.5) KOAc $Mg(OAc)_2$ DTT	250 mM 50 mM 50 mM 2 mM 1 mM
Sample binding buffer	Tris HCl (pH 7.4) NaCl NP40	25 mM 125 mM 0.1% (v/v)
3 x Sample buffer	Tris HCl (pH 6.8) SDS Glycerol Bromophenol blue	0.2 M 6% (w/v) 30% (v/v) 0.2% (w/v)
End of table		

2.1.10 List of antibiotics

Table 2.10: List of antibiotics

Antibiotic	**Working concentration (μg/mL)**	**Source**	**Catalogue number**
Blasticidin	15	Merck	203351
Hygromycin Gold	150	Merck	400052
Penicillin-Streptomycin	0.1 (units/mL); 100	Thermo Fisher Scientific	15140122
Puromycin	2	Thermo Fisher Scientific	A1113803
Zeocin	10	Thermo Fisher Scientific	R25001

2.2 Protocols

2.2.1 Cell Culture Methods

2.2.1.1 Maintenance of human cell lines

All cell lines were cultured at 37°C with a humidified, 5% CO_2 (v/v) atmosphere in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 Units/mL penicillin and 100 μg/mL streptomycin. Cells were maintained in 10 cm plates and passaged when they reached a confluency of 80–90%. To passage cells, cells were washed once in 1x PBS. Subsequently, cells were detached using 1 mL of trypsin. Cells were then seeded at 10–20% confluency.

For experiments requiring stringent control of cell numbers, cells were counted using the NucleoCounter NC-3000 and seeded accordingly. All cell lines were monitored for mycoplasma contamination every three months by the departmental Sir William Dunn School in-house service using the MycoAlert Mycoplasma Detection Kit (Lonza).

2.2.1.2 Cryopreservation of human cell lines

Cells grown in 10 cm dishes were detached using trypsin, resuspended in 9 mL media, spun down at 300 $\times$ g and the supernatant was discarded. Subsequently, cells were resuspended in 1.5 mL DMEM. 1.5 mL FBS with 20% DMSO was added dropwise. Cells were split equally into three cryovials and frozen at -80°C using a Corning CoolCell Freezing System. Cells were subsequently stored at -80°C for short term use or transferred to liquid nitrogen for long term storage.

2.2.1.3 Generating stable cell lines using the Flp-In system

Stable cell lines expressing proteins of interest fused to APEX2 were generated using the Flp-In T-REx system in HEK293 cells according to the manufacturer's instructions. Briefly, 2×10^6 cells were seeded per 60 mm plate. The following day, 25 μg of the pcDNA5-APEX2-fusion protein plasmid and 225 μg of Flp recombinase plasmid (pOG44) were transfected into the cells using the Mirus

Trans*IT*-LT1, according to the manufacturer's instructions. Media was subsequently changed 24 h post transfection. 48 h post transfections, cells were transferred to a 15 cm dish and put on 200 μg/mL Hygromycin Gold selection. Only cells in which the pcDNA5/FRT/TO plasmid was integrated would survive. After approximately 10 days of selection, single clones were picked, as well as a pool of polyclonal cells. Cells were expanded and frozen as described above. For the NLS-APEX2 and early Sec61β-APEX cell lines, polyclonal populations were used. For all other cell lines, single clones were selected and put on zeocin counter-selection. Cells which had the pcDNA5/FRT/TO plasmid integrated in the Flp-In target site, will be zeocin sensitive. Clones which had normal cell morphology and growth rate, were hygromycin resistant and zeocin sensitive were used for further validation.

2.2.1.4 Generating of CRISPR/Cas9-mediated knockout cells

For CRISPR/Cas9-mediated knockout 2×10^5 cells were seeded in a 24-well plate. The next day, cells were transfected using Mirus Trans*IT*-LT1, according to the manufacturer's protocol. On the next day cells were put on 2 μg/mL puromycin selection. 48 h later cells were expanded using standard medium without puromycin. To generate monoclonal cell lines, single cells were sorted using the BD FACSAria III. Knockout status of monoclonal cell lines was confirmed using western blotting and immunofluorescence microscopy.

2.2.2 Molecular biology methods

2.2.2.1 Restriction-enzyme based cloning

pcDNA3 APEX2-NES was obtained from Addgene. pcDNA5/FRT/TO was obtained from Thermo Fisher Scientific. The BAF, LAP2β, Nup153$_{1-144}$ and emerin sequences were obtained from IDT as gene blocks. Climp63 and Rtn4a containing plasmids were a kind gift from Gia Voeltz. LBR, SUN2 and Sec61β containing plasmids existed in the lab. Lap2β* is a truncation that lacks the DNA and LEM domain, and was described previously by Anderson and colleagues[141]. APEX2 was PCR amplified from the pcDNA3 APEX2-NES vector, whilst removing the NES signal. N-terminally and C-terminally tagged constructs were created by digesting the pcDNA5/FRT/TO vector with KpnI and XhoI, and subsequently recombining it by Gibson according to the manufacturer's instructions with FLAG-APEX2-POI or POI-APEX2-FLAG for N- and C-terminally tagged constructs, respectively. The APEX2-NLS cell lines already existed in the lab.

2.2.2.2 Designing gRNAs for CRISPR/Cas9-mediated knockout

CRISPR gRNAs were designed using the Benchling software using the default settings[142]. gRNAs with the highest off target score where prioritised and three gRNAs chosen per gene. gRNAs generation for lamin was restricted to the first 566 amino acids, to ensure knockout for both lamin A and C. As gRNAs were expressed from U6 promoter, sequences had to start with a guanine. If sequences started with a guanine naturally, no extra steps were done. If a guanine was in the 2nd position of the sequence, the first nucleotide was deleted. If no guanine was in close proximity, one was added at the start of the sequence. ZMPSTE24 gRNAs were then cloned into the RP418 vector, LMNA/C gRNAs in the RP-557 vector.

2.2.2.3 Polymerase chain reaction

Target DNA (∼10 ng) was amplified from plasmids using the Q5 High-Fidelity 2X Master Mix, according to the manufacturer's instructions. Annealing temperatures were determined using the NEB Tm calculator https://tmcalculator.neb.

com/#!/main. If required, gradient annealing temperatures were used for PCR optimisations. PCR reactions were performed in a Bio-Rad C1000 Touch Thermo Cycler.

2.2.2.4 Agarose gel electrophoresis

1% agarose (w/v) was heated in 1 × TAE. Upon cooling, SYBR Safe DNA Gel Stain was added at a dilution of 1:10,000. DNA was mixed in a 5:1 ratio with Gel Loading Dye. Samples were loaded onto the gel with, according to need, either a 1 kb ladder or a 100 bp ladder to determine DNA product size. Samples were run at 110 V in 1× TAE buffer.

2.2.2.5 Purification of DNA from agarose

DNA bands, visualised by UV light, were excised from the agarose gel. DNA was subsequently purified using the GeneJET PCR purification kit, according to manufacturer's instructions. DNA was eluted in 30 μL elution buffer.

2.2.2.6 Determining the nucleic acid concentration

DNA concentration and purity were determined by the absorbance at 260 nm and 280 nm respectively, using a spectrophotometer (IMPLEN, NanoPhotometer N60).

2.2.2.7 Gibson ligation

Gibson ligations were performed to clone the DNA inserts into the digested, purified vectors using the Gibson Assembly Master Mix according to manufacturer's instructions.

2.2.2.8 Transformation of competent *E. coli*

Turbo competent *E. coli* (lab stock) were thawed on ice. 2 μL of the Gibson transfection mix was added to 50 μL bacterial suspension and bacteria were incubated on ice for 15 min. Afterwards, bacteria were heat-shocked at 42°C for 1 min, and then placed on ice again for another 2 min. Subsequently 250 μL SOC media was added to the cells, and cells were allowed to grow in suspension

for 45–60 min at 37°C. Bacteria were then plated on LB-agar plates containing 100 μg/mL ampicillin. Bacterial cells were left to grow over night at 37°C. The following day, colonies were picked and inoculated in 3 mL LB with 100 μg/mL ampicillin for a minimum of 6 h at 37°C.

2.2.2.9 Small-scale preparation of plasmid DNA

Minipreps were carried out according to the manufacturer's instructions with the GeneJET Plasmid Miniprep Kit, using 1.5 mL of bacterial culture and eluting DNA in 30 μL elution buffer. All constructs generated were validated by Sanger sequencing (Source Bioscience).

2.2.3 Protein biochemistry methods

2.2.3.1 Proximity-biotinylation assay

PDB were performed as described by Hung and colleagues[143]. To summarise, Flp-In T-Rex HEK293 cells stably expressing the transgene of interest were seeded to achieve 80–90% confluency 48 h after seeding. 24 h after seeding, APEX fusion-protein expression was induced through addition of 100 ng/mL tetracycline into the culture media. Another 24 h later, proximity-biotinylation reactions were performed as follows. Biotin-phenol was added to cells to obtain a working concentration of 500 μM and cells were incubated at 37°C, 5% CO_2 (v/v) for 30 min. Subsequently, H_2O_2 was added at a concentration of 1 μM to start the reaction. After 60 s, the media containing the biotin-phenol and H_2O_2 was aspirated and cells were washed three times in quencher solution to stop the biotinylation reaction. After the final quenching step, cells were either lysed for further analysis by western blot or proteomics, or processed for immunofluorescence microscopy.

2.2.3.2 Preparation of whole-cell lysate

To prepare cell lysates for western blotting, cells were placed on ice and washed once in ice-cold PBS. Subsequently, RIPA or RIPA+ buffer (for biotinylation reactions) was added directly to the plate, and cells were incubated on ice for 20 min. Cells were then collected using a cell scraper and transferred to a 1.5 mL microcentrifuge tube. Prior to immunoprecipitations (IPs), cell lysates were cleared by centrifugation at 13,000 × g, 15 min, 4°C. The supernatant was carefully collected and transferred to a new tube.

2.2.3.3 SDS-polyacrylamide gel electrophoresis (PAGE)

Proteins were resolved on Criterion TGX 4–20% gels (Bio-Rad). For this, samples were mixed to contain a final concentration of 1 × sample buffer and 100 μg/mL DTT. Samples were heated at 50°C for 10 min in a ThermoMixer. Samples were subsequently loaded on a gel, aside the Precision Protein Dual Color marker.

Gels were run in 1 × Tris/Glycine/SDS buffer in Criterion Cells at a constant current of 0.06 or 0.12 A, for one or two gels respectively, until the loading dye started running out of the gel.

2.2.3.4 Western blotting and detection

Proteins were transferred to a PVDF membrane using the Bio-Rad Trans-Blot Turbo System. For this, the PVDF membrane was activated in methanol for 1 min. Subsequently, the membrane and stacking paper were equilibrated in 1 × Trans-Blot Turbo Buffer. Blotting-sandwiches were then assembled, as detailed in the machine's instruction guide. Proteins were transferred using the midi-gel, high molecular weight settings. Membranes were then blocked in 5% milk or 5% BSA. Primary antibody incubation was done at 4°C overnight or 1 h at RT in 1% milk or BSA, respectively. Blots were subsequently washed at least three times for 10 min in PBS-T and incubated with the appropriate secondary antibody for at least 1 h at RT. Protein bands were subsequently visualised using Western Lighting ECL Pro and detected using the Amersham Imager 600.

2.2.3.5 Immunofluorescence microscopy after biotinylation

Cells were plated in Nunc Lab-Tek II CC2 Chamber Slides so they would reach ∼40% confluency on day of the biotinylation reaction. Biotinylation reactions were performed as described in section 3.1.2. After the final wash in quenching buffer, cells were fixed in 4% PFA-PBS for 10 min at RT. Subsequently, cells were washed three times in PBS, before being incubated in 50 mM glycine containing PBS for 4 min. Cells were rinsed in PBS three more times and put in blocking buffer (0.1% saponin, 1% BSA in PBS) for 30 min. Primary antibody in blocking buffer was added to cells for 1 h at RT. Subsequently, cells were washed three times in blocking buffer, before being incubated with secondary antibody in blocking buffer for 1 h at room temperature (RT) protected from light. Unbound secondary antibody was washed off in three PBS washes and cells were mounted with Fluoromount-G Mounting Medium with DAPI. Cells were visualised using the Fluorescence Microscope (ZEISS), 100 × oil immersion objective.

2.2.3.6 Differential membrane permeabilisation

Differential membrane permeabilisation was performed as described previously[78]. Briefly, HEK or HeLa cells were seeded to be ∼40% confluent on the day of processing. To prepare the cells for immunofluorescence microscopy, cells were washed once in PBS and subsequently fixed in 4% PFA-PBS for 10 min at RT. Cells were then washed three times in PBS. Permeabilisation was performed using 0.001% digitonin for 14 min at 4°C for the soft permeabilisation, or following an additional 5 min permeabilisation using 0.1% Trition-X-100; 0.02% SDS in PBS at RT for the harsh permeabilising. Antibody staining was then performed as described in chapter 2.2.3.5, using 2% BSA/PBS instead of blocking buffer.

2.2.3.7 Sucrose gradient fractionation

Subcellular fractionation to enrich for nuclear membranes was performed as described previously[52] and as visualised in figure 3.17. Cells were washed three times in PBS and 1 × in HB buffer. Subsequently, 4 mL HB+ buffer was added per 15 cm dish, and cells were incubated on ice for 15 min. After 15 min, cells were scraped off plates using a cell scraper and lysed with 30 strokes in a dounce homogenizer to reach a lysis efficiency of > 90%, as determined by trypan blue staining. Subsequently, the whole cell lysate was layered on top of 1 mL 0.8 M sucrose in HB and centrifuged at 700 × g for 10 min at 4°C in a TX-150 swinging bucket rotor on a soft decelerate setting, creating a post-nuclear supernatant (PNS) top layer and a crude nuclear membrane pellet. The crude nuclear pellet was further processed by resuspending it in 6 mL of 1.8 M sucrose in HB buffer and layering it on top of 1 mL of 2 M sucrose in a 13.2 mL tube. HB buffer was subsequently layered over the loaded sample to fill the tubes. Samples were centrifuged at 210,000 × g for 1 h at 4°C with no brake in a SW41Ti rotor. Nuclear pellets were then re-suspended in 500 μL HB buffer with 1 mM $CaCl_2$ and treated with 100 ku/mL micrococcal nuclease (NEB, #M0247S) for 15 min at 37°C. Nuclei were put on ice and NaCl added to a final concentration of 500 mM. Digested nuclei were then layered on top of a 1 mL shelf of 0.8 M

sucrose in HB and centrifuged at 2,800 × *g* for 10 min at 4°C and stopped using the soft deceleration setting in the TX-150 swinging bucket rotor. The fraction above the 0.8 M sucrose comprising the "nuclear content" was collected. The pellet, containing the nuclear envelope was resolubilised in 100 μL HB buffer with 1% SDS. To enrich for cytoplasmic membranes, after the first low speed spin described in the protocol above, the post-nuclear supernatant was centrifuged in a TLA100.3 rotor at 50,000 *g* for 1 h at 4°C. Membranes were subsequently further processed as described by Avci *et al.*[144]. In brief, membranes were resolubilised in 4 mL RM buffer. Subsequently, 2 mL of 1.5 M KOAc, 150 mM EDTA buffer was added dropwise, resulting in a final concentration of 500 mM KOAc and 50 mM EDTA. Samples were incubated on ice for 15 min and afterwards layered in 13.2 mL Beckmann tubes on top of a 6 mL high-salt sucrose cushion. Samples were centrifuged at 100,000 *g* for 1 h at 4°C. Pellets containing the cytoplasmic membranes were resuspended in 100 HB buffer with 1% SDS.

2.2.3.8 Immunoprecipitation

For immunoprecipitations, the cytoplasmic membrane or nuclear envelope fractions, which were resolubilised in 100 μl HB buffer with 1% SDS, were topped up with 1.7 mL sample binding (SB) buffer, to dilute the SDS to 0.05%. Samples were then centrifuged for 15 min at 13,500 × *g*. Next, 25 μl of Pierce Streptavidin Magnetic Beads per sample were washed twice in sample binding buffer. 25 μL of beads were incubated with each sample at 4°C overnight. The next day, the unbound fraction was collected and beads were washed as followed: three times in SB buffer, once in 1 M NaCl in SB buffer, twice in 2 M urea in SB buffer and finally three times in SB buffer. After the final wash, proteins were either eluted for western blotting or the samples were used for proteomic analysis. To elute samples for western blotting and in-solution digestion, all beads were incubated in 1 x SDS elution buffer with 2 mM biotin and 100 mM DTT for 15 min at 65°C. To prepare beads for proteomics analysis by on-bead digest, 20% of the beads were collected and bound protein eluted (as specified for the individual

experiments), whereas the rest of the beads were further processed for Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.

2.2.4 Proteomic analysis

2.2.4.1 On-bead trypsin digest

Streptavidin magnetic beads were washed twice in 300 μL 50 μM ammonium bicarbonate (AB) solution. In the meantime, 20 μg lysophilized trypsin (Promega, #v5111) was resuspended in 100 μl resuspension buffer to generate a 0.2 μg/μL trypsin solution. After two wash steps, beads were resuspended once more in 300 μl 50 mM AB solution and 0.4 μl of trypsin solution was added per sample. Samples were incubated at 900 rpm in a ThermoMixer at 37°C overnight. On the next day, formic acid was added to a final volume of 1%. Peptides were then further processed by C18 purification, as described in chapter 2.2.4.3

2.2.4.2 In-solution trypsin digest

Elutions from streptavidin IPs were topped up with ultrapure water to a final volume of 200 μL. Next, 600 μL methanol was added to the samples, as well as 150 μL chloroform and samples were vortexed. Another 450 μL of ultrapure water was added, and samples were vortexed again, before being centrifuged at 13,500 $\times$ g for 1 min at RT. This results in phase-separation of two layers: an aqueous top layer and a chloroform bottom layer. The upper phase was discarded and the sample was topped up with another 450 μL of methanol. Samples were vortexed again and centrifuged at 13,500 g at RT for 2 min. The supernatant was then carefully removed. Pellets were resuspended in 200 μL of ultrapure water and the procedure was repeated once more. After taking of the supernatant the second time, the residual methanol was allowed to evaporate. Proteins were then digested as follows: pellets were resuspended in 50 μL of 6M urea by vortexing and sonication. Subsequently, 250 μL of ultrapure water was added to reduce the urea concentration to $<$ 1 M. Trpysin was added at a final concentration of 0.2 μg/μL and proteins were digested at 37°C overnight. On the next day, formic acid was added to a final concentration of 1%. Peptides were then further processed by C18 purification.

2.2.4.3 C18 Sep-Pak purification of peptides

For this, SOLA HRP SPE cartridges (Thermo Fisher Scientific, #60409-001) were equilibrated with 500 μl of solution A (65% CH_3CN, 35% H_2O, 0.1% TFA) and subsequently flushed in 1 mL of solution B (98% ultrapure water, 2% CH_3CN, 0.1% TFA). The peptide mix was then loaded onto the column and the column was once more washed with solution A. Subsequently, purified peptides were eluted in 600 μl of solution B and dried in a speed vac.

2.2.4.4 LC-MS/MS Settings

Peptides were injected into a LC-MS system comprised of a Dionex Ultimate 3000 nano LC connected to a Thermo Q-Exactive. Peptides were separated on a 50 cm long EasySpray column (ES803; Thermo Fisher Scientific) with a 75 μm inner diameter using a 60 minute gradient of 2 to 35% acetonitrile in 0.1% formic acid and 5% DMSO at a flow rate of 250 nL/min. MS1 spectra were acquired with a resolution of 70,000 and AGC target of 3×10^6 ions for a maximum injection time of 100 ms. The top 15 most abundant peak ions were fragmented after isolation with a mass window of 1.6 Th at a resolution of 17,500 with a maximum injection time of 128 ms. Normalized collision energy was 28% (HCD).

2.2.4.5 MaxQuant Analysis

Raw data processing was performed using the MaxQuant Software version 1.6.3.4[145]. In brief, proteins were identified against the human reference proteome (v2019.09) together with the MaxQuant integrated contaminant file. The search was performed for trypsin/P cleavage with a maximum of two missed cleavages. Variable modifications were set for methionine oxidation (M) and protein N-terminal acetylation. The instrument type was set for orbitrap and the default settings were used, with the exception of the first search peptides tolerance that was increased to 50 ppm. Cut-offs for peptide and protein identification were set at 0.01 FDR against a revert decoy library.

2.2.4.6 Perseus Analysis

Where stated, data analysis were performed using Perseus 1.6.1.4.0.[146]. Briefly, FASTA sequences of all proteins identified in the MaxQuant search were retrieved from Uniprot[147] and transmembrane helix predictions were run using TMHMM-2.0[148,149]. Data was subsequently uploaded in Perseus. Analysis was done identical either with raw intensities or LFQ intensities[150], as specified. Proteins were filtered out if the MaxQuant search classified them as potential contaminants, reverse or only identified by one site. Intensities were subsequently log_2 transformed. Histograms were used to assess whether intensities followed a normal distribution. Replicates were grouped and filtering applied for valid values according to groups as specified for each bait protein. Subsequently, missing values were replaced from normal distribution (width 0.3, down shift 1.8). Experimental quality and clustering of replicates was assessed using multi-scatter plots with Pearson correlation, hierarchical clustering and principal component analysis. For pairwise-comparisons of two probes, two-sided t-tests were performed ($s0 = 0.1$, FDR $= 0.05$, Number of randomisations $= 250$). Data was subsequently visualised using various scatter plots, profile plot, and PCA analysis.

For the profile plot analysis tool, each protein had the three fold change (FC) values associated. Of note, some proteins were not detected in individual probes, which resulted in issues in the analysis tool. Therefore missing values were imputed to enable optimal detection of INM proteins by the tool, as follows: for the Nup153_{1-144} probe, missing values were imputed with a constant value equal to the smallest FC observed, in an assumption that a protein not labelled by this bait is unlikely to be an INM. In contrast, in the Rtn4a comparison, missing values were replaced with a value equal to the largest FC observed, as a lack of the identified protein likely indicates that the protein is not a INM protein. Missing values in the Sec61β sample were imputed equal to 0, as no conclusions can be drawn on whether a protein may or may not localise to the INM.

2.2.4.7 R analysis

Data analysis was performed using custom scripts from Tiaan Heunis. Briefly, contaminants, reverse sequences and peptides only identified by one site were filtered out using the package 'tidyr'. Raw intensities were subsequently log_2 transformed. The quality and variance of the sample was assessed using box plots of the log_2 intensities, heat maps created using the 'pheatmap' packages, as well as principal component analysis (PCA). Replicates were grouped and filtering applied for valid values according to groups as specified for each bait. Variance stabilisation normalisation was applied using the 'vsn' package[151]. Subsequently, missing values were imputed using the 'imputeLCMD' package[152]. Hypothesis testing for pairwise comparisons was performed using the 'limma' package[153]. Cut-offs of $1 < log_2(FC) < -1, p.adj. \leq 0.05$ were used to determine significantly enriched and depleted proteins. For enrichment analysis between baits, whenever possible samples from the same facility were analysed together to minimise batch effects created from different mass analysers (for an overview of replicates see appendix table A.1). For example, for the pairwise comparison Climp63 versus $Nup153_{1-144}$ replicates 1–3 were used. In contrast, for the pairwise comparison $Nup153_{1-144}$ wt vs *LMNA* KO, the wt replicates 4–9 were used.

Let us choose for ourselves our path in life, and let us try to strew that path with flowers.

— Émilie du Châtelet

3

Developing an assay to study the protein composition of ER domains

Contents

3.1 Introduction 45
3.1.1 An overview over ER domains 45
3.1.2 Studying organelles by proximity biotinylation 46
3.2 Rational 48
3.3 Results 49
3.3.1 Designing and cloning APEX2-fusion constructs . . . 49
3.3.2 Choosing a model cell line 53
3.3.3 Generating stable FlpIn293 APEX2-fusion cells lines . 55
3.3.4 Testing and optimising protein expression 55
3.3.5 Investigating the role of proteasomal degradation on protein levels 57
3.3.6 Validating APEX2-PDB of different bait proteins . . . 59
3.3.7 Quality control of the APEX2-PDB system 60
3.3.8 Determining quenching efficiency 61
3.3.9 Validating APEX2-fusion protein localisation 65
3.3.10 Optimising sample preparation for LC-MS/MS 72
3.3.11 Enriching for membrane proteins 75
3.4 Discussion 81
3.4.1 Cloning strategy and cell line generation 81
3.4.2 LC-MS/MS sample preparation optimisation 85
3.4.3 Coupling proximity-dependent biotinylation to fractionation 86
3.4.4 Concluding remarks 88

3.1 Introduction

3.1.1 An overview over ER domains

The ER consists of various continuous, but distinct domains. The INM and ONM together form the NE. Thereby the INM contains a distinct set of proteins, many of which interact with chromatin or the nuclear lamina and are crucial for the versatile functions of the INM[41]. Stretching away from the NE into the cytoplasm, the ER forms a network made of flat membrane sheets and long tubules interconnected by three-way junctions. ER sheets are in the close periphery of the nucleus and are involved in the biogenesis of membrane and secretory proteins. In contrast, stretching far out into the cytoplasm are the dynamic and branched ER tubules. The functions of ER tubules include the formation of contact sides with other organelles, which play important roles in transfer of molecules, as well as in the regulation of organellar positioning, fission and inheritance[2,4]. Yet, for other ER functions like lipid homeostasis and calcium level regulation, it is not clear whether these functions happen widespread throughout the ER or whether they also take place in distinct ER domains. Confocal microscopy analysis in multiple cell lines indicated that whereas some ER proteins are uniformly spread over the ER, others are enriched either in ER tubules or ER sheets[23]. Proteins enriched in the ER sheets included p180, as well as components of the Sec61, TRAP and OST complex and the SRP receptor. In contrast, ER tubules were enriched in reticulons, which play a role in bending and shaping the ER membrane. Whereas data from these imaging experiments gave a first indication that protein enrichment in tubules and sheets exists, these studies were focused on investigating the localisation of individual proteins. However, our understanding of the global protein composition in these distinct domains is still limited.

3.1.2 Studying organelles by proximity biotinylation

One approach to understand organelles and their functions is by analysing their composition, in particular their proteins. Over the years, a method of choice for investigating the protein composition of membrane-bound organelles was biochemical fractionation coupled to immunoblotting or proteomic analysis. Yet, there are various disadvantages associated with this methodology:

1. Cell lysis and fractionation often disrupt organelle integrity
2. Certain organelles, or subcompartments of organelles, cannot be isolated by biochemical fractionation (e.g. the inner and outer membrane of the NE cannot be separated)
3. There may be contamination between different fractions due to contact-sites between organelles, as well as due to similar densities

Excitingly, newly discovered PDB assays, as given an overview in chapter 1.8, are powerful new tools to study protein compositions in organelles and overcome some of the issues associated with biochemical fractionation coupled to proteomic analysis.

The main aim of this project was to tailor a PDB assay to probe the protein composition of different ER domains, such as ER tubules, ER sheets and the INM. One of the perks of a PDB assay is that it works in living, intact cells. This offers the exciting opportunity to use PDB assays to investigate how disease phenotypes, cell cycle progression or external stimuli such as drug treatments affect protein composition changes over time. The fast enzyme kinetics of the APEX2 peroxidase allows labeling of whole organelles in the range of a few seconds to a minute, making it ideally suited to investigate dynamic processes by providing a snapshot of protein composition at a very distinct time point[134–136]. In contrast, BioID has labelling times of ∼18 h, making it unsuitable for studying dynamic changes[128]. Of note, turboID and miniTurbo ID were only developed in 2018, and were therefore not available at the start of this project[129]. Additionally,

their labelling time is still in the range of 10 minutes, making it considerably faster than the original BioID system, but still unsuitable for investigating short-lived and very dynamic processes.

A further point to consider in the design of the project was that INM proteins need to pass by the NPC to reach the INM. The narrow peripheral channel of the NPC thereby creates a size limit of < 60 kDa to the cytoplasmic/nucleoplasmic domain of INM proteins. It is therefore crucial to ensure fusing APEX2 to the cytoplasmic domain of the bait proteins doesn't exceed the 60 kDa size limit, as this will prevent proteins from reaching the INM[73,77,78,154,155]. APEX2 has a molecular weight of 27 kDa. As this is the same size as GFP, and as vast amounts of literature exists on tagging INM proteins with GFP for visualisation, the general assumption can be made that if GFP-tagging doesn't effect INM targeting, the same can be expected of APEX2-tagging.

3.2 Rational

Considering the ER is the largest organelle in mammalian cells and the wide-range of functions the ER has, one might hypothesis that different ER-membrane proteins will be enriched in different ER domains. Notably, among the various ER domains, the INM has the most distinctive protein composition, yet the close spatial proximity of the inner and outer nuclear membrane makes it difficult to determine in which of the two membranes a protein is placed. Besides the NE, it is also know that certain proteins are enriched in specialised ER subdomains such as ER tubules, ER sheets or various ER-organelle contact sites. Yet, due to the physical continuity of the membrane between these domains, and the large amount of proteins localising to the ER, a global overview of how proteins distribute along the ER is still lacking.

PDB coupled to proteomic analysis is a powerful tool to investigate protein localisation *in situ* at high-throughput. APEX2, is a second generation engineered ascorbate peroxidase, that can be genetically targeted to a cellular compartment of interest[133–136]. Upon addition of H_2O_2 and in the presence of biotin-phenol, APEX2 tags surrounding proteins with biotin-phenoxyl radicals. This process happens *in situ*, and due to the fast kinetics, the labeling time ranges in the scale of seconds. The labeling radius is estimated at around ∼20 nm, with the highest labelling closest to the APEX2, and labelling efficiency reducing the further away a protein is[123]. Subsequent lysis of the cells, followed by isolation of the biotinylated protein and their identification, enables to obtain spatial-temporal snapshot of the bait protein surrounding.

This project aimed to use stable FlpIn293 cells expressing APEX2-fusion proteins localising to various domains of the ER (INM, ER sheets and tubules) under a tetracycline-inducible promoter to gain a coherent understanding of protein compositions along different sections of the ER. In addition to creating a static understanding of the ER domains, the APEX2-ER system could provide a useful tool to study various perturbation to cells, as well as investigate dynamic changes in the protein composition in response to different stimuli.

3.3 Results

3.3.1 Designing and cloning APEX2-fusion constructs

To enable targeting of APEX2 to different ER domains, it needs to be expressed as a fusion protein. For the design of the expression strategy the Flp-In T-Rex system was chosen, in which the host cell line contains an integrated FRT site that enables homologous recombination. This allows easy, isogenic, stable integration of genes under a tetracycline (tet)-inducible promoter. The Flp-In T-Rex system requires cloning of proteins of interest (POIs) in the multiple cloning site of the pcDNA5/FRT/TO vector. A range of bait proteins were chosen for APEX2 targeting to ER tubules, ER sheets, as well as the INM. An overview over the chosen bait proteins can be found in table 3.1 and their topology can be viewed in figure 3.1.

Table 3.1: List of constructs.

Construct	Tag	Localisation	Biological function
BAF	N-terminus	INM (soluble) + cytoplasm	Plays a role in NE assembly, chromatin organisation and gene expression.
Climp63	N-terminus	ER (sheet)	Mediates anchoring of ER and MTs.
Emerin	N-terminus	INM	Stabilizes and promotes the formation of a nuclear actin cortical network. Involved in linking centrosomes to the NE.
Lap2β	N-terminus	INM	Involved in nuclear lamina assembly and structural organization of the NE.
Lap2β*	N-terminus	INM	Lap2β lacking the DNA and LEM domains, but still contains the lamin-interacting and the transmembrane domains. Localises to the INM[141].
LBR	C-terminus	INM	Sterol reductase activity and involved in anchoring the lamina and heterochromatin to the NE.
NLS	C-terminus	Nuclear (soluble)	No function. Additionally contains a GFP-tag at the very C-terminus.

Continuation of table 3.1			
Construct	**Tag**	**Localisation**	**Biological function**
$Nup153_{1-144}$	C-terminus	INM (membrane associated)	Component of the nuclear basket of the NPC. Involved in trafficking between the cytoplasm and the nucleus. The truncation contains the minimal NE-targeting cassette and binds to the INM via an amphipathic helix. Compared to the full length protein, this truncation acts as a generic INM motif and shows no enrichment at the NPC[53].
Rtn4a	C-terminus	ER (tubules)	Involved in ER curvature and ER tubule formation.
Sec61β	N-terminus	ER + NE	Component of SEC61 channel-forming translocon complex. Necessary for protein translocation into the ER.
SUN2	N-terminus	INM	Component of the LINC complex. Involved in connecting the nuclear lamina and the cytoskeleton.

The APEX2-fusion proteins were designed to consist of a bait protein required for targeting, the APEX2 peroxidase, as well as a FLAG-tag to enable easy visualisation by western blot (WB) and immunofluorescence (IF) microscopy. A visual representation summarizing the cloning strategy can be seen in figure 3.2. Proteins were tagged either C- or N-terminally, according to the following two aspects:

1. Ensure that the APEX2 enzyme faces the cytoplasm/nucleoplasm

2. When information was available, chose termini where a tag of a similar size (e.g. GFP) was shown not to interfere with protein localization and/or function

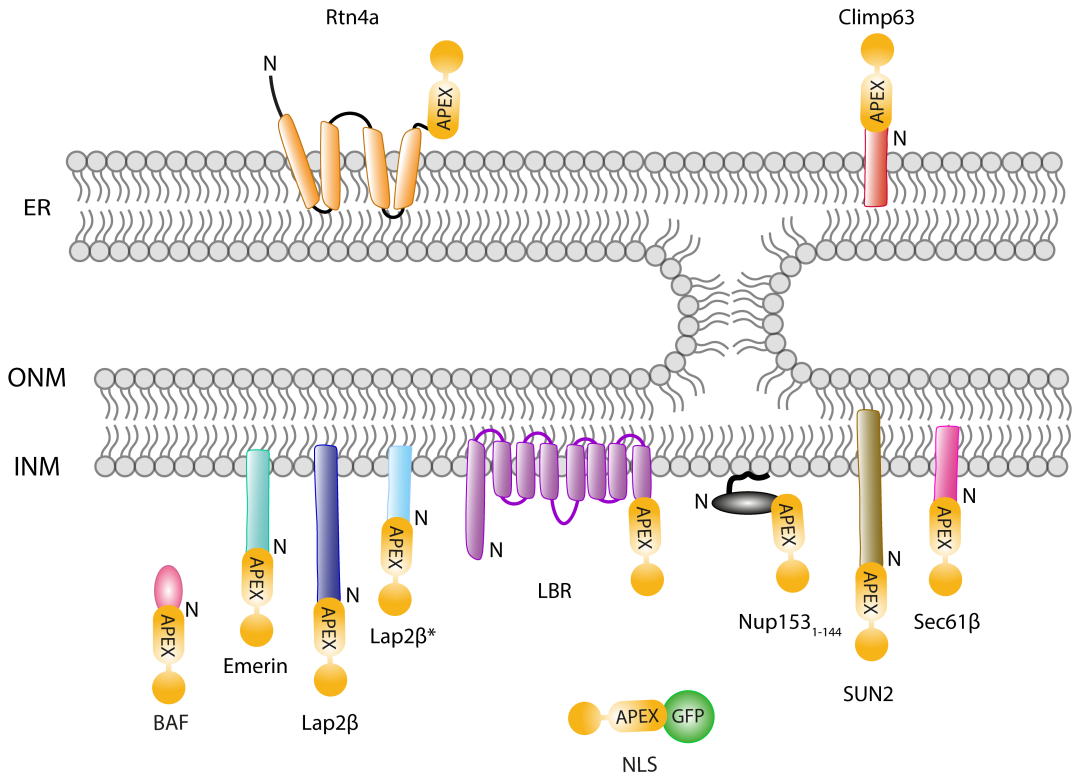


Figure 3.1: Topology of APEX2-bait proteins. The phospholipid bilayer is depicted in grey, consisting of the continuous INM, ONM and the bulk ER. FLAG-APEX2 (yellow) was fused to different bait proteins to for targeting either to the INM, the ER or the nucleus. The N-terminus of each protein is marked with an "N". BAF and NLS are soluble nuclear proteins. Nup153$_{1-144}$ is associated to the INM through an amphipathic helix. Emerin, Lap2β, Lap2β*, SUN2, Sec61β and Climp63 are type-II membrane proteins. LBR is a multispanning membrane protein. Rtn4a is known to adopt a hairpin-like structure. Detailed information about each bait and their physiological function can be found in table 3.1.

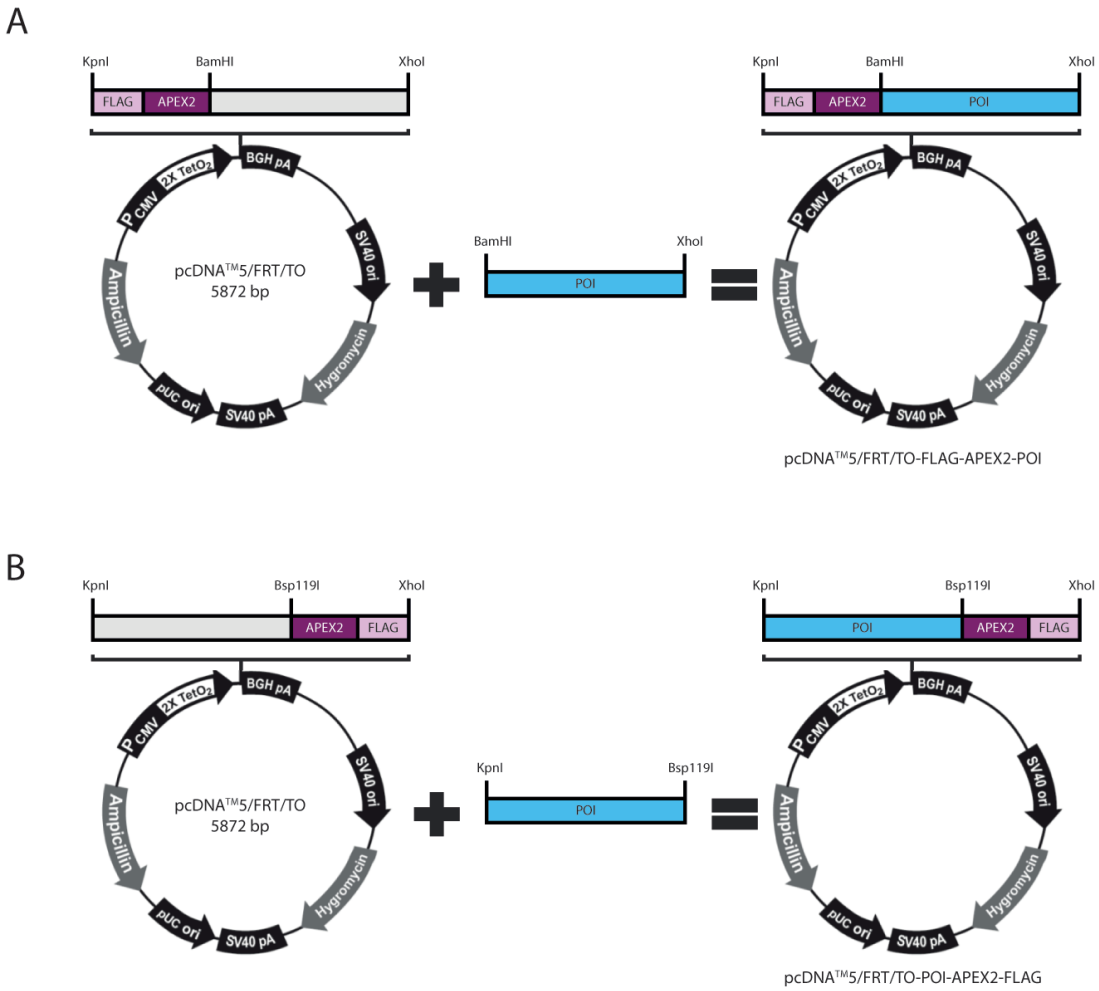


Figure 3.2: Cloning strategy. Illustration of the cloning strategy to create N- or C-terminally tagged FLAG-APEX2 fusion proteins. For the first plasmid generated for this project, FLAG-APEX2 was cloned into the pcDNA5 vector simultaneously with the protein of interest (POI) using Gibson assembly. For all subsequent fusion proteins, only the bait sequences were replaced with each other. pcDNA5/FRT/TO plasmid maps were obtained from the Thermo Fisher pcDNA5/TO User Guide.

3.3.2 Choosing a model cell line

Having cloned the APEX2-fusion proteins, the next step was to choose a model cell line. The large majority of studies using the APEX2-PDB system used the Flp-In T-REx HEK293 (FlpIn293) cell line. Yet, another interesting option would be T-REx HeLa cells. HeLa cells are very versatile in their application and are favourable over FlpIn293 cells for imaging based experiments. Additionally, HeLa cells are routinely used in cell cycle synchronisation assays, which could be useful for potential future applications of the assay, such as studying cell cycle dependent changes in the protein composition of ER domains. Yet, little data is available on the functionality of the APEX2-PDB system in HeLa cells.

Therefore, the first step was to test expression and activity of APEX2-fusion proteins in T-REx HeLa cells. For this, the APEX2-FLAG-Sec61β construct was used. Polyclonal FlpIn293 cells expressing APEX2-Sec61β existed in the lab, whereas polyclonal T-REx HeLa cells expressing the constructs were created as described in chapter 2.2.1.3.

Expression of FLAG-APEX2-Sec61β was induced in both FlpIn293 and HeLa cells using different concentrations of tetracycline and assessed by α-FLAG WB (figure 3.3A). Similar protein expression levels were observed, as judged by protein band intensity, between the two cell lines at all three tested tetracycline concentrations.

To test APEX2 labelling efficiency, biotinylation reactions were performed and labelling quenched after 60 s. Cells were subsequently lysed in RIPA+ buffer and streptavidin IPs were performed to enrich for biotinylated proteins. Conditions in which either tetracycline or H_2O_2 were omitted were used as negative controls. α-FLAG and streptavidin blots were done to investigate APEX2-Sec61β expression and biotinylation efficiency respectively (figure 3.3B & C). Surprisingly, whereas protein expression was comparable in the FlpIn293 and HeLa cell lines, there was a strong difference observed in the amount of biotinylated protein detected. This indicated that the APEX2 system is much less active in HeLa compared to FlpIn293 cells.

Whereas no literature was found to support this hypothesis, direct communication with the Ting Lab that established the APEX2-PDB system supported these findings.

Concluding, these results indicated that T-REx HeLa cells are not optimal for investigating the protein composition of ER domains using an APEX2-PDB system, thus the FlpIn293 cell line was chosen for all future PDB experiments.

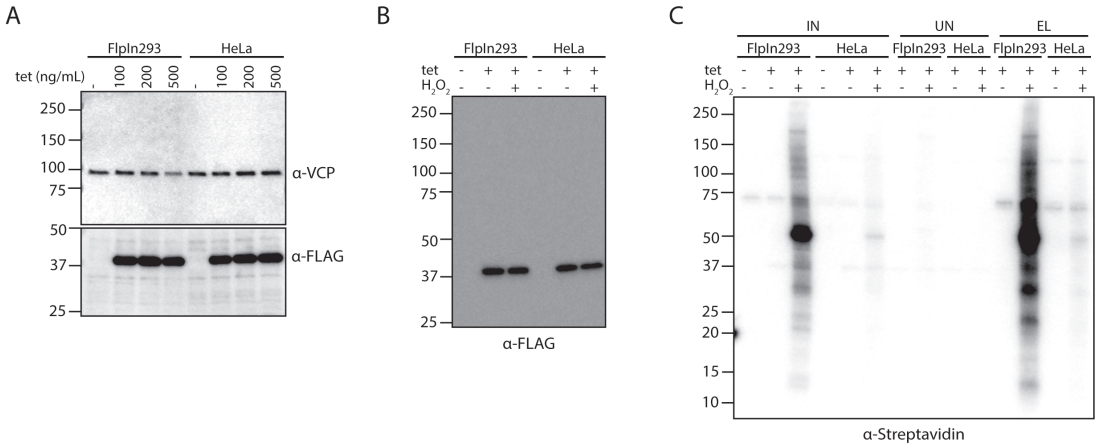


Figure 3.3: APEX2 PDB in Flp-In T-REx HEK293 & T-REx HeLa cell lines. a) Anti-FLAG WB for WCL originating from FlpIn293 or T-REx HeLa cell lines encoding the FLAG-APEX2-Sec61β. Protein expression was induced using 100, 200 or 500 ng/mL tetracycline. For the non-induction control, tetracycline was omitted. VCP was used as a loading control. b) α-FLAG WB for WCL lysate. Cells were induced with 500 ng/mL tetracycline. Control cells lines either had the tetracycline omitted, or the H_2O_2. c) Streptavidin WB on the input (IN), unbound (UN) or elution (EL) fraction of IPs. IN & UN: 3.5 % of sample. EL: 16 %.

3.3.3 Generating stable FlpIn293 APEX2-fusion cells lines

After deciding for a model cell line, stable FlpIn293 cell lines expressing various APEX2-fusion proteins were created using the Flp-In T-REx system (figure 3.4). The Flp-In T-REx system allows the stable integration of a gene of interest into a specific genomic Flp recombination target (FRT) site in the cellular genome by homologous recombination. The system relies on the co-transfection of two plasmids: the pcDNA5/FRT/TO plasmid containing a FRT site and the gene sequence of interest, as well as the pOG44 plasmid that constitutively expresses the Flp recombinase. FlpIn293 cells with a single, stably integrated FRT site in a transcriptionally active locus on their genome are commercially available. Successful recombination of the expression vector into the genomic FRT site will render cell lines hygromycin resistant and zeocin sensitive. Hygromycin selection and zeocin counter-selection can therefore be used to obtain cells in which the plasmid sequence for the fusion protein was integrated into the FRT site. Single cell clones were picked and chosen for further validation, if they met the following criteria:

1. Similar growth rate to the parental cell line
2. Non-aberrant morphology, as assessed by microscopy (Inverted microscope, ZEISS Primovert, 100 x magnification)
3. Hygromycin resistance
4. Zeocin sensitivity

3.3.4 Testing and optimising protein expression

Once the APEX2-FlpIn293 single cell clones were generated, protein expression for all constructs had to be tested and optimal conditions determined for induction of the tet-inducible promoter. Therefore, cells were grown in a range of tetracycline concentrations (0 ng/mL, 100 ng/mL, 500 ng/mL, 1000 ng/mL) and protein expression was determined by anti-FLAG WB (figure 3.5).

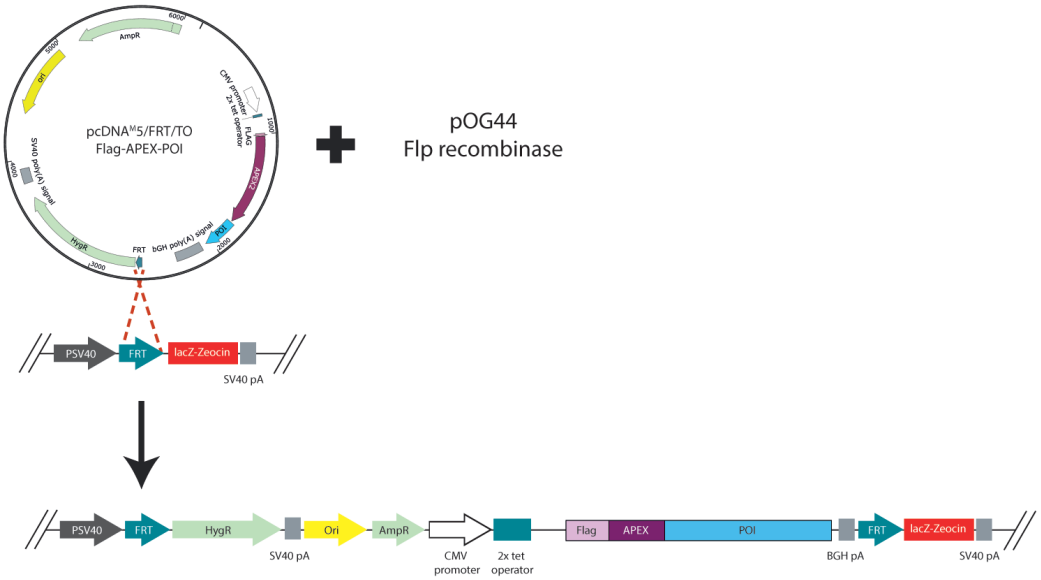


Figure 3.4: Integration of the Flp-In T-REx vector. The FRT/TO FlpIn system is based on a site-directed recombination technology. The Flp recombinase, constitutively expressed from the pOG44 plasmid, results in the recombination of sequences between the short flippase recognition target (FRT) sites present in the genome of the cell line, as well as the pcDNA5 vector. Various different cell lines, including Flp-In™ T-REx™-293 cells, containing a single integrated FRT site at a transcriptionally active genomic locus, are commercially available. This technique enables targeted integration of a gene sequence of interest in the same locus in each cell. Successful integration of the gene sequence in the FRT site renders cells hygromycin resistant and zeocin sensitive. The tet repressor represses transcription in absence of tetracycline.

For none on the constructs the tetracycline concentration affected protein expression levels, indicating that there is no linear relationship between the tetracycline concentration and the amount of protein detected. Instead the tet-promoter follows a strict on/off mechanism. Curiously, whereas all genes were expressed under the same promoter, there was strong variance in protein levels observed between different constructs, ranging from low levels for NLS-APEX2 to very strong levels as seen for APEX2-Climp63. Of note, whereas there was high variability between different baits, very little difference was observed between single cell clones expressing the same constructs (data now shown). Various reasons, such as differences in protein translation or degradation, as well as varying mRNA levels, could be responsible for these differences.

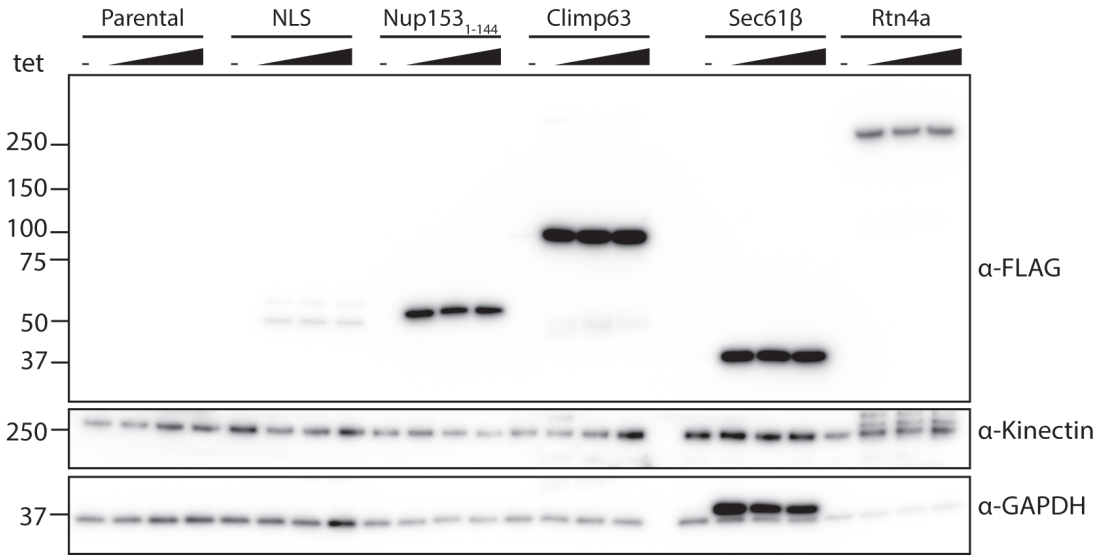


Figure 3.5: Titration of tetracycline concentration of APEX2-fusion protein expression. Anti-FLAG WBs from WCL originated from FlpIn293 cell lines expressing APEX2-fusion proteins subjected to tetracycline titrations for 24 h. Tetracycline conditions used were 100 ng/mL, 500 ng/mL and 1000 ng/mL. Unstimulated cells were grown in the absence of tetracycline. FLAG bands appear at the expected molecular weight for all constructs. Kinectin and GAPDH blots are used as loading controls. Top band for Rnt4a in the Kinectin blot corresponds to the FLAG band. Top band for Sec61β in the GAPDH blot corresponds to the FLAG band.

3.3.5 Investigating the role of proteasomal degradation on protein levels

One possibility that could explain the variance observed in protein expression levels, is that different bait proteins might be turned over at different rates. To test this possibility, cells were treated with CB-5083, a p97 inhibitor that interferes with protein extraction from the membrane, a necessary step for subsequent ubiquitin-proteasomal degradation. For this, protein expression was induced for 12 h. Subsequently, the control and the proteasome inhibition group were treated with DMSO or 2.5 μM CB-5083 for 4 h, respectively.

WB analysis showed no difference in protein levels for any of the baits upon treatment with CB-5083. In contrast, the well established ERAD substrate SQLE, used as a positive control, showed protein stabilisation upon CB-5083 treatment.

Concluding, these results indicated that the variability observed in bait protein expression levels is not caused by differences in proteasomal turnover rates (figure

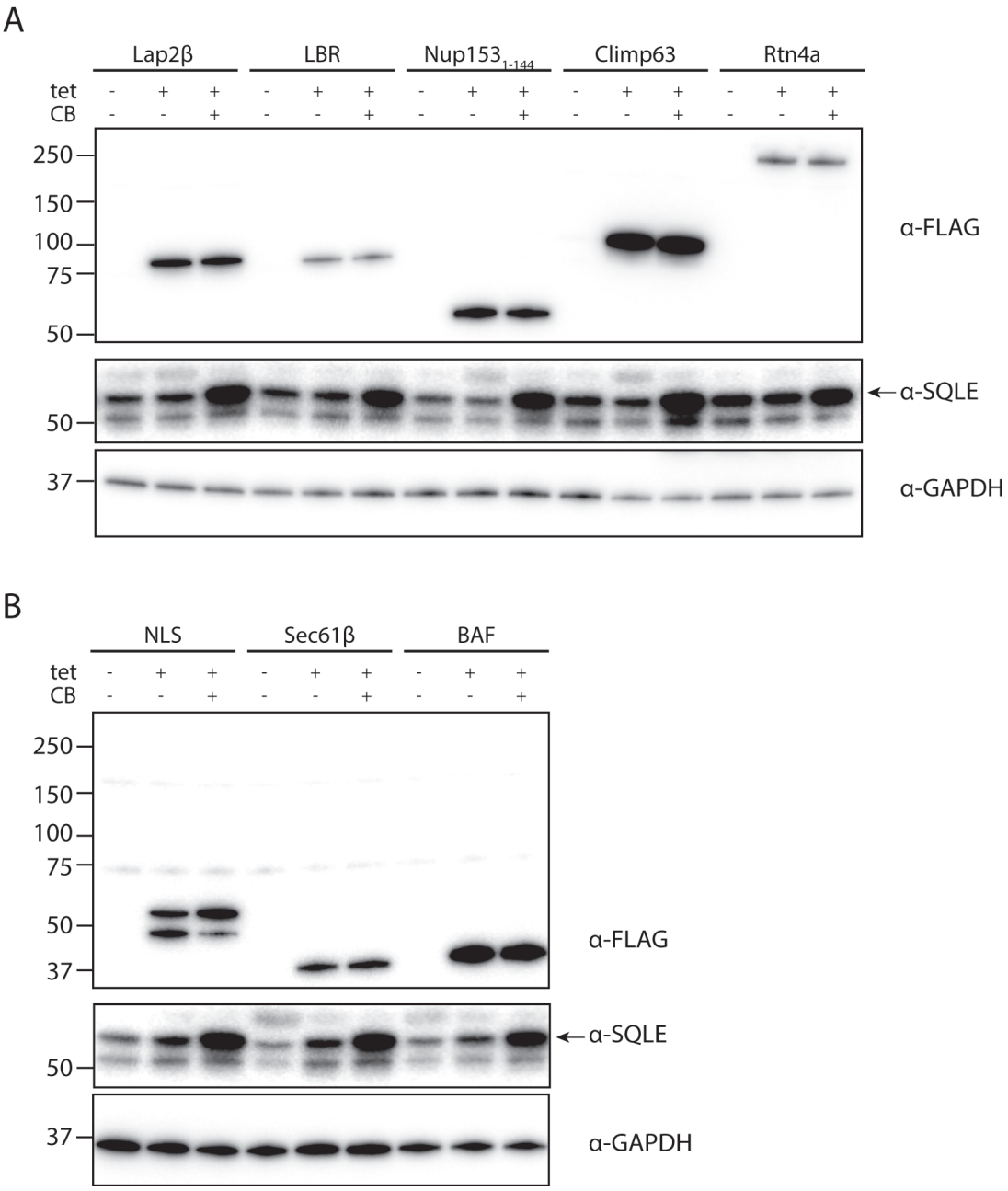


Figure 3.6: APEX2-fusion protein stability assay. Cell lines were grown either in presence of absence of 100 ng/mL tetracycline for 16 h. Cells expressing the fusion proteins were then either treated with DMSO or the p97 inhibitor CB-5083 (2.5 μM) for 4 h. Cell lysates were subjected to SDS-PAGE followed by immunoblotting. GAPDH acts as a loading control. SQLE acts as a positive control for CB-5083 treatment.

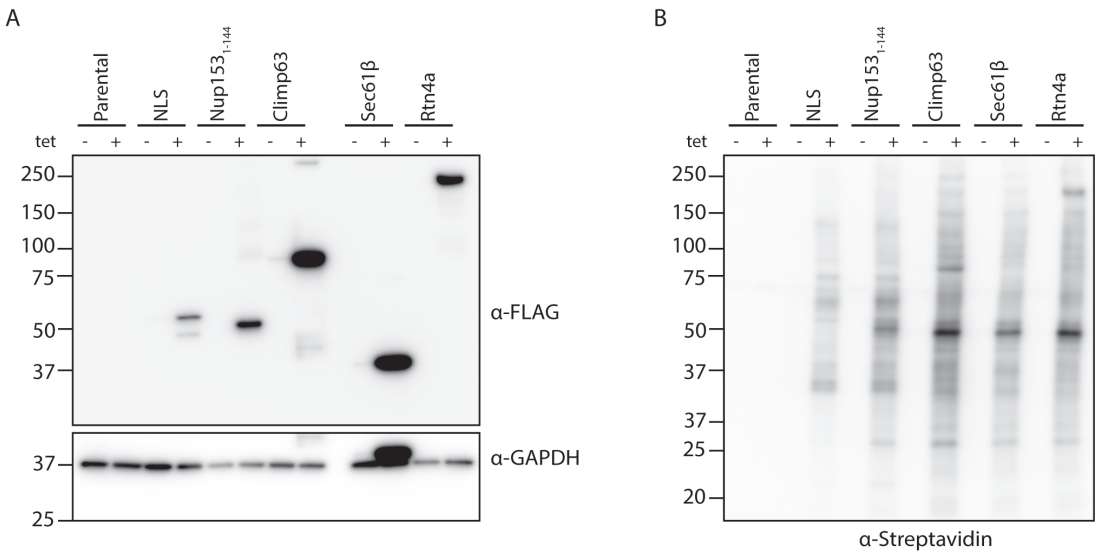


Figure 3.7: Expression and activity of APEX2 fusion protein. Anti-FLAG (A) and streptavidin (B) WBs from WCL lysate originated from FlpIn293 cell lines encoding APEX2-fusion proteins. APEX2-fusion protein expression was induced using 500 ng/mL of tetracycline. Unstimulated cells were grown in the absence of tetracycline. GAPDH serves as a loading control. Top band for Sec61β in the GAPDH blot corresponds to the FLAG band.

3.6). Indeed, all proteins tested were shown to be stable and unaffected by inhibition of proteasomal degradation over a 4 h time span.

3.3.6 Validating APEX2-PDB of different bait proteins

After validating that all baits were expressed, the next step was to assess the activity of the different APEX2-fusion proteins. For this experiment, wild type (wt) FlpIn293 cells and cell lines expressing different bait proteins were grown either in presence or absence of tetracycline. PDB reactions were performed, cells lysed and anti-FLAG and streptavidin WB were run using whole cell lysate (WCL) to test for expression and activity, respectively (figure 3.7). In line with previous tetracycline titrations (figure 3.5), the results of this experiment showed that expression levels varied widely between constructs (figure 3.7A). Additionally a strong positive correlation was apparent between protein expression levels and the amount of biotinylated proteins detected by anti-streptavidin WB, with highly expressed proteins leading to strong biotinylation levels, while lower expressed proteins resulted in less biotinylated proteins (figure 3.7B).

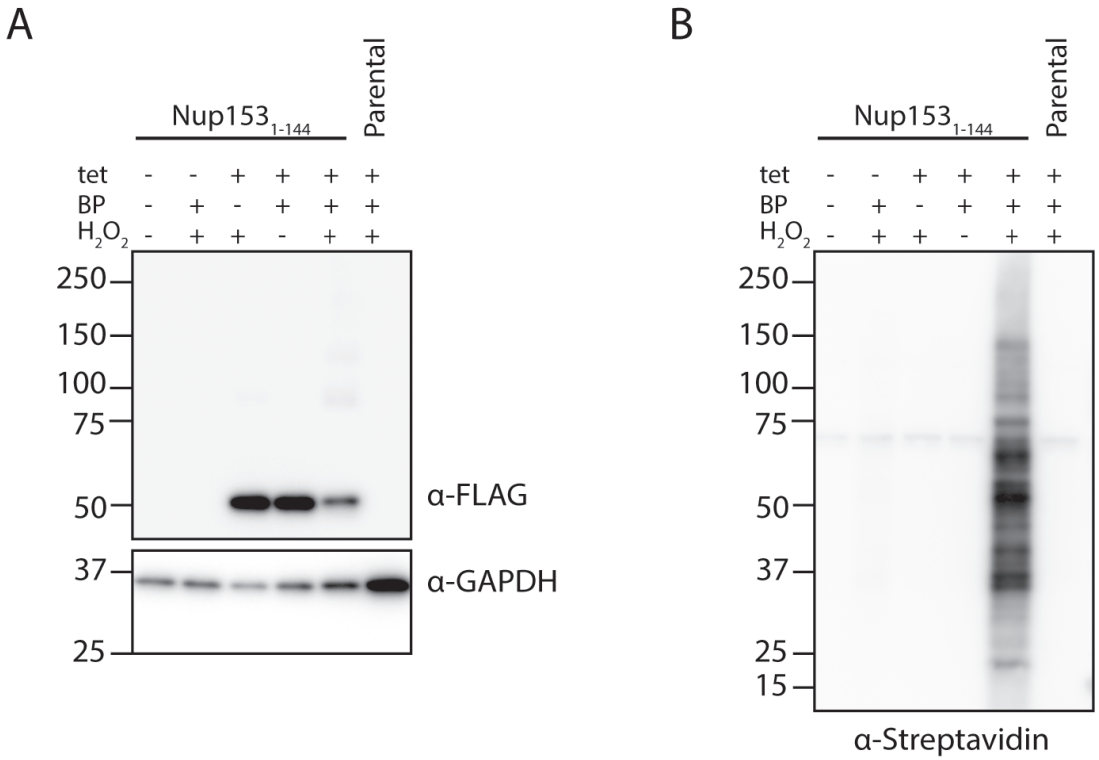


Figure 3.8: Validating the APEX2 proximity dependent biotinylation system. Anti-FLAG and streptavidin WBs from WCL lysate originated from a WT FlpIn293 cell line or a Nup153$_{1-144}$-APEX2-FLAG cell line. Cells were grown either in presence of absence of tetracycline. Biotinylation reactions were performed as described, or omitting either BP or H_2O_2. GAPDH was used as a loading control.

3.3.7 Quality control of the APEX2-PDB system

Next, quality control experiments were performed to test the dependency of the APEX2-biotinylation system on its individual components (tet, BP and H_2O_2). Data is shown for the FlpIn293 Nup$_{1-144}$-APEX2 cell line, though results are representative for all baits. For the experiment, WT FlpIn293 cells used as a negative control and Nup$_{1-144}$-APEX2 cells, were either treated with all three components (tet, BP and H_2O_2) needed for the PDB, or with one of them omitted. APEX2-biotinylation was performed, cells lysed and anti-FLAG and streptavidin WBs were run to test for expression and activity, respectively (figure 3.8).

In absence of tetracycline, no expression of protein was observed, indicating that there was no leakage from the promoter. Additionally, as expected in samples in which tetracycline was added, but either BP or H_2O_2 was omitted, the

$Nup153_{1-144}$-APEX2 protein was expressed but no biotinylation was observed. This data indicated that the APEX2-PDB under a tet-inducible promoter of the Flp-In T-REx cells worked well, and that there was no leakage, and no background biotinylation, if any of the components needed for the reaction was omitted.

3.3.8 Determining quenching efficiency

APEX2-labelling is performed *in situ*. The peroxidase reaction is quenched with antioxidants present in the quenching buffer and after the third quenching buffer wash, cells get lysed in RIPA+ buffer. Thereby the PDB assay relies on the assumption that quenching is complete and no labelling happens post lysis.

In order to test this assumption and investigate whether quenching is complete an experiment was set up following the illustrated workflow of figure 3.9A. In this experimental setup, if quenching was complete, no labelling would happen in the mixed lysates. This in turn means that $Nup153_{1-144}$ would not get biotinylated and would be lost in the unbound fraction of a streptavidin IP.

The exact setup was as followed: two different cell lines, Climp63 and $Nup153_{1-144}$, were grown in triplicates and as an individual sample. Bait protein expression was induced using tetracycline for both the FlpIn293-Climp63 and FlpIn293-$Nup153_{1-144}$ cells. 16 h post-induction, BP was added to two Climp63 replicates (samples 2 & 3), and omitted from one (sample 1). Biotinylation reactions were then performed as described previously. After 60 s of labelling, the sample in which the BP was omitted, and one of the BP replicates, were washed three times in quenching buffer (samples denoted 1 & 3). For sample 2, the reaction was not quenched. Subsequently all three samples were lysed and each mixed independently at a 1 : 1 ratio with cell lysates from cells expressing $Nup153_{1-144}$-APEX2. Streptavidin IPs were performed on the samples. Additionally to the three Climp63 bait samples, there was one reciprocal sample in which the biotinylation reaction was performed in the $Nup153_{1-144}$ and it was subsequently mixed with the Climp63 lysate (figure 3.9B).

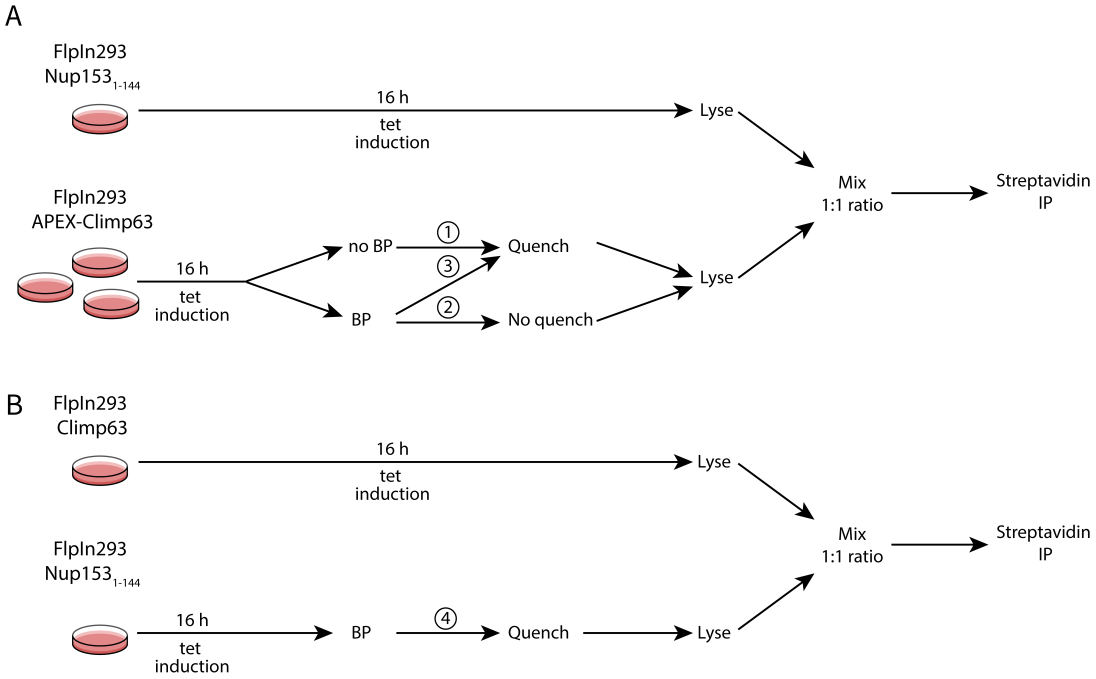


Figure 3.9: Investigating quenching efficiency. a) Expression of APEX2-fusion protein expression was induced using 100 ng/mL tet for 16 h. Two dishes with the Climp63 cells were preincubated with BP for 30 min, whereas in one dish BP was omitted. Biotinylation reactions were performed in the three samples independently. In two of the samples, one with and one without BP addition, the biotinylation reactions were quenched. In sample number 3 no quenching was done. All three Climp63 samples, as well as the $Nup153_{1-144}$ cells were lysed separately. Then each Climp63 sample was mixed at a 1:1 ratio with the $Nup153_{1-144}$ cells. Streptavidin IPs were performed on all three samples. Results were analysed by western blot. b) Sample number 4 was the reciprocal experiment to sample 2, meaning biotinylation reaction and quenching was performed in the $Nup153_{1-144}$ cell line, and the lysate was subsequently mixed with lysate from Climp63 cells.

If quenching of the biotinylation reaction would be complete, no labelling would take place in the mixed cell lysates, resulting in its loss in the unbound fraction of a streptavidin IP. In contrast, if quenching was incomplete, labelling would still take place in the mixed lysates, resulting for the Climp63 samples in labelling of $Nup153_{1-144}$, in which case the $Nup153_{1-144}$-APEX2-FLAG construct would be visible in the elution of the streptavidin IP.

Streptavidin and FLAG-blots were used for analysis (figure 3.10). As expected, no biotinylation was observed in the negative control in which BP was omitted (sample 1), whereas the strongest biotinylation is observed in the sample in which no quenching washes were performed (figure 3.10A, sample 2). In line with

previous results, for the reciprocal sample in which biotinylation was performed in the Nup153$_{1-144}$ (sample 4), biotinylation was observed at a level lower than for the Climp63 (sample 3). Surprisingly, the streptavidin IP of the positive control (sample 2) was not successful and most biotinylated protein was lost in the unbound section, as seen in figure 3.10A. This is most likely as the BP in the positive control was not washed out and therefore the structural similarity of BP to biotin prevented binding of the biotinylated sample to the beads. This means that the non-quenching condition could not be used as a positive control for the IP and post-lysis labeling.

Importantly, the IP for both the Climp63 and the Nup153$_{1-144}$ worked, with little of the biotinylated proteins lost in the unbound fraction. As expected, in the input (IN) fractions of all conditions two bands were visible in the anti-FLAG WB (figure 3.10B). The top band corresponded to the FLAG-APEX2-Climp63 construct, the bottom band to the Nup153$_{1-144}$-APEX2-FLAG protein. As anticipated the Climp63 sample (sample 3) only contained the FLAG-APEX2-Climp63 protein in the elution fraction, but not the Nup153$_{1-144}$-APEX2-FLAG sample, indicating that quenching worked well. In line with these results, the reciprocal Nup153$_{1-144}$ sample (sample 4) showed immunoprecipitation of the Nup153$_{1-144}$ probe, but only very low amounts of FLAG-APEX2-Climp63. Together, these results indicated that quenching works efficiently and that only very minor amounts of labelling took place post lysis.

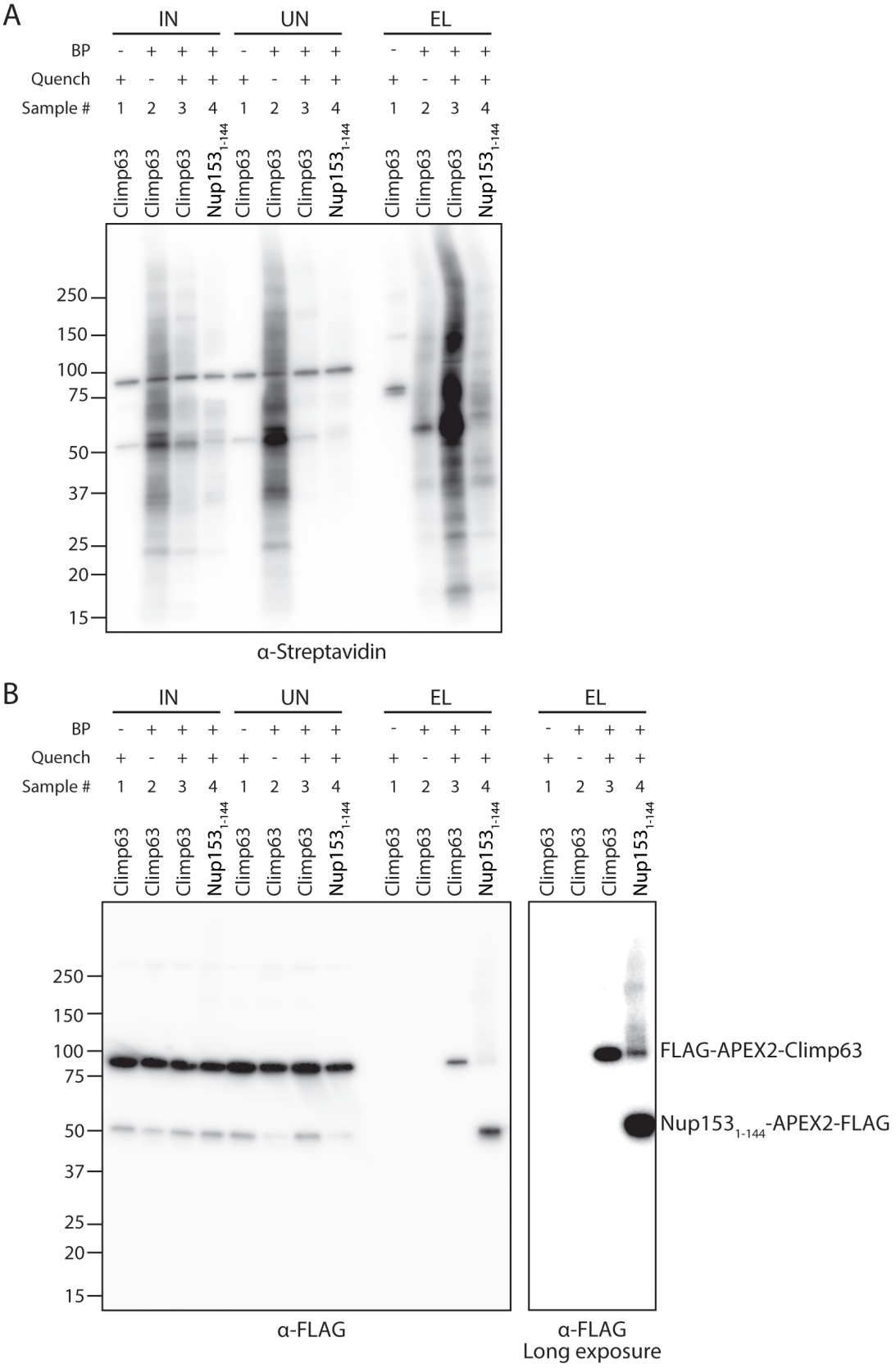


Figure 3.10: Investigating quenching efficiency. a) Streptavidin and b) FLAG WB of input (IN), unbound (UN) or elution fractions of streptavidin IPs of cell lines expressing APEX2-Climp63 and $Nup153_{1-144}$-APEX2. Experiments were performed as described in figure 3.9, to which also the sample numbers refer.

3.3.9 Validating APEX2-fusion protein localisation

Overexpression and tagging of proteins can lead to changes in their localisation. It is therefore crucial for PDB assays to determine whether the bait proteins localise to their normal position within a cell. To address this issue and validate protein localisation, IF microscopy experiments were performed. Thereby anti-FLAG antibody was used to visualise the bait proteins, whereas Alexa Fluor 488/555 streptavidin was used to visualise the biotinylated proteins within the cells. Calnexin and Bap31 were used as ER markers and DAPI staining was used to visualize the nucleus.

For the constructs localising to the peripheral ER (Climp63, Rtn4a and Sec61β), the FLAG signal was only visible after tetracycline induction and showed good correlation with the ER marker (figure 3.11). For Climp63 and Rtn4a, this ER co-localisation was also seen for the biotinylated proteins, though slightly more diffuse. Additionally, some streptavidin signal was seen inside the nucleus for the Climp63 sample (figure 3.11A & B). In line with the Sec61β bait localising both to the bulk ER and the INM, biotinylation correlated with both the ER and the nucleus (figure 3.11C).

Unlike the INM localisation anticipated for the emerin bait, the FLAG and streptavidin staining of emerin correlated with the ER marker (figure 3.12A). Whereas it is known that emerin also localises to the ONM and the peripheral ER[156,157], the majority of the protein is usually seen at the INM. This indicated that either the tagging or the overexpression of emerin stopped its trafficking to, or enrichment at, the INM.

Nup153$_{1-144}$ is a truncation of the Nup153 nucleoporin that is part of the nuclear basket. The N-terminal truncation contains the minimal NE-targeting cassette that binds to the INM through its alpha helix. Unlike the full length protein, it was shown not to enrich at the NPCs, but acted as a generic minimal INM targeting motif[53]. Yet, whereas the FLAG signal of Nup153$_{1-144}$ generally

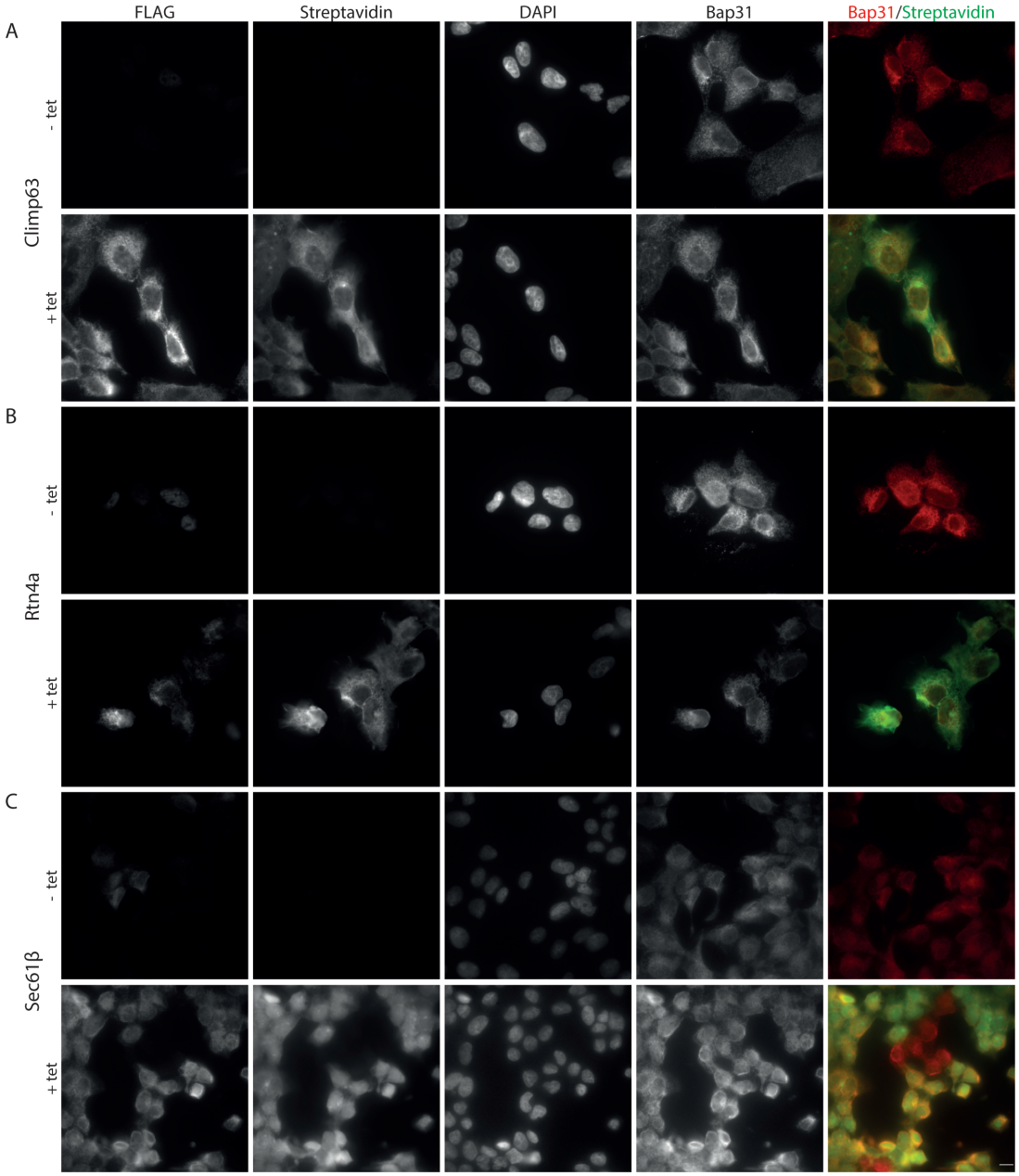


Figure 3.11: Testing the localisation of ER APEX2-bait proteins. IF microscopy images of FlpIn293 cells expressing FLAG-APEX2 fusion proteins. Anti-Bap31 and Anti-FLAG antibodies were used in combination with their corresponding Alexa Fluor secondary antibodies for staining. Alexa Fluor 488 streptavidin was used for visualising biotinylated protein. Nuclei were stained using DAPI. a) Climp63 bait; b) Rtn4a bait; c) Sec61β bait. Scale bar: 10 μm.

correlated with the nuclear rim, in our data some foci in the nucleus showed increased signal. Biotinylated protein was shown to be enriched in the nucleus (figure 3.12B).

For unknown reason, FLAG signal could not be detected for any of the other INM baits upon addition of tetracycline. This is shown for SUN2, representative of all INMs (figure 3.12C). Whereas some signal was detected, this is unspecific background as no changes were observed upon tetracycline induction, which has been shown by WB to successfully induce protein expression (chapter 3.3.6). Various trouble shooting attempts were undertaken to tackle this issue, trialling different FLAG antibodies and optimising staining conditions. As all of them were unsuccessful in resulting in FLAG staining, the streptavidin signal, which is directly dependent on the localisation of the bait, was used as an indirect method to validate protein localisation instead. Intriguingly, all INM baits resulted in biotinylated protein throughout the nucleus rather than just along the nuclear rim, indicating that INM probes modify soluble proteins in the nucleoplasm. Additionally, whereas various baits localised to the INM, subtle differences could be observed between them.

For the FlpIn293-SUN2 cells biotinylation was observed in the nucleus. Yet, a strong heterogeneity existed between cells, with only a minority showing successful PDB even though the imaged cells were derived from a single cell clone (figure 3.12C).

BAF is a 10 kDa, soluble protein that has been shown to localise both to the cytoplasm and as well as the nucleus, where it is enriched at the NE[158–161]. As anticipated, a strong nuclear streptavidin signal was observed in some cells, whereas in other cells the signal was more diffuse throughout the cell (figure 3.13A).

Unlike all other baits, NLS-FLAG-APEX2-GFP could be visualised directly due to its fluorescent tag, and both the GFP signal and the biotinylation signal were present in the nucleus as anticipated (figure 3.13B).

For the Lap2β construct, the biotinylation pattern was localised to the nucleus as expected (figure 3.14A).

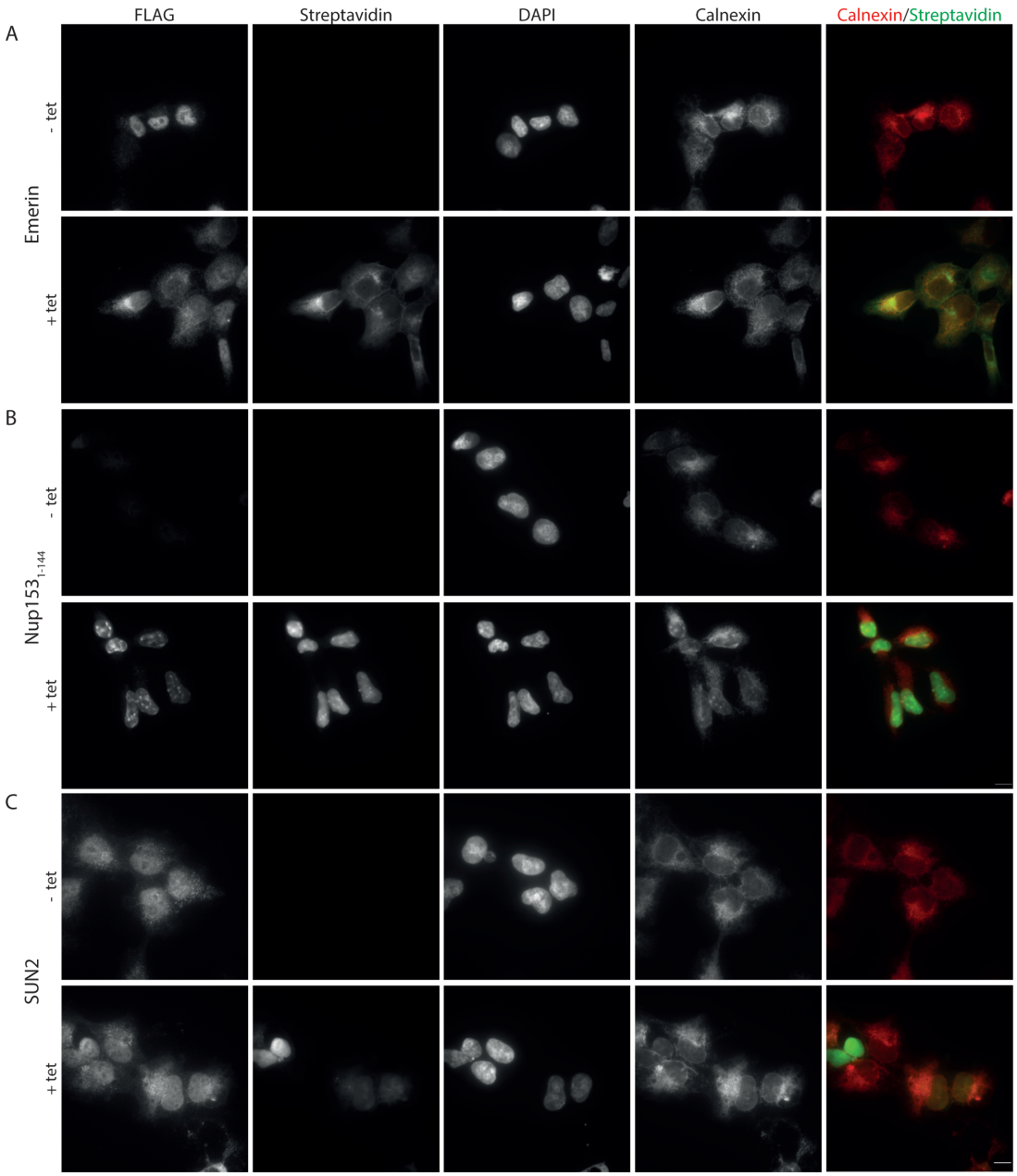


Figure 3.12: Testing the localisation of INM APEX2-bait proteins. IF microscopy images of FLpIn293 cells expressing FLAG-APEX2 fusion proteins. Anti-Calnexin and Anti-FLAG were used in combination with their corresponding Alexa Fluor secondary antibodies for staining. Alexa Fluor 488 streptavidin was used for visualising biotinylated protein. Nuclei were stained using DAPI. a) Emerin bait; b) Nup153_{1-144} bait; c) SUN2 bait. Scale bar: 10 μm.

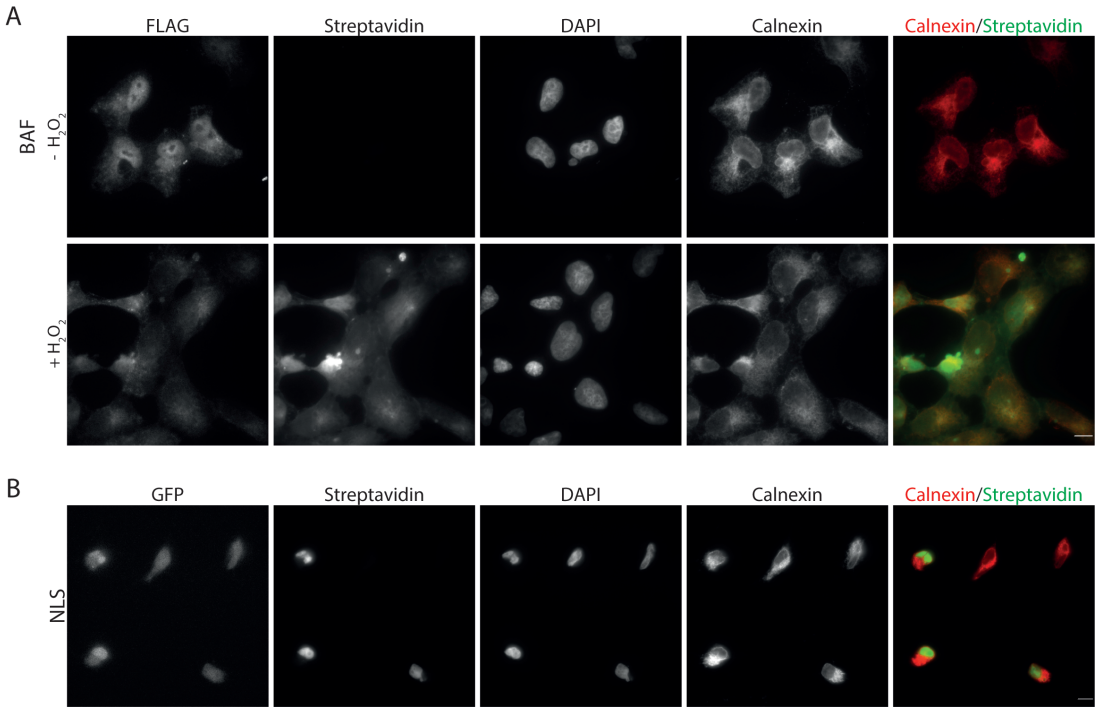


Figure 3.13: Testing the localisation of soluble APEX2-bait proteins. a) IF microscopy images of FlpIn293 cells expressing FLAG-APEX2-BAF fusion proteins. Anti-Calnexin, Anti-FLAG and streptavidin antibodies were used in combination with their corresponding Alexa Fluor secondary antibody for staining. Nuclei were stained using DAPI. Scale bar: 10 μm. b) IF microscopy images of FLpIn293 cells expressing NLS-FLAG-APEX2-GFP. Anti-Calnexin was used in combination with its corresponding Alexa Fluor secondary antibodies for staining. Alexa Fluor 555 streptavidin was used for visualising biotinylated protein. GFP was measured directly. streptavidin-conjugated. a) Lap2β bait; b) Lap2β* bait; c) LBR. Scale bar: 10 μm.

Lap2β* contains a N-terminal truncation compared to the WT protein. It retains the lamin-interacting and transmembrane domain, but lacks the DNA and LEM domain. In overexpressed conditions the protein was shown to localises to the NE without affecting NE formation[141]. While the protein was successfully expressed, biotinylation was not observed in the nucleus as expected, but instead in foci just outside the nucleus, indicating mislocalisation of the bait (figure 3.14B).

For the LBR construct streptavidin labeling was observed both in the nucleus, as well as throughout the ER. Additionally, some heterogeneity was observed, both in the amount of streptavidin signal detected, as well as the localisation seeming more nuclear in some cells and more ER localised in others (figure 3.14C).

Overall, all ER baits behaved as their endogenous proteins, localising to the bulk ER. In contrast, this localisation data demonstrated the complexity of targeting APEX2 to the INM. Firstly, validation of fusion protein localisation using the FLAG-tag failed for all INM proteins, whereas it was successful for generic ER proteins, as well as our soluble probes. Additionally, tagging of proteins or overexpression for some construct prevented them from getting enriched at the INM, as seen for emerin, LBR and the Lap2β*. Other cell lines like SUN2 and BAF showed large amount of heterogeneity in the localisation and/or intensity of biotinylated protein. To move forward, the following five constructs covering the different regions of the ER and the cell were chosen: Climp63 (ER sheets), NLS (nuclear), Nup153$_{1-144}$ (INM), Rtn4a (ER tubules) and Sec61β (generic ER).

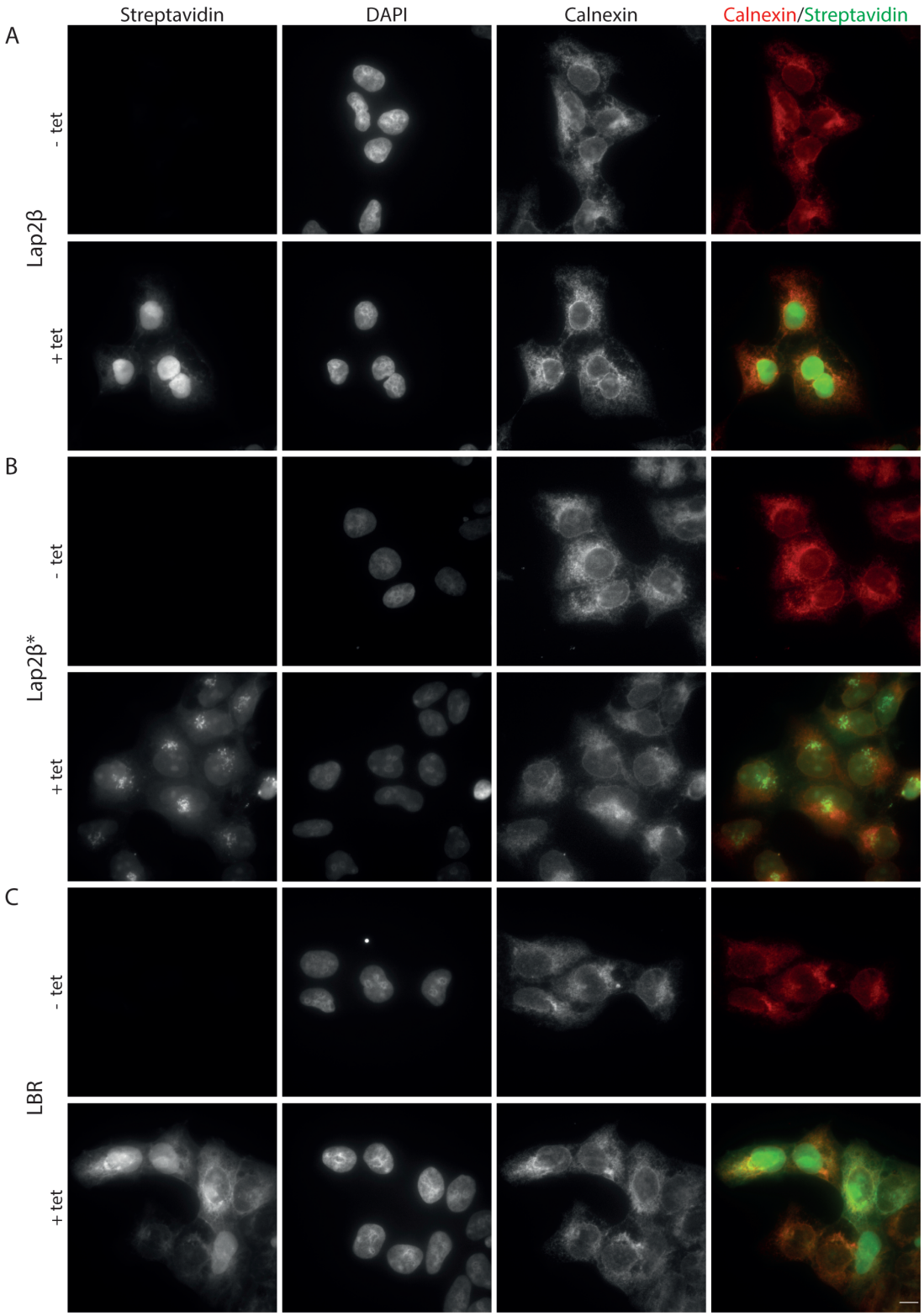


Figure 3.14: Testing the localisation of INM APEX2-bait proteins. IF microscopy images of FlpIn293 cells expressing FLAG-APEX2 fusion proteins. Anti-Calnexin and was used in combination with its corresponding Alexa Fluor secondary antibody for staining. Alexa Fluor 488 streptavidin was used for visualising biotinylated protein. Nuclei were stained using DAPI. Scale bar: 10 μm.

3.3.10 Optimising sample preparation for LC-MS/MS

Having validated expression, activity and localisation of all bait proteins, the next step was to optimise a protocol to prepare samples for proteomic analysis. After the PDB reactions, streptavidin IPs are used to enrich for biotinylated proteins. In the original manuscript[143], biotinylated proteins get subsequently eluted from the beads by boiling the beads in SDS protein loading buffer in the presence of free biotin. Whereas this method successfully worked in the elution of proteins from the beads as shown by previous WBs (e.g. figure 3.10), issues were encountered in the downstream processing of samples by proteomic analysis, as residual SDS and biotin affected the quality of protein detection and led to clogging of columns and spray needles. Therefore, I conducted experiments to optimise a protocol to ensure good protein recovery after IPs, while improving the compatibility with downstream processing by mass spectrometry analysis. For this, various methods were trialled to reduce the SDS content in the peptide mix for proteomic analysis:

1. Reducing the amount of SDS in the elution buffer
2. SDS/biotin elution followed by methanol-chloroform precipitation
3. On-bead digestion to circumvent the need for SDS elution

Various optimisation steps trying to reduce the amount of SDS in the elution buffer were done in the lab prior to the start of this project. For this, varying percentages of SDS in the elution buffer (0.5%, 1% and 3%) in combination with 2 mM biotin and 20 mM DTT were trialled. Elution efficiency was tested by WBs and proteomic analysis. Little difference was observed between the 1% and 3% conditions, yet less proteins and in particular less membrane proteins were detected when the SDS concentration was reduced to 0.5% in the elution buffer. Therefore, an overall reduction from 1% to 3% of SDS was possible without impacting sample quality, however further SDS removal steps were still required prior to proteomic analysis. This was done by processing elutions of samples using filter aided sample preparation (FASP)[162]. While this is a widely

established technique for proteomic sample preparation, some issues persisted with the quality of individual samples during proteomic runs.

Therefore, two other methods, SDS/biotin elution followed by methanol-chloroform precipitation and on-bead digest, were trialled to obtain better quality samples. Proximity-dependent biotinylation was performed in FlpIn293 cells expressing APEX2-Lap2β. Cells were lysed and biotinylated proteins purified by streptavidin pulldown. After the last wash, beads were split in two samples. On half of the samples on-bead digestion was performed. For the other half, proteins were eluted in 3% SDS with 2 mM biotin and 20 mM DTT. Subsequently, methanol-chloroform precipitation was performed for the latter sample, followed by in-solution digestion. After trypsin digestion, both samples were further processed by C18 cleanup, analysed by mass spectrometry and the results compared. Most importantly, no issues were encountered in the mass spectrometry analysis and good quality samples were obtained for both methods.

Proteins were filtered for identification by > 1 significant sequence. When comparing the on-bead versus in-solution data set, nearly identical amounts of proteins were identified in both samples. Thereby the overlap of all proteins that were found in both samples was around 70%, whereas roughly $\sim$15% of proteins were just identified in either of the two samples (figure 3.15A).

Interestingly, independently of the proteomic sample preparation method, even though the APEX2-Lap2β bait is a membrane protein, the majority of the proteins identified were soluble. Indeed, only $\sim$10% of all proteins identified where predicted to be transmembrane proteins by TMHMM2[148,149]. The overlap of these predicted transmembrane domain containing proteins between the two samples again was $\sim$70% (figure 3.15B).

Finally, both data sets were searched against list composed of what we consider well-defined organellar proteins (table A.2). These include mostly membrane proteins of the ER and INM, as well as soluble proteins expected to co-purify with the INM, such are nucleoporins or lamins. Marker proteins for Golgi, plasma membrane and mitochondria were also included. Here, the overlap of the marker

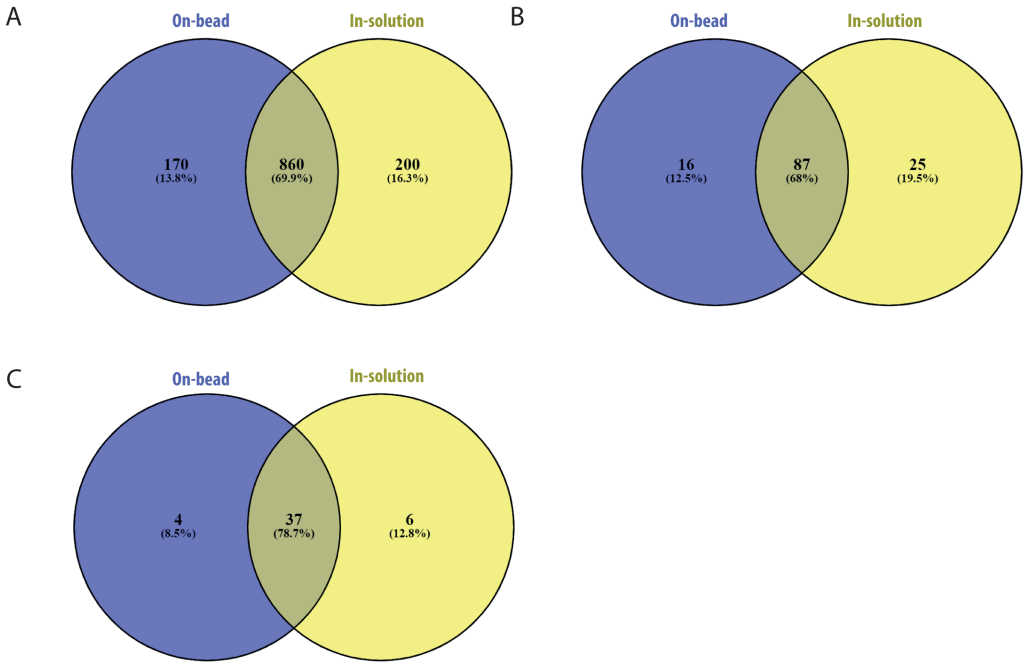


Figure 3.15: On-bead versus in-solution trpysin proteomic sample preparation. Comparison for on-bead versus in-solution trypsin digest. All samples are filtered for > 1 number of significant sequences and Uniprot IDs used to analyse the overlap of identified proteins a) Overlap between all proteins identified. b) Overlap of predicted membrane proteins by TMHMM 2.0[149] only. c) Overlap between proteins identified that are part of the marker protein list (table A.2).

proteins identified in the two data sets was just under 80% (figure 3.15C). Slightly more marker proteins were identified in the in-solution compared to the on-bead data set, yet these proteins were soluble nucleoporins (Nup50 and Nup85), as well as generic ER proteins (Sec13 and TMEM33). In contrast, the proteins only identified in the on-bead data set included the soluble Lap2α, as well as the well characterised INM protein SUN2.

Taken together, whereas there are some differences in the proteins identified by on-bead versus in-solution trypsin digest, overall similar amounts of proteins were identified by both methods and the overlap between the individual proteins identified was around 70%. Most proteins identified only in one sample or the other were identified with low significant sequence numbers. In particular as only one sample rather than several replicates were run per conditions, some variance between samples was expected. Considering the high similarity between samples,

I decided to option for the on-bead digest method for all future experiments, as it eliminates any risk of residual SDS in samples and is much less labour intense.

3.3.11 Enriching for membrane proteins

While optimising a sample preparation protocol for proteomic analysis, it became apparent that even when using a membrane-bound bait, the majority of the proteins identified by PDB (∼90%) were soluble. Underrepresentation of membrane proteins when investigating protein composition of organelles using PDB may be one of the caveats of this technique compared to biochemical fractionation assays. Low identification of membrane proteins might be due incomplete resolubilisation of membranes prior to streptavidin IPs. Additionally, hydrophobic proteins such as membrane proteins may have less lysine and arginine residues. As trypsin cuts peptide chains at the carboxyl side of lysines and arginines, this would give rise to larger peptides, which have unfavourable characteristics for identification by proteomic analysis. Lastly, lipid membranes pose a barrier to the biotin-phenoxyl radicals, resulting in the masking of luminal and transmembrane segments of ER and INM proteins from PDB-dependent labelling.

Yet, PDB also offers certain advantages compared to subcellular fractionation as it might be able to distinguish between

1. Inner nuclear membrane versus outer nuclear membrane;

2. ER tubules versus ER sheets.

We hypothesised that coupling PDB with subcellular fractionation, as visualised in the workflow in figure 3.16, could minimise the disadvantages associated with either of the two techniques used independently.

To investigate this hypothesis, I adapted a well established protocol[52] used for biochemical fractionation to enrich either for the NE or the bulk ER (figure 3.17). To check whether the fractionation worked, samples were taken after each centrifugation step, run on WB and probed for various well-defined bulk

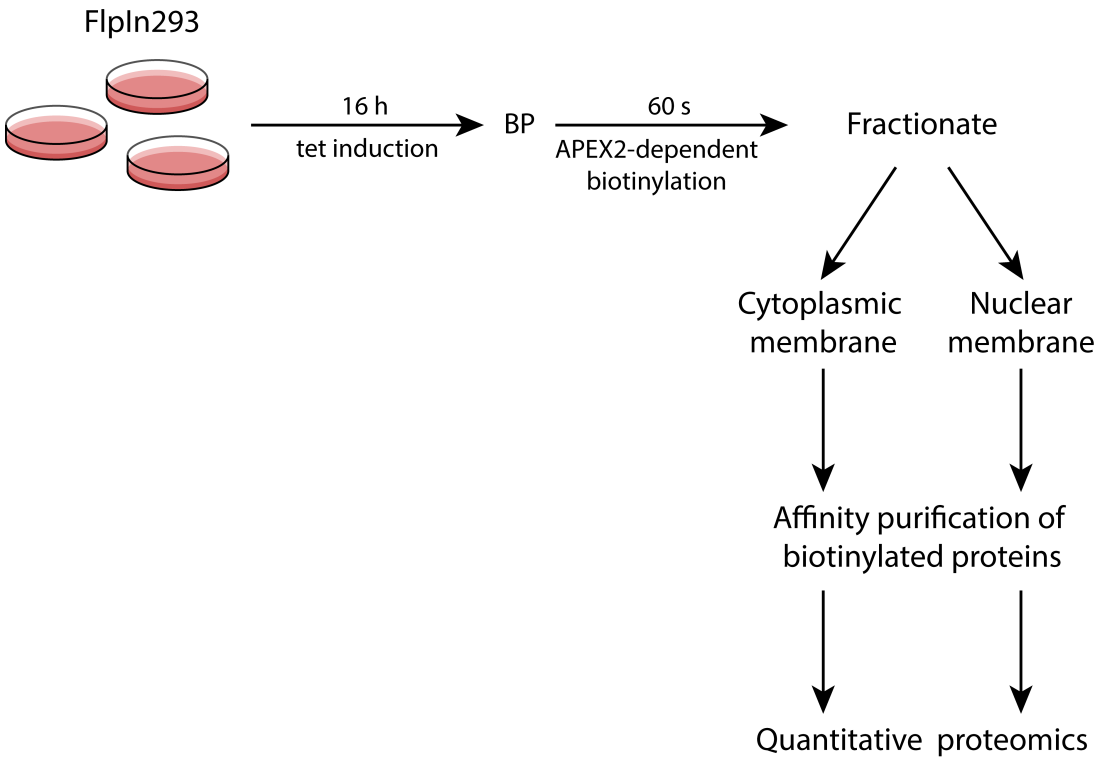


Figure 3.16: Coupling proximity-dependent biotinylation to subcellular fractionation. Expression of APEX2-fusion protein expression was induced in FlpIn293 cells using 100 ng/mL of tetracycline for 16 h. Cells were then incubated in media with 500 μM BP in complete medium for 30 min. Subsequently, H_2O_2 was added at a final concentration of 1 mM. After 60 s, the labeling solution was taken of and cells were washed 3 x in quenching buffer followed by one wash in HB buffer. Cells were subsequently incubated in HB+ buffer on ice for 30 min, scraped down and lysed in a douncer. Subcellular fractionation is then used to enrich independently for cytoplasmic membranes and nuclear membranes. Membranes got resolubilised and streptavidin IPs performed. Elutions of the the IPs were then used for western blotting and/or quantitative proteomics.

ER, INM and nuclear proteins (3.18). The post-nuclear supernatant (PNS) sample is an intermediate step when enriching for cytoplasmic membranes (CMs), whereas the low-speed pellet (LSP) and nuclei samples are intermediate samples collected during the process of enriching for the nuclear envelope (NE) and nuclear content (NC).

As expected, proteins known to be enriched in the bulk ER, such as calnexin and kinectin, were enriched in the CM fraction, whereas the well defined INM proteins SUN2 and LBR were enriched in the NE fraction. Yet, low levels of ER marker proteins could still be detected in the NE fraction and vice versa.

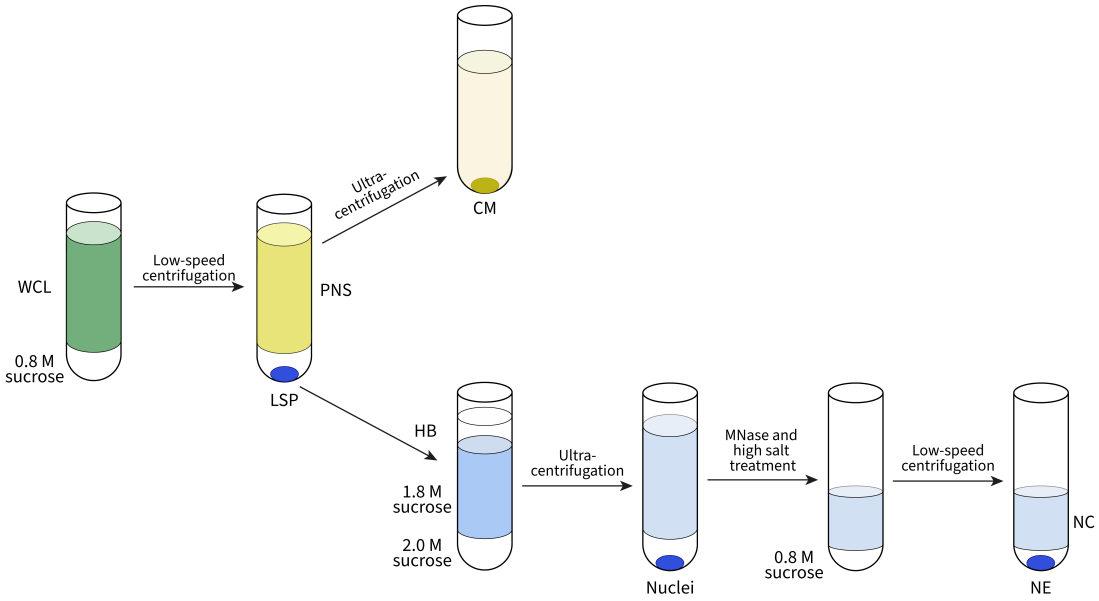


Figure 3.17: Subcellular fractionation to enrich for ER and NE. WCL lysate was layered on top of 1 mL of of 0.8 M sucrose in HB and centrifuged at 700 × g, 10 min, 4 °C. The PNS was subsequently centrifuged at 100,000 g, 1 h, 4 °C to obtain a pellet enriched for CM. The LSP was resuspended in 6 mL 1.8 M sucrose in HB buffer and layered on top of 1 mL 2.0 M sucrose and centrifuged at 210,000 × g, 1 h, 4 °C and stopped without break. The nuclei were subsequently resuspended in HB buffer, treated with MNase and high salt and centrifuged once more on top of a 1 mL cushion of 0.8 M sucrose to obtain a soluble fraction containing the nuclear contents and a pellet containing the NE. NE and CM fractions were then re-solubilised and used for analysis by western blotting or quantitative proteomics.

This is expected, as subcellular fractionation is a method to enrich for specific fractions, but does not result in full separation of organelles. Additionally, some INM proteins will always be detected in the bulk ER, as this is where they get synthesised before they shuttle to the INM. Emerin, another well characterised INM protein, was also found both in the CM and the NE fraction. Yet, compared to LBR and SUN2, it seemed to be present at similar levels in the two fractions. Having lower enrichment in the INM compared to the bulk ER for emerin compared to SUN2 and LBR is in line with previous findings[157]. Lastly, as expected, the soluble protein histone H3, is mostly present in the NC and GAPDH is found in the PNS fraction and co-purifies with the CMs.

These results indicated that the protocol works well in separating soluble proteins and membrane proteins, while enriching either for the bulk ER or the NE.

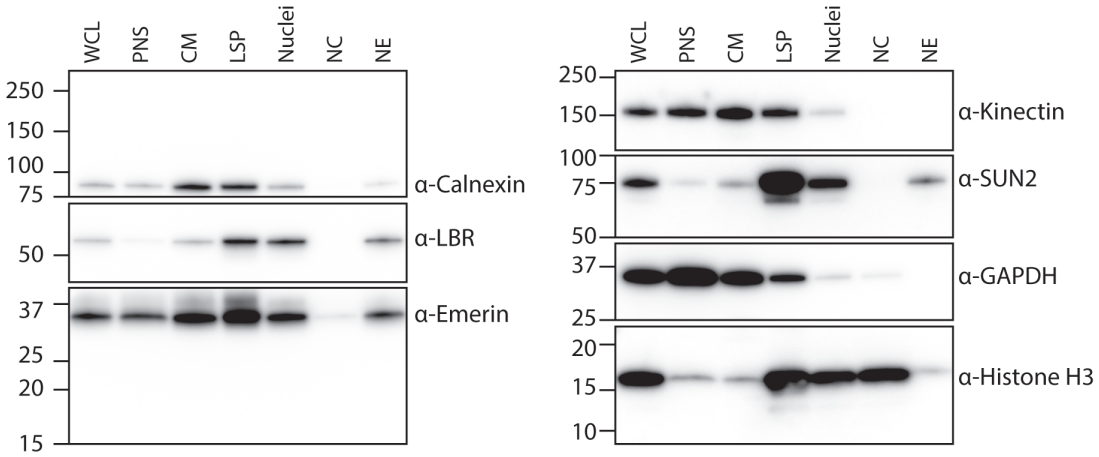


Figure 3.18: Cell fractionation enriches for known INM proteins. WBs of various fractions obtained during subcellular fractionation blotted for proteins with well-characterised subcellular localisation. ER Markers: calnexin, kinectin. INM markers: LBR, emerin, SUN2. Nuclear markers: Histone H3. Generic soluble marker: GAPDH. WCL: whole cell lysate, PNS: post-nuclear supernatant, CM: cytoplasmic membranes, LSP: low-speed pellet, NC: nuclear content, NE: nuclear envelope.

Yet, it also highlights that some cross-contamination exists between fractions.

Having both a PDB, as well as a subcellular fractionation protocol in place, the next step was the combination of the two (figure 3.16). For this, three different APEX2-fusion cell lines were used: FlpIn293-NLS, Nup153$_{1-144}$ and Climp63. As a negative control the NLS-APEX2 cell line was used, but no BP was added.

After APEX2-dependent biotinylation and subcellular fractionation, cell membranes were resolubilised and WBs run to test whether this leads to enrichment of biotinylation in the expected fractions (figure 3.19). FLAG-signals were detected in all WCL fractions at the expected molecular weight (figure 3.19A), indicating successful bait protein expression. Biotinylation could be detected in all samples apart from the negative control which had omitted the BP (figure 3.19B). To test whether the established subcellular fractionation was still possible after APEX2-mediated PDB, fractions were blotted with marker proteins for the bulk ER (kinectin), the INM (SUN2), a generic soluble protein (GAPDH) as well as a nuclear marker (histone H3) (figure 3.19C). All proteins were enriched in the anticipiated fractions, though there was some variance of protein levels detected

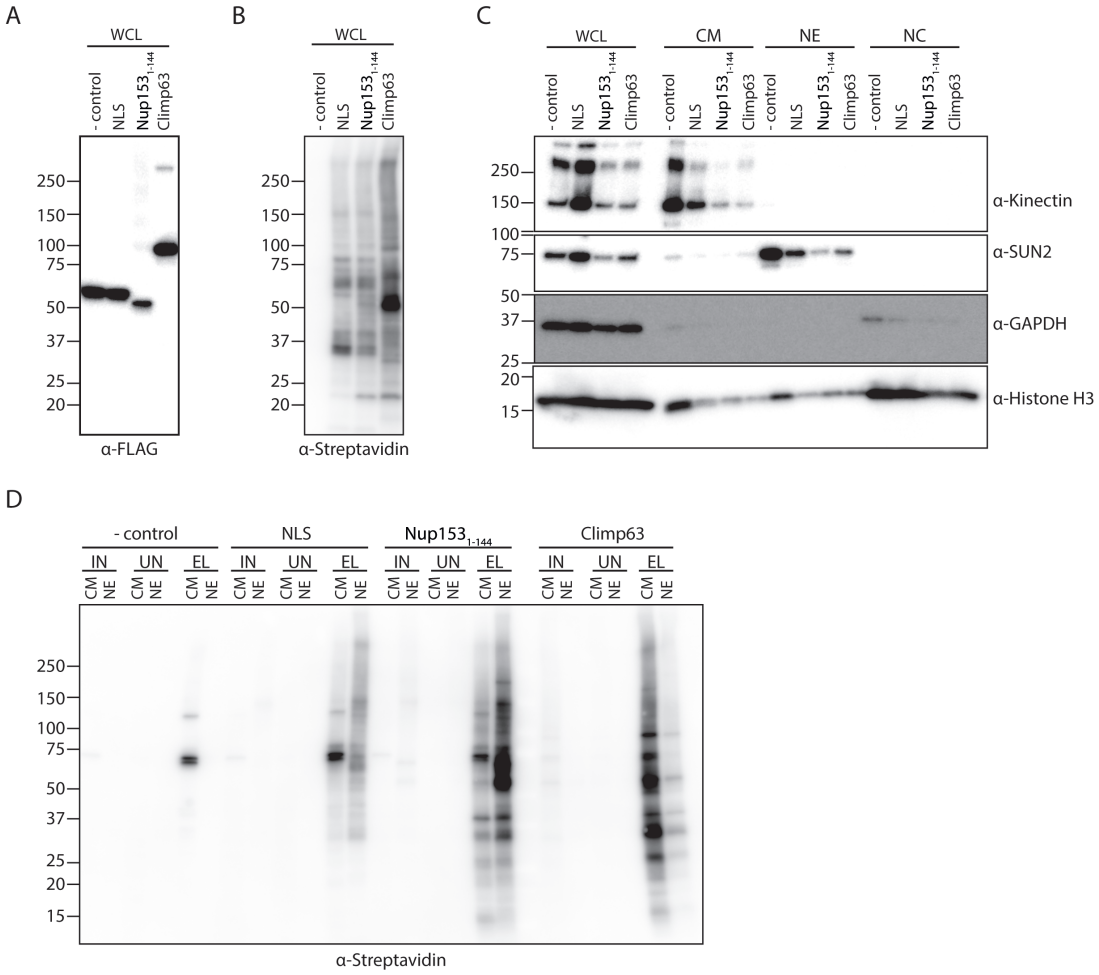


Figure 3.19: Coupling proximity-biotinylation to cellular fractionation. a & b) Anti-FLAG and streptavidin WBs for WCL originating from FlpIn293 cells encoding APEX2-fusion proteins, subjected to tetracycline induction. c) WBs using sample either from the WCL, or from the CM, NE and NC fractions obtained during subcellular localisation. Blotted for well-characterised proteins to test for successful fractionation. WCL: 0.21 % of total sample, CM/NE/NC: 1.7 % of total fraction. d) Streptavidin WB of a streptavidin pulldown from the CM and NE fraction for different probes. IN: input (0.42 %), UN: unbound (0.42 %), EL: elution (6.6 %).

for each of the four probes. This data indicated that biotinylation reactions do not interfere with a subsequent fractionation.

Next, streptavidin IPs were performed on the resolublised CM and NE fractions. Proteins were subsequently eluted from the beads and WBs were run on the input, unbound and elution fraction to test if the biotinylated proteins would enrich in the CM or NE fraction, depending on the bait used (figure 3.19D). Indeed, while biotinylation was observed in both fractions for all three probes, for

the NLS and $Nup153_{1-144}$ cell lines most of the streptavidin signal was detected in the NE fraction, whereas for the FlpIn293-Climp63 cells the majority of the biotinylation was detected in the CM fraction. These results indicate that coupling the APEX2-PDB assay to a subcellular fractionation step worked convincingly.

3.4 Discussion

3.4.1 Cloning strategy and cell line generation

Understanding the protein composition of an organelle, or distinct domains within an organelle, provides the basis for all functional and mechanistic studies. Due to the continuity of the membrane-system, studying the protein compositions of different ER domains has been extremely challenging to date. Yet, exciting advances in proteomic analysis and new PDB-assays may offer a solution to this, as they circumvent the need for separating domains of interest by fractionation. Therefore, the main objective of this project was to develop a PDB-assay to investigate the protein composition of distinct ER domains, namely ER tubules, sheets, as well as the INM.

The Flp-In system allows stable integration and expression of a gene of interest in a single locus by homologous recombination. This minimizes off-target effects and ensures isogenic cell lines. Additionally, the tet-inducible promoter enables temporal control of gene expression, thereby minimizing disruption to cells. Various FLAG-APEX2 fusion proteins were cloned. Well-characterised bait proteins, known to localise to the INM, ER sheets or the ER tubules were chosen. Additionally, a NLS-APEX2-GFP construct was already available in the lab and was used as part of this project. A FLAG-tag was integrated in all constructs to enable simple identification of protein expression and localisation by WB and IF, respectively. Of note, whereas the bait proteins are known to enrich in certain ER domains, this localisation is never absolute. Indeed, all INM proteins are synthesised in the ER. Additionally Climp63, which is a traditional ER sheet marker protein, has recently been identified to reside in low amounts in ER matrices, that form under certain condition from ER tubules close to the cell periphery[9]. In contrast, the ER tubule protein Rtn4, was recently identified to stabilise ER nanoholes in otherwise flat ER sheets[25].

The FlpIn system can be used in combination with various cell lines. Both HeLa and HEK293 cell lines were trialled for this project. The results of these experiments indicated that, for unknown reasons, the enzymatic activity of APEX2 was much lower in HeLa compared to HEK293 cells (chapter 3.3.2). As good enzymatic activity of the peroxidase is the core characteristic required for PDB-based studies, stable cell lines were created for all bait constructs using the FlpIn293 cell line.

Interestingly, whereas all proteins were expressed from the same promoter, strong variability in protein expression levels of the different constructs was seen by WB (figure 3.7). Thereby, a positive correlation between bait protein expression levels and the amount of biotinylated proteins could be observed in the established cell lines. This difference of protein level was not due to proteasomal degradation, as bait proteins were not stabilised by the addition of the p97 inhibitor CB (chapter 3.3.5). Protein expression levels could also not be matched to each other using different concentrations of tetracycline, as protein expression from the tet-inducible promoter followed a strict on/off pattern (chapter 3.3.4). As proteomic analysis is a semi-quantitative technique that relies on the comparison of two or more biological samples, and as protein expression levels were linked to biotinylation levels observed, this poses the risk of proteins being skewed towards the strongly expressed baits. As no solution could be found to align protein expression between baits experimentally, this will require normalisation in the downstream bioinformatic analysis of the proteomic data.

For PDB assays, it is absolutely crucial that the enzyme responsible for biotinylation is localised correctly. A FLAG-tag was part of each fusion protein. IF experiments were performed to validate the localisation of the bait proteins, as well as to investigate the cellular localisation of the labelled proteins. Of note is that the elaborate structure of the ER makes it prone to artefacts and subtle changes in structure upon fixation. To circumvent this risk, morphological studies of the ER are often performed in live cells. As IF experiments performed in this

chapter were used solely to validate the localisation of the bait, whereas the data collection to investigate the protein composition of distinct ER domains will be carried out in live cells using the PDB assay followed by proteomics, small changes in ER morphology upon fixation are neglectable. Additionally fixing cells stops the diffusion of the biotinylated proteins and enables the visualisation the original localisation at which the biotinylation reaction took place. Unfortunately, whereas the FLAG-tag gave a very clear signal by WB, it was not possible to visualize any of the INM constructs (apart from the soluble $Nup153_{1-144}$) by IF using the FLAG-tag. Several trouble-shooting approaches were undertaken to solve this issue, including optimisation steps such a testing different FLAG antibodies and antibody concentrations, as well as trialling different cell permeabilisation protocols. Yet, none of these attempts resulted in a visible FLAG signal for any of the constructs by IF. In contrast to this, visualisation of all bulk ER constructs (Climp63, Rtn4a and Sec61β) worked well. One possible explanations for this could be the inaccessibility of FLAG-tag at the INM due to the nuclear lamina network that is lining the INM.

As none of the optimisations led to an improvement of visualizing the construct, an indirect approach of using the biotinylation pattern observed by IF was used as a method to assess localisation of the fusion proteins. For Lap2β, BAF, LBR and SUN2 strong biotinylation could be seen within the nucleus. Surprisingly, the labeling was spread through the whole nucleus, rather than giving a nuclear rim staining. In these experiments, PDB labelling was performed first and cells were fixed subsequently. Changing the order, so fixing cells prior to biotinylation, prevented labelling of the whole nucleus and instead resulted in a clear nuclear rim staining (data not shown). This indicated that labelling of soluble nuclear proteins in proximity of the INM and their subsequent rapid diffusion results in the labeling of the whole nucleus prior to fixation. This data was reflected later in the proteins identified by LC-MS/MS, which showed that 90% of all labelled proteins were soluble. Additionally, these findings are in line with published APEX2 data using a lamin B1 bait, which also showed that diffusion of proteins post labelling,

but prior to fixation is responsible for the diffuse nuclear signal observed[163]. Yet, it is surprising that the diffusion of soluble proteins prior to fixation was observed so strongly in the nucleus, whereas for the ER probes the biotinylation signal showed good co-localisation with an ER marker rather than being spread through the whole cytoplasm. Another surprising observation was that even though single cell clones were selected for all cell lines, some cell lines like SUN2 and LBR showed strong heterogeneity in their biotinylation signal. Another finding was, that whereas for most APEX2-bait constructs the biotinylated proteins were observed in the anticipated localisation, for some baits such as Lap2β* and emerin, biotinylated proteins were detected in unexpected localisations. This indicated that the bait proteins might mislocalise. One possible reason for this could be that the FLAG-APEX2 tag interferes with translocation past the NPC to reach the INM. Using a smaller tag might circumvent this issue and there are different options available to reduce the size of the tag. One possibility would be to use rapamycin-dependent targeting of APEX2. The FRB-FKBP system is a rapamycin-inducible dimerization system[164]. Indeed, the FRB-FKBP system has been used to determine binding partners of VAPB specifically at the INM using APEX2[54]. Interestingly, this approach not only allows the reduction of the size of the tag, but may be a good solution to study proteins that are mostly in the ER, but are simultaneously present in the INM, albeit at a lower abundance. Using these proteins with a NLS-FKBP12-APEX2 could allow focusing on the small subpool of protein in the nucleus, while neglecting the bulk ER based pool. An alternative approach to reduce the size of the cytosolic domain of the fusion protein would be to work with a split APEX2[165]. The split APEX2 system consists of two parts: a 200 amino-acid N-terminal fragment and a 50 amino-acid C-terminal fragment. The two parts, named AP and EX, are inactive on their own, but are functional once they are in close proximity and complement each other. Lastly, combing the split APEX system with the FRB-FKBP system could even make this an inducible PDB-system.

Of note, both of these techniques will require extensive optimisation, as single pilot experiments trialling either of the two systems at the INM proved unsuccessful in my hands.

To conclude, a range of APEX2-bait proteins were designed to target to different ER domains. Cell lines were generated and subjected to stringent quality control experiments, where they were tested for cell growth and morphology, as well as bait protein expression, activity and localisation. Following these quality control experiments, the following five 293FlpIn bait cell lines, targeting all different ER domains, were chosen for proteomic analysis to investigate the protein composition of the ER domains.

- Climp63 (ER sheets)
- NLS (Nuclear)
- $Nup153_{1-144}$ (INM)
- Rtn4a (ER tubules)
- Sec61β (generic ER)

3.4.2 LC-MS/MS sample preparation optimisation

Biotin-streptavidin interactions have a very high binding affinity ($K_d \sim 10^{-14}$ M). This allows easy purification of biotinylated substrates using streptavidin IPs and enables the use of stringent washes, including detergents, high salt or denaturing reagents, without disrupting the binding. Yet, in turn this also means that harsh conditions are required for eluting samples from the streptavidin-beads. Generally, this is done by boiling the samples in presence of 3% SDS and free biotin. Whereas these are the standard conditions proposed for sample preparation[143], biotin and detergents can lead to problems in downstream processing of samples by proteomic analysis. Indeed, in our case these elution conditions resulted in inconsistent sample quality and in some cases clogging of the mass spectrometers.

To circumvent this problem, various optimisation steps in the samples processing for proteomic analysis were performed. Firstly, SDS concentrations in the elution buffer were lowered from originally 3% to 1% without impacting the proteins identified by mass spectrometry. Yet, problems in downstream processing persisted even if samples were eluted in 1% SDS and were further processed by FASP[162].

Therefore two other options were trialled. Firstly, eluting the samples in 3% SDS, followed by methanol-chloroform precipitation and subsequent in-solution trpysin digestion. In parallel, an on-bead digest was performed, thereby circumventing the requirement for eluting samples from the beads completely. Both samples were then further processed using the same C18 purification. Proteomic data sets of both methods were compared to each other. Analysis of the overall number of proteins, the number of TMD-containing proteins and the number of marker proteins, demonstrated relatively little difference between the two samples (chapter 3.3.10). Most importantly, when overlaying the not just the number of proteins identified, but the individual proteins, the overlap was still around 70%. The 30% of proteins only found in one of the samples were thereby often proteins that were identified with low numbers of significant sequences. As this experiment was performed without replicates, the results might be partly explained by normal variation between the samples. We hypothesis that running the samples in replicates would result in even higher similarities between the samples. Due to these strong overlap between the two techniques, and the much easier sample preparation by on-bead digest, all further sample preparations for proteomic analysis were carried out using the on-bead digest protocol.

3.4.3 Coupling proximity-dependent biotinylation to fractionation

So far, the majority of APEX2-based PDB studies were used to investigate membrane-bound organelles, yet with no particular focus on identifying membrane

proteins. In contrast, the focus of this project was to identify TM or membrane-associated proteins, and in particular to understand whether different regions of the ER, even though part of a continuous membrane system, have distinct protein compositions.

Initial proteomic data sets used to identify optimal conditions for proteomic sample preparation, showed that only ∼10% of the identified proteins of the PDB assay were predicted membrane proteins, even when using a membrane bound bait (chapter 3.3.10). This is in line with data from other data sets using PDB to investigate the INM[55]. Membrane proteins might be underrepresented due to a variety of reasons. First, protein parts embedded in the membrane, as well as in the perinuclear space and ER lumen, are inaccessible to PDB labelling. Secondly, membranes may not get fully resolubilies prior to streptavidin IPs. Thirdly, the hydrophobic nature of membrane proteins often makes them sparse in arginine and lysine residues. As these two amino acids are required for trypsin digest, this could result in larger peptides and make these peptides less favourable for identification by LC-MS/MS. Together, these reasons pose the risk that membrane proteins, even if in close proximity of the bait, will not be identified. One possible solution to tackle this problem is to use subcellular fractionation to enrich for membrane proteins, while simultaneously reducing sample complexity. Indeed, prior to PDB assays subcellular fractionation assays coupled to shotgun proteomics was a commonly applied method to study the proteome of organelles. We hypothesised that coupling APEX2-dependent PDB to fractionations can minimise the disadvantages associated with each of these individual techniques. This would result in an assay that could differentiate between ER domains that can not be separated by fractionation, such as the INM, ONM, ER tubules and ER sheets, while still enriching for membrane proteins compared to standardised PDB-assays.

Focusing on the technical aspects, some minor adjustments had to be made to the PDB protocol, as subcellular fractionation required lysis in HB buffer, followed by cell lysis through douncing, instead of lysis in RIPA+ buffer as described in

the original protocol[143]. The composition of the wash buffers of the subsequent streptavidin IPs was changed accordingly from RIPA-based washes to sample binding buffer (SBB)-based washes. Quality control experiments were performed to test that PDB reactions and consecutive quenching steps did not interfere with subcellular fractionation. Indeed, WB analysis showed that PDB did not have an impact on subsequent fractions, as assessed by enrichment of various marker proteins in the respective fractions (chapter 3.3.11). Combined, these results are a strong indication that coupling PDB to subcellular fractionation is a working methodology to enrich for biotinylated membrane proteins.

3.4.4 Concluding remarks

To conclude this chapter, FlpIn293 cell lines expressing FLAG-APEX2 fusion proteins targeting to distinct compartments, such as ER tubules, sheets and the INM, were generated. Expression and activity of FLAG-APEX2 bait proteins were validated by α-FLAG and streptavidin WB, respectively. Localisation of bait proteins was validated directly using FLAG staining. In cases were this was not successful, bait protein localisation was determined indirectly using the localisation of the biotinylated proteins in the cell, observed by streptavidin. Protocols were optimised for proteomic sample preparation using on-bead trypsin digest. Finally, a new method was developed in which PDB is coupled to subcellular fractionation of the NE or CM to enrich for membrane proteins. The next chapter discusses the analysis of the proteomic data sets, as well as the attempt to validate some of the *bona fide* proteins identified by LC-MS/MS.

Lernen ist Erfahrung. Alles andere ist einfach nur Informationen.

Learning is experience. Everything else is just information.

— Albert Einstein

4

Determining the protein composition of ER domains

Contents

4.1 Rational 90
4.2 Results 92
4.2.1 A proteomic data set to study ER protein domains . . 92
4.2.2 Optimising data analysis for PDB-fraction LC-MS/MS data 93
4.2.3 Investigating the INM proteome using enrichment analysis 104
4.2.4 Designing a tool to identify putative INM proteins . . 105
4.2.5 Investigating protein compositions of ER sheets and tubules 110
4.2.6 An overview of methods to validate INM protein localisation 114
4.2.7 Implementing a differential permeabilisation assay . . 116
4.2.8 Validating putative INM proteins 121
4.2.9 Additional interesting putative INM proteins 127
4.2.10 Validating ER tubule proteins 134
4.3 Discussion 139
4.3.1 Optimising a workflow to investigate the protein composition of distinct ER domains 139
4.3.2 Deciphering the proteomic composition of the ER . . 142
4.3.3 Proteomic characterisation of the INM 143
4.3.4 Validating putative INM, ER sheet and ER tubule proteins 148
4.3.5 Interesting observations to follow up 149
4.3.6 Concluding remarks 151

4.1 Rational

PROXIMITY-DEPENDENT biotinylation assays combined with mass spectrometry are a powerful tool to investigate protein complexes and map organelles. Thereby, protein ligases or peroxidases are targeted to their desired location in a cell through creation of fusion proteins. Upon protein expression and addition of enzyme specific substrates, the enzymes covalently label proteins in close proximity with a biotin residue. Labelled proteins can then be purified by immunoprecipitation and subsequently be analysed by mass spectrometry. Mass spectrometry is an extremely sensitive and accurate technique to identify proteins. For the analysis, proteins first get digested into peptides using trypsin. Subsequently, peptides get ionized and their intensity is measured, producing MS1 data. The detected precursor ions get fragmented and those fragments are detected in a second mass analyser to obtain the MS2 data.

Whereas PDB coupled to mass spectrometry is an incredibly powerful tool, a single bait protein can lead to the identification of thousands of proteins. The majority of these proteins are thereby background and include endogenously biotinylated proteins, abundant proteins that are detected independent of the bait and proteins that nonspecifically bind during the immunoprecipitation experiment.

To distinguish between proteins truly in the vicinity of the bait protein and false positives, PDB assays rely on stringent controls. Controls can include samples in which no biotinylation reactions were performed to identify unspecific binders and endogenously biotinylated proteins. Yet, more commonly conditions are used in which the ligase or peroxidase gets promiscuously expressed or targeted to a different localisation in the cell than that of interest. Subsequently, an enrichment of the bait against the control experiments is determined. This is commonly done using pairwise-enrichment analysis or affinity-purification mass spectrometry tools such as SAINT[166,167] or ComPASS[168]. Thereby, both the controls and the

optimal data analysis method need to be optimised for the type of enzyme chosen, as well as the organelle or protein complex studied and the quality of the data set.

To this aim, this chapter will focus on the optimisation of a data analysis method to investigate the protein composition of different ER domains. Subsequently, using the optimised method, proteins enriched in the INM, ER tubules and ER sheets will be identified. Finally, several of the *bona fide* targets of specific ER domains will be discussed and further validated.

4.2 Results

4.2.1 A proteomic data set to study ER protein domains

Proteomic data was collected for five baits: Climp63, NLS, Nup153$_{1-144}$, Rtn4a and Sec61β. These probes were chosen to cover all parts of the ER, as shown in figure 4.1. Additionally, a negative control was included, which consisted of the NLS-APEX2 cell line, but did not have BP added to the reaction, resulting in no PDB. Each lysate was then fractionated to enrich for the NE or the CM. To have sufficient statistical power for robust data analysis, each bait data set was obtained at least in triplicate. The exception is the negative control, for which only one replicate was collected. An overview over the replicates of each bait can be found in appendix table A.1.

After filtering out contaminants, reverse sequences and proteins only identified by one side, the overall data set contained 5126 unique identified proteins. The amount of identified proteins per individual sample thereby ranged from as little as 215 to as high as 2800. Transmembrane domains were predicted using the TMHMM2.0 software[148,149] and of the 5126 identified proteins 20% were membrane proteins. This is roughly twice the amount seen in other PDB studies of the INM, as well as in the data set obtained in chapter 3.3.10, in which PDB was not coupled to fractionation[55].

To test the robustness of the assay, scatter plots were generated between all samples and the Pearson correlations determined. Good reproducibility was observed between replicates in the fractions in which labelling was expected, with Pearson correlations ranging from 0.78 to 0.94. In comparison, Pearson correlations between replicates of fractions in which no labelling was expected was overall lower, dropping to as low as 0.51 for probes such as NLS CM.

Overall, the number of identified proteins, the high percentage of membrane proteins and the high Pearson correlations observed between replicates indicated that the data set collected was of good quality and that the assay was robust.

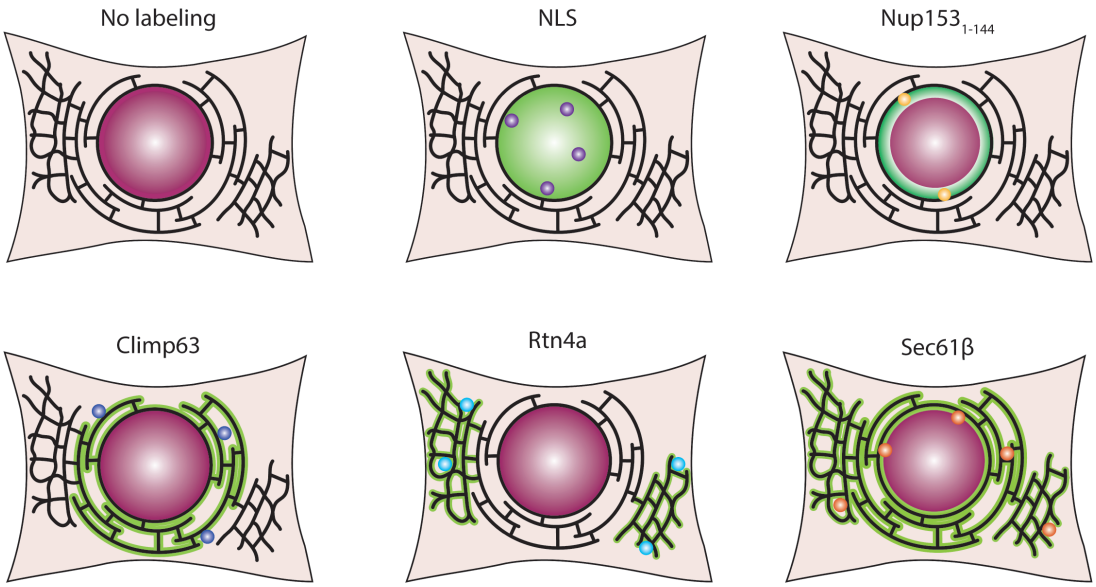


Figure 4.1: Bait proteins target to all domains of the ER. Five different APEX2-bait proteins (small circles) were chosen for PDB analysis to cover all known ER domains, including the nucleus, INM, ER sheets and ER tubules. The negative control consisted of a NLS cell line in which no BP was added. Green indicates the expected localisation of biotinylated protein for each corresponding bait.

4.2.2 Optimising data analysis for PDB-fraction LC-MS/MS data

The next step was to develop a data analysis pipeline for the large data set obtained. Generally, analysis for PDB assays is done by calculating an enrichment score of every protein identified by a bait protein against a background control. As this project is focused on determining whether proteins are enriched in a specific ER domain, the enrichment score of a bait is calculated against another domain of the ER, to determine, for example, if a protein is enriched in the INM compared to the bulk ER. For optimisation of the pairwise-comparison data analysis conditions, the $Nup153_{1-144}$ NE versus Climp63 CM sample was chosen, as it was expected to separate the INM from the bulk ER domain. This can be easily validated by Gene Ontology (GO) enrichment analysis, as well as using a list of well-characterised proteins known to localise to the INM or to the bulk ER. For this, a customized table with marker proteins was created, which can be found in appendix table A.2. Of note, for acquisition of the complete proteomics data set, different facilities had to be used and samples were run in

separate runs, leading to batch effects and variance between replicates. This is also the case for the two chosen samples used for optimisation, therefore allowing us to find the most suitable ways for normalisation.

For the data analysis, three different methods were trialled:

1. Perseus analysis by student's t-test using raw intensities
2. Perseus analysis by student's t-test using label-free quantification (LFQ) intensities
3. R analysis by linear models for microarray (limma) using raw intensities

LFQ intensities are based on an intensity determination and normalization procedure for label-free proteomic data in MaxQuant[146,169]. The principle relies on a delayed normalization and LFQ values are obtained directly from the MaxQuant searches[150]. LFQ intensities are routinely used in proteomic analysis.

Having obtained both the raw intensity values and the LFQ values from the MaxQuant search, data analysis methods (1) and (2) were pursued in Perseus[146,169]. Thereby, the same analysis was done in parallel using (1) raw intensities and (2) LFQ intensities. Briefly, data was filtered for proteins that were detected in all three replicates (3/3 valid values) in either one or both of $Nup153_{1-144}$ NE and Climp63 CM. Subsequently, missing values were imputed from a normal distribution (width: 0.3, down shift: 1.8).

For the R analysis (3), raw intensities were used and proteins were filtered for 3/3 valid values in either one or both of $Nup153_{1-144}$ NE and Climp63 CM. Subsequently variance stabilisation normalisation (vsn) was applied as a way to normalise data and obtain a constant signal-to-noise ratio[151]. Lastly, missing values were imputed by left-censored missing data (LCMD) imputation using the standard settings of the 'imputeLCMD' package[152].

After running the three analysis, the output of each the data analysis method was assessed based on the following criteria:

1. How closely do replicates cluster together, as determined by principal component analysis (PCA)

2. Do well-characterised marker proteins for the INM or the bulk ER show enrichment in the expected fraction

3. Using Gene Ontology cellular component (GO CC) enrichment analysis, do expected classification enrich in each sample

To investigate how well each data analysis method is in clustering the three replicates together, PCA analysis was performed for the $Nup153_{1-144}$ NE and Climp63 CM samples (figure 4.2). All three data analysis methods resulted in the three replicates of the same probe clustering together. Yet, the Perseus LFQ intensity analysis and the R-based analysis showed stronger separation between baits, while simultaneously showing higher clustering of replicates of the same bait. This data indicated that the two latter methods are better suited in data normalisation and imputation for our samples.

The next step was to assess how well each analysis can place well characterised ER or INM proteins in the expected fraction. For this a list with well characterised ER, INM and other organellar marker proteins was created (appendix table A.2). Next, statistical analysis was performed to investigate whether specific proteins are enriched in the $Nup153_{1-144}$ NE or the Climp63 CM fraction. For the Perseus analysis, two-sided student's t-tests were used ($s0 = 0.1, p < 0.05$). In R, statistical testing was done using the default settings in the 'limma' package[153]. limma testing is comparable to a t-test, but uses a standard error that is calculated using a linear method based on the Bayesian model. Originally developed for the analysis of microarray and RNA-sequencing data, it is nowadays routinely used in proteomic analysis. Results of all three analysis were visualised by volcano plots (figures 4.3, 4.4, 4.5). Overall, all three data analysis methods worked well in placing the INM marker proteins in the $Nup153_{1-144}$ fraction, whereas most ER marker proteins were enriched in the Climp63 CM fraction (figures 4.3B, 4.4B,

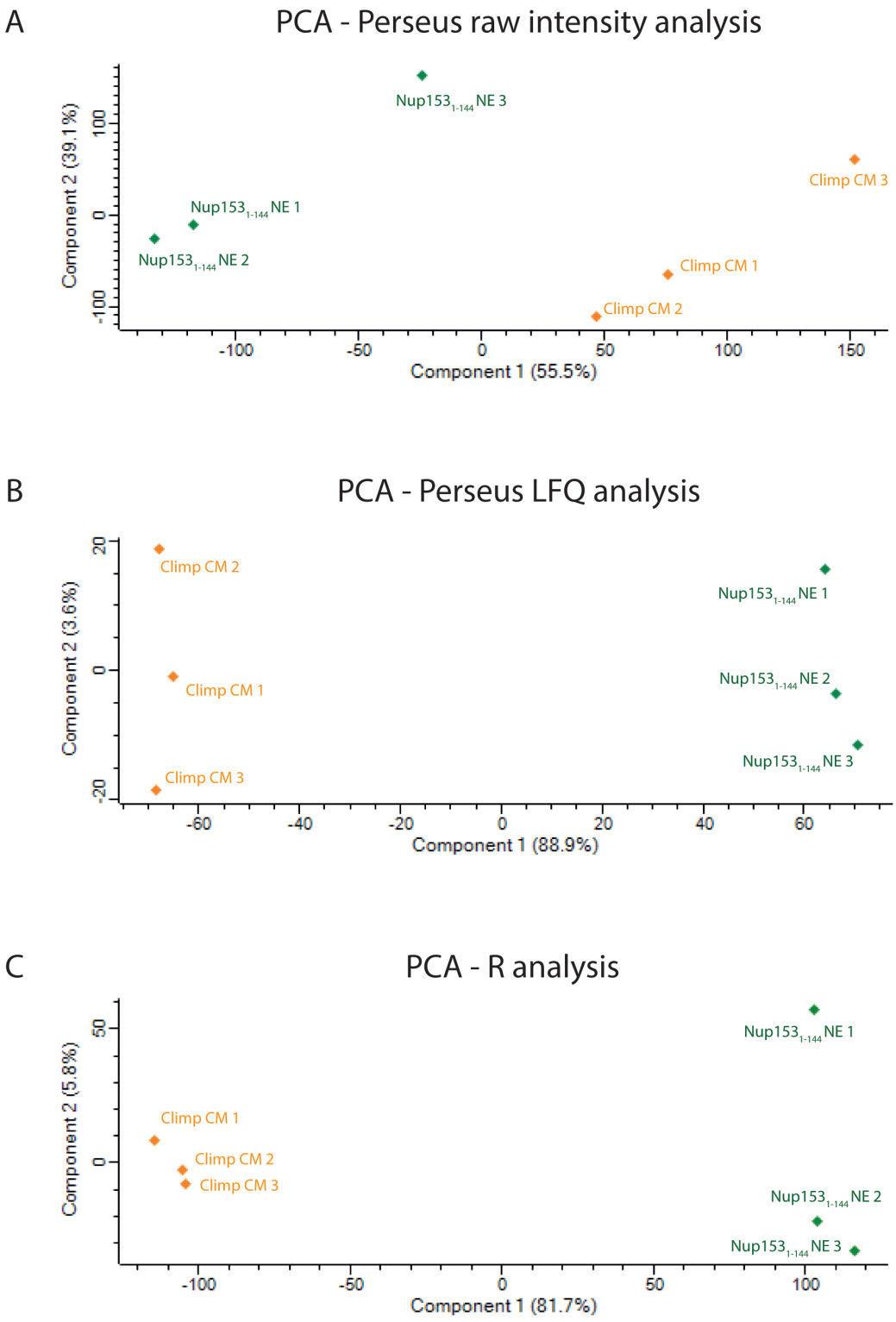


Figure 4.2: Using PCA analysis to assess different data analysis methods. 2D PCA analysis of $Nup153_{1-144}$ NE and the Climp63 CM samples. Each bait consisted of three independent replicates and data was filtered for 3/3 valid values in at least on group. PCA plot based on data analysis from A) Raw intensities using Perseus missing value imputation from normal distribution (width: 0.3, down shift: 1.8). B) LFQ intensities using Perseus missing value imputation from normal distribution (width: 0.3, down shift: 1.8) C) R analysis using variance stabilisation normalisation and LCMD imputation.

4.5B). This showed high specificity, as overall membrane proteins were strongly enriched in the Climp63 CM sample (figures 4.3A, 4.4A, 4.5A). Intriguingly, when working with raw intensities combined with student's t-test, not a single protein was considered significantly enriched or depleted using a $s0 = 0.1, p < 0.05$ cut-off. (4.3B), most likely due to the large variance observed between replicates (figure 4.2). In contrast, both the Perseus-based student t-test analysis using LFQ values and the R-based analysis using limma, not only successfully resulted in near perfect separation of the INM and ER marker proteins into their corresponding sample (figures 4.4B, 4.5B), but also classified many proteins as significantly enriched either in the $Nup153_{1-144}$ NE or Climp63 CM sample.

To gain a better understanding of differences in the results of the t-test analysis in Perseus using LFQ values versus limma analysis in R, the overlay of unique proteins considered significantly enriched or depleted by either of the two methods was determined (figure 4.6). The overlap of proteins considered either significantly enriched or depleted by both analysis methods was 72% and 63%, respectively (figure 4.6A & B). When analysing in more depth, not just focusing on the significant proteins, it became apparent that the difference in the results arose mainly from the fact that more unique proteins were discovered in the R-based analysis. This can be explained as more missing values existed in the LFQ intensities compared to the raw intensity, which led to loss of information after filtering for valid values (figure 4.6). Indeed, comparing filtering for 3/3 valid values in $Nup153_{1-144}$ NE or Climp63 CM resulted in 2030 unique proteins when using raw intensities, and 1310 unique proteins when working with LFQ intensities. Thus, while limma and t-test analysis led to similar results and high overlap as seen in figure 4.6, the R analysis based on raw intensities has increased information compared to the Perseus based LFQ intensity analysis.

Both the PCA and tracing marker protein enrichment indicated that t-test analysis using LFQ values in Perseus and limma based analysis in R are most suited for our needs. Yet, using a small list of manually selected marker proteins

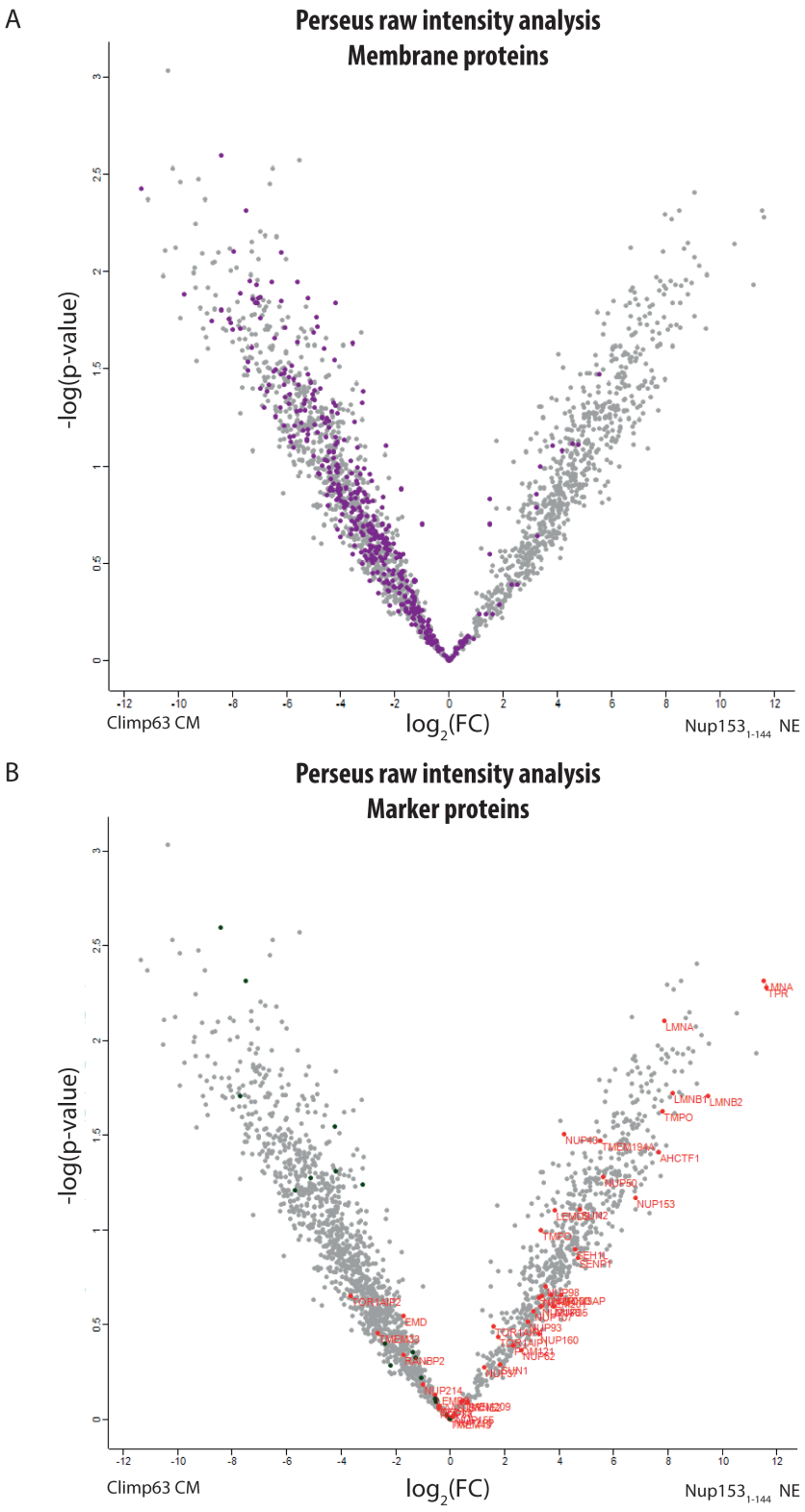


Figure 4.3: Investigating Perseus t-test analysis of raw intensities values using ER and INM marker protein enrichment. Pairwise-comparison for protein enrichment. Student t-test ($s0 = 0.1, p < 0.05$) of Nup153$_{1-144}$ NE vs Climp63 CM in Perseus using raw intensities. a) Purple – predicted TM proteins. b) orange - significantly enriched in Climp63 CM, blue – significantly enriched in Nup153$_{1-144}$ NE, red – INM marker proteins, green – ER marker proteins.

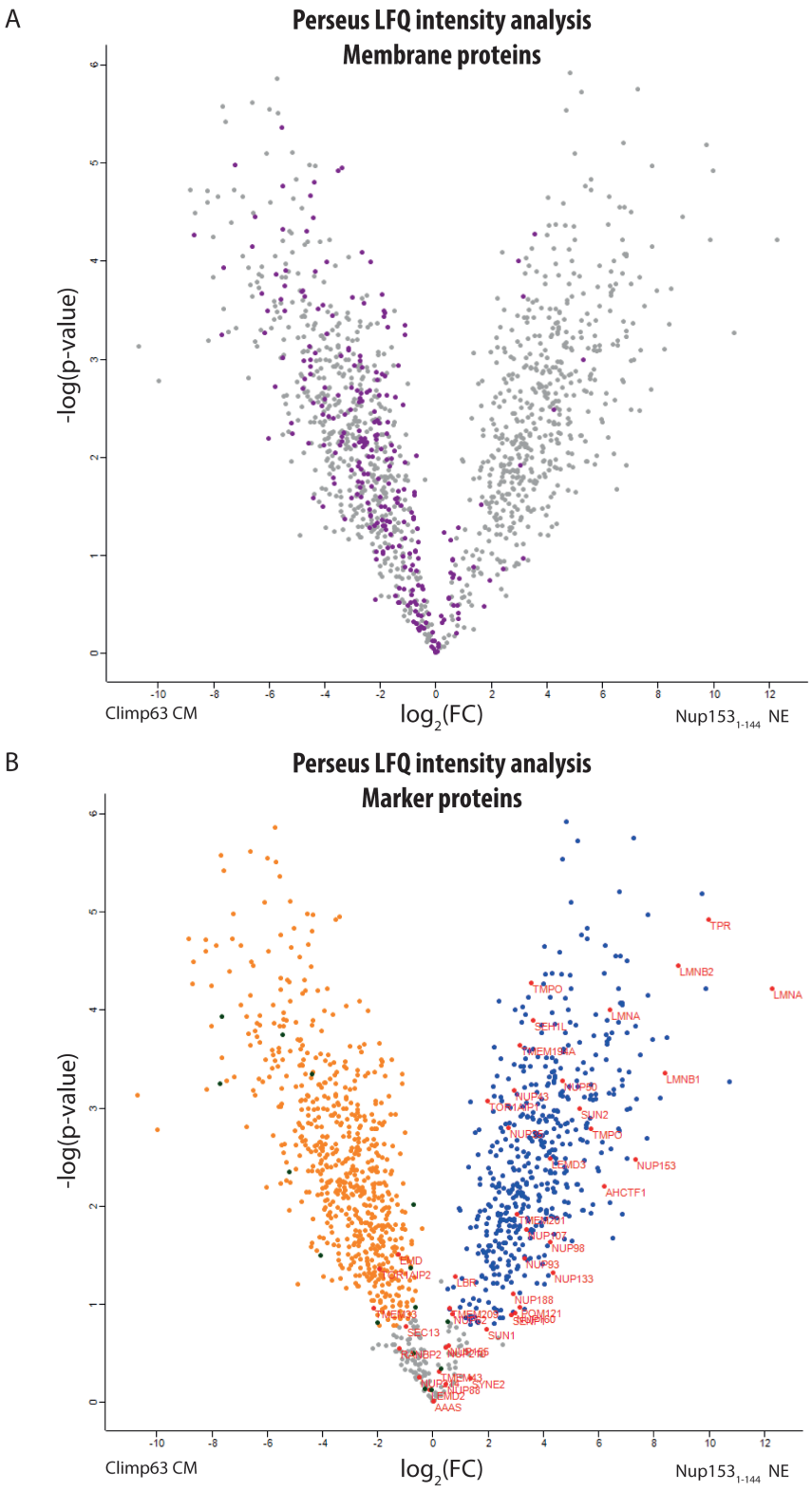


Figure 4.4: Investigating Perseus t-test analysis of LFQ intensities values using ER and INM marker protein enrichment. Pairwise-comparison for protein enrichment. Student t-test ($s0 = 0.1, p < 0.05$) of Nup153$_{1-144}$ NE vs Climp63 CM in Perseus using LFQ intensities. a) Purple – predicted TM proteins. b) orange - significantly enriched in Climp63 CM, blue – significantly enriched in Nup153$_{1-144}$ NE, red – INM marker proteins, green – ER marker proteins.

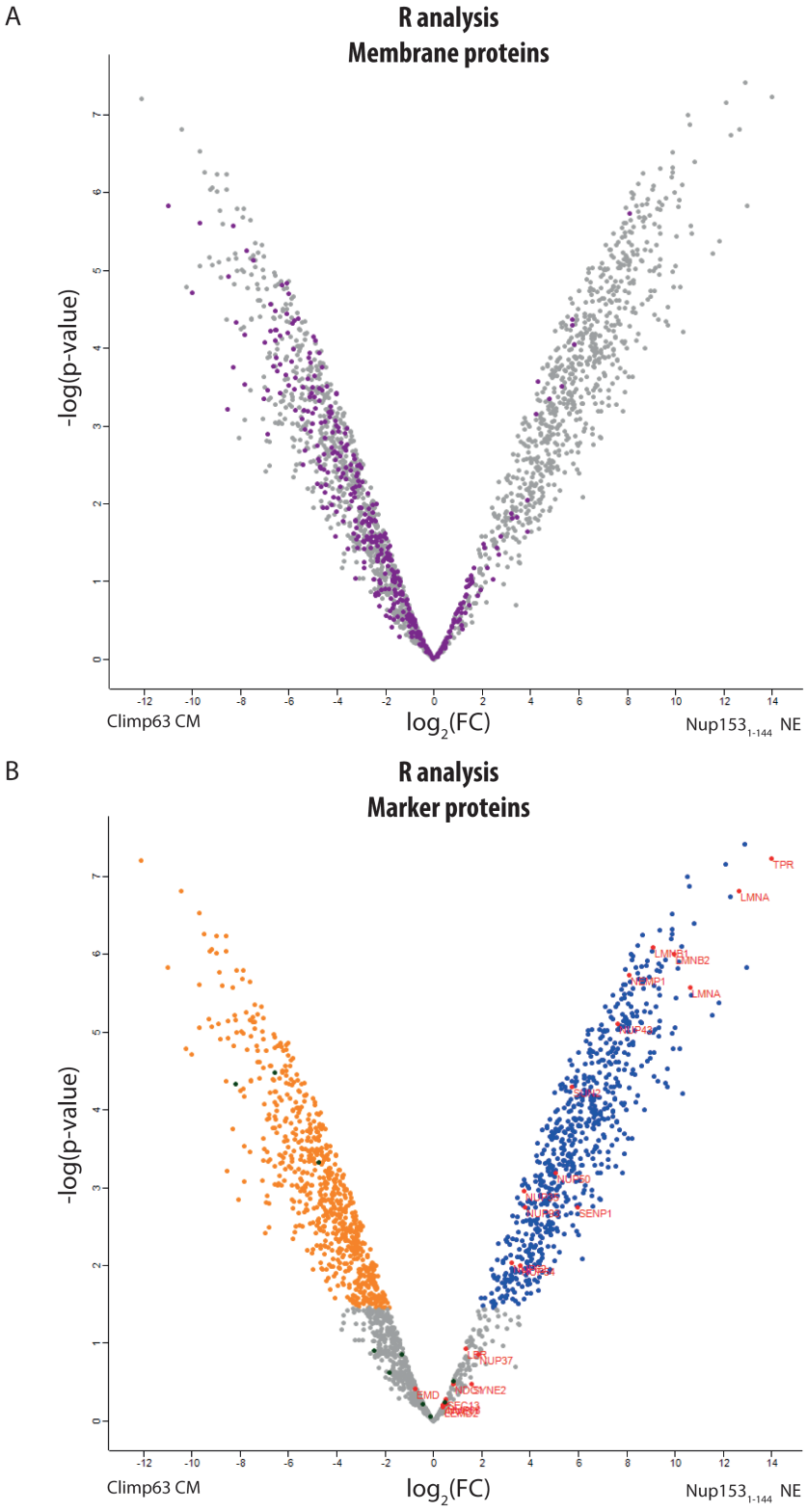


Figure 4.5: Investigating R-based Limma analysis of raw intensities values using ER and INM marker protein enrichment. Pairwise-comparison for protein enrichment. Limma analysis ($-1 < FC < 1; adj.p - value < 0.05$) of Nup153_{1-144} NE vs Climp63 CM in R using raw intensities. a) Purple – predicted TM proteins. b) orange - significantly enriched in Climp63 CM, blue – significantly enriched in Nup153_{1-144} NE, red – INM marker proteins, green – ER marker proteins.

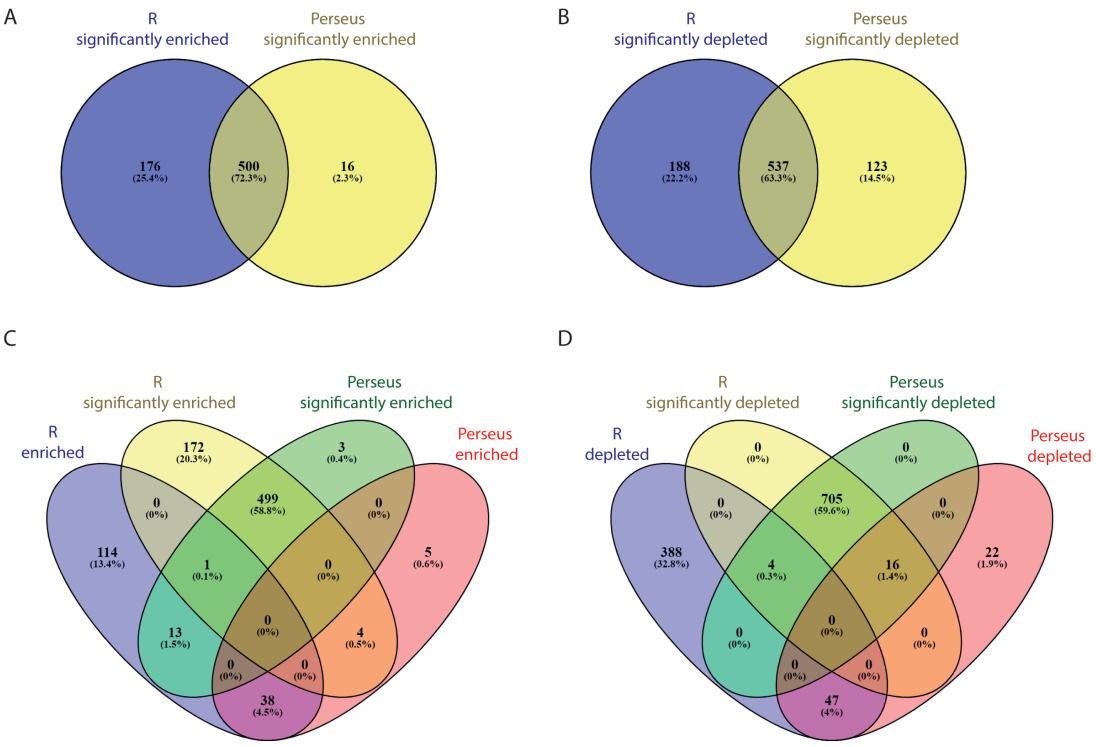


Figure 4.6: Comparing statistical testing using R-based limma and Perseus-based student t-test analysis. Comparison of statistical testing for the pairwise comparison of $Nup153_{1-144}$ NE vs Climp63 CM done in Perseus using student t-test analysis ($s0 = 0.1, p < 0.05$) or in R using limma ($-1 < FC < 1$; adj. p-value < 0.05). Enrichment indicates that proteins are enriched at in the $Nup153_{1-144}$ NE, whereas depletion indicates that proteins are enriched in the Climp63 CM sample. Venn diagrams show overlap between the two statistical analysis methods. a) Overlap of proteins significantly enriched in $Nup153_{1-144}$ NE by R-based limma (blue) and Perseus-based t-test (yellow). b) Overlap of proteins significantly enriched in Climp63 CM by R-based limma (blue) and t-test (yellow). c) Overlap of proteins significantly, as well as insignificantly enriched in $Nup153_{1-144}$ NE by R-based limma and Perseus-based t-test (yellow, blue, green and red respectively). d) Overlap of proteins significantly, as well as insignificantly, enriched by R-based limma and Perseus-based t-test in Climp63 (yellow, blue, green and red respectively).

may create a bias. Therefore, to assess the capability of each analysis method in an unbiased way, Gene Ontology enrichment analysis was performed for cellular compartment (CC) classification using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool[170,171]. GO CC enrichment was determined for all proteins considered significantly enriched in the $Nup153_{1-144}$ NE or Climp63 CM sample by both analysis methods. Outcomes were filtered for a Benjamini-Hochberg q-value < 0.05. The top 20 classifications, as determined

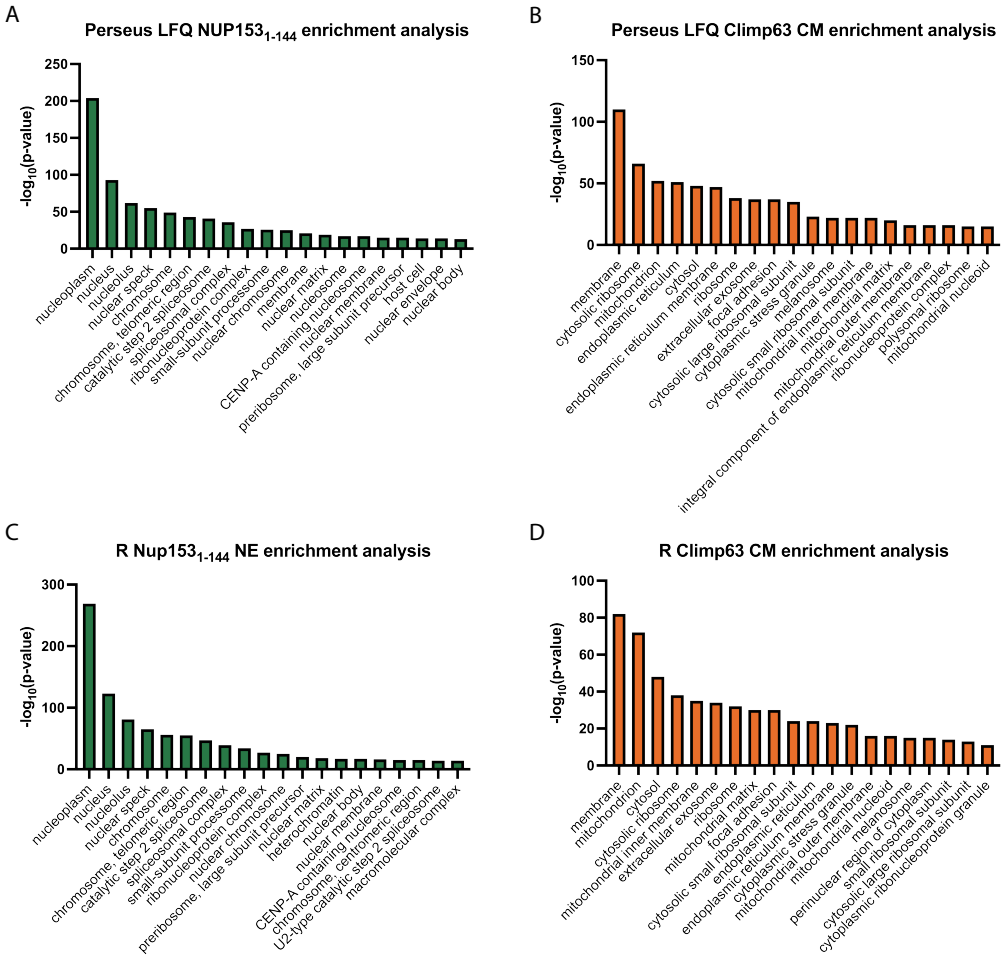


Figure 4.7: Gene Ontology cellular compartment classification of the Nup153$_{1-144}$ NE and Climp63 CM enriched proteins.. GO CC enrichment for all proteins enriched in the Nup153$_{1-144}$ NE or Climp63 CM samples over three replicates. Filtered for Benjamini–Hochberg q-value < 0.05. Top 20 hits shown classed by $-log_{10}(p - value)$.

by $-log_{10}$(p-value), can be seen in figure 4.7.

Results of GO CC enrichment analysis for both analysis methods were similar and demonstrated a strong enrichment for nuclear terms, such as the nucleoplasm, nuclear speckle, chromosomes and the nuclear envelope for the Nup153$_{1-144}$ NE sample. The fact that the top rated GO CC terms are for soluble cellular components is in line with ∼80% of the identified proteins being soluble proteins. In comparison, the GO enrichment analysis of the Climp63 CM sample using both data analysis methods revealed an enrichment in mitochondrial, ER and cytoplasmic cellular components. Thus, using an unbiased approach, both

analysis methods successfully demonstrated protein enrichment in line with their bait protein.

To conclude, the PCA analysis, the identification of marker proteins in the correct fraction and the GO CC enrichment analysis indicated that the LFQ based Perseus analysis and the raw intensity based R analysis are comparably well suited for analysis of our proteomic data set. Analogous results were obtained with different probes indicating this approach is suitable to investigate our data set. To avoid loss of information observed when working with LFQ intensities, the R-based analysis was chosen for all further data analysis steps. To summarize, the three important analysis steps in the R workflow are

- **Normalisation.** Variance stabilisation normalisation was applied as a way to normalise and obtain a constant signal-to-noise ratio[151]. This methodology also takes into account the problem of machine calibration, which is especially beneficial considering the batch effect expected due to processing samples at different times on different machines.

- **Imputation of missing values.** The standard settings of the 'imputeLCMD' package was used to replace missing values[152].

- **Statistical testing**. Statistical testing in R was performed using the 'limma' package[153]. Whereas this works comparable to a t-test, it uses a standard error that is calculated using a linear method based on the Bayesian model.

4.2.3 Investigating the INM proteome using enrichment analysis

Having optimised a data analysis pipeline, the processed data can now be used to identify *bona fide* proteins enriched in each of the ER domains. To identify putative INM proteins the $Nup153_{1-144}$ NE versus Climp63 CM pairwise comparison, that was just used to set up the analysis pipeline, was utilised. In this enrichment analysis, the $Nup153_{1-144}$ NE was shown to have strong enrichment for nuclear terms in the $Nup153_{1-144}$ NE sample by GO CC enrichment analysis (figure 4.7), as well as the enrichment of all our INM marker proteins in the $Nup153_{1-144}$ NE sample (figure 4.5). Besides the INM marker proteins, 6 additional TM proteins were shown to be significantly enriched in the $Nup153_{1-144}$ NE fraction (RENR, ASPH, AR6L6IP, DCAKD, TMEM109, ALG9). Three of these proteins (ARL6IP6, ASPH, TMEM109), that were also identified to be enriched at the INM using additional pairwise comparison such as Rtn4a CM vs $Nup153_{1-144}$, will be discussed in more detail in chapter 4.2.8.

Another promising pairwise comparison to identify novel INM proteins is the Sec61β NE versus CM comparison. The Sec61β probe is a generic ER probe that was designed to label proteins in the bulk ER and in the INM. To check that the bait behaved as expected, GO CC enrichment analysis was performed (figure 4.8). As anticipated, the Sec61β NE fraction was strongly enriched for nuclear proteins, both soluble and membrane-bound. In comparison the Sec61β CM fraction showed high enrichment for mitochondria and ER proteins. The presence of large amounts of mitochondrial proteins may be explained by the extensive number of contact sites that exist between the ER and the mitochondria[2]. Next, the separation of our ER and INM marker proteins in the two fractions was tested (figure 4.9). Indeed, nearly all marker proteins were found in the expected fraction. Thereby, overall membrane proteins were strongly enriched in the Sec61β CM fraction, in line with a stronger biotinylation signal observed for the Sec61β in the CM, compared to the NE fraction. Of note, while most well defined ER proteins

are enriched in the Sec61β CM fraction, these proteins are still identified in the Sec61β NE fraction, albeit with lower intensities. This may suggest that a large fraction of the ER membrane proteins traffic through the INM, as observed in yeast[56] and in line with a diffusion-retention based INM targeting mechanism as discussed in chapter 1.4. To identify novel INM proteins, significantly enriched proteins in the Sec61β NE fraction were filtered for membrane proteins. In total 19 membrane proteins were significantly enriched in the Sec61β NE fraction, of which 10 were part of our INM marker list, as well as two nucleoporins not included in our marker list. This resulted in the identification of 7 *bona fide* INM proteins (DSC1, CDSN, UBP48, PIGS, TMX4, KMT5A, SMPD4). Of these, the majority were identified with very low peptides numbers. One of the proteins identified with higher confidence was SMPD4, which will be discussed further in chapter 4.2.9.

To conclude, our data set coupled to pairwise enrichment analysis is a great resource to identify putative INM proteins. Note that, whereas only two pairwise-comparisons were discussed here in detail, we collected data sets of two fractions per bait, with 5 different baits overall. This allows 45 different pairwise-comparisons to obtain a road map of protein enrichment in different fractions of the ER, from the tubules that stretch out to the plasma membrane to the INM. The complete proteomic data will be deposited publicly in the future.

4.2.4 Designing a tool to identify putative INM proteins

Whereas pairwise enrichment analysis worked well in investigating the INM and ER proteome, these analyses result in an independent fold change (FC) and p-value for each comparison. With the large number of pairwise comparisons possible in our data set, we wanted to create a tool to combine the data of various independent comparisons to better identify *bona fide* INM proteins. To do this, the FC enrichment data for the NE versus CM fraction of three baits ($Nup153_{1-144}$, Sec61β and Rtn4a) was determined independently. This resulted in three FC enrichment values associated with each protein. Of note, some proteins

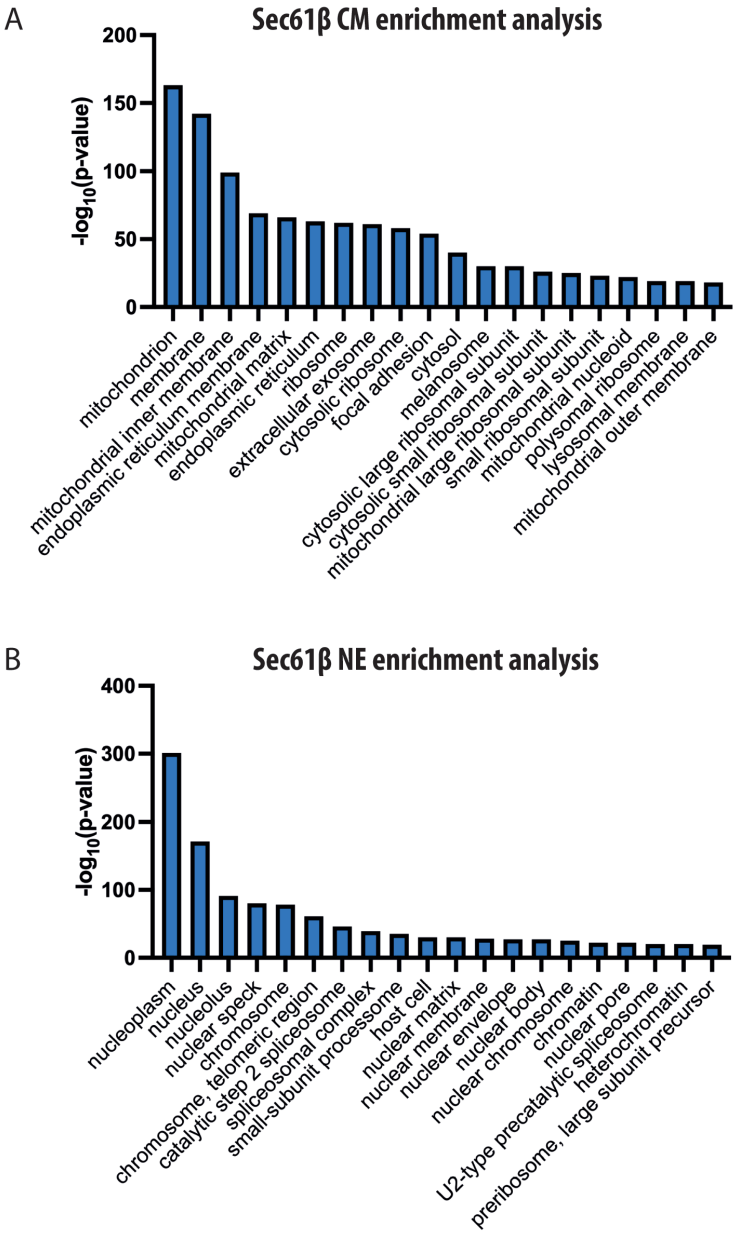


Figure 4.8: Investigating the INM and ER protein composition by GO-enrichment analysis using the Sec61β bait. GO CC enrichment over six replicates. Filtered for Benjamini–Hochberg q-value < 0.05. Top 20 hits shown classed by $-log_{10}(p - value)$. Top hit in NE enrichment analysis showed perfect enrichment with $p = 0$. To enable plotting, the value was replaced with a value close to $0(2.11e^{-301})$. A) Proteins enriched in the Sec61β CM versus NE samples. B) Proteins enriched in the Sec61β NE versus CM samples.

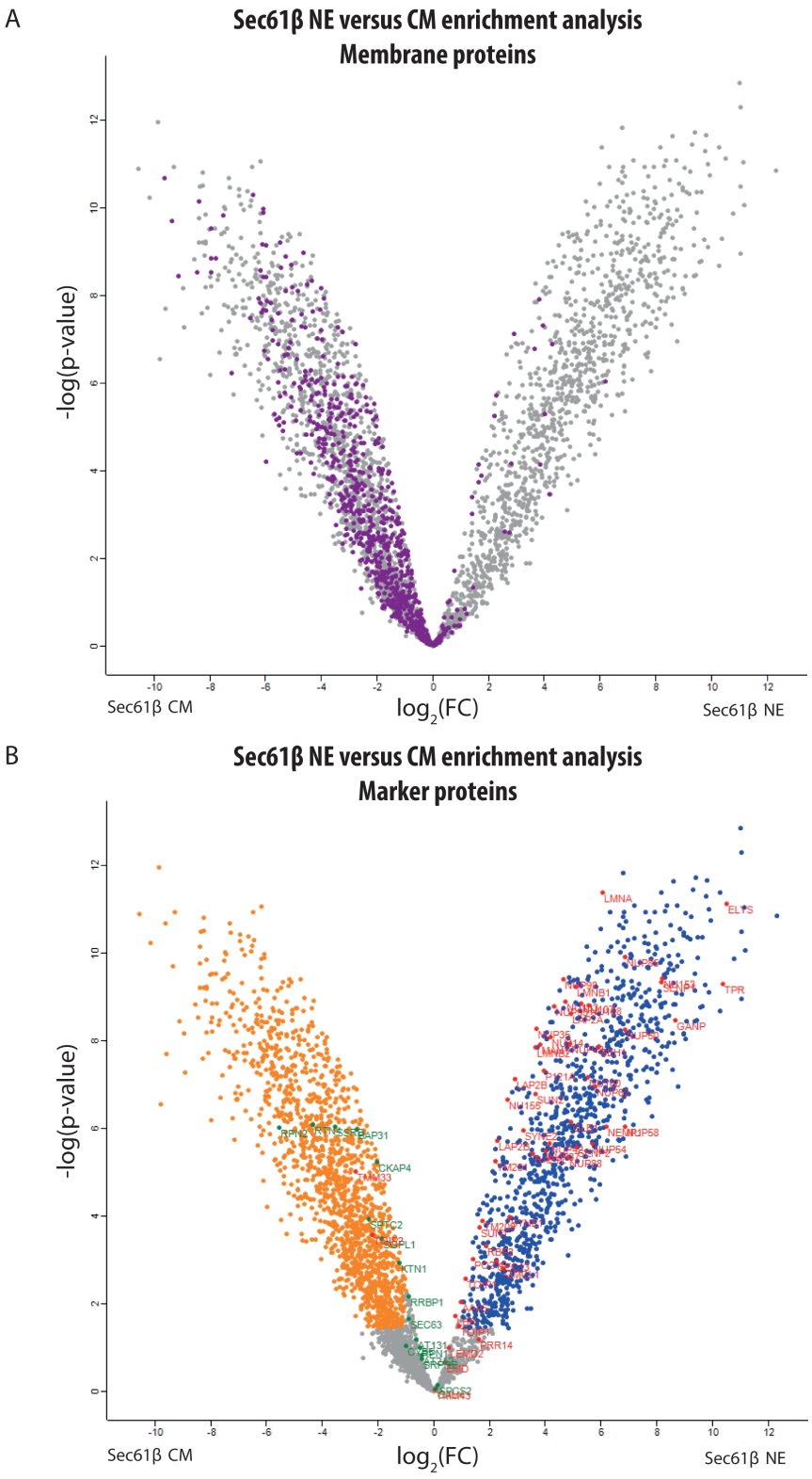


Figure 4.9: Investigating the INM and ER protein composition by Limma analysis using the Sec61β bait. Pairwise-comparison for protein enrichment in the INM or ER by analysing protein enrichment in Sec61β NE or CM, respectively. Limma analysis of raw intensities of Sec61β NE vs CM in R ($-1 < FC < 1; adj.pvalue < 0.05$). A) Purple – predicted TM proteins. B) orange - significantly enriched in Sec61β CM, blue – significantly enriched in Sec61β NE, red – INM marker proteins, green – ER marker proteins.

were not detected for certain baits, resulting in issues in the analysis tool. In these cases missing values were imputed as detailed in chapter 2.2.4.6. Next, we concatenate the three values associated to each protein in an interactive tool (using the profile plot tool within Perseus). This tool enables users to choose any protein as a reference protein, and determine a predefined number of proteins with the most similar behaviour to this reference protein across the three probes.

We hypothesise, that proteins with similar enrichment scores across the various probes are likely to have conserved localisations in the cell. We first tested whether marker proteins for different organelles showed comparable behaviour over the three probes (figure 4.10A). As expected, marker proteins for different cellular localisations (INM, NPC, ER, mitochondria) followed different patterns over the three baits. Next, we tested using one INM protein as a reference protein whether the tool was capable of identifying other INM proteins. For this, SUN2 was chosen as a reference protein and the 49 proteins showing the most similar behaviour over the three enrichments, as determined by Pearson correlation, were identified (figure 4.10B). Among the 49 proteins identified, we found 8 INM proteins, as well as three ER marker proteins (SRPRB, Climp63, AT131). Interestingly, whereas SRPRB, Climp63 and AT131 are all well characterised ER proteins, none of them has a large enough cytoplasmic domain to stop them from diffusing to the INM, indicating once more that ER proteins may readily traffic through the ER. Another interesting observation was that over 70% of the 49 proteins determined to be most similar to SUN2 were membrane proteins. This is astonishing, considering the overall percentage of membrane proteins identified in the data set was ∼20%. This suggests that while many soluble proteins are enriched in individual baits they do not follow the same enrichment pattern as membrane proteins combining the NE/CM enrichment data over three bait proteins used here. This may make this tool useful in identifying membrane-associated proteins that are not picked up by TMHMM2.0 membrane prediction due to the lack of a TMD, but may still behave similar to a membrane protein. A table containing the proteins identified identified as most similar to SUN2 can be found in appendix table A.3.

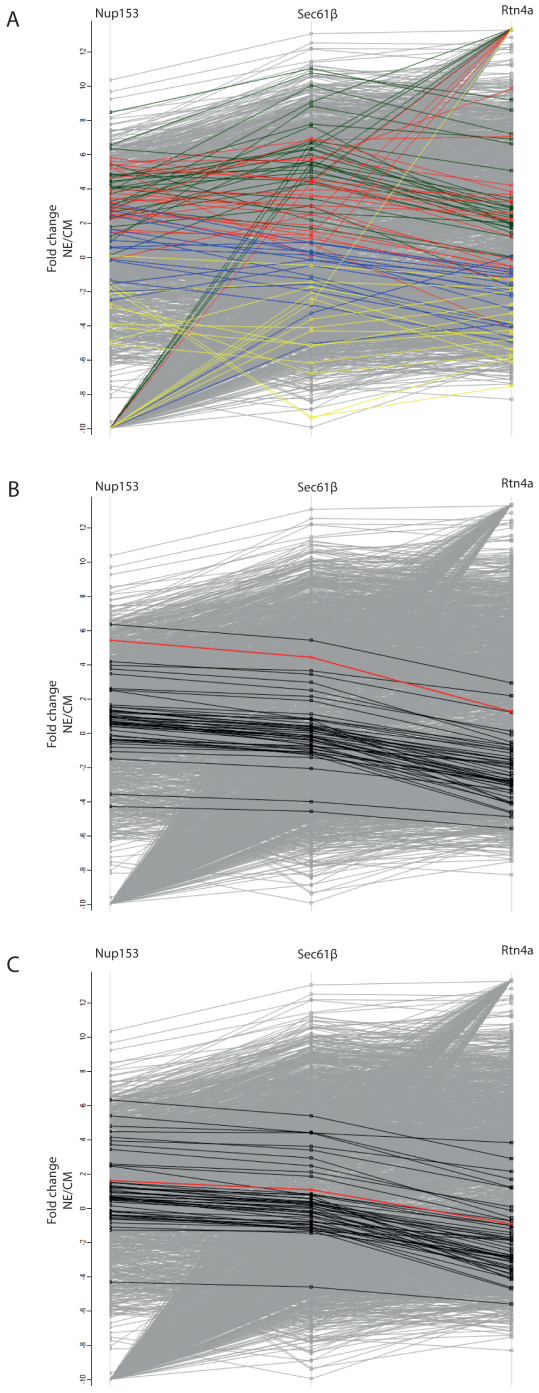


Figure 4.10: Combining FC analysis across multiple probes to identify *bona fide* INM proteins.. a) Fold changes of the NE versus CM fractions of the three probes Nup153_{1-144} (left), Sec61β (middle), Rtn4a (right). Marker proteins are visualized in colour. Red: INM, green: NPC, blue: ER, Yellow: mitochondria. b) Red: SUN2 reference profile. Black: 49 proteins with closest Pearson correlation to the SUN2 reference profile. c) Red: Emerin reference profile. Black: 49 proteins with closest Pearson correlation to the emerin reference profile.

To test the system with a second INM protein, the analysis was repeated using emerin as a reference protein (figure 4.10C). Whereas emerin is also a well defined INM protein, its enrichment is less strong and besides the INM it was also shown to localise to the bulk ER[172]. In line with that, the FC enrichment in the NE fractions is lower compared to SUN2. Searching the top 49 profile plots by Pearson correlation, four other INM/NPC marker proteins were identified, as well as various ER proteins. Additionally, 70% of the proteins with similar profile plots were again membrane proteins. An overview over all of the identified proteins can be found in appendix table A.4.

To conclude, this method is a powerful tool to combine information from various independent pairwise comparisons. Using the three NE/CM FC scores of $Nup153_{1-144}$, Sec61β and Rtn4a, the method works well in separating different organellar marker proteins from each other (figure 4.10A). Additionally, using SUN2 as a reference protein, the tool successfully identified other INM proteins, as well as three ER proteins that, due to their size, may readily diffuse into the INM. It was particularly interesting to see that using a TM reference protein, over 70% of proteins showing the highest similarity by Pearson correlation over the 3 probes, were also membrane proteins. This is in stark contrast to an overall membrane protein identification rate of ~20%. This makes the tool valuable in identifying *bona fide* INM proteins, some of which will be discussed later in chapter 4.2.8. Of note, while we focused on FCs between NE and CM fractions within baits, this tool can be used for combining FC information from any kind of comparison.

4.2.5 Investigating protein compositions of ER sheets and tubules

Two of our APEX2-fusion proteins, Climp63 and Rtn4a, are enriched in ER sheets and tubules, respectively[23]. This makes them suitable to study differences in the protein composition of the two domains. We therefore performed pairwise enrichment analysis of the Rtn4a CM versus the Climp63 CM samples. Overall, the fold changes observed in protein composition between ER tubules and ER

sheets were much more subtle compared to the changes observed between the ER sheets and the INM. This is in line with the idea that ER tubules, sheets and the ONM have a similar protein composition, whereas the INM has a distinct proteome. Yet, subtle differences could still be observed between the data sets collected using the Climp63 or the Rtn4a bait. To assess the quality of our data set, we first checked how well-defined ER marker proteins behave in this comparison (figure 4.11B). As our proteins are generic ER proteins and not known to enrich in either ER tubules nor ER sheets, they were found spread over both two samples. Additionally, both probes labelled roughly equal amounts of membrane proteins (figure 4.11A). As ER tubules are known to form contact sites with other organelles, we also checked whether organellar markers from the plasma membrane, the Golgi and the mitochondria are enriched in one of the baits (figure 4.11B). Intriguingly, there seemed to be a trend that Golgi marker proteins were more equally split between the two baits, whereas mitochondrial marker proteins seemed slightly skewed towards the Rtn4a CM sample. Indeed, whereas GO enrichment analysis did not result in any significant differences between the two data sets with a Benjamini-Hochberg cut-off of $p < 0.05$, the proteins just below the threshold for the Rtn4a CM sample were mostly GO CC terms associated with mitochondria. This seemed intuitive as by IF and EM microscopy the Golgi is generally localised in closer proximity to the nucleus, and therefore also ER sheets, whereas mitochondria are spread throughout the cytoplasm.

Next, the localisation of proteins known to enrich in ER sheets or tubules was analysed (figure 4.12). When comparing our data set with the well characterised tubule and sheet proteins from Shibata *et al.*[23], for some of the proteins we observed a similar trend, yet for many proteins we found discrepancies. RRBP1, shown to be enriched in ER sheets in IF microscopy experiments by Shibata *et al.*, was also enriched in the ER sheets for our data set, yet just below the significance threshold. Additionally, their data indicated the enrichment of the signal recognition particle (SRP) receptor (SRPR) and the OST complex in ER sheets. In comparison to that, in our data set SRPRα and SRPRβ show no

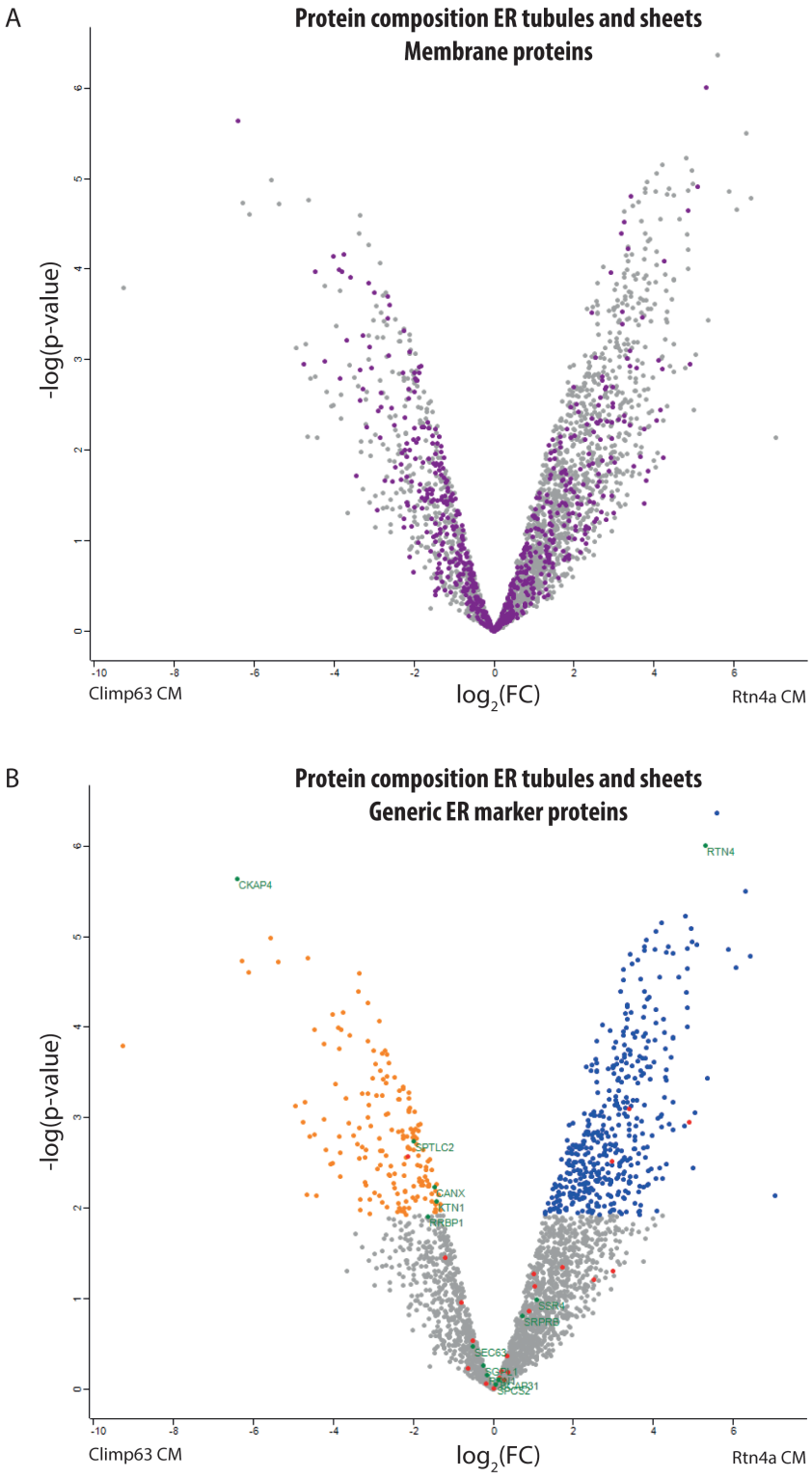


Figure 4.11: Investigating the protein composition of ER tubules and ER sheets. Limma analysis of raw intensities of Rtn4a CM versus Climp63 CM in R ($-1 < FC < 1; adj.pvalue < 0.05$). a) Purple – predicted TM proteins. b) Orange - significantly enriched in Climp63 CM, blue – significantly enriched in Rtn4a CM, green – ER marker proteins, red – Golgi marker proteins.

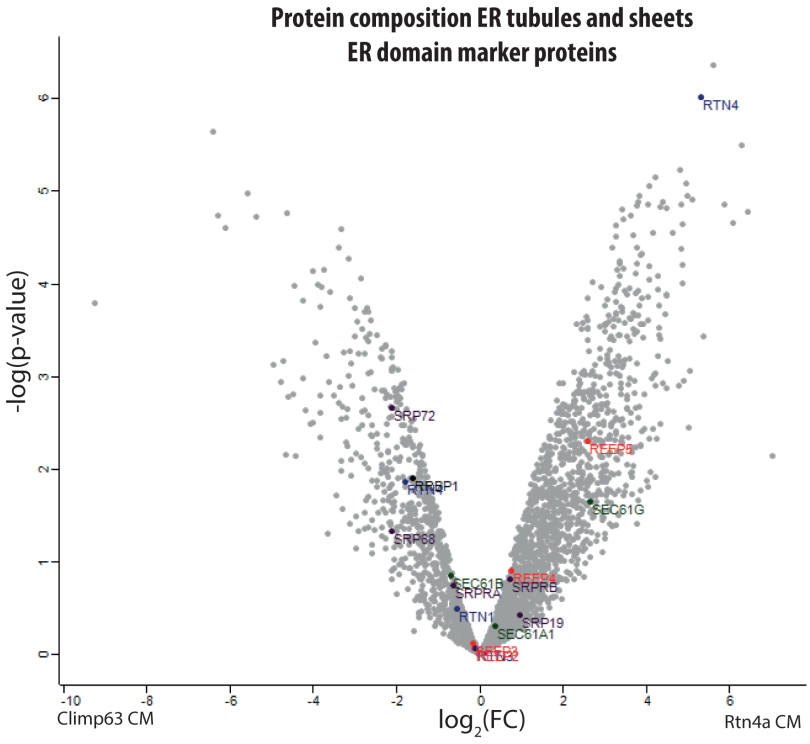


Figure 4.12: Investigating the protein composition of ER tubules and ER sheets using tubular and sheet marker proteins. Limma analysis of raw intensities of Rtn4a CM versus Climp63 CM in R ($-1 < FC < 1; adj.pvalue < 0.05$). Marker proteins for ER tubules and sheets, as described by Shibata *et al.*[23].

enrichment for a specific ER domain, and the SRP subunits even show a slight skew towards ER tubules. Additionally, IF data from their paper suggests that Sec61β is enriched in the ER sheets. As this protein is part of the Sec61 translocon, they hypothesised that the whole translocon is enriched in the ER sheets. Yet, in our data Sec61β, Sec61α and Sec61γ, were neither significantly enriched in the tubules, nor in the sheets. The same was true for subunits of the OST complex, with the only exception being OST48, which showed significant enrichment in ER sheets. REEP proteins and reticulons were previously shown to be enriched in the ER tubules. When looking at REEP proteins in our Rtn4a versus Climp63 CM enrichment analysis, the data was less clear: whereas Rtn4a and REEP5 were significantly enriched in the Rtn4a bait data set, REEP2, REEP3, REEP4, Rtn1 and Rtn3 were not significantly enriched for either of the two baits.

To conclude, some of the identified, enriched proteins were in line with previous

findings, such as REEP5, whereas for other proteins results were less clear. Yet, some of these findings are interesting to follow up. We focused on the further validation of the RNF185-Membralin complex, which was identified in our lab group as a novel ERAD complex[173]. This complex was enriched in the CM fraction of the Rtn4a bait compare to the Climp63 bait.

4.2.6 An overview of methods to validate INM protein localisation

Having identified *bona fide* proteins enriched in the INM, ER tubules and ER sheets, the next step was to validate some of these findings further. Thereby specific challenges are associated with investigating whether proteins are localised at the INM or ONM, one being that the perinuclear space separating the INM from the ONM is only $\sim$30 – 50 nm[174,175]. Additionally, the prevalent model for protein targeting to the INM is a diffusion-retention based mechanism, by which proteins are synthesised in the ER, from where they need to pass the NPC to reach the INM. Once proteins have reached the INM, INM proteins interact with nuclear component such as DNA, the nuclear lamina or adaptor proteins, resulting in INM proteins to become less mobile and therefore getting enriched in the INM[72–74,76,77,176–178]. This mechanism creates two difficulties when working with INM proteins. Firstly, tagging proteins with large cytoplasmic domains will create problems with the $<$ 60 kDa size limit of the NPC and stop proteins from reaching the INM. Secondly, overexpression of INM proteins can potentially lead to a saturation of binding partners in the nucleus, thereby stopping proteins from being retained at the INM.

A range of available techniques exist to investigate whether a protein localises to the INM or ONM[179]. The gold standard to determine the exact localisation of a protein is EM. Yet, the drawbacks of this technique are its low throughput, high costs, as well as the need to work in fixed cells. Chemically induced dimerisation assays are another tool that can be used to validate INM localisation. There are several reviews looking at chemically induced dimerisation tools[180–182]. Having a

soluble NLS-FRB protein, in combination with a putative INM protein with a FKBP domain, should show a clear nuclear rim signal of the soluble NLS-FKBP protein upon induction of dimerisation using rapamycin[54]. Yet, this technique is again dependent on tagging the putative INM protein and potentially overexpressing it. Additionally, the lamina network might intervene with dimerisation of the FRB-FKBP system. A third approach to validate localisation of proteins at the INM are bimolecular fluorescence complementation (BiFC) assays. BiFC relies on the complementation of two fragments of fluorescent proteins, that only emit a signal upon complementing each other. Complementation requires the two parts to be in close proximity to each other, making BiFC a common tool to study protein interactions. Various different split fluorescent molecules nowadays exist[183–185]. Similar to the FRB-FKBP system, one part of the BiFC system could be targeted to the nucleus by either fusing it to a soluble NLS signal, or by putting it on a well-characterised INM protein. The other part could then be put on putative INM proteins. If these proteins indeed enrich at the INM, a fluorescent signal should be visible. Again, tagging of proteins is required, and therefore the drawbacks of this technique are identical to the chemically induced dimerisation assays.

Another method that can be used to distinguish between INM and ONM localisation of proteins is differential membrane permeabilisation. Digitonin has a very high affinity for binding cholesterol, which is found in large amounts in the cell membrane, but in much lower degrees in the NE[186]. Due to its preferred interaction with cholesterol, treatment with low concentrations of digitonin results in selective perforation of the cell membrane, while leaving all other major organellar membranes intact[187]. In contrast, Triton X-100 and SDS permeabilise all cellular membranes. This enables the creation of an assay, in which fixed cells are permeabilised either with digitonin only, or with SDS/Triton X-100 (TX). Using fluorescent labelled antibodies, proteins in the INM will only be accessible in cells permeabilised with SDS/TX, whereas proteins in the ONM will be accessible also in presence of digitonin alone. This approach has been used

in the past to validate localisation of proteins in the INM versus the ONM[78]. A major drawback of this technique is that an antibody suitable for IF needs to be available for the proteins of interest and that titration of the correct amount of digitonin to use for each cell type can be tedious. Yet, this method prevents the two major drawbacks that can lead to mislocalisation of INM proteins: the need for overexpression and the need to tag a protein.

Concluding, one can say that each method has advantages, but also drawbacks. After carefully weighing out the advantages and disadvantages of each technique, we decided to option for a differential membrane permeabilisation approach to try and validate some of the putative INM proteins identified.

4.2.7 Implementing a differential permeabilisation assay

The first step was to optimise the differential permeabilisation assay for FlpIn293 cell lines. Available protocols from literature were optimised for HeLa cells, so the first step was to validate the method was working in my hands in HeLa cells. For the validation, two different marker proteins were used: LBR and LMNA, known to be enriched in the INM and in the nuclear lamina, respectively. HeLa cells were fixed and subsequently either no membrane permeabilisation, soft permeabilisation using digitonin or hash permeabilisation using SDS/TX after the initial digitonin permeabilisation were performed (figure 4.13). As expected, for the LMNA antibody no labelling was observed in the no-permeabilisation and digitonin condition, whereas upon harsh permeabilization, strong labelling could be observed along the nuclear rim and in the nucleus. For the LBR antibody, weak labelling could be observed in the no permeabilisation conditions. Upon permeabilisation with digitonin, a signal was observed to overlay with the Bap31 ER marker. This is in line with LBR localising to the ER, as well as the INM. Upon harsh permeabilisation, a very strong signal could be observed at the nuclear rim and in the nucleus, whereas the signal in the bulk ER disappeared. Note that weak staining by anti-Bap31 could be observed even in the conditions with no permeabilisation, suggesting that the plasma membrane must be porous to

some degree to allow the Bap31 antibody to reach the ER. Overall, this data indicated that the differential permeabilisation assay works well in HeLa cells.

As all proteomics experiments were performed in FlpIn293 cells, and all putative INM proteins were identified in that cell line, the next step was to establish a differential permeabilisation protocol for FlpIn293 cells. Using the same conditions (0.001% digitionin, 14 min) for the FlpIn293 as for the HeLa cells (figure 4.14) resulted in a nuclear rim signal for LMNA. Neither reducing the permeabilisation time (figure 4.14) nor the concentration of digitonin (figure 4.15), nor a combination of the two resulted in loss of the LMNA signal. This indicated that the FlpIn293 cell membrane, as well as the NE, are permeable to antibodies even in absence of a permeabilisation step. Therefore, differential permeabilisation is not a viable method in FlpIn293 cells.

To conclude, I was unable to adapt a differential permeabilisation protocol to determine localisation of proteins at the INM in FlpIn293 cells. Yet, the technique works well in HeLa cells. Therefore all validation experiments for INM proteins were done in HeLa cells.

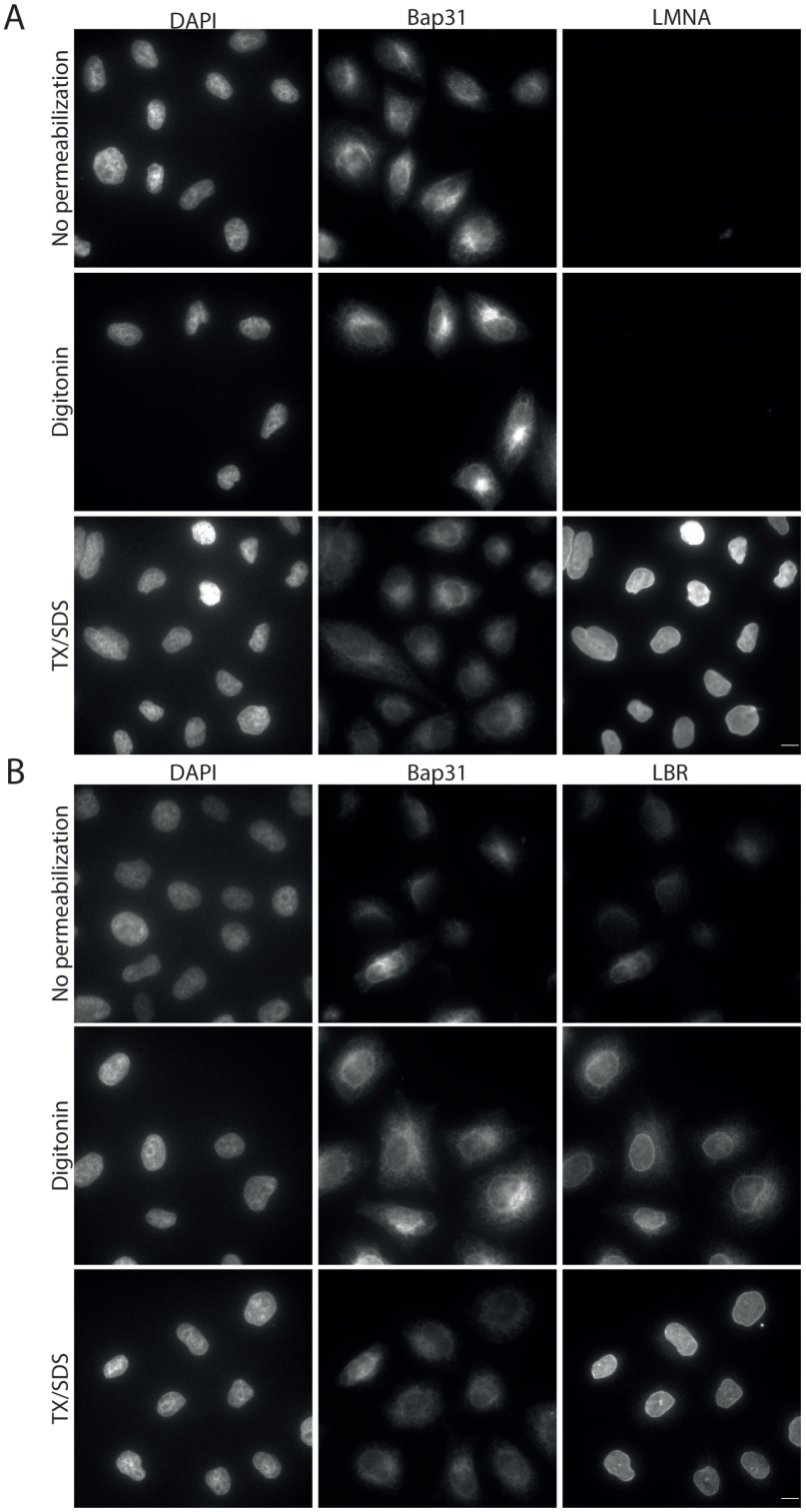


Figure 4.13: Differential membrane permeabilisation in HeLa cells. IF microscopy images of HeLa cells that were fixed and either not permeabilised, or permeabilised with either digitonin or digitonin followed by Triton X-100/SDS (TX/SDS). Nuclei were stained using DAPI. Scale bar: 10 μm. Immunostaining was performed using anti-Bap31 and a) anti-LMNA in combination with their corresponding Alexa Fluor secondary antibodies. b) and anti-LBR in combination with their corresponding Alexa Fluor secondary antibodies.

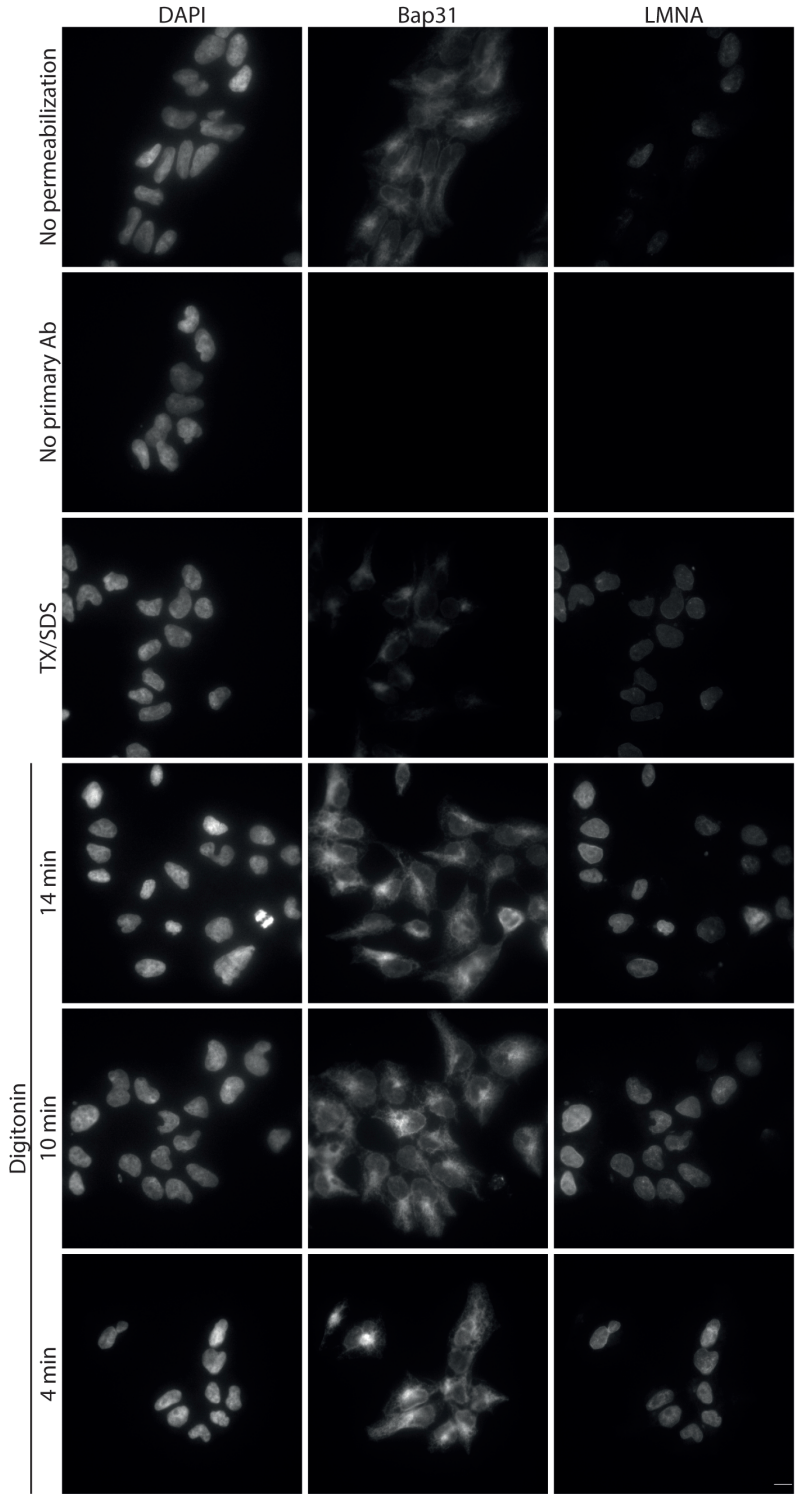


Figure 4.14: Optimising the duration of differential membrane permeabilisation in FlpIn293 cells. IF microscopy images of FlpIn293 cells that were either not permeabilised, permeabilised for different duration with digitonin or digitonin followed by Triton X-100/SDS (TX/SDS). Nuclei were stained using DAPI. Scale bar: 10 μm. Immunostaining was performed using anti-Bap31 and anti-LMNA in combination with their corresponding Alexa Fluor secondary antibodies.

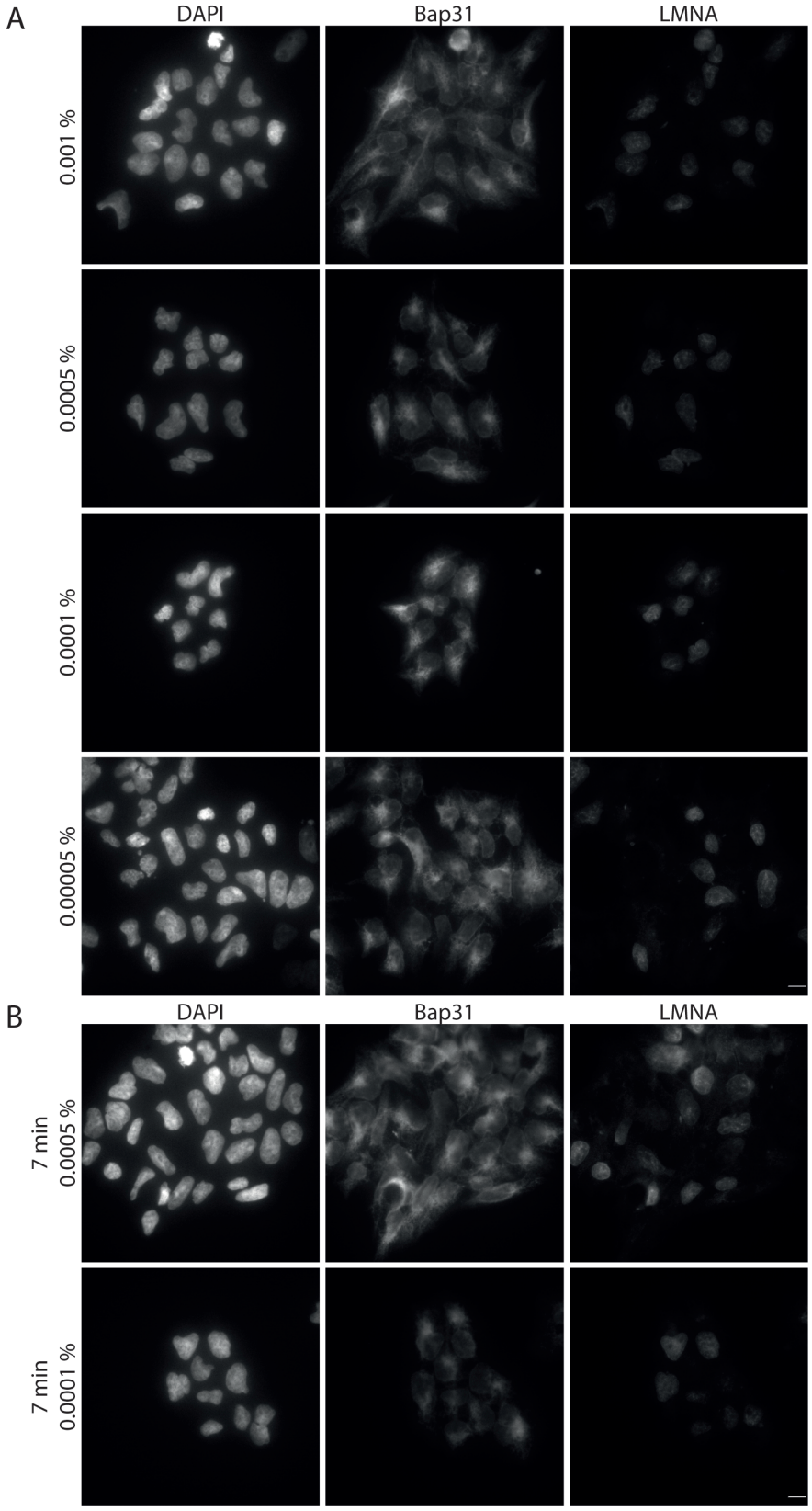


Figure 4.15: Optimising differential membrane permeabilisation conditions in FlpIn293 cells. IF microscopy images of FlpIn293. Immunostaining was performed using anti-Bap31 and anti-LMNA in combination with their corresponding Alexa Fluor secondary antibodies. Nuclei were stained using DAPI. Scale bar: 10 μm. a) Cells were permeabilised using different concentrations of digitonin for 14 min. b) Cells were permeabilised using different concentrations of digitionin for 7 min.

4.2.8 Validating putative INM proteins

A list of putative INM candidate proteins was created, with a focus on the following pairwise comparisons: $Nup153_{1-144}$ NE vs Climp63 CM; $Nup153_{1-144}$ NE vs Rtn4a CM and Sec61β NE vs Sec61β CM. Proteins were chosen as putative INM proteins if they fulfilled the following four criteria

- significantly enriched in the NE vs CM fraction in one of these comparisons. Priority was given to proteins enriched in more than one comparisons.
- predicted to have at least one TMD
- identified with more than one peptide in the relevant samples
- promising antibodies were available for IF

Additionally, SUN2 and emerin were used to identify further *bona fide* INM proteins using the profile plot tool (4.2.4).

4.2.8.1 ASPH

Aspartyl/asparaginyl beta-hydroxylase (ASPH) is a 85 kDa protein that was enriched in the $Nup153_{1-144}$ NE versus Climp63 CM comparison. It is a type-II membrane protein with a large ER luminal domain. Its function is the hydroxylation of β-carbons of specific aspartly and asparaginyl residues in noncanoncical EGDF substrates. Mutations in ASPH have been associated with various diseases such as the extremely rare Traboulsi syndrome, as well as tumorigenesis and metastasis[188]. For the differential membrane permeabilisation analysis, an antibody was used for which the exact epitope sequence is proprietary, yet the antibody was raised against a peptide comprising amino acids 250 - 350, which refers to the ER luminal part of the protein. In line with that, in the soft permeabilisation condition most cells did not show any staining for anti-ASPH (figure 4.16A). In contrast, upon permeabilisation with TX/SDS, a strong anti-ASPH signal could be observed to overlay with the Bap31 ER marker, as well as a the nuclear rim. This indicated that the protein may be present both at the INM and the ER.

4.2.8.2 ELOVL5

Elongation of very long chain fatty acid protein 5 (ELOVL5) is one of seven elongases (ELOVL1-7) in mammals[189]. Very long chain fatty acids (VLCFA) are fatty acids with more than 20 carbon atoms, which get synthesised in the ER[190]. The synthesis of fatty acid elongation consists of four steps: condensation, reduction, dehydration, followed by a second reduction step (figure 4.19). Members of the ELOVL family are thereby involved in the first and rate-limiting step in the fatty acid elongation cycle[191]. ELOVL5 specifically is involved in the elongation of C18-CoA and C20-CoA[189] and was shown to be enriched in the $Nup153_{1-144}$ NE versus Rtn4a CM comparison. It has a size of 35 kDa and 7 predicted TMH. This makes the protein small enough to diffuse into the INM. When validating the protein, a strong loss in Bap31 signal could be observed, indicating that harsh permeabilisation also affect proteins localised at the ER. This adds an extra layer of complexity in the interpretation interpreting data from differential permeabilisation experiments. Besides no strong difference could be observed between the soft and harsh permeabilization method, indicating that this method is not suitable to determine if ELOVL5 is enriched at the INM (figure 4.16B). Yet, a clear nuclear rim signal indicative of INM localisation could be observed again in addition to the ER signal, for both the harsh and soft permeabilisation conditions, indicating that the protein may be localised both in the bulk ER and the INM. Of note, besides ELOVL5, various other proteins involved in VLCFA synthesis were found at the INM. For example, HACD2 involved in the dehydration reaction of VLCFA synthesis, and TER involved in the final reduction step, were also present in the INM data sets, and will be discussed in sections 4.2.9.4 and 4.2.9.5. Additionally, VLCFAs act as building block for sphingolipids and glycerophospholipids[189], and whereas synthesis of sphingolipids in the INM membrane has not yet been shown, enzymes involved in sphingolipid synthesis have been identified previously in nuclear extracts isolated from different cell types[192].

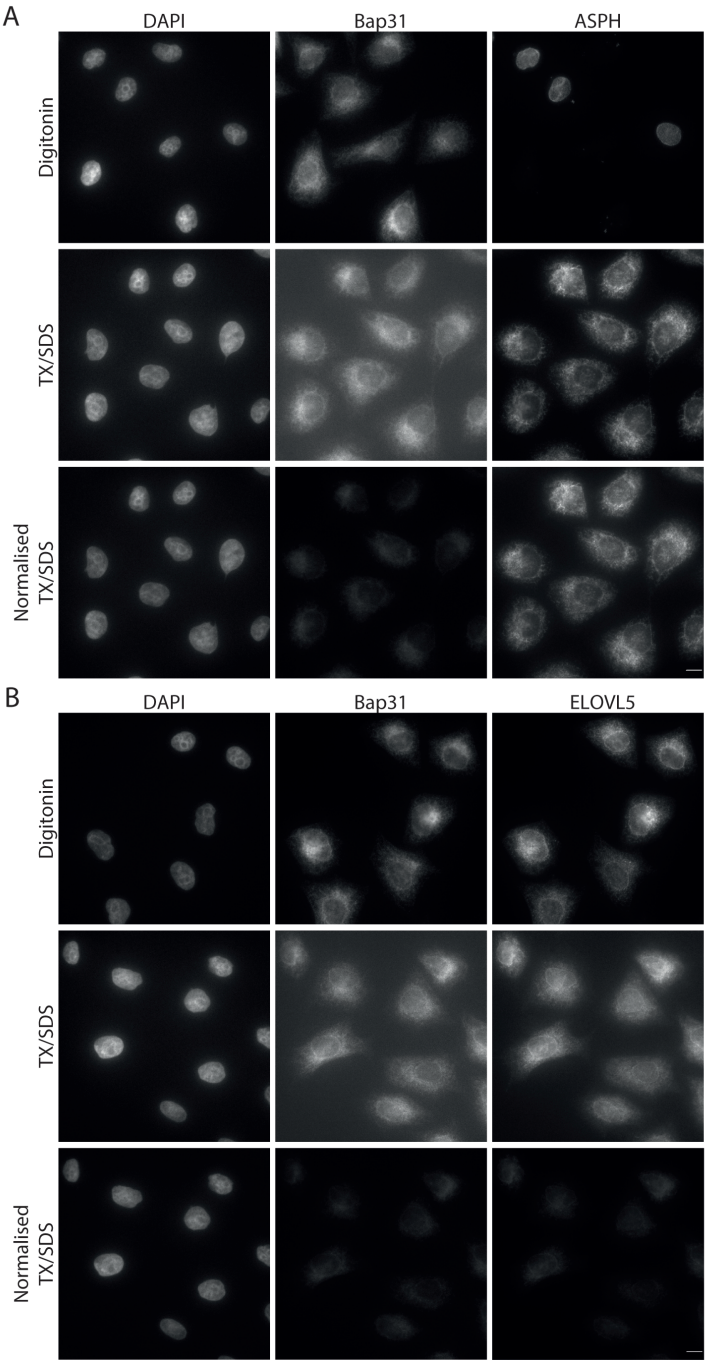


Figure 4.16: Validating the localisation of ASPH and ELOVL5. Differential membrane permeabilisation in HeLa cells. IF microscopy images of HeLa cells that were permeabilised with either digitonin alone or digitonin followed by Triton X-100/SDS (TX/SDS). Nuclei were stained using DAPI, the ER was stained using anti-Bap31 in combination with its corresponding Alexa Fluor secondary antibody. The normalised TX/SDS colimns were normalised against the average intensity of their respective digitonin permeabilisation panel. a) Anti-ASPH immunostaining was performed in combination with its corresponding Alexa Fluor secondary antibody. b) Anti-ELOVL5 immunostaining was performed in combination with its corresponding Alexa Fluor secondary antibody. Scale bar: 10 μm.

4.2.8.3 PGRMC1 and PGRCM2

Membrane-associated progesterone receptor component (PGRMC) 1 and 2, together with neudesin (NENF) and neuferricin (CYB5D2), form a group of membrane-associated progesterone receptors (MAPR). MAPRs share a homologous CYB5-like heme/steroid-binding domain. PGRCM1 and 2 have a mass of 22 and 24 kDa, respectively. Both are predicted to have one TMD in proximity of their N-terminus. Additionally, interaction data by colocalisation, pull-down assays and *in situ* PDB indicates that PGRCM1 and PGRCM2 form a complex[193].

Most IF studies have found PGRMC1 localised at the ER and mitochondria[194], but there is also evidence suggesting the protein also localises to the nucleus[195]. In line with this idea, PGRCM1 was shown to be enriched in our $Nup153_{1-144}$ NE versus Rtn4a CM comparison. PGRMC1 is involved in a wide-range of functions including cholesterol/steroid biosynthesis and metabolism, iron homeostasis and heme trafficking, as well as promotion of autophagy, regulation of cell cycle progression and cell proliferation, as summarised by a review from Ryu *et al.*[196].

PGRMC2 has been less well studied, but sequence similarity analysis demonstrates a 64.9% homology with PGRMC1. While not considered significantly enriched at the INM according to any individual pairwise comparison, PGRMC2 did get identified in the top 50 hits most similar to both emerin and SUN2 using the profile plot tool (chapter 4.2.4). Interestingly, imaging data from Parker *et al.* also showed a strong nuclear rim[197] for PGRMC2 and identified the protein as an interaction partner of the nucleoporin ALADIN[198]. Coherent with this localisation, it was shown that both PGRMC1 and PGRMC2, play a role in cell cycle regulation. Depletion of either of the two proteins in rat granulosa cells resulted in increased mitotic entry, where cells then accumulate in metaphase and subsequently underwent apoptosis. Conversely, overexpression of either of the two proteins led to inhibition of cells entering mitosis[193].

When validating whether PGRCM1 and PGRCM2 actually enrich to the INM, no clear change in localisation could be observed for PGRCM1 upon harsh permeabilisation. Yet, a clear nuclear rim could be observed both in the soft

and the harsh permeabilisation condition, indicative of INM localisation (figure 4.17A). The signal for the PGRCM2 antibody was quite faint in the digitonin condition, however upon addition of TX/SDS this was completely lost (figure 4.17B). Yet, again a nuclear rim can be observed for the majority of the cells in soft permeabilisation conditions.

Again, whereas the differential permeabilisation did not result in a clear answer on whether PGRMC1 and PGRMC2 localise to the INM, the IF signal at the nuclear rim is indicating that the proteins are present at the NE. Additionally, previous localisation data of PGRMC1, as well as its interaction with the nucleoporin ALADIN suggests localisation in the INM. Lastly, localisation at the INM would also be in line with the findings that PGRMC1 and PGRMC2 play a role in cell cycle progression.

4.2.8.4 TMEM109

TMEM109 was shown to be enriched at the INM by both the $Nup153_{1-144}$ NE versus Rtn4a CM, as well as the $Nup153_{1-144}$ NE versus Climp63 CM comparison. The protein, also known as Mitsugumin-23, is poorly characterised. It is 26 kDa in size and has 5 predicted TMDs[147,149]. First identified in rabbit cells, TMEM109 was shown to be present in a wide-range of different tissues and IF microscopy showed its localisation in the sacroplasmic reticulum membrane, the ER and along the nuclear rim[199]. Further studies showed that TMEM109 forms a large bowl-shaped complex, with a putative central pore that might act as a Zn^{2+} voltage-dependent, cation-conducting channel[200,201]. Additionally, knockdown of TMEM109 increased the sensitivity of cells towards ultraviolet C, triggering cell death through an unknown mechanism[202]. Little data is available for the human protein, which shares of 88% sequence similarity with the rabbit (*Oryctolagus cuniculus*) homolog. IF microscopy was used to validate the presence of TMEM109 at the INM. Unfortunately, while the TMEM109 antibody worked well in the soft permeabilisation conditions, the signal was completely lost upon permeabilisation

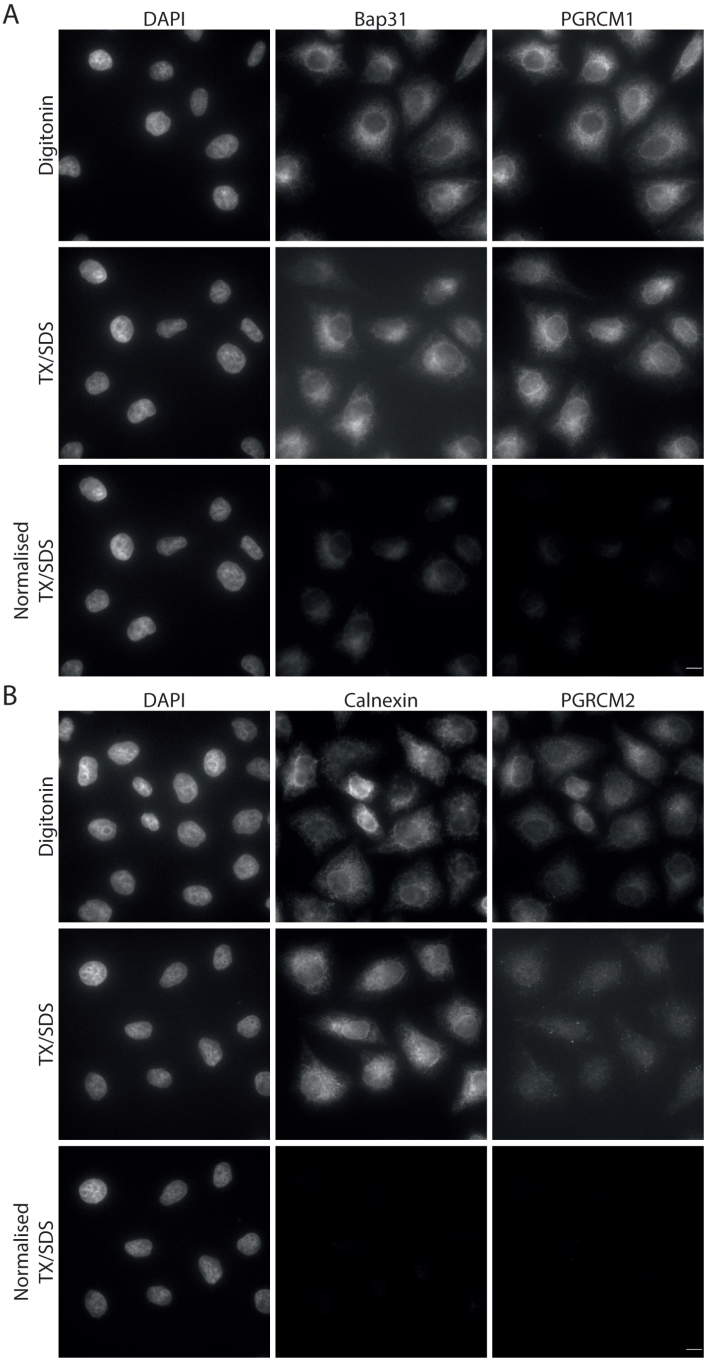


Figure 4.17: Validating the localisation of PGRMC1 and PGRMC2. Differential membrane permeabilisation in HeLa cells. IF microscopy images of HeLa cells that were permeabilised with either digitonin alone or digitonin followed by Triton X-100/SDS (TX/SDS). Nuclei were stained using DAPI. The normalised TX/SDS columns were normalised against the average intensity of their respective digitonin permeabilisation panel. a) Immunostaining was performed using anti-Bap31 and anti-PGRMC1 in combination with their corresponding Alexa Fluor secondary antibodies. b) Immunostaining was performed using anti-Bap31 and anti-PGRMC2 in combination with their corresponding Alexa Fluor secondary antibodies. Scale bar: 10 μm.

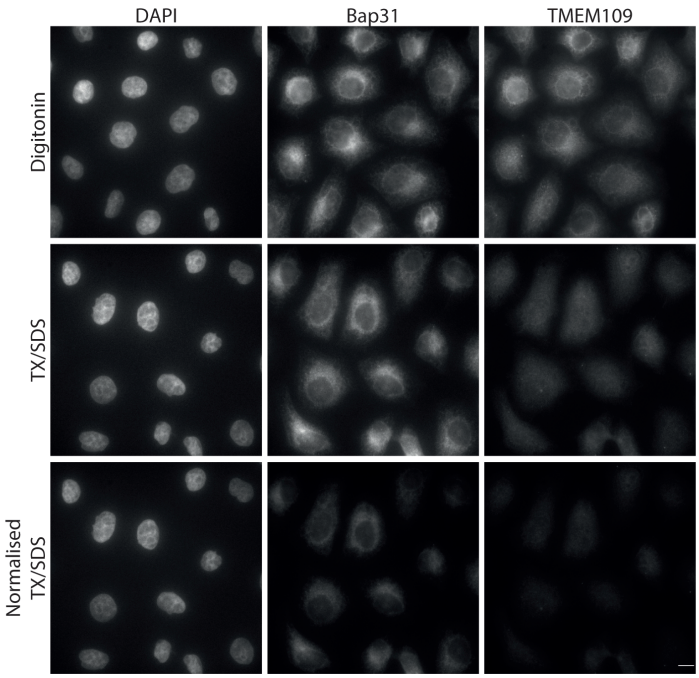


Figure 4.18: Validating the localisation of TMEM109. Differential membrane permeabilisation in HeLa cells. IF microscopy images of HeLa cells that were permeabilised with either digitonin alone or digitonin followed by Triton X-100/SDS (TX/SDS). Nuclei were stained using DAPI. The normalised TX/SDS columns were normalised against the average intensity of their respective digitonin permeabilisation panel. Immunostaining was performed using anti-Bap31 and anti-TMEM109 in combination with their corresponding Alexa Fluor secondary antibodies. Scale bar: 10 μm.

with SDS/TX (figure 4.18). Yet, a clear signal around the nucleus can be observed in soft permeabilisation conditions, indicative of INM localisation.

4.2.9 Additional interesting putative INM proteins

Similarly to the proteins validated in the previous section (chapter 4.2.6), the putative INM proteins discussed in this chapter were identified to be enriched at the INM by one (or several) of the pairwise comparisons or due to similarity to SUN2 or emerin behaviour across the various comparisons, as assessed using the profile plot tool. Yet, due to time constraints or the lack of good antibodies available for IF microscopy, no attempts were made to validate these proteins further. Yet, they are proteins that would be interesting to follow up on in the future.

4.2.9.1 CERS6

To date, 6 mammalian ceramide synthase (CerS) have been described (CerS1-6). They are involved in the synthesis of ceramide, a key intermediate in sphingolipid synthesis, which is one of the main membrane lipids in eukaryotic cells. Each CerS is highly specific towards the acyl-CoA chain length used for N-acylation, whereby CerS6 works on C14 and C16 acyl substrates[203]. Additionally, qPCR of 14 mouse tissues showed that CerS1-6 are expressed at different levels. CerS2 was expressed at the highest level in kidney cells, followed by CerS6. CerS4 and 5 seem to be expressed at low levels, whereas CerS1 and CerS3 are nearly undetectable[204]. In our proteomic data set CerS1, 2, 5 and 6 were detected, whereas CerS3 and 4 are absent. Interestingly, CerS2, which showed by far the highest mRNA expression levels in kidney cells, is identified with less peptides than CerS5 and CerS6. Of note, for neither of the CerS proteins, peptides were identified in every single replicate. Yet, there was an enrichment of CerS6 in the NE fraction, which we cannot see for any of the other CerS proteins. Intriguingly, whereas previous imaging studies of ectopically-expressed CerS proteins indicated that all CerS proteins localise to the ER, another another imaging based study in human SW480 cells demonstrated a strong enrichment of FLAG-CerS6 in perinuclear region of the ER[203,205]. Additionally, the authors of the paper show that RNAi knockdown of CerS6 results in a specific and significant decrease of C_{16}-ceramide, which inhibits tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-induced apoptosis. Curiously, whereas caspase-3 was activated normally in cells with reduced levels of CerS6, caspase-3 failed to be translocated to the nucleus. The combined perinuclear localisation and the effect of CerS6 knockdown on caspase-3 nuclear translocation, leads to the speculation that CerS6 might be involved in the regulation of the nuclear permeability. This functional data would be in line with our findings of CerS6 being enriched in the INM. Yet, due to the lack of specific antibodies we were unable to carry out any validation steps to unequivocally validate the localisation of CerS6 at the INM.

4.2.9.2 MLF2

Myeloid leukemia factor 2 (MLF2) is a 28 kDa protein, that showed enrichment at the INM in both the $Nup153_{1-144}$ NE versus Rtn4a CM, as well as the $Nup153_{1-144}$ NE versus Climp63 CM comparison. Whereas it is predicted to contain one TMD by TMHMM2.0[149], experimental data indicated that MLF2 is a soluble protein that enriches in the lumen of blebs of the NE[206]. Interestingly, cells with a Torsin ATPase deficiency, often display NE blebs, containing FG-nucleoporins and ubiquitin. NE bleb formation is observed in a cell cycle dependent manner, with blebs appearing rapidly and synchronously after NE reformation after mitosis. Whereas our proteomic data set is in asynchronous cells, this data is still generally in line with an enrichment of MLF2 in the NE fraction.

4.2.9.3 ARL6IP6

ADP-ribosylation factor-like protein 6-interacting protein 6, also known as ARL6IP6, is a 24 kDa protein with three predicted TMHs that was enriched in the $Nup153_{1-144}$ NE sample versus both Rtn4a and Climp63 CM comparison. Interestingly, it has been identified in another recent proteomic study as an INM protein[52]. Protein localisation was subsequently validated using lentiviral transduction of a V5-ARL6IP6. Thereby V5-ARL6IP6 was shown to be concentrated strongly at the NE, especially in densely grown cultures. However, at lower cell density V5-ARL6IP6 seemed to be more spread throughout the whole ER in C3H cells, U2OS cells, as well as differentiated adipocytes and myocytes. Interestingly, existing interaction data also indicates a putative interaction of ARL6IP6 with CHMP4A[207], a subunit of the ESCRT-III complex involved in NE re-sealing after mitosis or NE repair after rupture.

4.2.9.4 HACD2

Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2 (HACD2) is a 28 kDa protein with 4 - 6 predicted TMDs. In mammals, there are four HACD isoforms (1-4). Their role is the dehydration of 3-hydroxyacyl-CoA to 2,3-trans-enoyl-CoA,

which is the third step in the VLCFA elongation circle (figure 4.19). HACD2, like all other HACDs, is localised to the ER, with a strong nuclear rim signal indicative of NE localisation[208]. Whereas HACD1 is known to be mainly expressed in the heart, and in lower degrees in skeletal muscles, HACD2 and HACD3 mRNA is expressed ubiquitously in all tissues[208]. Co-IP experiments indicate that HACD2 can interact with different ELOVLs, including ELOVL5, which interestingly also seems to be enriched in our data set at the INM (4.2.8.2). Our proteomic data set indicates that HACD2 is enriched at the INM. Combined with its biological function and its interaction partners, this would be an interesting protein to follow up on and validate its putative localisation at the INM more thoroughly.

4.2.9.5 TER

Very-long-chain enoyl-CoA reductase (TER) is a 36 kDa protein with four predicted TMDs. TER catalyses the final step in the VLFA elongation cycle (figure 4.19), in which it reduces *trans*-2-enoyl-CoA into acyl-CoA, which can either enter a new cycle of elongation or be processed otherwise. TER has been shown to be ubiquitously expressed in all tissues, localising to the ER[210]. Additionally to its function in VLFA synthesis, TER was also shown to play a role in the sphingosine 1-phosphate (S1P) metabolic pathway, where it is involved in the saturation step leading to the conversion of trans-2-hexadecenoyl-CoA to palmitoyl-CoA[211]. Interestingly, the yeast ortholog of TER, called TSC13, was shown to be enriched at the nucleus-vacuole junction (NVJ), where it was shown to be involved in piecemeal microautophagy of the nucleus and vesicle formation under nutrient stress condition[212,213]. Whereas sequence similarity between TER and its yeast ortholog was only 32%[214,215], this might still be an indication of functions involved at the NE. The combination of our proteomic data, together with previous reports demonstrating the localisation of the yeast ortholog TSC13 at the NVJ, as well as the fact that various other proteins of the FA elongation cycle are enriched at the INM, make this an interesting protein to follow up.

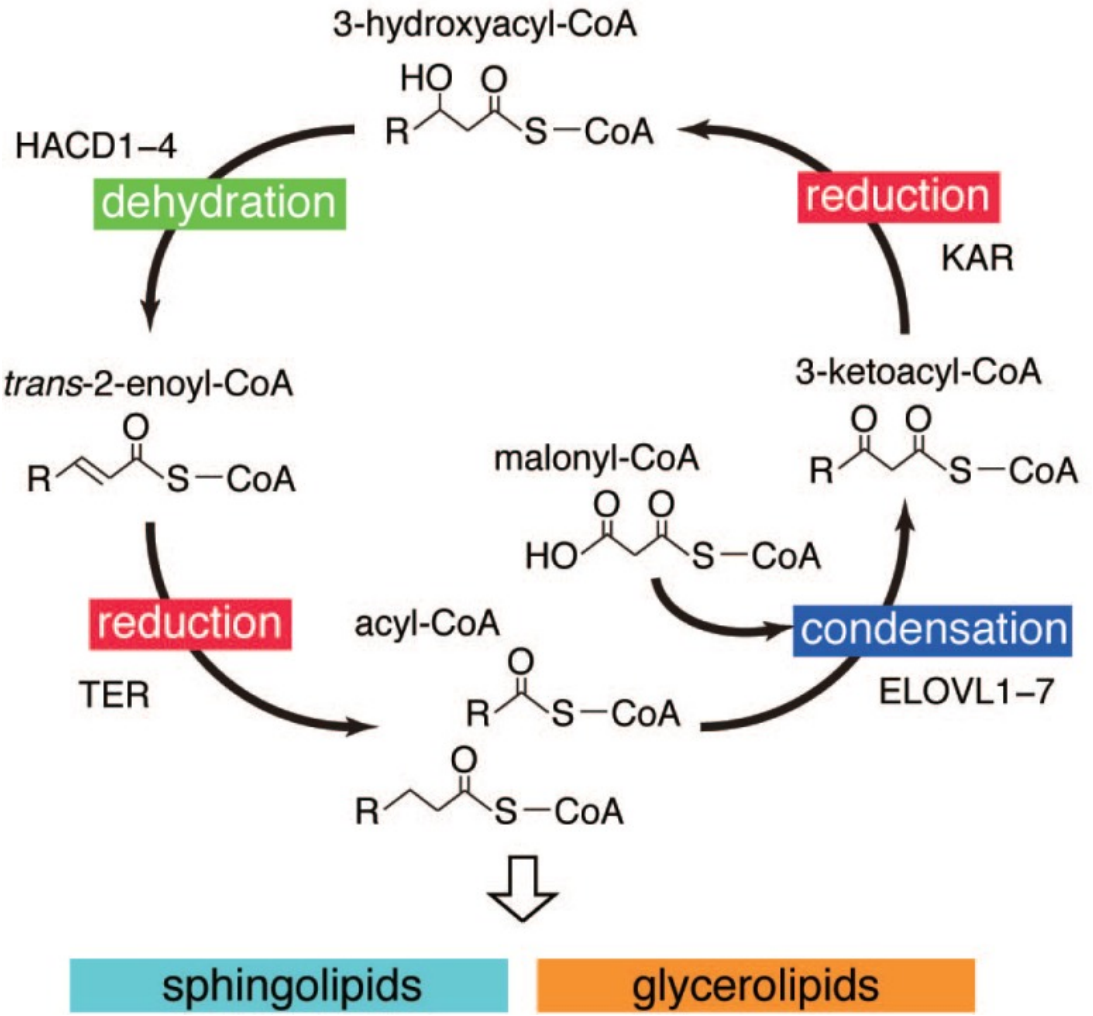


Figure 4.19: Mammalian FA elongation cycle. The FA elongation cycle consists of four steps: condensation, reduction, dehydration and reduction. The enzymes involved in each step are illustrated. Each cycle leads to the extension of the Acyl-CoA by two additional carbon units. Figure from Kihara, 2014[209].

4.2.9.6 SMPD4

Sphingomyelin phosphodiesterase 4 (SMPD4) is a 98 kDa protein, with one predicted TMD close to its C-terminus, indicative of being a C-tail anchored integral membrane protein. It is one of four neutral sphingomyelinases (nSMases) in mammals (SMPD2, SMPD3, SMPD4 and SMPD5). Yet, SMPD4 does not show any sequence homology to any of the other nSMases[214–218]. SMPD4 was shown to be highly expressed in muscle cells of humans and mice, as well as in differentiated C2C12 myotubes[219]. SMPDs are involved in catalysing shingomyelin to form

ceramide. Sphingolipids are a diverse group of lipids that play a crucial role as structural components of cell membranes, and as potent signalling molecules[220]. SMPD4 was shown to be dependent on Mg^{2+} and Mn^{2+} for activity, and is enzymatically involved in processing of sphingomyelin. Additionally, SMPD4 showed minor but significant enzymatic activity toward phosphatidylethanolamine (PE), yet no activity towards phosphatidylcholine (PC) and phosphatidylserine (PS) in HEK293 and MCF7 cells[221]. Bi-allelic loss-of-function variants of SMPD4 were found in children of 12 unrelated families, which show developmental disorders characterised by microcephaly and congenital arthogryposis[222]. To understand the molecular function of SMPD4, gene expression profiling in two of the affected individuals was undertaken. GO-term analysis showed that most significantly deregulated genes belonged to the protein secretory pathway. EM microscopy demonstrated altered ER appearance with dilated cisternae, as well as increased lysosomes and late autophagic vacuoles[222]. Interestingly, whereas expression of a HA-tagged SMPD4 showed co-localisation in HeLa cells with the ER, as well as the Golgi and *trans*-Golgi region[221], various studies have also indicated the presence of SMPD4 at the NE[51,222,223]. Indeed, new proximity ligation mapping using miniTurbo-tagged SMPD4 demonstrated biotinylation of various nucleoporins, such as NUP35, NDC1, ALADIN, NUP155 and NUP85[224]. Interestingly, these proteins are all part of the inner or outer ring, or the transmembrane nucleoporins. EM data suggests that interphase NPC insertion happens through an inside-out extrusion of the INM[225]. Additionally, ceramids are known to create packaging defects in the lipid bilayer. This opens the possibility to speculation that the presence of SMPD4 could play a role in locally increasing the concentration of ceramide to facilitate NPC insertion. Yet, how SMPD4 with its large cytoplasmic domain could reach the INM is unclear. Nevertheless, the enrichment of the SMPD4 in our INM data set, as well as the interaction data with nucleoporins and its importance in certain diseases, make SMPD4 an interesting protein to validate further.

4.2.9.7 REEP4

Receptor expression-enhancing proteins (REEPs) belong to the DP1/Yop1p family and in mammalian cells includes 6 proteins (REEP1-6). REEPs can be further divided, according to their sequence similarity, into two subfamilies, REEP1-4, and REEP5/6 (figure 4.20A & B). REEPs are abundant proteins, that localise to the ER tubules and the edges of ER sheets[7]. Indeed, they share a similar transmembrane hairpin architecture than reticulons and also play a role in ER shaping[7]. Besides their generic ER localisation, REEP1 has also been found in proximity to mitochondria, indicating that the different REEP proteins might have specialised spatial roles within the cell[226,227]. In our data set, we identified REEP2-5. Interestingly, our data indicates that REEP4, but none of the other REEP proteins, is enriched at the INM. In line with that data, REEP4 has previously been identified to play a role in proper NE architecture, and co-depletion of REEP3 and REEP4 was shown to result in defects in NE structure, namely nuclear reticula structures that contained NPCs[228]. The role of REEP3/4 was cell cycle dependent and connected to stop the attachment of ER membrane to the metaphase chromosomes. Additionally, REEP3/4 were shown to play a role in generating the highly tubular ER morphology during mitosis[229]. Recently, it was demonstrated that REEP4 plays a role in promoting nuclear pore complex formation, in particular at the end of mitosis, and that the recruitment of REEP4 is dependent on the nucleoporin ELYS[15]. Whether ELYS thereby directly interacts with REEP4, and which domain of REEP4 is required for this targeting, remains to be elucidated. The high sequence conservation in the N-terminus of the REEP proteins thereby makes it tempting to speculate that the interaction domain is located towards the C-terminus of the protein (figure 4.20C), as this is the region differing most between the REEP proteins. As NPCs disassemble during mitosis and the INM gets integrated into the bulk ER, this opens up the possibility that REEP4 plays a cell-cycle dependent dual role. During metaphase, REEP4 is involved in ensuring that the ER membrane does not associate to the chromatin and is involved in creating a high curvature,

tubule-rich ER morphology[228,229]. At the end of mitosis, REEP4 translocates to the newly forming nucleus, where it plays a vital role in introducing high membrane curvature to in the NE, enabling NPC insertion[15].

4.2.10 Validating ER tubule proteins

4.2.10.1 The RNF185-Membralin complex

One intriguing finding in the data set is that the RNF185-Membralin (MBRL) complex is enriched in the Rtn4a CM sample compared to Climp63 CM. Membralin and RNF185 were identified in our lab as components of an ubiquitin-ligase complex involved in ERAD of a subset of membrane proteins[173]. Proteomic analysis subsequently identified TMUB1 and TMUB2 as additional interaction partners. Besides being an integral part of the complex, TMUB2 levels were also shown to be controlled by RNF185 and MBRL. Additionally to degradation of the model substrate CYP51A1TM, the RNF185-MBRL complex was shown to control levels of endogenous CYP51A1. Besides the RNF185-MBRL complex itself, protein degradation is also dependent on p97, as well as the UBE2K and UBE2D3 ubiquitin-conjugating enzymes.

Curiously, our proteomic data indicates that MBRL is significantly upregulated in the ER sheets. Additionally, RNF185 and TMUB2 are also strongly upregulated in ER sheets, though they fall just below the significant threshold. Of note is that RNF5, which is more than 70% identical to RNF185[230] seems to be equally distributed between ER sheets and tubules. Additionally, no enrichment in a particular ER domain can be observed for other ERAD factors such as VCP/p97, AMFR/gp7, Nedd4, Syvn1, March5, Trim25.

As overexpression of proteins can result in changes in the protein localisation, and might also affect ER morphology depending on the function of the protein, the first attempt in validating the localisation of RNF185, MBRL, TMUB1 and TMUB2 was by IF. As none of the antibodies used in the lab against these proteins was previously used by IF, the first step was validating the antibodies by testing them in HeLa wild-type and knockout cells. Unfortunately, neither the

A

	REEP4	REEP1	REEP2	REEP3	REEP5	REEP6
sp|Q9H6H4|REEP4		63.59	52.67	55.56	33.57	28.22
sp|Q9H902|REEP1	63.59		67.34	64.95	32.86	29.57
sp|Q9BRK0|REEP2	52.67	67.34		53.85	33.57	25.77
sp|Q6NUK4|REEP3	55.56	64.95	53.85		27.14	24.54
sp|Q00765|REEP5	33.57	32.86	33.57	27.14		54.26
sp|Q96HR9|REEP6	28.22	29.37	25.77	24.54	54.26	

% identity

B



C

```
sp|Q9H6H4|REEP4_HUMAN      ------------------------------------------------MVSWMICRLVVL 12
sp|Q6NUK4|REEP3_HUMAN      ------------------------------------------------MVSWMISRAVVL 12
sp|Q9H902|REEP1_HUMAN      ------------------------------------------------MVSWIISRLVVL 12
sp|Q9BRK0|REEP2_HUMAN      ------------------------------------------------MVSWIISRLVVL 12
sp|Q00765|REEP5_HUMAN      MSAAMRERFDRFLHEKNCMTDLLAKLEAKTGVNRSFIALGVIGLVALYLVFGYGASLLCN 60
sp|Q96HR9|REEP6_HUMAN      -MDGLRQRVEHFLEQRNLVTEVLGALEAKTGVEKRYLAAGAVTLLSLYLLFGYGASLLCN 59
                                                                  ::    .  :

sp|Q9H6H4|REEP4_HUMAN      VFGMLCPAYASYKAVKTKNIREYVRWMMYWIVFALFMAAEIVTDIFISWFPFYYEIKMAF 72
sp|Q6NUK4|REEP3_HUMAN      VFGMLYPAYYSYKAVKTKNVKEYVRWMMYWIVFALYTVIETVADQTVAWFPLYYELKIAF 72
sp|Q9H902|REEP1_HUMAN      IFGTLYPAYYSYKAVKSKDIKEYVKWMMYWIIFALFTTAETFTDIFLCWFPFYYELKIAF 72
sp|Q9BRK0|REEP2_HUMAN      IFGTLYPAYSSYKAVKTKNVKEYVKWMMYWIVFAFFTTAETLTDIVLSWFPFYFELKIAF 72
sp|Q00765|REEP5_HUMAN      LIGFGYPAYISIKAIESPNKEDDTQWLTYWVVYGVFSIAEFFSDIFLSWFPFYYMLKCGF 120
sp|Q96HR9|REEP6_HUMAN      LIGFVYPAYASIKAIESPSKDDDTVWLTYWVVYALFGLAEFFSDLLLSWFPFYYVGKCAF 119
                           ::*   *** * **::: .  : . *: **:::..:   * .:*  :.***:*:  * .*

sp|Q9H6H4|REEP4_HUMAN      VLWLLSPYT-KGASLLYRKFVHPSLSRHEKEIDAYIVQAKERSYETVLSFGKRGLNIAAS 131
sp|Q6NUK4|REEP3_HUMAN      VIWLLSPYT-KGASLIYRKFLHPLLSSKEREIDDYIVQAKERGYETMVNFGRQGLNLAAT 131
sp|Q9H902|REEP1_HUMAN      VAWLLSPYT-KGSSLLYRKFVHPTLSSKEKEIDDCLVQAKDRSYDALVHFGKRGLNVAAT 131
sp|Q9BRK0|REEP2_HUMAN      VIWLLSPYT-KGSSVLYRKFVHPTLSNKEKEIDEYITQARDKSYETMMRVGKRGLNLAAN 131
sp|Q00765|REEP5_HUMAN      LLWCMAPSPSNGAELLYKRIIRPFFLKHESQMDSVVKDLKDKAKETADAITKEAKKATVN 180
sp|Q96HR9|REEP6_HUMAN      LLFCMAPRPWNGALMLYQRVVRPLFLRHHGAVDRIMNDLSGRALDAAAGITRNVLQVLAR 179
                           : : ::*   :*: ::*::.::* :  :.  :*  : :   :. ::   . :.  :  .

sp|Q9H6H4|REEP4_HUMAN      AAVQAATKSQGALAGRLRSFSMQDLRSISDAPAPAYHDPLYLEDQVSHRRPPIGYRAG-- 189
sp|Q6NUK4|REEP3_HUMAN      AAVTAAVKSQGAITERLRSFSMHDLTTIQGDEPVG-QRPYQPLPEAKKKSKPAPSESAGY 190
sp|Q9H902|REEP1_HUMAN      AAVMAASKGQGALSERLRSFSMQDLTTIRGDGAPAPS---GPPPPGSG-------RASG- 180
sp|Q9BRK0|REEP2_HUMAN      AAVTAAA--KGVLSEKLRSFSMQDLTLIRDEDALPLQRPDGRLRPSPG-------SLLDT 182
sp|Q00765|REEP5_HUMAN      -----------LLGEEK--KST---------------------------------------- 189
sp|Q96HR9|REEP6_HUMAN      SRAGITP--VAVAGPST--PLEADL------------KPSQTPQPKDK------------ 211

sp|Q9H6H4|REEP4_HUMAN      GLQDSDTEDECWSDTEA-VPRA------PARPREKPLIRSQSLRVVKRKPPVREGTSRSL 242
sp|Q6NUK4|REEP3_HUMAN      GIPLKDGDEK-TDE-EAEGPY-----SDNEMLTHKGLRRSQSMK--------------SV 229
sp|Q9H902|REEP1_HUMAN      ----KHGQPK-MSRSASESASSSGTA---------------------------------- 201
sp|Q9BRK0|REEP2_HUMAN      -IEDLGDDPA-LSLRSSTNPADSRTEASEDDMGDKAPKRAKPIKKA---PKAEPLASKTL 237
sp|Q00765|REEP5_HUMAN      ------------------------------------------------------------ 189
sp|Q96HR9|REEP6_HUMAN      ------------------------------------------------------------ 211

sp|Q9H6H4|REEP4_HUMAN      KVRTRKKTVPSDVDS-----------   257
sp|Q6NUK4|REEP3_HUMAN      KTTKGRKEVRYGSLKYKVKKRPQVYF   255
sp|Q9H902|REEP1_HUMAN      --------------------------   201
sp|Q9BRK0|REEP2_HUMAN      KTRPKKKTSGGGDSA-----------   252
sp|Q00765|REEP5_HUMAN      --------------------------   189
sp|Q96HR9|REEP6_HUMAN      --------------------------   211
```

Figure 4.20: The REEP protein family. a) Percentage identity of protein conservation between REEP family proteins. b) Phylogenetic tree of the REEP protein family. c) Sequence alignment of REEP4 with the other REEP family proteins.

TMUB antibodies, nor the MBRL antibody worked in IF microscopy (data not shown), whereas the RNF185 antibody gave a very faint signal that seemed to overlay with the ER.

Another option is visualising endogenously-tagged protein. In the lab, we have both HeLa MBRL-mNeonGreen (mNG) and MBRL-Halo cell lines. In a first preliminary IF experiment, no signal could be detected by for mNG. This did not change when trying to amplify the signal using a anti-mNG antibody, most likely due to the low levels of endogenously expressed MBRL. The next step would be to attempt visualisation using the MBRL-Halo.

The exact physiological function of the RNF185/MBRL complex is not yet understood. Yet, one might speculate that if the complex is enriched in ER tubules, that knocking it out might result in phenotypic changes in ER tubules but not sheets. To test this, the morphology of ER sheets and tubules was assessed in both wild type, as well as cells in which different components of the complex were knocked out. ER sheets were visualised using a Climp63 antibody (figure 4.21), while ER tubules were visualised using Rtn4a antibody (figure 4.22). As a generic ER marker Bap31 was used.

Unfortunately, no striking phenotype was visible when analysing the ER morphology in HeLa wt cells, compared to MBRL, RNF185 and TMUB1/2 KO cells. To check for more subtle differences, an automated image processing method quantifying the overlap between Climp63 or Rtn4a with Bap31 in the different cell lines, is still outstanding.

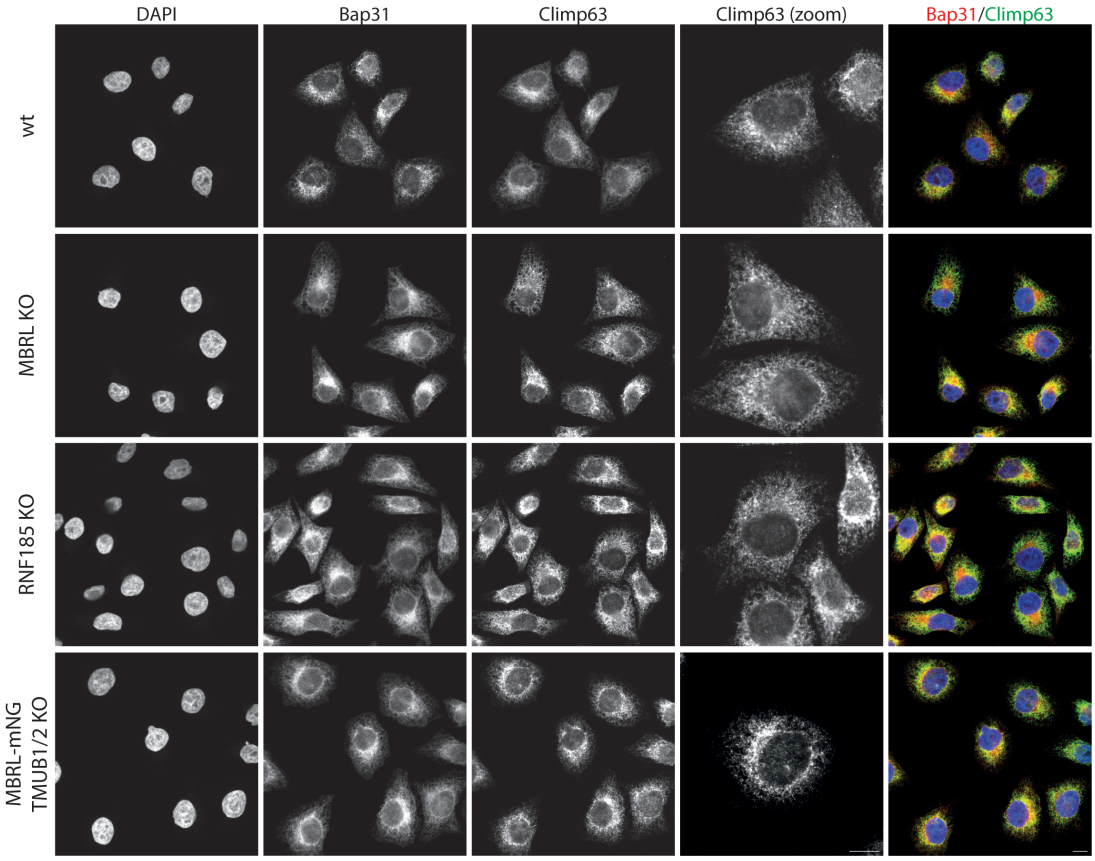


Figure 4.21: ER morphology analysis of the MBRL-RNF185 complex. IF microscopy images of HeLa cells. Anti-Bap31 and Anti-Climp63 antibodies were used in combination with their corresponding Alexa Fluor secondary antibodies for staining to investigate ER morphology. Nuclei were stained using DAPI. Scale bar: 10 μm.

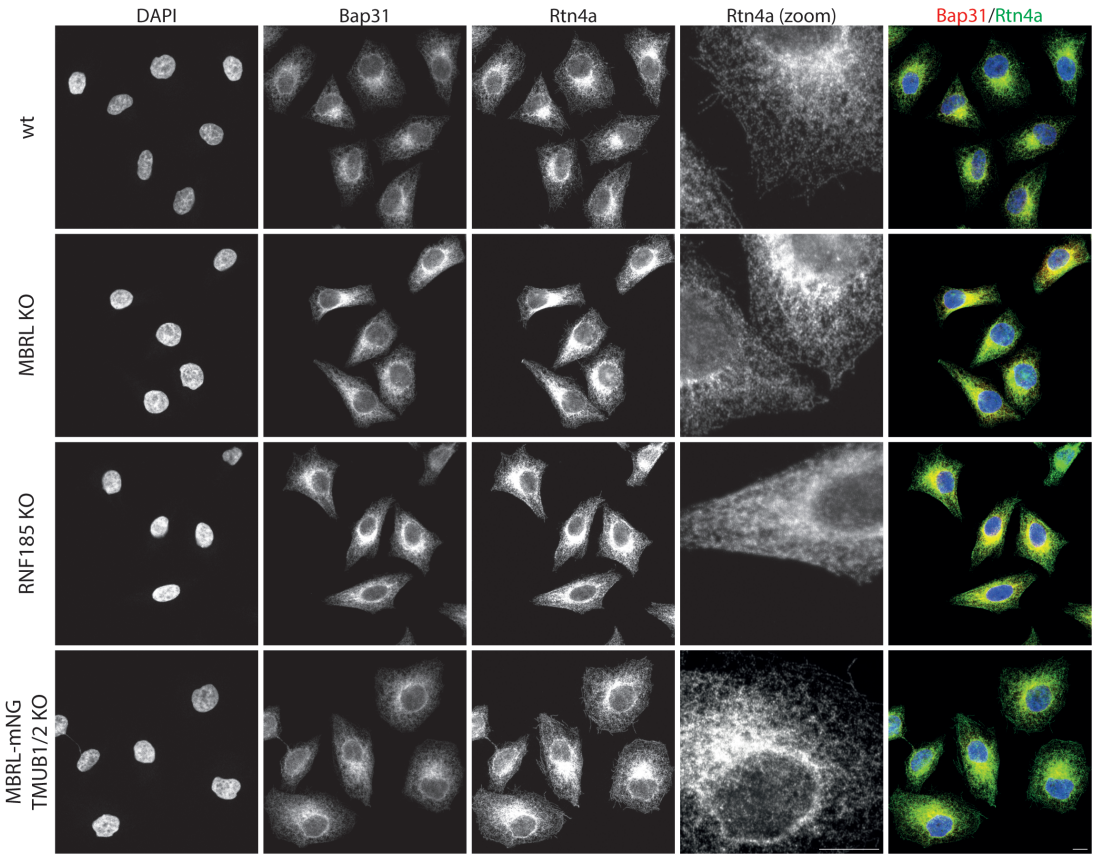


Figure 4.22: ER morphology analysis of the MBRL-RNF185 complex. IF microscopy images of HeLa cells. Anti-Bap31 and Anti-Rtn4a antibodies were used in combination with their corresponding Alexa Fluor secondary antibodies for staining to investigate ER morphology. Nuclei were stained using DAPI. Scale bar: 10 μm.

4.3 Discussion

4.3.1 Optimising a workflow to investigate the protein composition of distinct ER domains

Since the development of PDB systems for proteomic analysis, they have become a popular tool to probe different cellular environments *in situ*. Indeed, during the course of this project various studies using different PBD systems, mostly BioID, were published, investigating the NPC, the nuclear lamina and the INM[53–55,139,140,163,231]. However, these studies were based on individual proteins, rather than a combination of different baits to systematical analyses and annotated distinct ER domains coherently. To investigate the protein composition of distinct ER domains, differences in protein composition in one part of the ER have to be determined in comparison to another part of the ER. This means that each of the Climp63, $Nup153_{1-144}$, Rtn4a and Sec61β baits acts as a control for all other baits. As imaging, biochemical fractionation and proteomic data indicated that INM baits label large amounts of soluble proteins (chapter 3.3), a soluble NLS-APEX2 bait was used as an additional control. Lastly a negative control in which no biotinylation reaction was performed was used to identify proteins which non-specifically bind to the beads, as well as common contaminants.

Whereas incredibly powerful, PDB assays often lead to extremely large data sets, which are for BioID in the range of hundreds of identified proteins, while APEX2 experiments often detect thousands of proteins. Indeed, our data set to investigate ER domains using 5 different baits comprised 5126 unique proteins. The large amount of proteins identified makes data analysis extremely challenging and requires stringent filtering to create a number of candidate proteins that are possible to validate further. Additionally, working with such large amounts of identified proteins bears a high risk of false positives. This means that rigorous statistical testing will need to be applied. Another limitation of proteomic analysis is that the dynamic range in such complex data sets limits the detection of low expressed proteins, as they may me masked by abundant proteins. To gain meaningful information out of the collected proteomic data set, it is therefore

crucial to optimise a data analysis pipeline. Besides the large number of proteins identified, an additionally challenge in analysing PDB data is that bait proteins are often expressed at different levels. According to our western blots, bait expression levels are the main factor determining the amount of proteins that get biotinylated, and thus that will get identified by LC-MS/MS (chapter 3.3.6). This bears the risk that the identified proteins will be skewed towards strongly expressed baits. If feasible, expression levels of different baits should be comparable. Yet, if this is not possible, as in the case of this project, additional normalisation steps have to be integrated into the data analysis to normalise for those differences. Lastly, an additional hurdle in this particular data set was that, due to external circumstances, proteomic samples were run on different days and on different mass spectrometers, as detailed in appendix table A.1. This poses an additional risk of batch effects and therefore required carefully monitored normalisation.

To determine an optimal data analysis approach, three different methods were trialled side by side. As the INM is known to have a protein composition distinct from the bulk ER, and as there are many well-characterised INM proteins, the $Nup153_{1-144}$ NE versus Climp63 CM comparison was chosen as a surrogate to optimise data analysis. The success of each data analysis method was assessed by various criteria, including how well replicates cluster together, GO-term enrichment analysis and the division of well-characterised INM and ER proteins in their corresponding sample (chapter 4.2.2). Our analysis indicated that two approaches, a Perseus-based analysis using LFQ intensities and a R-based approach using raw intensities, worked comparably well and demonstrated good separation between the INM and ER domains. In the end, we optioned for the R-based limma analysis, as this circumvents loss of information associated to working with LFQ intensities, enables slightly better control over individual parameters during data analysis, and as this was easier to apply to a large number of pairwise-comparisons, as required for our data set.

Besides the individual pairwise-comparisons, we also developed a method of combining the FC information across various probes using the profile plot

analysis tool (figure 4.10). The assay was validated using a list of well-defined marker proteins of different ER domains, as well as of the Golgi and mitochondria. Thereby, similar behaviours could be observed across the three probes for proteins known to localise to the same compartment within cells. Additionally, determining the 49 proteins that exhibited a comparable behaviour across the three probes as either emerin or SUN2, resulted in the identification of many INM marker proteins, indicating the validity of this methodology. Besides known INM proteins, a list of putative INM proteins was identified. Of note, 70% of the proteins with the highest similarity as defined by Pearson correlation were membrane proteins, indicating the potential of this tool in probing a membranous environment. Another interesting point to make is that for this particular analysis, the fold change was always determined between the NE and CM fraction within a bait. This is helpful, as it minimises the variance in the sample, both for the sample processing as well as for the LC-MS/MS analysis, as the CM and NE fraction of proteins were always processed together. This creates an exciting method to combine information from across various baits, while minimising the impact of variability between probes. One point to note when looking at the profile plots in Perseus is, that using a fixed number of protein plots to be determined (e.g. 50), will always lead to 50 proteins being identified, even if the lower ranking ones might have quite different characteristics to the reference plot. It would therefore be an option to create a tool that automatically integrates a cut-off for the Pearson correlation, which is currently not possible in the Perseus integrated profile plot tool.

To conclude, a PDB-fractionation assay using five different bait proteins was used to collect a data set to investigate the protein composition of distinct ER domains. The data set contained over 5000 unique proteins and is a rich resource for enhancing our understanding of the protein organisation within the ER. To interpret the data, an analysis pipeline was optimised. Having this tool in place, the focus can now shift to the biological questions and the identification of proteins enriched in each ER domain. Additionally, it is of note that the sample preparation method coupled to the data analysis pipeline can now easily

be adapted to analyse protein compositions of other membrane-bound organelles, when adjusting the bait and control proteins.

4.3.2 Deciphering the proteomic composition of the ER

Considering the wide-range of functions the ER undertakes in a cell, having a complete map of its proteome would be extremely valuable. Therefore, various different approach were taken in the past two decades to solve this question. One approach to investigate the ER protein composition of cells from different organelles was differential fractionations coupled to proteomic analysis[137]. Whereas this methodology gave first insights into the protein composition of the ER of different cell types, fractionations are enrichment techniques and don't create homogenous organellar ER fractions. This is inparticular true for the ER that forms contact sites with various other organelles. This creates the difficulty of having to distinguish between putative ER proteins and potential contaminants. Additionally, whereas some intraorganellar separation in rough or the smooth ER is possible due different different densities, it is not possible to separate ER tubules from ER sheets and investigate these distinct ER domains[137].

A newly evolving tool to study the ER protein composition are PDB assays coupled to mass spectrometry analysis. Targeting APEX to the ER using an artificial ER targeting construct consisting of the N-terminal 27 amino acids of P450 oxidase 2C1 resulted in the identification of 5270 proteins[138]. After applying SILAC cut-offs, this resulted in a putative ER proteome containing 634 proteins, including many well characterised ER proteins. Whereas this is an interesting data set, this study does not differentiate between different ER domains such as ER tubules and ER sheets.

BioID data from more physiological constructs were obtained as part of an effort that created 234 BioID bait proteins to map the complete human proteome in FlpIn293 cells[139]. After quality control for expression and localisation, data sets were collected for 192 baits covering 32 different subcellular compartments. Of these, 30 baits were ER membrane localised and another 7 were ER luminal.

Interestingly, this included some of the proteins which have been shown to be enriched in the ER tubules (REEP5), as well as in the ER sheets (CKAP4). And indeed, whereas many interaction partners were shared between the two constructs, CKAP4 had stronger contacts with proteins previously associated to ER sheets (e.g. Sec63) and ONM proteins, whereas REEP5 showed stronger contacts with proteins associated to ER tubules such as RTN4 and RTN3[139].

Besides proteomic studies, imaging studies looking at endogenous proteins and overexpressed tagged proteins were used to obtain a better understanding of ER organisation. It thereby became apparent that certain ER proteins preferentially localise to ER tubules, whereas others are enriched in ER sheets[12,23]. Recent advances in imaging technologies, as well as assays such as bimolecular fluorescence complementation (BiFC), also gave an insight into ER contact sites, indicating that different proteins are enriched at contact sites with different organelles[2].

Interstingly, none of the proteomic studies was designed to investigate differences in the protein composition of the ER tubules versus the ER sheets. Instead, all studies determined a generic ER proteome. In contrast, imaging studies were set out to specifically investigate whether distinct proteins may enrich in a distinct ER domain. Yet, these imaging studies only follow individual proteins and don't give a global overview of protein composition. This is different from the data set collected here using the Climp63, Rtn4a and Sec61β baits, making this data set a novelty.

4.3.3 Proteomic characterisation of the INM

Determining the INM proteome is intrinsically difficult due to the continuity of the INM with the rest of the ER. The first INM proteome was determined from rat liver cells using subtractive proteomics[51]. For this approach, subcellular fractionation was used to enrich either the nuclei or microsomal membranes. Subsequently samples from the microsomal membranes and the nuclei were analysed independently by multidimensional protein identification analysis (MudPIT). Proteins identified in the microsomal membrane fraction were subtracted from

the nuclei fraction, resulting in the identification 13 known NE proteins, as well as 67 uncharacterised open reading frames with predicted TMDs. Validation of 8 of these proteins showed all of them localising to the INM.

A comparison of the INM proteins identified by Schirmer *et al.*[51] with our INM data set obtained from the $Nup153_{1-144}$ NE versus Climp63 CM pairwise-comparison (chapter 4.2.3) showed a surprisingly low overlap of 3.7%. This was mostly driven as our data set contain a five-fold higher amount of putative INM proteins. Reasons for this could be the extensive progress made in mass spectrometry over the past two decades, as well as the fact that our data analysis method does not discard proteins that are present in both the INM and the ER fraction. Additionally, two different methodologies were used to obtain the data sets, with Schirmer *et al.* using biochemical fractionation only, whereas our data set was obtained using a combination of PDB and fractionation. Lastly, the use of two different cell types could be a driver leading to further differences.

This would be in line with findings from a study by Cheng *et al.* who determine the INM proteome of murine mesenchymal stem cells and its differentiated adipocytes and myocytes by fractionation coupled to proteomic analysis[52]. Besides identifying and validating a range of interesting *bona fide* INM proteins, one important finding of the study was that the overlap of unique proteins identified between the three cell types was only ∼60%, indicating that significant variance may exist in the INM protein composition of different cell types.

Whereas subcellular fractionation coupled to LC-MS/MS was successfully applied in these publications to identify many new INM proteins, there are various caveats associated with the technique. Firstly, using this methodology, proteins may be overlooked, that are present both in the INM and the bulk ER, resulting in false negatives. Additionally, fractionation is unable to separate the INM and ONM.

To circumvent the disadvantages associated with biochemical fractionation assays, a whole range of papers was published using different PDB assays to investigate the INM, the NPC or the nuclear lamina and obtain a more complete

INM proteome[53–55,139,140,163,231,232]. The first paper using this methodology used BioID-tagging of various NPC components in FlpIn293 cells to study NPC architecture[127]. As BioID has a small radius of 10 nm, it is well-suited to identify direct interaction partners and indeed identified mostly nucleoporins. BioID was also used to investigate the difference between the two isoforms lamin A and lamin C in LMNA$^{-/-}$ mouse adult fibroblasts cells[233], as well as lamin A, SUN2 and the INM using a minimal INM targeting motif (Nup153$_{1-144}$)[53], identifying a range of interaction partners.

Another approach to investigate the nuclear lamina, with a particular focus on lamina-associated domains (LADs) of DNA, was the use of a BioID fused to Lap2β and the design of a chromatin-directed m6A-tracer BioID[140]. This is an interesting methodology to study the INM proteome, but also its interplay with the DNA. Interestingly, this study compared the proteome identified by their Lap2β-BioID construct with the three previously mentioned BioID data sets. Surprisingly the overlap between proteins identified and considered significant by the previous studies and their own Lap2β and chromatin-directed BioID data set was just around 2%. This discrepancy will arise out of a combination of factors, including technical factors such as differences in sample preparation and methodology, as well as biological differences, such as different lamin domains. Comparing the BioID-Lap2β data set with our own INM proteome showed a similarly low overlap of just 2%. Interestingly, the overlap of our data set with the m6A-tracer bait, which targets the BioID to lamina-associated domains, was much higher at around 6.9%.

A global BioID-based map of the human cell also contained 38 nuclear baits targeting to chromatin, the NE, the nuclear lamina, the nuclear body, the nucleolus, the nucleoplasm and NPCs[139]. Whereas this is an incredible effort in mapping the nucleus of cells, unfortunately there are only three baits of the NE: EMD, SYNE3 and TMPO. Emerin is known to also be present in the ER, something reflected in the BioID labelling data obtained here. SYNE3 is known

to localises to the ONM. TMPO in contrast is an interesting INM protein and identifies many of the well-characterised INM proteins.

Additionally to probing human and murine cell INM protein composition, BioID has also been used to investigate the INM of *Arabidopsis* plant cells. Thereby bait proteins were created using various nucleoporins, as well as the conserved SUN1 protein, and NEAP1, a protein without known human orthologues. This resulted in the identification of some well characterised INM proteins, and also identified 200 *bona fida* plant NETs (PNETs)[234]. Interestingly, among the proteins identified at the INM, were some plant ubiquitin regulatory X (UBX) domain containing (PUX) proteins, that seem to be critical regulators of the INMAD pathway in *Arabidopsis*[235], showing that BioID targeting to the INM can result in the identification of INMAD components.

Whereas the majority of research to investigate the INM proteome used BioID-based PDB, there are also some studies using APEX2 and BASU, a BirA* based system derived from *Bacilluar subtilis*[236].

Firstly, an APEX2-lamin-B1 fusion protein was used to investigate the nuclear lamina in FlpIn293 cells and its associated LADs[163].Thereby the APEX system was used to map proteins, RNA and DNA. This resulted in the identification of many well-charactised INM proteins. Additionally the RNA and DNA data indicates that the nuclear lamina interactome has a bias for long 3' UTR RNAs. Interestingly this paper also reiterates that APEX is active in fixed cells. The authors make use of this, by fixing cells and using FACS to sort for cells at different stages during the cell cycle. Post-sorting, the APEX reaction is then undertaken, showing an interesting approach in using APEX2-PDB in synchronised, fixed cells. This study also determined the overlap between their own nuclear lamina proteomic data set, compared data sets obtained in other studies. Thereby great variability was found between data sets, with overlap ranging from as little as 3% to a maximum of 23%. Determining the overlap of our own INM data set showed an overlap of 14.4%, which is the second highest of all the studies. This indicates a good reliability of our data set.

Another methodology developed for PDB is biotinylation by antibody recognition (BAR)[232]. This method was used to investigate the interactome of lamin A/C and resulted in the identification of over 500 putative INM proteins. The overlap with our own INM proteome was thereby 12.2%.

Another interesting study combined BASU-PDB labelling with chemical cross linking to investigate the INM proteome. For this, the authors created HEK293 cell lines stably expressing BASU-LAP1 proteins. Biotinylation reactions were performed, cells lysed and proteins captured on streptavidin beads. On-beads DSS-dependent cross-liking was performed. Proteins were then digested and peptides identified by LC-MS/MS. Comparison between the PDB-LC-MS/MS data with the PDB-XL-LC-MS/MS data showed that cross-linking increased the number of well-characterised INM proteins identified in the data set. Cross-linking data also showed the strong interconnectivity of the proteins working at the INM. Indeed, all but four INM proteins identified (SUN2, LEMD2, Zmpste24 and nurim) were interconnected either directly, or indirectly, with one another among the INM.

Lastly, two studies used APEX2-PDB in combination with rapamycin-dependent targeting and SILAC labelling (RAPIDS: rapamycin- and APEX-dependent identification of proteins by SILAC) to investigate the cellular environments of emerin and VAPB respectively[54,55]. Thereby emerin and VAPB were created as mCherry-FRB fusion proteins and APEX2-GFP-FKBP12 was either expressed as a cytosolic control or in addition with a NLS signal. In the case of VAPB, this enabled probing the VAPB binding partners specifically at the INM, even though only a small part of the total cellular pool of VAPB localised to the INM. Interestingly, the top identification partners proteins were EMD and TMEM43, which in other data sets had the lowest enrichment score of any of the known INM proteins in terms of INM enrichment[52]. Yet, whereas this is a very interesting methodology, both studies only identified a very low number of putative INM proteins (9 and 11 significant proteins, respectively).

4.3.4 Validating putative INM, ER sheet and ER tubule proteins

Besides the well-characterised marker proteins, our assay pinpoints to many additional proteins that might be enriched at the INM, ER tubules or ER sheets. The next step after identifying these proteins is to validate some of them. As expected, a stronger enrichment was observed between the INM and the bulk ER, whereas differences between ER tubules and sheets were more subtle, and therefore the focus for validating proteins was put on *bona fide* INM proteins. Yet, validating INM proteins comes with its own challenges, as the close proximity of INM and ONM makes it impossible to determine which of the two membranes a protein localises to by conventional microscopy. Additionally, both tagging and overexpression of INM proteins have been shown to result in mislocalisation. Nevertheless, a range of methods exist for validating INM localisation, as discussed in chapter 4.2.6. As the majority of these techniques rely on tagging and/or overexpression, the method of choice for validating *bona fide* INM protein localisation was a differential membrane permeabilisation approach. Differential permeabilisation either just permeabilises the plasma membrane under soft permeabilisation conditions, or the plasma membrane as well as internal organellar membranes using harsher permeabilisation conditions. Yet, various issues were encountered during the validation. Firstly, we were unable to establish a differential permeabilisation assay in the FlpIn293 bait cell lines. Instead, we optioned for a HeLa cell line for validation, building on the assumption that INM protein localisation is comparable between these two human cell lines. Yet, a change of cell lines always makes negative results more challenging to interpret. Additionally, whereas striking differences could be seen between the harsh and soft permeabilisation in HeLa cells for the two control proteins LBR and lamin A/C (figure 4.13), no striking differences could be observed between harsh and soft permeabilisation conditions for any of the *bona fide* INM proteins. Yet, a clear nuclear rim was visible for most putative INM proteins. This is indicative, but not unequivocal, of INM localisation[23]. Therefore, additional validation would

be required. The next imminent step would be to perform quantification on the differential permeabilisation images obtained, to pick up more subtle differences. If unsuccessful, the final step would be to trial some of the other techniques mentioned in chapter 4.2.6, such as BiFC assays, to validate INM localisation.

Concluding, the IF data of the *bona fide* INM proteins showed an overlay with the ER marker, as well as a clear nuclear rim signal. This indicates that the proteins localise both to the ER, as well as the INM, which is in line with the proteomic data. Our proteomic data is further backed up by a number of recent, independent studies that identified that various of the *bona fide* INM proteins have a small cellular pool at the INM. Together, these data support the findings of our proteomic study and suggest that the *bona fide* INM proteins identified by proteomic analysis are true finds and localise to the INM. Yet, another validation method is needed to unambiguously show that the candidate proteins do localise to the INM. This would be the next follow up step for this project.

4.3.5 Interesting observations to follow up

Many proteins shown to be enriched in the INM, ER tubules and ER sheets were discussed in chapters 4.2.5 and 4.2.8, and attempts were made to validate these results. Besides validating the localisation of individual proteins further, it would also be intriguing to investigate whether their enrichment in a particular ER domain is connected to a (patho-)physiological function. As many proteins require specific binding partners or play a role in a more elaborate pathway, we investigated whether certain protein complexes or pathways were upregulated in certain ER domains. Thereby two compelling observations were made.

Firstly, three out of four enzymes involved in the elongation cycle of FA are significantly enriched in the INM. The one protein not shown to be significantly enriched in the INM is KAR (Very-long-chain 3-oxoacyl-CoA reductase). Yet, this protein was still identified and enriched in the INM, it just falls short of the significance threshold. VLCFA can be build into sphingolipids, which have previously been described to play important signaling and regulatory roles in

nuclei[237]. Additionally, they act as a building block for ceramide, and CerS6, one of the proteins involved in this process, has also been identified to be enriched in the INM data set. Interestingly, there is evidence in *S. cerevisiae* linking VLCFA synthesis to the structure and function of the nuclear membrane-pore complex[238]. Tomography electron microscopy images show that *acc1-7-1* mutants, which are impaired in VLCFA synthesis, have an aberrant morphology, which generally seems to be caused by disturbances of the INM[238]. Additionally, *acc1-7-1* often have clusters of vesicles adjacent to NPCs. A follow up paper showed that VLCFA supplementation to *acc1*Δ cells did not lead to a rescue of lethality[239]. Additionally, treatment with cerulenin, an inhibitor of FA synthetase in WT cells, did not result to comparable NE perturbations as seen in the acc1^{ts} mutant. Indeed, the effect of NE morphology seems not to be related to ceramide or GPI-anchor synthesis, but might be due to a C_{26}-substituted PI lipid and its properties in stabilising highly curved membrane structures[239]. Curiously, the C_{26}-substituted PI only accounts for $\sim$1% of all the PI species in the cells. Whereas these overall levels are very low, one might speculate that synthesis at the INM close to NPCs could lead to locally increased levels, that may be sufficiently high to exert stabilising functions on the highly curved pore membrane. Together this data indicated that local VLCFA synthesis might be important for NPC insertion in yeast due to membrane remodelling. Taken this into account, our PDB biotinylation data could be a first indication that VLCFA synthesis in mammals could take place locally at the INM. Yet, further analysis is required to determine if depletion of VLCFA also leads to aberrant NPC membrane phenotypes in mammalian cells. Additionally, further experiments would be required to validate the localisation of the enzymes involved in VLCFA synthesis at the INM and to determine whether proteins are functionally active at the INM.

The second observation that would be intriguing to investigate further was that RNF185, MBRL and TMUB2 showed strong enrichment in the Rtn4a sample, indicating their preferential localisation to ER tubules. These proteins, together with TMUB1, form an ubiquitin-ligase complex involved in ERAD. The lack

of TMUB1 enrichment in tubules may be explained as TMUB1/2 are essential but redundant. Therefore, even in absence of TMUB1 the complex would be functional. To date, no ubiquitin ligase complex is known to be enriched in ER sheets or tubules. In line with that, our proteomic data indicated that other ERAD ligases such as MARCH5 or AMFR, are equally identified by both the Rtn4a and the Climp63 bait. Curiously, enrichment in ER tubules was also not found for RNF5, which is a paralogue of RNF185, with more than 70% sequence similarity. The next step to follow up on these observations would be to validate the localisation of the RNF185-MBRL complex by IF microscopy. As no good antibodies are available for these proteins, cell lines in which components of the complex are endogenously tagged would need to be generated. Expressing proteins at endogenous levels may result in difficulties in their detection by IF. Therefore, the choice of tag is crucial and should be suited to visualise low expressing proteins, such as the HaloTag or GFP, for which the GFP-booster technology is available. Another point to take into consideration is that fixing cells can result in the distortion of the elaborate structure of the ER and result in artefacts, as discussed in chapter 3.4.1. One possibility to circumvent this, would be to do set up live-cell imaging experiments. A further advantage of live-cell imaging would be that this would enable the investigation of protein localisation in the context of the cell cycle or external stimuli, which is of interest, due to the highly dynamic nature of the ER. Functional assays would be more difficult to pursue. One hypothesis was that if the complex is enriched in ER tubules, a knockout might result in phenotypic changes in the ER tubules. Yet, IF experiments using ER tubules and ER sheet marker proteins did not support this hypothesis and no changes could be observed in ER morphology.

4.3.6 Concluding remarks

To conclude, LC-MS/MS analysis was successfully carried out for NE and CM fractions of five APEX2-bait proteins and an extensive data set was obtained to enhance our understanding of ER organisation. A R-based data analysis

pipeline was optimised that entails VSN, LCMD imputation and limma analysis. The optimised data analysis pipeline was subsequently used to identify proteins enriched in the INM, ER tubules or ER sheets and a list of *bona fide* proteins for each of the domains was curated. Differential permeabilisation experiments coupled to IF microscopy was chosen to validate localisation of these proteins. Thereby most results were indicative of the expected localisation, yet more experiments are needed to be certain. How this assay can be used to determine changes in ER protein composition under different conditions will be discussed in the next chapter.

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

— Marie Curie

5

Investigating the INM protein composition in disease phenotypes

Contents

5.1 Rational 154
5.2 Results 156
5.2.1 Generating *LMNA* and *ZMPSTE24* KO cell lines . . . 156
5.2.2 Validating *LMNA* and *ZMPSTE24* KO cell lines by WB 156
5.2.3 Validating PDB localisation in *LMNA* KO cells using IF 157
5.2.4 Validating subcellular fractionation of nuclei from *LMNA KO* cells 158
5.2.5 Proteomic analysis of *LMNA* KO cells 159
5.3 Discussion 165
5.3.1 Adapting our PDB assay to study laminopathies . . . 165
5.3.2 Depletion of A-type lamins does not result in global changes of the INM protein composition 166

5.1 Rational

Lamins are type V intermediate filament proteins that assemble into a meshwork that lines the nucleoplasmic side of the INM[43]. Whereas lamins do not possess any enzymatic activity, they act as structural support of the NE. Additionally, lamins were shown to play a role in chromatin organisation, as well as in influencing transcription and replication[43]. Nuclear lamins fall into two classes: A-type and B-type lamins. A-type lamins include lamin A and lamin C, which are alternative splice variants of the *LMNA* gene, that share the same 566 amino acid N-terminus, but differ in their C-terminus[104]. Lamin A and lamin C1 are ubiquitously expressed in somatic cells, whereas lamin C2 is found in the testis[43]. Soon after synthesis, lamin A undergoes complex post-translational processing (figure 5.1). Thereby, the N-terminus of pre-lamin A gets farnesylated at the CaaX motif. The three residues of the farnesylated tail then get proteolytically removed by the ZMPSTE24 protease. Subsequently, the C-terminal tail gets methylated by ICMT before being cleaved off again in a final ZMPSTE24-mediated cleavage step[43,104].

Astonishingly, there are more than 70 different *LMNA* mutations linked to over a dozen skeletal and cardiac muscle diseases, including the autosomal dominant form of Emery-Dreifuss muscular dystrophy (EMDM), Limb-Grindle Muscular dystrophy 1B and HGPS[101,102]. Among the 70 mutations are various nonsense and missense mutations, which lead to severe depletion or complete absence of lamin A/C[240,241]. Mouse models played a crucial role in enhancing our mechanistic understanding of these diseases. Mice experiment showed that LMNA$^{-/-}$ mice are born without any visible abnormalities, but develop postnatal growth defects and muscular dystrophies[112]. On a cellular level, mouse embryonic fibroblast (MEF) cells of LMNA$^{-/-}$ mice demonstrated NE perturbations and their nuclei were less contractile than their WT counterparts[108,112]. Additionally, *LMNA* knockout (KO) was shown to affect both the targeting rate of certain INM proteins, as well as the protein enrichment of specific NETs at the INM at steady-state[70,178].

Besides mutations resulting in the absence or depletion of A-type lamins, a frequent mutation in the *LMNA* gene associated with the HGPS syndrome is

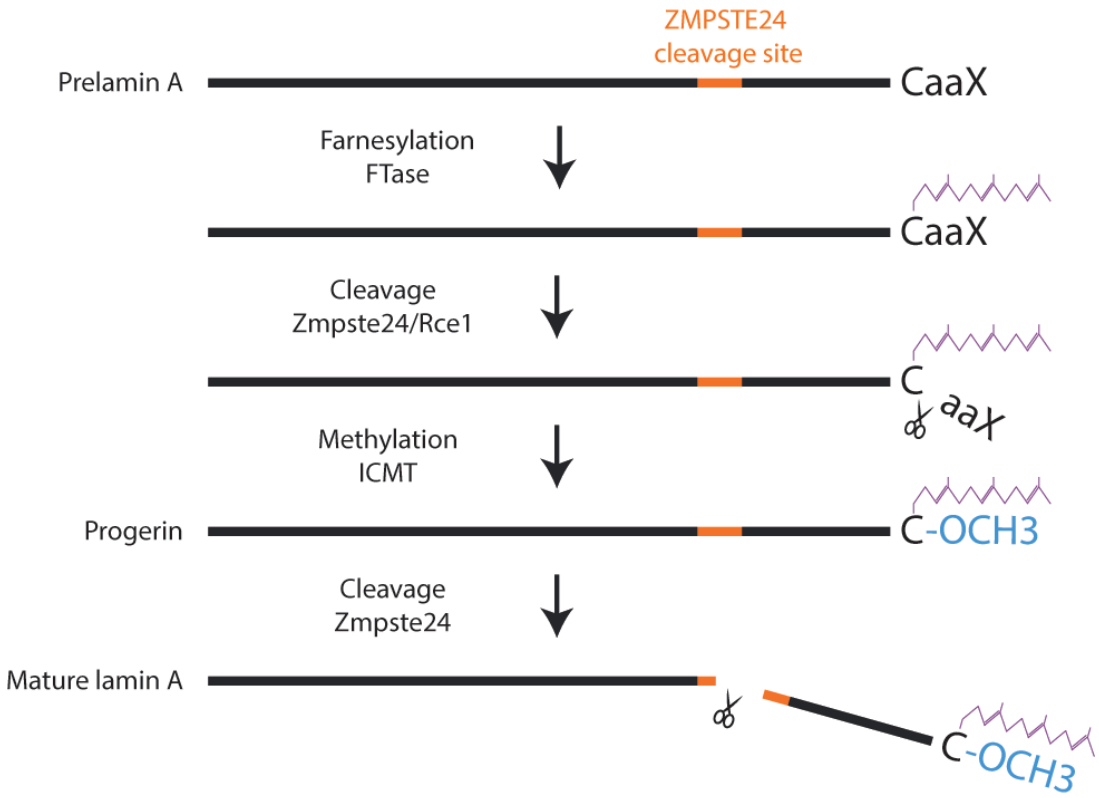


Figure 5.1: Lamin A post-translational processing. Lamin A undergoes complex multi-step post-translational processing. In a first step, the C-terminal CaaX motif of lamin A gets farnesylated by FTase. Subsequently the terminal three aaX residues get protolytically cleaved. The C-terminal tail then gets methylated by ICMT. This results in progerin, an incompletely processed lamin A form that, if accumulated in cells, is associated to disease. In a final step, the farnesylated and methylated C-terminal tail gets cleaved by Zmpste24 to generate mature lamin A.

the G608G mutation. This mutation leads to the activation of a cryptic splice site in *LMNA*, resulting in a 50 amino acid C-terminal truncation[43,104]. This truncated form of lamin A, often referred to as progerin, looses the internal splice site required for the final cleavage by ZMPSTE24 to form mature lamin. Accumulation of progerin leads to morphological changes such as increased lobulation of the NE, thickening of the lamina, decreased heterochromatin and clustering of NPCs[105]. Additionally, cells exhibited increased oxidative stress and mitochondrial dysfunction[242]. A similar disease phenotype as seen in *LMNA* G608G patients, can be observed in patients lacking ZMPSTE24. Indicative that this is due to the function of ZMPSTE24 in processing of lamin A, heterozygosity for *LMNA* deficiency eliminates many of the disease-like phenotypes

in *ZMSTE24* KO mice[243].

To improve our understanding of laminopathies, we used our newly developed and validated PDB-fractionation assay to investigate the INM proteome in *LMNA* and *ZMPSTE24* KO cells. Understanding whether global changes to the proteome arise in disease associated phenotypes may be a first step in gaining a more mechanistically understanding of the diseases.

5.2 Results

5.2.1 Generating *LMNA* and *ZMPSTE24* KO cell lines

To investigate if known mutations linked to disease phenotypes result in global changes to the INM proteome, cell lines mimicking these phenotypes were generated. For this, two knockouts were chosen: *LMNA* KO and *ZMPSTE24* KO. KO cell lines were generated using CRISPR-Cas9 technology in the following FlpIn293 APEX2-bait cells lines: parental, NLS, BAF, $Nup153_{1-144}$ and Climp63. After virus transduction of the gRNAs, cells were put on 2 μg/mL puromycin selection for two days, before letting them recover in full media. Subsequently, single cell clones were obtained by plating them into 96-well plates using a FACS machine.

5.2.2 Validating *LMNA* and *ZMPSTE24* KO cell lines by WB

Western blot analysis was used to validate whether KO of *LMNA* and *ZMPSTE24* was successful. To test for *LMNA* KO, a lamin A/C antibody was used and four monoclonal cell lines were tested for each bait protein (figure 5.2). The antibody resulted in a strong band at the expected size for lamin C, whereas lamin A was hardly detectable. *LMNA* KO resulted in a complete loss of signal in nearly all single cell clones indicating the successes of *LMNA* KO (figure 5.2).

To validate the *ZMPSTE24* KO various antibodies were trialled. As no good commercial antibody was available for ZMPSTE24, an indirect validation approach was chosen. In *ZMSTE24* KO cells, the final proteolytic cleavage

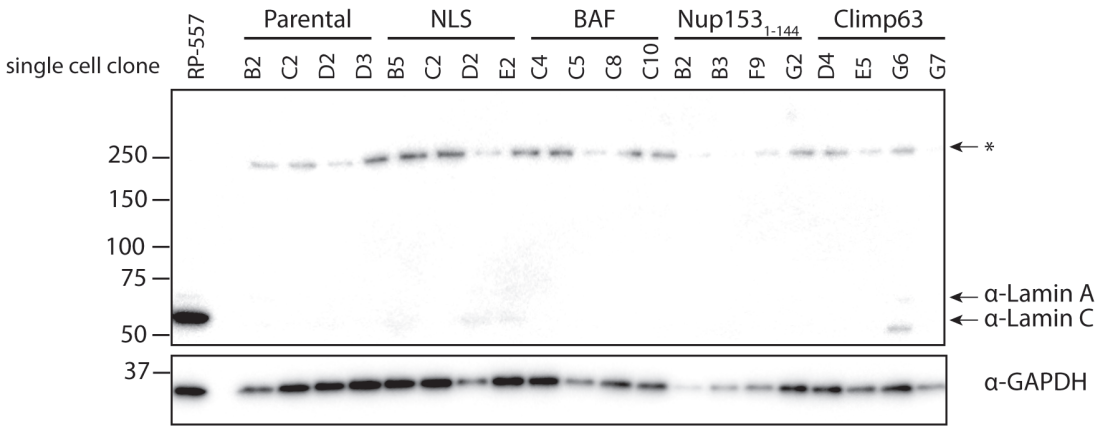


Figure 5.2: Validating the *LMNA* KO cell lines by WB. Anti-LMNA/C WB from WCL lysate originated from single cell *LMNA* KO clones in FlpIn293 parental or APEX2-fusion protein backgrounds. RP-557 is the empty vector negative control. GAPDH was used as a loading control. The asterisk (*) indicates a uncharacterised band.

to convert pre-lamin to mature lamin A will not take place. This will result in pre-lamin accumulation which can be visualised using a pre-lamin antibody. Alternatively, accumulation of pre-lamin A can be visualised using lamin A/C antibody, as pre-lamin has a molecular weight of 74 kDa, compared to the 72 kDa size of mature lamin A. Yet, no signal could be observed by WB using the pre-lamin antibody and no size shift was observable using the lamin A/C antibody (data not shown). This means that *ZMPSTE24* KO could not be validated by WB and therefore the *ZMPSTE24* KO cells were not used for any further analysis.

5.2.3 Validating PDB localisation in *LMNA* KO cells using IF

The combination of Climp63 and Nup153$_{1-144}$ APEX2-bait cell lines was previously shown to be a powerful tool to investigate the protein composition of the INM (chapter 4.2.3). Therefore, the respective *LMNA* KO cells were further validated by IF. As expected, *LMNA* KO did not interfere with bait protein localisation and Climp63 labelling was still observed in the ER, whereas APEX2-Nup153$_{1-144}$ labelling resulted in biotinylated protein in the nucleus (figure 5.3). Of note, data from other studies showed that *LMNA* KO can lead to aberrant nuclear

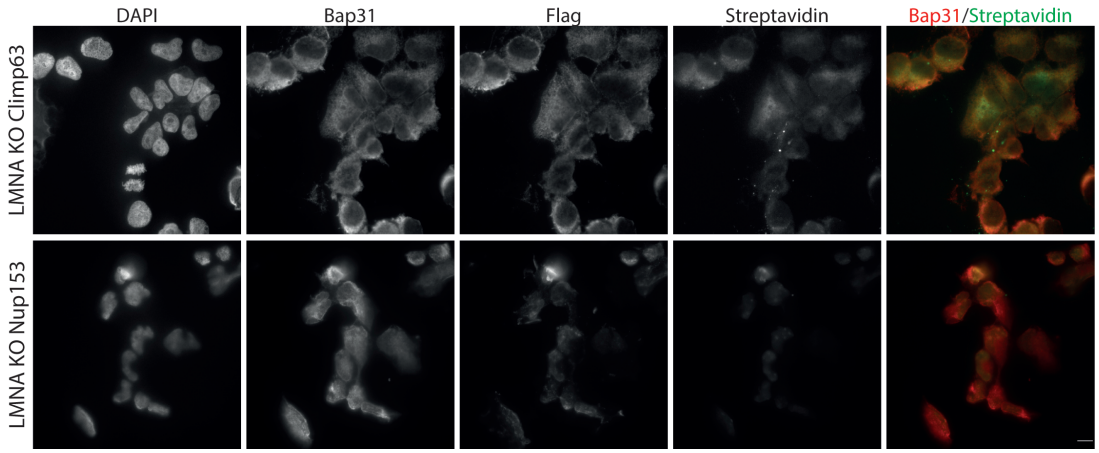


Figure 5.3: Validating PDB in the LMNA KO cell lines. IF microscopy images of *LMNA* KO FlpIn293 cells expressing Climp63-APEX2-Flag or Flag-APEX2-Nup153$_{1-144}$. Anti-Bap31 and anti-FLAG antibodies were used in combination with their corresponding Alexa Fluor secondary antibodies for staining. Alexa Fluor-488 streptavidin was used for visualising biotinylated protein. Nuclei were stained using DAPI. Scale bar: 10 μm.

morphology in the form of elongated nuclei[106,244,245]. Yet, this was not apparent for the FlpIn293 cells investigated here, as nuclei were comparable to WT cells.

5.2.4 Validating subcellular fractionation of nuclei from *LMNA KO* cells

Nuclei from *LMNA* KO cells have previously been shown to be more prone to rupture[106]. Therefore, the next step was to assess whether this would interfere with subcellular fractionations. For this, fractionations were performed as described previously (chapter 2.2.3.7). FlpIn293 APEX2-Sec61β cells were used as a control and all cell lines were analysed in duplicates. Fractionations were assessed by WB using well-characterised markers proteins for each fraction (figure 5.4). No striking differences could be observed in fractionations of the control compared to the *LMNA* KO cell lines. These results indicated that, even though prior data suggests that nuclei of *LMNA* KO cell lines are more prone to rupture, this does not interfere with subcellular fractionation to enrich for the NE and CM.

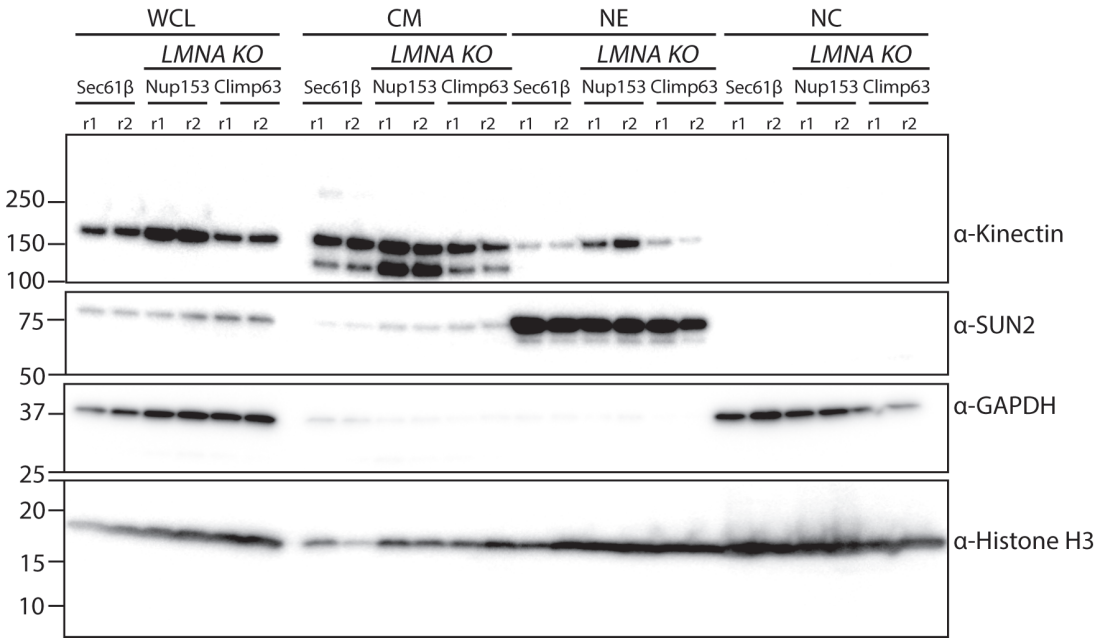


Figure 5.4: Validating subcellular fractionation in *LMNA* KO cells. WBs using WCL or CM, NE and NC fractions obtained during subcellular fractionation were blotted for proteins with well-characterised subcellular localisation. Kinectin (ER marker), SUN2 (INM marker), Histone H3 (nuclear marker), GAPDH (generic soluble marker). Sec61β was used as a reference control cell line. All samples were performed in two biological replicates (r1 and r2). WCL: 0.21% of total sample, CM/NE/NC: 1.7% of total fraction.

5.2.5 Proteomic analysis of *LMNA* KO cells

Having validated that *LMNA* KO was successful and that APEX2-bait proteins in the FlpIn293 *LMNA* KO cell lines were expressed, active and localised as in the WT cells, the next step was to prepare samples for proteomic analysis. The proteomic data sets for the WT Climp63 and Nup153$_{1-144}$ cell lines already existed (chapter 4.2), whereas the data sets for the *LMNA* KO cells were collected in quadruplicates. Subsequently, the raw data of the WT and *LMNA* KO cells were searched in a combined MaxQuant run and the data was processed further in R as described in chapter 4.2.2. Limma enrichment analysis of Nup153$_{1-144}$ NE WT versus *LMNA* KO was then used to investigate whether depletion of A-type lamins leads to changes in the INM protein composition (figure 5.5), whereas the Climp63 CM WT versus *LMNA* KO comparison was used as a control (figure 5.6A and B).

As expected, some of the most highly enriched proteins in the Nup153$_{1-144}$ NE

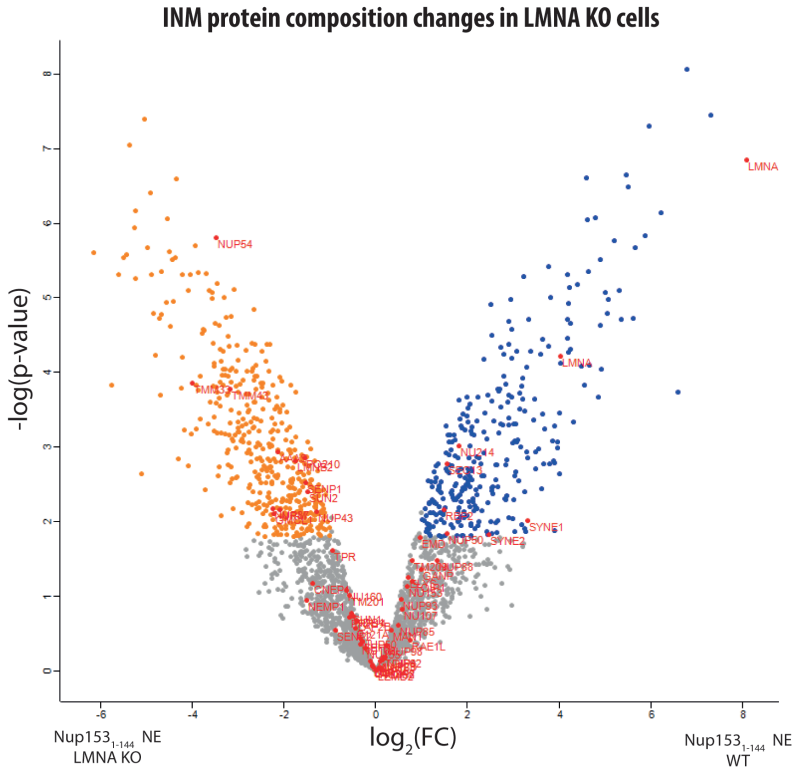


Figure 5.5: Investigation global changes in the protein composition of the INM of *LMNA* KO cells. Limma analysis using raw intensities of FlpIn293 WT versus LMNA KO cells ($-1 < FC < 1; adj.pvalue < 0.05$). $Nup153_{1-144}$ WT NE versus $Nup153_{1-144}$ *LMNA* KO NE. Orange - significantly enriched in $Nup153_{1-144}$ *LMNA* KO NE, blue – significantly enriched in $Nup153_{1-144}$ WT NE, red – INM marker proteins.

WT samples were lamin A and C, indicating once more the success of *LMNA* KO.

Next, using the marker protein list (table A.2) we investigated whether depletion of A-type lamin affects the enrichment of other proteins at the INM. Thereby no significant changes were observed for the majority of the INM marker proteins between the $Nup153_{1-144}$ NE WT and *LMNA* KO conditions (figure 5.5). As a control Climp63 CM data was compared for WT versus *LMNA* KO cells. In line with depletion of A-type lamins not affecting the surrounding of the Climp63 bait, little difference in both INM and ER marker proteins was observed between the two conditions (figure 5.6A and B).

Yet, whereas most proteins were unaffected by *LMNA* KO, differences for some proteins could be observed. One interesting observation was that lamin

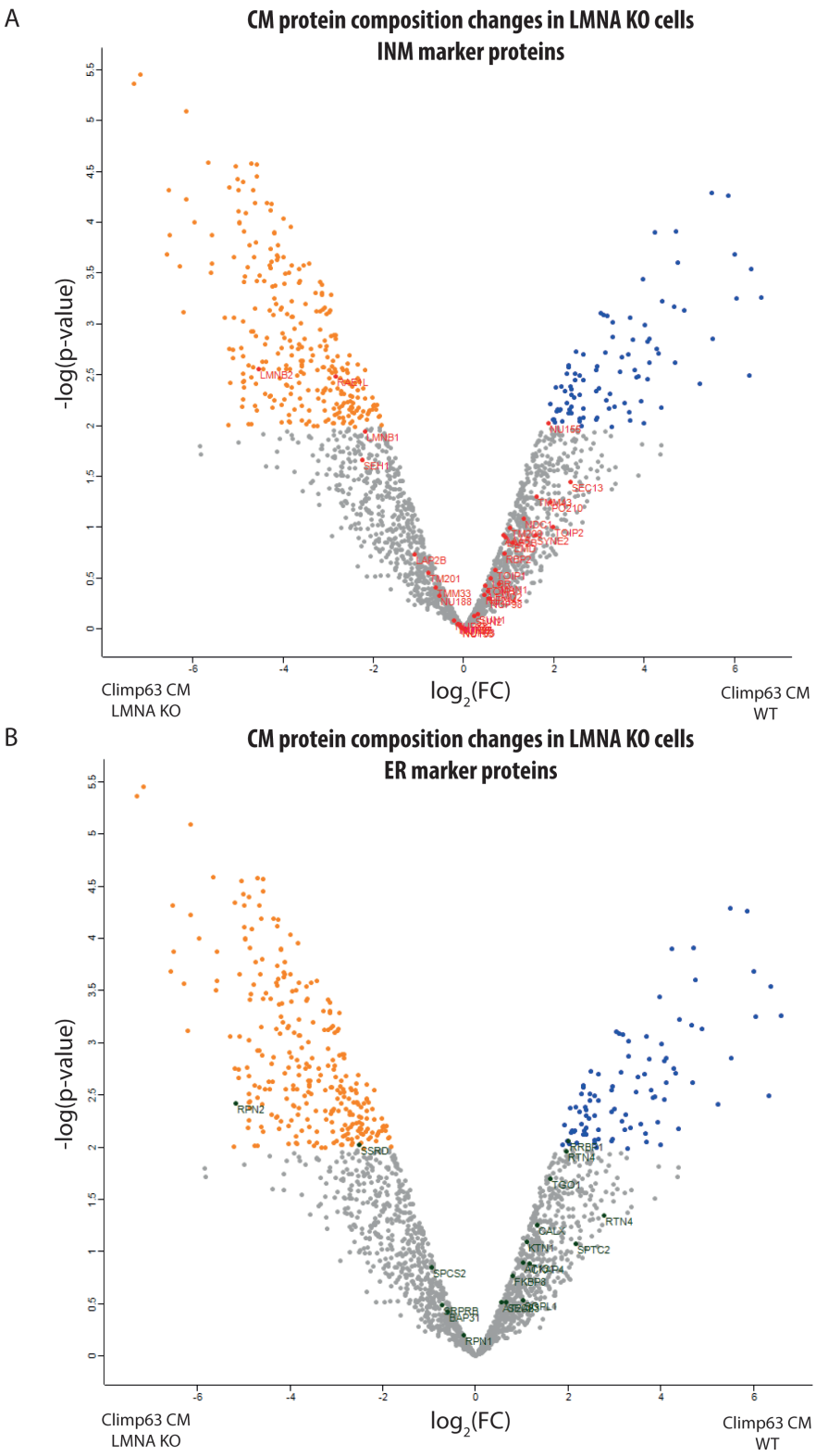


Figure 5.6: Investigating global changes in the protein composition of the bulk ER of *LMNA* KO cells. limma analysis using raw intensities of FlpIn293 WT versus LMNA KO cells ($-1 < FC < 1; adj.pvalue < 0.05$). a) Climp63 WT CM versus Climp63 *LMNA* KO CM. Orange - significantly enriched in Climp63 *LMNA* KO NE, blue – significantly enriched in Climp63 WT NE, red – INM marker proteins. b) Climp63 WT CM versus Climp63 *LMNA* KO CM. Orange - significantly enriched in Climp63 *LMNA* KO NE, blue – significantly enriched in Climp63 WT NE, green – ER marker proteins.

B2 was enriched significantly in the nuclei of $Nup153_{1-144}$ *LMNA* KO cells. A- and B-type lamins collectively form the nuclear lamina. Thereby, lamin A and C are alternative splice variance of the *LMNA* gene, whereas lamin B1 and B2 are expressed from two separate genes. Nevertheless, A- and B-type lamins share around 70% sequence similarity and are redundant in some of their functions[109]. Thus, one hypothesis to explain the enrichment of lamin B2 at the INM of *LMNA* KO cells could be that cells upregulate lamin B2 expression to compensate for the lack of lamin A/C. Yet, whereas lamin B2 was enriched in the KO cells, this was not the case for lamin B1, which was found at similar levels in the $Nup153_{1-144}$ NE WT and KO cells. Interestingly, both B-type lamins were also shown to be enriched in the Climp63 CM *LMNA* KO sample compared to WT cells. As lamins are synthesised in the cytoplasm before being shuttled to the NE where they assemble into the nuclear lamina meshwork, this may be an indication that B-type lamins are upregulated in *LMNA* KO cells (figure 5.6A).

Intriguingly, one function that is exclusive to A-type lamins is the recruitment of emerin to the INM and cells depleted of lamin A were shown to exhibit mislocalisation of emerin to the bulk ER[108,109]. In line with these results, emerin is shown to be enriched in the NE of WT compared to *LMNA* KO cells in our data set, though the enrichment is just below the significance threshold ($-1 < FC < 1; adj.pvalue < 0.05$).

Another interesting protein is TMEM43, also known as LUMA, which was found to be enriched in the *LMNA* KO cells. This protein is highly conserved among species, sharing over 90% sequence identity with the murine protein and mutations in this protein are associated to the EDMD disease. Previous immunoprecipitation experiments from mouse C2C12 cells indicated a direct interaction of TMEM43 with both A- and B-type lamins, as well as emerin. Additionally, RNA interference (RNAi) knockdown of *LMNA* in HeLa cells was shown to result in mislocalisation of TMEM43 to the ER[110]. Surprisingly, whereas we also see an effect of *LMNA* depletion in our cells on TMEM43 localisation, our data actually showed increased levels of TMEM43 at nuclei of *LMNA* KO

cells. One possible explanation for this could be that, whereas murine TMEM43 interacted with A-type lamins, it showed much stronger interaction for lamin B2[110]. Therefore, the enrichment of lamin B2 in *LMNA* KO cells could result in increased recruitment of TMEM43.

Another well-characterised INM protein that demonstrated enrichment in the nuclei of *LMNA* KO cells in our PDB data was SUN2. Previous data on the effect of *LMNA* KO on SUN2 protein localisation exists, but showed heterogeneity in the phenotypes observed. IF data of MEF cells isolated from $LMNA^{-/-}$ mice showed enrichment of SUN2 in the cytoplasm compared to WT MEFs, however only in a proportion of the cells[111]. Similar results were obtained in a study from another group, which classified the SUN2 localisation in $LMNA^{-/-}$ cells in four distribution patterns observed: cells in which (1) SUN2 was lost from one pole of the nucleus, (2) SUN2 localised to the cytoplasm, (3) SUN2 localised to both the NE and the cytoplasm, and (4) cells in which the SUN2 localisation was comparable to WT cells[246]. Interestingly, the localisation of SUN2 observed in *LMNA* KO MEFs could not be restored in rescue experiments were lamin A and C were reintroduced[111]. Furthermore, unlike in MEFs, depletion of A-type lamins by RNAi in HeLa cells only resulted in very minor changes in SUN2 localisation[111]. In contrast to this already ambiguous data, our data set indicates an enrichment rather than a depletion of SUN2 in *LMNA* KO compared to the WT cells at the INM.

In comparison to SUN2, the localisation of its paralogue SUN1 at the INM was shown to be independent of A-type lamins[111]. This was also reflected in our data set, where no enrichment of SUN1 was seen in either sample.

Some nucleoporins also seemed to be affected by *LMNA* KO. Nup214 and Nup50 were significantly enriched in WT cells, whereas Nup54, Nup210, AAAS, Nup43 and Nup37 were significantly enriched in *LMNA* KO cells (figure 5.5). Published data also shows some heterogeneity in the effect of *LMNA* KO on NPC localisation: in MEFs *LMNA* KO led to the loss of Nup153 from one pole of the NE[108]. The same was observed using the MAb414 antibody, that labels

the conserved FxFG repeats that various Nups contain to build the hydrophobic core of the NPCs[247]. Yet, studies by another group showed that *LMNA* KO only affected NPC localisation in differentiated cells, but not in ESCs[109]. The same study also demonstrated that NPC localisation was only affected by depletion of A-type, but not B-type lamins. Yet, sufficiently high levels of lamin B1 or lamin B2 alone could reverse the phenotype observed in *LMNA* KO cells and ensure correct NPC localisation within differentiated cells even in absence of A-type lamins.

To conclude, for most well-characterised INM proteins no change in protein level was observed at the NE between WT and *LMNA* KO cells. Yet, for a handful of proteins differences could be observed between the two conditions. Thereby most proteins followed similar trends as those observed in previous studies, yet some proteins like TMEM43 exhibited opposing trends, which might be related to species-specific differences.

5.3 Discussion

5.3.1 Adapting our PDB assay to study laminopathies

In the past two decades, an increasing number of different mutations in INM proteins and lamins were identified in cells from patients with diseases such as muscular dystrophies and progeroid syndromes. In spite of these findings, our understanding of the mechanisms on how different mutations of these genes lead to such a wide-range of diseases is still very primordial. Therefore, we decided to use our newly established PDB assay to investigate whether disease-related protein depletions lead to wider changes in the INM protein composition.

For this, CRISPR/Cas9-mediated technology was used to introduce *LMNA* and *ZMPSTE24* KOs in APEX2-bait cell lines. When good antibodies were available against the deleted protein, as in the case of A-type lamins, KOs could easily be validated by WB. In the absence of a good antibody, as in the case of ZMPSTE24, KO of genes would require validation by Sanger sequencing, next-generation sequencing or qPCR on the genomic level, or mass spectrometry at the proteomic level. Of note, the majority of mutations associated to diseases are point mutations, rather than depletions of protein. Whereas cell lines mimicking these conditions can be generated using knock-in technology, validation will require proteomic analysis or genomic DNA extraction coupled to sequencing, as immunoblotting would not be applicable for this validation.

One potential caveat of our PDB-fractionation based assay in studying nucleopathies is that it relies on subcellular fractionation of NE membranes, yet nuclei from cells affected by these diseases are often more prone to nuclear rupture. Therefore, quality control experiments were performed to investigate whether nuclei of *LMNA* KO cells could be successfully fractionated. Thereby no difference was observed in the enrichment of NE membranes from WT and *LMNA* KO cells, indicating that the depletion of A-type lamins does not interfere with nuclear fractionation in FlpIn293 cells.

To investigate whether *LMNA* KO cells demonstrate differences in their INM protein composition compared to WT cells, the data analysis pipeline optimised in chapter 4.2.2 was used. No adjustments were required for the data analysis methodology, as the same normalisation steps and pairwise-enrichment analysis used to compare data sets from different baits is applicable to compare WT versus *LMNA* KO samples. This once more demonstrates the versatility of the assay and data analysis pipeline generated in this project.

5.3.2 Depletion of A-type lamins does not result in global changes of the INM protein composition

Focusing on the enrichment analysis of WT versus *LMNA* KO cells, our data indicated that depletion of A-type lamins does not result in global changes in the protein composition of the INM. Indeed, the majority of INM marker proteins were unaffected by *LMNA* KO. In line with IF data from prior studies, this included proteins such as Lap2β, LBR and SUN1[108,111,178]. Yet, some proteins were either shown to be enriched or depleted at the INM upon depletion of A-type lamins. Thereby some of the trends observed in our data set, such as the enrichment of emerin in WT cells, were in line with IF data from previous studies[108,109]. Yet, for other proteins like TMEM43 we observed an opposing trend, as the protein was previously shown to be depleted in nuclei of *LMNA* KO cells, yet in our data set a strong enrichment could be observed. It is worth mentioning that available data mostly originates from other cell types and organisms, and that cell type-dependent variability has been observed in the behaviour of INM proteins upon *LMNA* deletion[111]. Indeed, understanding how mutations in ubiquitously expressed proteins such as lamins result in cell type-specific diseases is an open question in the field. Therefore, it would be interesting to follow up on the differences observed here between cell lines, to try and understand if they might play a role in the cell type-specific phenotypes observed in disease.

Another observation that is difficult to interpret is that whereas various nucleoporins seem to be affected by *LMNA* KO, some nucleoporins show en-

richment compared to WT cells, whereas other proteins are depleted. How and why different nucleoporins are affected differentially would require further validation and investigation.

Finally, whereas our PDB-fractionation based assay is a very powerful to obtain an overview over global protein changes, there are two caveats associated with it. Firstly, an enrichment or depletion in the a specific ER domain between two samples could always be due to two reasons: (1) because the overall protein expression is up- or downregulated, or (2) because proteins change their localisation. To differentiate between these possibilities, further experiments such as immunoblotting would be required to determine if there are any changes to the cellular protein levels. The second caveat is that whereas the assay will be able to pick up proteins that are enriched or depleted at the INM compared to the bulk ER, the assay in its current form is not capable of detecting changes of protein localisation within the INM. It is known from previous IF studies that some of the changes observed in the localisation of INM proteins upon *LMNA* KO are differences in the localisation of proteins within the NE itself, such as loss of NPC localisation at one nuclear pole[108].

To conclude, in this chapter we used our newly developed and validated PDB-fractionation based assay to investigate the INM proteome in *LMNA* KO cells. Thereby we showed that depletion of A-type lamins affected the nuclear enrichment of individual proteins, yet the majority of INM proteins were not affected. In addition to enhancing our understanding on how *LMNA* KO affects the localisation of other proteins, this study also acted as an example to show that the assay and its associated pipeline can easily be applied to study the protein composition of ER domains under different conditions.

One child, one teacher, one pen, and one book can change the world.

— Malala Yousafzai

6

Concluding remarks and future perspectives

Contents

6.1 Developing an assay to study the ER domains *in situ* 169
6.2 *Bona fide* INM, ER tubule and ER sheet proteins . 171
6.3 Using the PDB assay to investigate the protein composition of ER domains in various physiological conditions 174
6.4 Concluding remarks 176

6.1 Developing an assay to study the ER domains *in situ*

The ER is the largest membrane-bound organelle in cells and is involved in a range of functions, including synthesis of membrane and secretory proteins, calcium homeostasis and lipid synthesis[1]. Besides, the ER forms contact sites with other organelles, thereby enabling the exchange of macromolecules, ions and lipids, while also influencing organelle distribution and division[2,4]. Whereas some functions, such as contact site formation, are known to be predominately restricted to specific ER domains, for other functions like lipid synthesis it is unclear if these take place throughout the ER or whether they are confined to a specialised domain. As different functions are dependant on the presence of specialised proteins, the principal aim of this project was to develop an assay to investigate the protein composition of distinct ER domains. This can widen our understanding of ER organisation, while providing the foundation for subsequent functional and mechanistic studies.

The method of choice for assay development was an APEX2-based PDB system. To target APEX2 to the desired cellular compartment, it was fused to bait proteins localising to distinct ER domains and stable, tet-inducible FlpIn-293 bait cell lines were generated. A series of quality control experiments were performed to ensure that APEX2 was expressed, active and targeted correctly. Despite the success of generating cell lines in which APEX2 was targeted to distinct ER domains, bait proteins were still overexpressed which may trigger unwanted perturbations in cells or result in minor changes in protein localisation. To avoid problems associated with overexpression and mimic more physiological conditions, one strategy would be to tag proteins endogenously using CRISPR/Cas9 mediated knock-in technology or a CRISPR-Cas12a-assisted PCR tagging approach[248]. One potential issue that could thereby arise is that endogenous expression levels of bait proteins may be too low to ensure sufficient labelling for detection by

LC-MS/MS, yet the efficiency of the APEX2-PDB and the sensitivity of mass spectrometry make this rather unlikely.

Having generated the cell lines, various methods were trialled to optimise the sample preparation of PDB-labelled proteins for LC-MS/MS analysis. Thereby the observation was made, that even with a membrane-bound bait only around 10% of proteins identified by LC-MS/MS were membrane proteins. To tackle this issue and enrich for membrane proteins, an additional subcellular fractionation step was added between the PDB and immunoprecipitation (chapter 3.3.11). Using this technique, the amount of membrane proteins identified doubled from approximately 10% to 20%. Another advantage of coupling PDB to subcellular fractionation is the stringency of the assay, as proteins need to be both labelled by PDB and be enriched in the correct fraction to be identified by LC-MS/MS. This method could help identify proteins which are present at the INM, but at lower abundance than in the bulk ER. Indeed, previous results from a systematic split-GFP screen in yeast indicated that many proteins with well-defined functions in ER membranes and the cytoplasm traffic through the INM[56]. The fact that only a small fraction of the total protein thereby localises to the INM makes the detection at the INM extremely challenging. Whereas no comparable systematic analysis exists in human cells, the prevailing model of INM targeting by a diffusion-retention based mechanism relies on random trafficking of ER proteins through the INM (chapter 1.4). One could therefore speculate that in mammalian cells, similar to yeast, a large number of ER proteins reach the INM. Excitingly, a PDB-fractionation based assay coupled to LC-MS/MS analysis might pose a solution to detect and study INM pools of proteins, even if the relative abundance compared to their presence in the ER is low.

For the finalised data collection five different probes targeting to ER tubules, ER sheets, the INM as well as the nucleus were chosen. For each probe two fractions were collected (NE & CM) and data was obtained at least in triplicates. Overall, this led to the identification of 5126 unique proteins. Using this data set,

different workflows for the data analysis were tested and finally optimised. R based analysis using VSN and LCMD imputation, followed by limma analysis was fitted best for the project (chapter 4.2.2). The assay and analysis were validated by a range of methods, including PCA analysis to determine replicate clustering and GO enrichment analysis to test whether the ER and INM samples are enriched in the expected proteins. Overall the results of these experiments indicated that (1) the assay and its associated data analysis pipeline work very well in investigating the distinct ER domains and that (2) we obtained a very extensive data set to investigate the protein composition in different ER domains. Excitingly, the methodology of coupling PDB to subcellular fractionation to decipher membrane compartments can easily be adapted to study other organelles. For this, only the bait proteins and the fractionation need to be tailored to the organelle of interest, whereas the protein analysis pipeline be will not require any adjustment.

6.2 *Bona fide* INM, ER tubule and ER sheet proteins

Using the newly developed assay a range of proteins enriched in ER sheets, ER tubules and the INM were identified. As expected, the strongest difference in protein composition was thereby observed between the INM and the bulk ER, yet more subtle differences also existed between ER sheets and tubules. Besides well-characterised proteins for each of these domains, the list of identified proteins entailed a range of generic ER membrane proteins that were not previously shown to be enriched in a particular ER domain.

Some putative INM proteins were already discussed in chapters 4.2.7 and 4.2.9. Besides individual proteins being enriched at the INM, one compelling observation made was that all proteins required for VLCFA synthesis were identified at the INM. It is thereby interesting to note, that VLCFAs were shown to have functional roles both at the INM and in the nucleus. Data from liposomes indicates that VLCFAs reduce bilayer-to-hexagonal phase transition, indicating that they may play a role in stabilising highly curved membranes[249]. Additionally,

S. cerevisia VLCFA auxotrophs fail to insert NPCs correctly and show aberrant protuberances extend from the INM into the perinuclear space[238,239]. Combined with data demonstrating that NPC insertion starts at the INM and progresses in an inside-out extrusion of the NE, it is tempting to speculate that synthesis of VLCFAs at the INM would be a mechanism of cells to locally increase VLCFA concentration required for NPC insertion. Besides their role in NPC insertion, VLFCAs act as building blocks for sphingolipids, which have been associated with signaling and regulatory functions in nuclei[237]. Whereas no synthesis of VLCFAs was shown to take place at the INM to date, a range of novel studies based on IF data and lipid sensors provide evidence that the INM might be metabolically active and involved in lipid synthesis[47–50]. Thereby a DAG enrichment was observed at the INM compared to the bulk ER, and PA/DAG levels were shown to be higher in nuclear lipid dropslets compared to their cytoplasmic counterparts[47]. To follow up on the finding that all enzymes involved in VLCFA synthesis are present at the INM, besides validating the localisation of these proteins using imaging technology, it would be interesting to investigate whether there is any biological function associated with their presence at the INM. Yet, testing this would not be trivial. One possible approach could be to retain the proteins involved in VLCFA synthesis at the ER and stop their diffusion to the INM. This could be done using an anchor-away approach at the peripheral ER or by tagging proteins with large tags that would stop their translocation past the NPC. Subsequently, IF studies could be used to investigate whether depleting proteins from the INM affects NPC insertion or NE morphology. To ensure that any effect observed is dependent on the localisation of the enzymes involved in VLCFAs rather than changes in the cellular VLCFA concentration, the assay could be coupled with lipidomic analysis.

Another exciting observation to follow up on is that the RNF185/Membralin ERAD ubiquitin ligase complex was enriched in the Rtn4a CM sample compared to Climp63 CM. This is intriguing as so far none of the ERAD complexes were shown to be enriched in either ER tubules or sheets. In line with that, our

proteomic data indicated that other ERAD ligases such as RNF5 and AMFR, as well as the mitochondrial ligase MARCH5, were equally identified by both the Rtn4a and the Climp63 bait. Understanding how E3 ligases recognise their substrates is an ongoing area of research. A conceptually interesting idea thereby came from new findings of the Asi complex. The Asi complex in yeast was identified as the first E3 ubiquitin ligase complex exclusively localised to the INM[91,92]. Among its substrates are folded, but unassembled subunits of membrane complexes[93]. Spatially separating protein synthesis from quality control could be one mechanism cells employ to keep subunits of complexes at near stoichiometric levels, as orphaned subunits are more likely to diffuse past the NPC and reach the INM to get degraded by the Asi complex. Whereas highly speculative and currently not backed with experimental data, conceptually a similar mechanism could apply to an ERAD complex enriched in ER tubules, but depleted in the ER sheets where most of the ribosomes sit and proteins get synthesised. Whereas no size filter comparable to the NPC exists between ER tubules and ER sheets, the spatial segregation could still give proteins sufficient time to find their interaction partners and form mature protein complexes.

To follow up on this observation, the localisation of the RNF185/Membralin complex would need to be validated. Unfortunately, no good commercial antibodies for IF were available for any of the proteins. We next speculated that the enrichment of the RNF185/Membralin complex may indicate that it plays a role in ER tubule maintenance and that depletion of the complex may result in morphological changes in ER tubules. To test this, ER tubules and sheets were visualised in WT and KO cell lines, but no morphological differences could be observed, as discussed in chapter 4.2.10.1. In the future, it would be interesting to validate the localisation of proteins of the complex by endogenous-tagging. HeLa MBRL-mNeonGreen and HeLa MBRL-HaLo cells exist in the lab, whereas cell lines with endogenously-tagged RNF185, TMUB1 and TMUB2 would still need to be generated.

6.3 Using the PDB assay to investigate the protein composition of ER domains in various physiological conditions

As APEX2 PDB works *in situ* in living cells, the assay developed in this project could be a powerful tool to monitor how specific perturbations or cellular metabolic conditions affect protein composition of ER domains. To put this to test, we used the assay to investigate if *LMNA* and *ZMPSTE24* KO, both associated with laminopathies, lead to changes in the protein composition of the INM (chapter 5).

For this, CRISPR/Cas9 technology was deployed to generate KOs in the FlpIn-293 cells expressing different APEX2-bait proteins. *LMNA* KO was validated using WB, whereas *ZMPSTE24* KO validation by WB was unsuccessful due to the lack of a good antibody and would therefore require further validation by proteomic analysis. Enrichment analysis between $Nup153_{1-144}$ WT and *LMNA* KO cells was used to investigate changes in the INM protein composition. The results demonstrated that most INM proteins were unaffected by *LMNA* KO, yet a handful of INM proteins showed enrichment or depletion at the INM as a result of knocking out A-type lamins. Curiously, whereas our observed enrichments partially followed the trends reported in previous publications, as in the case for emerin, Lap2β and LBR, other proteins that were described to enrich in *LMNA* KO cells were depleted in our data set and vice versa. One explanation for this could be that the effect of *LMNA* KO on localisation of INM proteins differs between cell types and none of the published data comes from FlpIn-293 cells[111]. This variability between cell lines is coherent with the observation that, whereas lamins and INM proteins are expressed across most tissues, pathophysiological changes associated with their depletion or mutations often show tissue specificity. To gain a better understanding of how mutations in proteins ubiquitously expressed across cell types lead to tissue specific disease phenotypes, it would be interesting to deploy our PDB-fractionation assay in a range of different cell types, with a particular focus on muscle cells which are most often affected in the laminopathies. A potential issue that one could expect

to encounter during these experiments is that APEX2 activity was shown to vary widely between cell types (chapter 3.3.2). As the data analysis pipeline developed in this project was optimised to normalise for expression levels between different baits, this normalisation approach is also applicable for variability arising from differences in APEX2 activity between cell types, yet some activity of the APEX2 is still required for these experiments to be feasible.

Besides disease phenotypes, another intriguing application of the PDB assay would be to study if, and how, the protein composition of ER domains changes across the cell cycle. The most drastic changes to ER and NE morphology happen during cell division (chapter 1.7). Thereby the NPCs and nuclear lamina disassemble and the NE falls apart and gets integrated into the bulk ER, which is mostly tubular during mitosis[46,113,114]. At the end of mitosis, this process is reversed. Intriguingly, whereas the INM shows a homogeneous spread of proteins during interphase, after mitosis INM proteins and lamins are known to form distinct domains, defined as the core and non-core regions. How these regions are formed and whether there are distinct functions associated with them are unanswered questions in the field. One hypothesis is that the core and non-core regions originate from different parts of the ER. This idea relies on the speculation that specific INM proteins are enriched in different parts of the ER during mitosis, a hypothesis that is to date not backed with any data. Thereby transmembrane Nups could be more prone to be stored in tubules, which have increased fenestration compared to ER sheets[116]. For other proteins it is not clear what would determine their preference for a specific ER domain, yet factors could include presence or absence of interaction partners, membrane curvature or protein crowding.

Determining global changes in protein composition during mitosis is impossible using the conventional approach of subcellular fractionation coupled to LC-MS/MS, as no intact NE that could be fractionated exists. Additionally, mitosis in HeLa and other commonly used human adherent cell lines typically lasts only

40–60 min, with NE reformation itself taking as little as 10–12 min[141,250,251] and the rapidity of this process makes it intrinsically challenging to study distinct steps during this progress.

Coupling PDB to subcellular fractionation might pose an exciting opportunity to tackle these issues and study NEBD and re-assembly. For this, cells would have to be synchronised. Subsequently, PDB labelling would be performed at specific time points using different probes targeting to the NE. For answering questions on whether INM proteins are stored in distinct ER domains during mitosis, the ER could be fractionated straight after labelling. In contrast, to investigate NEBD or re-assembly, the NE would have to reform prior to fractionations. For this, cells would have to continue cell cycle progression after the PDB labelling reaction. Excitingly, data from preliminary experiments we performed indicated that cells do continue cycling after the PDB reaction and that the re-assembled NE can successfully be fractionated (data not shown).

Many open questions still remain on whether studying changes in the proteome during mitosis will be feasible: How do baits behave during mitosis, and how can this behaviour be validated? Do our artificial constructs behave the same way as endogenous substrates during mitosis? Is the spatial restriction of biotin-phenoxyl radicals sufficient to get resolution of different domains at the INM? Stringent quality control steps are required to determine if the assay will be suitable to study changes in ER domains during mitosis. Yet, if feasible, this methodology would offer an exciting opportunity to widen our understanding of NEBD and re-assembly in cells.

6.4 Concluding remarks

To conclude, a workflow consisting of APEX2-based PDB coupled to subcellular fractionation and LC-MS/MS analysis was developed to investigate the protein composition of distinct ER domains. The assay was validated using a range of different methods and a data analysis pipeline was optimised. Using five different baits targeting to ER tubules, sheets and the INM, a data set comprising

over 5000 proteins was collected. This data set is a great resource to enhance our understanding of ER organisation and was used to successfully identify a range of exciting *bona fide* proteins enriched in ER tubules, sheets and at the INM. Initial attempts were made to validate some of the findings, yet more work is required in this area. Lastly, the assay is an adaptable tool to study ER remodelling in conditions mimicking disease, ER stresses, physiological stimuli or in a cell cycle dependent manner. This was shown by our example investigating whether *LMNA* KO, known to be pathogenic, leads to global changes in the INM proteome. Thereby, most INM proteins were shown to be unaffected by depletion of A-type lamins, yet some proteins did demonstrate an enrichment or depletion at the INM. Besides investigating protein compositions of ER domains under different conditions, adapting the bait proteins and the subcellular fractionation, the assay and data analysis pipeline can be easily applied to study the protein composition of other membranes within cells.

References

1. Perkins, H. T. & Allan, V. Intertwined and finely balanced: Endoplasmic reticulum morphology, dynamics, function, and diseases. *Cells* **10,** 1–39 (2021).
2. Wu, H., Carvalho, P. & Voeltz, G. K. Here, there, and everywhere: The importance of ER membrane contact sites. *Science* **361** (2018).
3. Almeida, C. & Amaral, M. D. A central role of the endoplasmic reticulum in the cell emerges from its functional contact sites with multiple organelles. *Cellular and Molecular Life Sciences* **77,** 4729–4745 (2020).
4. Scorrano, L. *et al.* Coming together to define membrane contact sites. *Nature Communications* **10,** 1–11 (2019).
5. English, A. R. & Voeltz, G. K. Endoplasmic reticulum structure and interconnections with other organelles. *Cold Spring Harbor Perspectives in Biology* **5,** 1–16 (2013).
6. Friedman, J. R. & Voeltz, G. K. The ER in 3D: A multifunctional dynamic membrane network. *Trends in Cell Biology* **21,** 709–717 (2011).
7. Schwarz, D. S. & Blower, M. D. The endoplasmic reticulum: Structure, function and response to cellular signaling. *Cellular and Molecular Life Sciences* **73,** 79–94 (2016).
8. Shibata, Y., Voeltz, G. K. & Rapoport, T. A. Rough Sheets and Smooth Tubules. *Cell* **126,** 435–439 (2006).
9. Nixon-Abell, J. *et al.* Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. *Science* **354** (2016).
10. Gao, G., Zhu, C., Liu, E. & Nabi, I. R. Reticulon and CLIMP-63 regulate nanodomain organization of peripheral ER tubules. *PLoS Biology* **17,** 1–19 (2019).
11. Wang, B., Zhao, Z., Xiong, M., Yan, R. & Xu, K. The endoplasmic reticulum adopts two distinct forms of tubules. *bioRxiv* (2021).
12. Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* **124,** 573–586 (2006).
13. Wang, N. *et al.* Mechanism of membrane-curvature generationby ER-tubule shaping proteins. *Nature Communications* **12,** 1–15 (2021).
14. Hu, J. *et al.* Membrane proteins of the endoplasmic reticulum induce high-curvature tubules. *Science* **319,** 1247–1250 (2008).
15. Golchoubian, B. *et al.* Reticulon-like REEP4 at the inner nuclear membrane promotes nuclear pore complex formation. *The Journal of cell biology* **221** (2022).

16. Wang, S., Tukachinsky, H., Romano, F. B. & Rapoport, T. A. Cooperation of the ER-shaping proteins atlastin, lunapark, and reticulons to generate a tubular membrane network. *eLife* **5,** 1–29 (2016).

17. Hu, J. *et al.* A Class of Dynamin-like GTPases Involved in the Generation of the Tubular ER Network. *Cell* **138,** 549–561 (2009).

18. Orso, G. *et al.* Homotypic fusion of ER membranes requires the dynamin-like GTPase Atlastin. *Nature* **460,** 978–983 (2009).

19. Chen, S. *et al.* Lunapark stabilizes nascent three-way junctions in the endoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 418–423 (2015).

20. Wisesa, S., Yamamoto, Y. & Sakisaka, T. TMCC3 localizes at the three-way junctions for the proper tubular network of the endoplasmic reticulum. *Biochemical Journal* **476,** 3241–3260 (2019).

21. Betancourt-Solis, M. A., Desai, T. & McNew, J. A. The atlastin membrane anchor forms an intramembrane hairpin that does not span the phospholipid bilayer. *Journal of Biological Chemistry* **293,** 18514–18524 (2018).

22. Shemesh, T. *et al.* A model for the generation and interconversion of ER morphologies. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 5243–5251 (2014).

23. Shibata, Y. *et al.* Mechanisms determining the morphology of the peripheral ER. *Cell* **143,** 774–788 (2010).

24. Puhka, M., Vihinen, H., Joensuu, M. & Jokitalo, E. Endoplasmic reticulum remains continuous and undergoes sheet-to-tubule transformation during cell division in mammalian cells. *Journal of Cell Biology* **179,** 895–909 (2007).

25. Schroeder, L. K. *et al.* Dynamic nanoscale morphology of the ER surveyed by STED microscopy. *Journal of Cell Biology* **218,** 83–96 (2019).

26. Otsuka, S. *et al.* Postmitotic nuclear pore assembly proceeds by radial dilation of small membrane openings. *Nature Structural and Molecular Biology* **25,** 21–28 (2018).

27. Zheng, P. *et al.* ER proteins decipher the tubulin code to regulate organelle distribution. *Nature* **601** (2021).

28. Pendin, D., McNew, J. A. & Daga, A. Balancing ER dynamics: Shaping, bending, severing, and mending membranes. *Current Opinion in Cell Biology* **23,** 435–442 (2011).

29. West, M., Zurek, N., Hoenger, A. & Voeltz, G. K. A 3D analysis of yeast ER structure reveals how ER domains are organized by membrane curvature. *Journal of Cell Biology* **193,** 333–346 (2011).

30. Schuck, S., Prinz, W. A., Thorn, K. S., Voss, C. & Walter, P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *Journal of Cell Biology* **187,** 525–536 (2009).

31. Jensen, D. & Schekman, R. COPII-mediated vesicle formation at a glance. *Journal of Cell Science* **124,** 1–4 (2011).

32. Kurokawa, K. & Nakano, A. The ER exit sites are specialized ER zones for the transport of cargo proteins from the ER to the Golgi apparatus. *Journal of Biochemistry* **165,** 109–114 (2019).

33. Funato, K., Riezman, H. & Muñiz, M. Vesicular and non-vesicular lipid export from the ER to the secretory pathway. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1865,** 158453 (2020).

34. Gomez-Navarro, N. & Miller, E. A. COP-coated vesicles. *Current Biology* **26,** R54–R57 (2016).

35. Abcam. *Western blot protocol*

36. Carrasco, S. & Meyer, T. STIM proteins and the endoplasmic reticulum-plasma membrane junctions. *Annual Review of Biochemistry* **80,** 973–1000 (2011).

37. Vance, J. E. Phospholipid synthesis in a membrane fraction associated with mitochondria. *Journal of Biological Chemistry* **265,** 7248–7256 (1990).

38. Stone, S. J. & Vance, J. E. Phosphatidylserine synthase-1 and -2 are localized to mitochondria-associated membranes. *Journal of Biological Chemistry* **275,** 34534–34540 (2000).

39. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO reports* **18,** 1905–1921 (2017).

40. Buchwalter, A., Kaneshiro, J. M. & Hetzer, M. W. Coaching from the sidelines: the nuclear periphery in genome regulation. *Nature Reviews Genetics* **20,** 39–50 (2019).

41. Hetzer, M. W. The nuclear envelope. *Cold Spring Harbor perspectives in biology* **2,** 1–16 (2010).

42. Hampoelz, B., Andres-Pons, A., Kastritis, P. & Beck, M. Structure and Assembly of the Nuclear Pore Complex. *Annual Review of Biophysics* **48,** 515–536 (2019).

43. Burke, B. & Stewart, C. L. The nuclear lamins: Flexibility in function. *Nature Reviews Molecular Cell Biology* **14,** 13–24 (2013).

44. Mekhail, K. & Moazed, D. The nuclear envelope in genome organization, expression and stability. *Nature Reviews Molecular Cell Biology* **11,** 317–328 (2010).

45. Hatch, E. & Hetzer, M. Breaching the nuclear envelope in development and disease. *Journal of Cell Biology* **205,** 133–141 (2014).

46. Ungricht, R. & Kutay, U. Mechanisms and functions of nuclear envelope remodelling. *Nature Reviews Molecular Cell Biology* **18,** 229–245 (2017).

47. Romanauska, A. & Köhler, A. The Inner Nuclear Membrane Is a Metabolically Active Territory that Generates Nuclear Lipid Droplets. *Cell* **174,** 700–715 (2018).

48. Romanauska, A. & Köhler, A. Reprogrammed lipid metabolism protects inner nuclear membrane against unsaturated fat. *Developmental Cell* **56,** 2562–2578 (2021).

49. Sołtysik, K., Ohsaki, Y., Tatematsu, T., Cheng, J. & Fujimoto, T. Nuclear lipid droplets derive from a lipoprotein precursor and regulate phosphatidylcholine synthesis. *Nature Communications* **10** (2019).

50. Sołtysik, K. *et al.* Nuclear lipid droplets form in the inner nuclear membrane in a seipin-independent manner. *Journal of Cell Biology* **220** (2021).

51. Schirmer, E. C., Florens, L., Guan, T., Yates, J. R. & Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301,** 1380–1382 (2003).

52. Cheng, L.-c. C. *et al.* Identification of new transmembrane proteins concentrated at the nuclear envelope using organellar proteomics of mesenchymal cells. *Nucleus* **10,** 126–143 (2019).

53. Birendra KC *et al.* VRK2A is an A-type lamin–dependent nuclear envelope kinase that phosphorylates BAF. *Molecular Biology of the Cell* **28,** 2241–2250 (2017).

54. James, C. *et al.* Proteomic mapping by rapamycin-dependent targeting of APEX2 identifies binding partners of VAPB at the inner nuclear membrane. *Journal of Biological Chemistry* **294,** 1624–16254 (2019).

55. Müller, M., James, C., Lenz, C., Urlaub, H. & Kehlenbach, R. H. Probing the Environment of Emerin by Enhanced Ascorbate Peroxidase 2 (APEX2)-Mediated Proximity Labeling. *Cells* **9,** 605 (2020).

56. Smoyer, C. J. *et al.* Analysis of membrane proteins localizing to the inner nuclear envelope in living cells. *Journal of Cell Biology* **215,** 575–590 (2016).

57. Wilson, K. L. & Foisner, R. Lamin-binding Proteins. *Cold Spring Harbor Perspectives in Biology* **2** (2010).

58. Wagner, N. & Krohne, G. LEM-Domain Proteins: New Insights into Lamin-Interacting Proteins. *International Review of Cytology* **261,** 1–46 (2007).

59. Barton, L. J., Soshnev, A. A. & Geyer, P. K. Networking in the nucleus: A spotlight on LEM-domain proteins. *Current Opinion in Cell Biology* **34,** 1–8 (2015).

60. Emery, A. E. Emery-Dreifuss muscular dystrophy - A 40 year retrospective. *Neuromuscular Disorders* **10,** 228–232 (2000).

61. Worman, H. J. & Schirmer, E. C. Nuclear membrane diversity: underlying tissue-specific pathologies in disease? *Current Opinion in Cell Biology* **28,** 1304–1314 (2015).

62. Janin, A., Bauer, D., Ratti, F., Millat, G. & Méjat, A. Nuclear envelopathies: A complex LINC between nuclear envelope and pathology. *Orphanet Journal of Rare Diseases* **12,** 1–16 (2017).

63. Tsai, P. L., Zhao, C., Turner, E. & Schlieker, C. The Lamin B receptor is essential for cholesterol synthesis and perturbed by disease-causing mutations. *eLife* **5,** 1–26 (2016).

64. Sosa, B. A., Rothballer, A., Kutay, U. & Schwartz, T. U. LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins. *Cell* **149,** 1035–1047 (2012).

65. Chen, C. Y. *et al.* Accumulation of the inner nuclear envelope protein Sun1 is pathogenic in progeric and dystrophic laminopathies. *Cell* **149,** 565–577 (2012).

66. Lei, K. *et al.* Inner nuclear envelope proteins SUN1 and SUN2 play a prominent role in the DNA damage response. *Current Biology* **22,** 1609–1615 (2012).

67. Silver, P. A. How proteins enter the nucleus. *Cell* **64,** 489–497 (1991).

68. Cavazza, T. & Vernos, I. The RanGTP pathway: From nucleo-cytoplasmic transport to spindle assembly and beyond. *Frontiers in Cell and Developmental Biology* **3** (2016).

69. Komeili, A. & O'Shea, E. K. New perspectives on nuclear transport. *Annu. Rev. Genet.* **35,** 341–64 (2001).

70. Boni, A. *et al.* Live imaging and modeling of inner nuclear membrane targeting reveals its molecular requirements in mammalian cells. *Journal of Cell Biology* **209,** 705–720 (2015).

71. Katta, S. S., Smoyer, C. J. & Jaspersen, S. L. Destination: Inner nuclear membrane. *Trends in Cell Biology* **24,** 221–229 (2014).

72. Powell, L. & Burke, B. Internuclear exchange of an inner nuclear membrane protein (p55) in heterokaryons: In vivo evidence for the interaction of p55 with the nuclear lamina. *Journal of Cell Biology* **111,** 2225–2234 (1990).

73. Ungricht, R., Klann, M., Horvath, P. & Kutay, U. Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *Journal of Cell Biology* **209,** 687–704 (2015).

74. Zuleger, N. *et al.* System analysis shows distinct mechanisms and common principles of nuclear envelope protein dynamics. *Journal of Cell Biology* **193,** 109–123 (2011).

75. Laba, J. K., Steen, A. & Veenhoff, L. M. Traffic to the inner membrane of the nuclear envelope. *Current Opinion in Cell Biology* **28,** 36–45 (2014).

76. Pawar, S., Ungricht, R., Tiefenboeck, P., Leroux, J. C. & Kutay, U. Efficient protein targeting to the inner nuclear membrane requires Atlastin-dependent maintenance of ER topology. *eLife* **6,** 1–25 (2017).

77. Ohba, T., Schirmer, E. C., Nishimoto, T. & Gerace, L. Energy- and temperature-dependent transport of integral proteins to the inner nuclear membrane via the nuclear pore. *Journal of Cell Biology* **167,** 1051–1062 (2004).

78. Turgay, Y. *et al.* A classical NLS and the SUN domain contribute to the targeting of SUN2 to the inner nuclear membrane. *The EMBO Journal* **29,** 2262–2275 (2010).

79. Mudumbi, K. C. *et al.* Nucleoplasmic signals promote directed transmembrane protein import simultaneously via multiple channels of nuclear pores. *Nature Communications* **11,** 1–14 (2020).

80. Rempel, I. L. *et al.* Flexible and Extended Linker Domains Support Efficient Targeting of Heh2 to the Inner Nuclear Membrane. *Structure* **28,** 185–195 (2020).

81. King, M. C., Lusk, C. P. & Blobel, G. Karyopherin-mediated import of integral inner nuclear membrane proteins. *Nature* **442,** 1003–1007 (2006).

82. Braunagel, S. C., Williamson, S. T., Ding, Q., Wu, X. & Summers, M. D. Early sorting of inner nuclear membrane proteins is conserved. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 9307–9312 (2007).

83. Liu, D. *et al.* Truncated Isoforms of Kap60 Facilitate Trafficking of Heh2 to the Nuclear Envelope. *Traffic* **11,** 1506–1518 (2010).

84. Koch, B. & Yu, H.-g. Regulation of inner nuclear membrane associated protein degradation. *Nucleus* **10,** 169–180 (2019).

85. Buchwalter, A., Schulte, R., Tsai, H., Capitanio, J. & Hetzer, M. Selective clearance of the inner nuclear membrane protein emerin by vesicular transport during ER stress. *eLife* **8,** 1–29 (2019).

86. Coyaud, E. *et al.* BioID-based identification of Skp Cullin F-box (SCF)β-TrCP1/2 E3 Ligase Substrates. *Molecular and Cellular Proteomics* **14,** 1781–1795 (2015).

87. Dou, Z. *et al.* Autophagy mediates degradation of nuclear lamina. *Nature* **527,** 105–109 (2015).

88. Deng, M. & Hochstrasser, M. Spatially regulated ubiquitin ligation by an ER/nuclear membrane ligase. *Nature* **443,** 827–831 (2006).

89. Carvalho, P., Goder, V. & Rapoport, T. A. Distinct Ubiquitin-Ligase Complexes Define Convergent Pathways for the Degradation of ER Proteins. *Cell* **126,** 361–373 (2006).

90. Foresti, O., Ruggiano, A., Hannibal-Bach, H. K., Ejsing, C. S. & Carvalho, P. Sterol homeostasis requires regulated degradation of squalene monooxygenase by the ubiquitin ligase Doa10/Teb4. *eLife* **2013,** 1–17 (2013).

91. Foresti, O., Rodriguez-Vaello, V., Funaya, C. & Carvalho, P. Quality control of inner nuclear membrane proteins by the Asi complex. *Science* **346,** 751–755 (2014).

92. Khmelinskii, A. *et al.* Protein quality control at the inner nuclear membrane. *Nature* **516,** 410–413 (2014).

93. Natarajan, N. *et al.* Quality Control of Protein Complex Assembly by a Transmembrane Recognition Factor. *Molecular Cell* **77,** 1–12 (2020).

94. Smoyer, C. J., Smith, S. E., Gardner, J. M., Mccroskey, S. & Unruh, J. R. Distribution of Proteins at the Inner Nuclear Membrane Is Regulated by the Asi1 E3 Ligase in Saccharomyces cerevisiae. *Genetics* **211,** 1269–1282 (2019).

95. Koch, B. A., Jin, H., Tomko, J. & Yu, H. G. The anaphase-promoting complex regulates the degradation of the inner nuclear membrane protein Mps3 in budding yeast. *Journal of Cell Biology* **3,** 839–854 (2019).

96. Lee, C. W. *et al.* Selective autophagy degrades nuclear pore complexes. *Nature Cell Biology* **22,** 159–166 (2020).

97. Allegretti, M. *et al.* In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586,** 796–800 (2020).

98. Mochida, K. *et al.* Receptor-mediated selective autophagy degrades the endoplasmic reticulum and the nucleus. *Nature* **522,** 359–362 (2015).

99. Mochida, K. *et al.* Atg39 links and deforms the outer and inner nuclear membranes in selective autophagy of the nucleus. *Journal of Cell Biology* **221,** e202103178 (2022).

100. Chandra, S. *et al.* Atg39 selectively captures inner nuclear membrane into lumenal vesicles for delivery to the autophagosome. *Journal of Cell* **220** (2021).

101. Chi, Y. H., Chen, Z. J. & Jeang, K. T. The nuclear envelopathies and human diseases. *Journal of Biomedical Science* **16,** 1–8 (2009).

102. Roux, K. J. & Burke, B. Nuclear envelope defects in muscular dystrophy. *Biochimica et Biophysica Acta* **1772,** 118–127 (2007).

103. De Leeuw, R., Gruenbaum, Y. & Medalia, O. Nuclear Lamins: Thin Filaments with Major Functions. *Trends in Cell Biology* **28,** 34–45 (2018).

104. Carrero, D., Soria-Valles, C. & López-Otín, C. Hallmarks of progeroid syndromes: Lessons from mice and reprogrammed cells. *Disease Models and Mechanisms* **9,** 719–735 (2016).

105. Goldman, R. D. *et al.* Accumulation of mutant lamin A progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 8963–8968 (2004).

106. Earle, A. J. *et al.* Mutant lamins cause nuclear envelope rupture and DNA damage in skeletal muscle cells. *Nature Materials* **19,** 464–473 (2020).

107. Onoue, K. *et al.* Cardiomyocyte Proliferative Capacity Is Restricted in Mice With Lmna Mutation. *Frontiers in Cardiovascular Medicine* **8,** 1–13 (2021).

108. Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *Journal of Cell Biology* **147,** 913–919 (1999).

109. Guo, Y., Kim, Y., Shimi, T., Goldman, R. D. & Zheng, Y. Concentration-dependent lamin assembly and its roles in the localization of other nuclear proteins. *Molecular Biology of the Cell* **25,** 1287–1297 (2014).

110. Bengtsson, L. & Otto, H. LUMA interacts with emerin and influences its distribution at the inner nuclear membrane. *Journal of Cell Science* **121,** 536–548 (2008).

111. Crisp, M. *et al.* Coupling of the nucleus and cytoplasm: Role of the LINC complex. *Journal of Cell Biology* **172,** 41–53 (2006).

112. Van Loosdregt, I. A. *et al.* Lmna knockout mouse embryonic fibroblasts are less contractile than their wild-type counterparts. *Integrative Biology (United Kingdom)* **9,** 709–721 (2017).

113. Dey, G. *et al.* Closed mitosis requires local disassembly of the nuclear envelope. *Nature* **585,** 119–123 (2020).

114. Makarova, M. & Oliferenko, S. Mixing and matching nuclear envelope remodeling and spindle assembly strategies in the evolution of mitosis. *Current Opinion in Cell Biology* **41,** 43–50 (2016).

115. Schellhaus, A. K., De Magistris, P. & Antonin, W. Nuclear Reformation at the End of Mitosis. *Journal of Molecular Biology* **428,** 1962–1985 (2016).

116. Liu, S. & Pellman, D. The coordination of nuclear envelope assembly and chromosome segregation in metazoans. *Nucleus* **11,** 35–52 (2020).

117. Huguet, F., Flynn, S. & Vagnarelli, P. The Role of Phosphatases in Nuclear Envelope Disassembly and Reassembly and Their Relevance to Pathologies. *Cells* **8,** 687 (2019).

118. LaJoie, D. & Ullman, K. S. Coordinated events of nuclear assembly. *Current Opinion in Cell Biology* **46,** 39–45 (2017).

119. Gatta, A. T. & Carlton, J. G. The ESCRT-machinery: closing holes and expanding roles. *Current Opinion in Cell Biology* **59,** 121–132 (2019).

120. Kutay, U., Jühlen, R. & Antonin, W. Mitotic disassembly and reassembly of nuclear pore complexes. *Trends in Cell Biology* **31,** 1019–1033 (2021).

121. Kim, D. I. & Roux, K. J. Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends in Cell Biology* **26,** 804–817 (2016).

122. Varnaitė, R. & MacNeill, S. A. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics* **16,** 2503–2518 (2016).

123. Gingras, A. C., Abe, K. T. & Raught, B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Current Opinion in Chemical Biology* **48,** 44–54 (2019).

124. Samavarchi-Tehrani, P., Samson, R. & Gingras, A. C. Proximity dependent biotinylation: Key enzymes and adaptation to proteomics approaches. *Molecular and Cellular Proteomics* **19,** 757–773 (2020).

125. Majewski, Ł., Sobczak, M., Havrylov, S., Jóźwiak, J. & Rdowicz, M. J. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Science* **90,** 3043–3050 (2004).

126. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **196,** 801–810 (2012).

127. Kim, D. I. I. *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 2453–2461 (2014).

128. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell* **27,** 1188–1196 (2016).

129. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology* **36,** 880–898 (2018).

130. Cho, K. F. *et al.* Split-TurboID enables contact-dependent proximity labeling in cells. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 12143–12154 (2020).

131. De Munter, S. *et al.* Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Letters* **591,** 415–424 (2017).

132. Connolly, C. N., Futter, C. E., Gibson, A., Hopkins, C. R. & Cutler, D. F. Transport into and out of the Golgi complex studied by transfecting cells with cDNAs encoding horseradish peroxidase. *Journal of Cell Biology* **127,** 641–652 (1994).

133. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature Biotechnology* **30,** 1143–1148 (2012).

134. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods* **12,** 51–54 (2014).

135. Rhee, H. W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging. *Science* **339,** 1328–1331 (2013).

136. Hung, V. *et al.* Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Molecular Cell* **55,** 332–341 (2014).

137. Chen, X., Karnovsky, A., Sans, M. D., Andrews, P. C. & Williams, J. A. Molecular characterization of the endoplasmic reticulum: Insights from proteomic studies. *Proteomics* **10,** 4040–4052 (2010).

138. Hung, V. *et al.* Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *eLife* **6** (2017).

139. Go, C. D. *et al.* A proximity-dependent biotinylation map of a human cell. *Nature* **595,** 120–124 (2021).

140. Wong, X. *et al.* Mapping the micro-proteome of the nuclear lamina and lamina-associated domains. *Life Science Alliance* **4,** 1–20 (2020).

141. Anderson, D. J., Vargas, J. D., Hsiao, J. P. & Hetzer, M. W. Recruitment of functionally distinct membrane proteins to chromatin mediates nuclear envelope formation in vivo. *Journal of Cell Biology* **186,** 183–191 (2009).

142. Benchling, [S. *No Title* 2022.

143. Hung, V. *et al.* Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nature Protocols* **11,** 456–475 (2016).

144. Avci, D. *et al.* The intramembrane protease SPP impacts morphology of the endoplasmic reticulum by triggering degradation of morphogenic proteins. *Journal of Biological Chemistry* **294,** 2786–2800 (2019).

145. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* **11,** 2301–2319 (2016).

146. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13,** 731–740 (2016).

147. Consortium, T. U. UniProt: The universal protein knowledgebase. *Nucleic Acids Research* **45** (2017).

148. Sonnhammer, E. L. L. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequence. *Sixth Int. Conf. on Intelligent Systems for Molecular Biology,* 176–182 (2008).

149. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305,** 567–580 (2001).

150. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular and Cellular Proteomics* **13,** 2513–2526 (2014).

151. Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (2002).

152. Lazar, C. *imputeLCMD: a collection of methods for left-censored missing data imputation - R-package* 2015.

153. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43,** e47 (2015).

154. Soullam, B. & Worman, H. J. Signals and structural features involved in integral membrane protein targeting to the inner nuclear membrane. *The Journal of cell biology* **130,** 15–27 (1995).

155. Theerthagiri, G., Eisenhardt, N., Schwarz, H. & Antonin, W. The nucleoporin Nup188 controls passage of membrane proteins across the nuclear pore complex. *Journal of Cell Biology* **189,** 1129–1142 (2010).

156. Salpingidou, G., Smertenko, A., Hausmanowa-Petrucewicz, I., Hussey, P. J. & Hutchison, C. J. A novel role for the nuclear membrane protein emerin in association of the centrosome to the outer nuclear membrane. *Journal of Cell Biology* **178,** 897–904 (2007).

157. Berk, J. M., Tifft, K. E. & Wilson, K. L. The nuclear envelope LEM-domain protein emerin. *Nucleus* **4,** 298–314 (2013).

158. Shimi, T. *et al.* Dynamic interaction between BAF and emerin revealed by FRAP, FLIP, and FRET analyses in living HeLa cells. *Journal of Structural Biology* **147,** 31–41 (2004).

159. Haraguchi, T. *et al.* BAF is required for emerin assembly into the reforming nuclear envelope. *Journal of Cell Science* **114,** 4575–4585 (2001).

160. Margalit, A., Brachner, A., Gotzmann, J., Foisner, R. & Gruenbaum, Y. Barrier-to-autointegration factor - a BAFfling little protein. *Trends in Cell Biology* **17,** 202–208 (2007).

161. Furukawa, K. *et al.* Barrier-to-autointegration factor plays crucial roles in cell cycle progression and nuclear organization in Drosophila. *Journal of Cell Science* **116,** 3811–3823 (2003).

162. Ni, M. W. *et al.* Modified filter-aided sample preparation (FASP) method increases peptide and protein identifications for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **31,** 171–178 (2017).

163. Tran, J. R. *et al.* An APEX2 proximity ligation method for mapping interactions with the nuclear lamina. *The Journal of Cell Biology* **220** (2021).

164. Inobe, T. & Nukina, N. Rapamycin-induced oligomer formation system of FRB-FKBP fusion proteins. *Journal of Bioscience and Bioengineering* **122,** 40–46 (2016).

165. Han, Y. *et al.* Directed evolution of split APEX peroxidase. *ACS Chemical Biology* **14,** 619–635 (2019).

166. Choi, H. *et al.* SAINT: Probabilistic scoring of affinity purificationg-mass spectrometry data. *Nature Methods* **8,** 70–73 (2011).

167. Choi, H., Glatter, T., Gstaiger, M. & Nesvizhskii, A. I. SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification – mass spectrometry experiments. *Journal of Proteome Research* **11,** 2619–2624 (2012).

168. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **138,** 389–403 (2009).

169. Cox, J. & Mann, M. 1D and 2D annotation enrichment : a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13** (2012).

170. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37,** 1–13 (2008).

171. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

172. Moser, B., Basílio, J., Gotzmann, J., Brachner, A. & Foisner, R. Comparative Interactome Analysis of Emerin, MAN1 and LEM2 Reveals a Unique Role for LEM2 in Nucleotide Excision Repair. *Cells* **9,** 463 (2020).

173. Van de Weijer, M. L. *et al.* Quality Control of ER Membrane Proteins by the RNF185/Membralin Ubiquitin Ligase Complex. *Molecular Cell* **79,** 768–781 (2020).

174. Franke, W. W., Scheer, U., Krohne, G. & Jarasch, E. D. The nuclear envelope and the architecture of the nuclear periphery. *Journal of Cell Biology* **91** (1981).

175. Feldherr, C. M. & Akin, D. The permeability of the nuclear envelope in dividing and nondividing cell cultures. *Journal of Cell Biology* **111,** 1–8 (1990).

176. Wu, W., Lin, F. & Worman, H. J. Intracellular trafficking of MAN1, an integral protein of the nuclear envelope inner membrane. *Journal of Cell Science* **115,** 1361–1372 (2002).

177. Furukawa, K., Fritze, C. E. & Gerace, L. The major nuclear envelope targeting domain of LAP2 coincides with its lamin binding region but is distinct from its chromatin interaction domain. *Journal of Biological Chemistry* **273,** 4213–4219 (1998).

178. Östlund, C., Sullivan, T., Stewart, C. L. & Worman, H. J. Dependence of diffusional mobility of integral inner nuclear membrane proteins on A-type lamins. *Biochemistry* **45,** 1374–1382 (2006).

179. Tingey, M., Mudumbi, K. C., Schirmer, E. C. & Yang, W. Casting a Wider Net: Differentiating between Inner Nuclear Envelope and Outer Nuclear Envelope Transmembrane Proteins. *International Journal of Molecular Sciences* **20** (2019).

180. Kley, N. Fusing Visual and Clinical Information for Lung Tissue Classification in HRCT Data. *Chemistry and Biology* **11** (2004).

181. Kolos, J. M., Voll, A. M., Bauder, M. & Hausch, F. FKBP Ligands—Where We Are and Where to Go? *Frontiers in Pharmacology* **9** (2018).

182. Stanton, B. Z., Chory, E. J. & Crabtree, G. R. Chemically induced proximity in biology and medicine. *Science* **359** (2018).

183. Shyu, Y. J., Liu, H., Deng, X. & Hu, C. D. Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. *BioTechniques* **40,** 61–66 (2006).

184. Feng, S. *et al.* Improved split fluorescent proteins for endogenous protein labeling. *Nature Communications* **8,** 1–11 (2017).

185. Feng, S. *et al.* Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. *Communications Biology* **2,** 1–12 (2019).

186. Colbeau, A., Nachbaur, J. & Vignais, P. M. Enzymac characterization and lipid composition of rat liver subcellular membranes. *BBA - Biomembranes* **249,** 462–492 (1971).

187. Adam, S. A., Marr, R. S. & Gerace, L. Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. *Journal of Cell Biology* **111,** 807–816 (1990).

188. Zheng, W., Wang, X., Hu, J., Bai, B. & Zhu, H. Diverse molecular functions of aspartate β-hydroxylase in cancer (Review). *Oncology Reports* **44,** 2364–2372 (2020).

189. Kihara, A. Very long-chain fatty acids: Elongation, physiology and related disorders. *Journal of Biochemistry* **152,** 387–395 (2012).

190. Leonard, A. E., Pereira, S. L., Sprecher, H. & Huang, Y. S. Elongation of long-chain fatty acids. *Progress in Lipid Research* **43,** 36–54 (2004).

191. Jakobsson, A., Westerberg, R. & Jacobsson, A. Fatty acid elongases in mammals: Their regulation and roles in metabolism. *Progress in Lipid Research* **45,** 237–249 (2006).

192. Lucki, N. C. & Sewer, M. B. Nuclear Sphingolipid Metabolism Natasha. *Annu. Rev. Physiol.* **74,** 131–151 (2012).

193. Peluso, J. J., Griffin, D., Liu, X. & Horne, M. Progesterone receptor membrane component-1 (PGRMC1) and PGRMC-2 Interact to suppress entry into the cell cycle in spontaneously immortalized rat granulosa cells. *Biology of Reproduction* **91,** 1–12 (2014).

194. Xu, J. *et al.* Identification of the PGRMC1 protein complex as the putative sigma-2 receptor binding site. *Nature Communications* **2** (2011).

195. Peluso, J. J., Lodde, V. & Liu, X. Progesterone regulation of Progesterone Receptor Membrane Component 1 (PGRMC1) sumoylation and transcriptional activity in spontaneously immortalized granulosa cells. *Endocrinology* **153,** 3929–3939 (2012).

196. Ryu, C. S., Klein, K. & Zanger, U. M. Membrane associated progesterone receptors: Promiscuous proteins with pleiotropic functions - Focus on interactions with cytochromes P450. *Frontiers in Pharmacology* **8,** 1–8 (2017).

197. Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168,** 527–541 (2017).

198. Juhlen, R., Landgraf, D., Huebner, A. & Koehler, K. Identification of a novel putative interaction partner of the nucleoporin ALADIN. *Biology Open* **5,** 1697–1705 (2016).

199. Nishi, M., Komazaki, S., Iino, M., Kangawa, K. & Takeshima, H. Mitsugumin23, a novel transmembrane protein on endoplasmic reticulum and nuclear membranes. *FEBS Letters* **432,** 191–196 (1998).

200. Venturi, E. *et al.* Mitsugumin 23 forms a massive bowl-shaped assembly and cation-conducting channel. *Biochemistry* **50,** 2623–2632 (2011).

201. Reilly-O'Donnell, B. *et al.* Dysregulated Zn2 homeostasis impairs cardiac type-2 ryanodine receptor and mitsugumin 23 functions, leading to sarcoplasmic reticulum Ca2 leakage. *Journal of Biological Chemistry* **292,** 13361–13373 (2017).

202. Yamashita, A., Taniwaki, T., Kaikoi, Y. & Yamazaki, T. Protective role of the endoplasmic reticulum protein mitsugumin23 against ultraviolet C-induced cell death. *FEBS Letters* **587,** 1299–1303 (2013).

203. Levy, M. & Futerman, A. H. Mammalian ceramide synthases. *IUBMB Life* **62,** 347–356 (2010).

204. Laviad, E. L. *et al.* Characterization of ceramide synthase 2: Tissue distribution, substrate specificity, and inhibition by sphingosine 1-phosphate. *Journal of Biological Chemistry* **283,** 5677–5684 (2008).

205. White-Gilbertson, S. *et al.* Ceramide synthase 6 modulates TRAIL sensitivity and nuclear translocation of active caspase-3 in colon cancer cells. *Oncogene* **28,** 1132–1141 (2009).

206. Rampello, A. J. *et al.* Torsin ATPase deficiency leads to defects in nuclear pore biogenesis and sequestration of MLF2. *Journal of Cell Biology* **219** (2020).

207. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184,** 3022–3040 (2021).

208. Ikeda, M. *et al.* Characterization of four mammalian 3-hydroxyacyl-CoA dehydratases involved in very long-chain fatty acid synthesis. *FEBS Letters* **582,** 2435–2440 (2008).

209. Sassa, T. & Kihara, A. Metabolism of very long-chain fatty acids: Genes and pathophysiology. *Biomolecules and Therapeutics* **22,** 83–92 (2014).

210. Moon, Y. A. & Horton, J. D. Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade. *Journal of Biological Chemistry* **278,** 7335–7343 (2003).

211. Wakashima, T., Abe, K. & Kihara, A. Dual functions of the trans-2-enoyl-CoA reductase TER in the sphingosine 1-phosphate metabolic pathway and in fatty acid elongation. *Journal of Biological Chemistry* **289,** 24736–24748 (2014).

212. Kohlwein, S. D. *et al.* Tsc13p Is Required for Fatty Acid Elongation and Localizes to a Novel Structure at the Nuclear-Vacuolar Interface in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **21,** 109–125 (2001).

213. Kvam, E., Gable, K., Dunn, T. M. & Goldfarb, D. S. Targeting of Tsc13p to Nucleus–Vacuole Junctions: A Role for Very-Long-Chain Fatty Acids in the Biogenesis of Microautophagic Vesicles. *Molecular Biology of the Cell* **16,** 3987–3998 (2005).

214. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic acids research* **36,** 5–9 (2008).

215. Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic acids research* **41,** 29–33 (2013).

216. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–410 (1990).

217. States, D. J. & Gish, W. Combined Use of Sequence Similarity and Codon Bias for Coding Region Identification. *Journal of Computational Biology* **1,** 39–50 (1994).

218. Acland, A. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **41,** 8–20 (2013).

219. Moylan, J. S. *et al.* Neutral sphingomyelinase-3 mediates TNF-stimulated oxidant activity in skeletal muscle. *Redox Biology* **2,** 910–920 (2014).

220. Breslow, D. K. Sphingolipid homeostasis in the endoplasmic reticulum and beyond. *Cold Spring Harbor Perspectives in Biology* **5,** 1–16 (2013).

221. Krut, O., Wiegmann, K., Kashkar, H., Yazdanpanah, B. & Krönke, M. Novel tumor necrosis factor-responsive mammalian neutral sphingomyelinase-3 is a C-tail-anchored protein. *Journal of Biological Chemistry* **281,** 13784–13793 (2006).

222. Magini, P. *et al.* Loss of SMPD4 Causes a Developmental Disorder Characterized by Microcephaly and Congenital Arthrogryposis. *American Journal of Human Genetics* **105,** 689–705 (2019).

223. Dreger, M. *et al.* Nuclear envelope proteomics: Novel integral membrane proteins of the inner nuclear membrane. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 11943–11948 (2001).

224. Piët, A. C., Post, M., Dekkers, D., Demmers, J. A. & Fornerod, M. Proximity Ligation Mapping of Microcephaly Associated SMPD4 Shows Association with Components of the Nuclear Pore Membrane. *Cells* **11** (2022).

225. Otsuka, S. *et al.* Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *eLife* **5,** 1–23 (2016).

226. Züchner, S. *et al.* Mutations in the novel mitochondrial protein REEP1 cause hereditary spastic paraplegia type 31. *American Journal of Human Genetics* **79,** 365–369 (2006).

227. Lim, Y., Cho, I. T., Schoel, L. J., Cho, G. & Golden, J. A. Hereditary spastic paraplegia-linked REEP1 modulates endoplasmic reticulum/mitochondria contacts. *Annals of Neurology* **78,** 679–696 (2015).

228. Schlaitz, A.-l., Thompson, J., Wong, C. C. L., Iii, J. R. Y. & Heald, R. Short Article Clearance from Metaphase Chromatin and Proper Nuclear Envelope Architecture. *Developmental Cell* **26,** 315–323 (2013).

229. Kumar, D., Golchoubian, B., Belevich, I., Jokitalo, E. & Schlaitz, A. L. REEP3 and REEP4 determine the tubular morphology of the endoplasmic reticulum during mitosis. *Molecular Biology of the Cell* **30,** 1377–1389 (2019).

230. El Khouri, E., Le Pavec, G., Toledano, M. B. & Delaunay-Moisan, A. RNF185 is a novel E3 ligase of endoplasmic reticulum-associated degradation (ERAD) that targets cystic fibrosis transmembrane conductance regulator (CFTR). *Journal of Biological Chemistry* **288,** 31177–31191 (2013).

231. Liu, C. H. *et al.* Combining Proximity Labeling and Cross-Linking Mass Spectrometry for Proteomic Dissection of Nuclear Envelope Interactome. *Journal of Proteome Research* **19,** 1109–1118 (2020).

232. Bar, D. Z. *et al.* Biotinylation by antibody recognition - A method for proximity labeling. *Physiology & behavior* **176,** 139–148 (2017).

233. Xie, W. *et al.* A-type Lamins Form Distinct Filamentous Networks with Differential Nuclear Pore Complex Associations. *Current Biology* **26,** 2651–2658 (2016).

234. Tang, Y., Huang, A. & Gu, Y. Global profiling of plant nuclear membrane proteome in Arabidopsis. *Nature Plants* **6,** 838–847 (2020).

235. Huang, A. *et al.* Proximity labeling proteomics reveals critical regulators for inner nuclear membrane protein degradation in plants. *Nature Communications* **11,** 1–10 (2020).

236. Ramanathan, M. *et al.* RN A-protein interaction detection in living cells. *Nature Methods* **15,** 207–212 (2018).

237. Ledeen, R. W. & Wu, G. Sphingolipids of the nucleus and their role in nuclear signaling. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1761,** 588–598 (2006).

238. Schneiter, R. *et al.* A yeast acetyl coenzyme A carboxylase mutant links very-long-chain fatty acid synthesis to the structure and function of the nuclear membrane-pore complex. *Molecular and Cellular Biology* **16,** 7161–7172 (1996).

239. Schneiter, R. *et al.* Identification and biophysical characterization of a very-long-chain-fatty- acid-substituted phosphatidylinositol in yeast subcellular membranes. *Biochemical Journal* **381,** 941–949 (2004).

240. Van Engelen, B. G. *et al.* The lethal phenotype of a homozygous nonsense mutation in the lamin A/C gene. *Neurology* **64,** 374–376 (2005).

241. Kato, K. *et al.* LMNA missense mutation causes nonsense- mediated mRNA decay and severe dilated cardiomyopathy. *Circulation: Genomic and Precision Medicine,* 435–443 (2020).

242. Sieprath, T. *et al.* Sustained accumulation of prelamin A and depletion of lamin A/C both cause oxidative stress and mitochondrial dysfunction but induce different cell fates. *Nucleus* **6,** 236–246 (2015).

243. Fong, L. G. *et al.* Heterozygosity for Lmna deficiency eliminates the progeria-like phenotypes in Zmpste24-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 18111–18116 (2004).

244. Bertrand, A. T. *et al.* Lamin A/C Assembly Defects in LMNA-Congenital Muscular Dystrophy Is Responsible for the Increased Severity of the Disease Compared with Emery-Dreifuss Muscular Dystrophy. *Cells* **9,** 1–17 (2020).

245. Nikolova, V. *et al.* Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *Journal of Clinical Investigation* **113,** 357–369 (2004).

246. Liang, Y., Chiu, P. H., Yip, K. Y. & Chan, S. Y. Subcellular localization of SUN2 is regulated by lamin a and Rab5. *PLoS ONE* **6** (2011).

247. Kubben, N. *et al.* Identification of differential protein interactors of lamin a and progerin. *Nucleus* **1,** 513–525 (2010).

248. Fueller, J. *et al.* CRISPR-Cas12a-assisted PCR tagging of mammalian genes. *Journal of Cell Biology* **219** (2020).

249. Erdbrügger, P. & Fröhlich, F. The role of very long chain fatty acids in yeast physiology and human diseases. *Biological Chemistry* **402,** 25–38 (2020).

250. Ma, H. T. & Poon, R. Y. C. Synchronization of HeLa cells. *Methods in Molecular Biology* **761,** 151–161 (2011).

251. Lu, L., Ladinsky, M. S. & Kirchhausen, T. Formation of the postmitotic nuclear envelope from extended ER cisternae precedes nuclear pore assembly. *Journal of Cell Biology* **194,** 425–440 (2011).

Appendices

A
Appendix 1

A.1 Proteomic data collection

Table A.1: Proteomic data collection. APF - Advanced proteomics facility Oxford; TDI - Target Discovery Institute Oxford; CfMSA - Center for MS Analytics, Kaiserslautern. Biological replicates: biologically distinct samples; biochemical replicates: samples split after fractionation but prior to IP.

Bait	**Replicate**	**Facility**	**Collection date**	**Note**
Climp63, NLS, $Nup153_{1-144}$	1	TDI	07/2019	-
Climp63, NLS, $Nup153_{1-144}$	2	TDI	09/2019	-
Climp63, NLS, $Nup153_{1-144}$, negative control	3	APF	10/2019	Negative control was NLS sample without BP added
Climp63 *LMNA* KO, $Nup153_{1-144}$ *LMNA* KO	1–4	CfMSA	12/2020	Two biological replicates, each split further into two biochemical replicates.
$Nup153_{1-144}$	4–5	CfMSA	07/2020	Negative control of a cell cycle synchronisation experiment.

Continuation of table A.2				
Bait	**Replicate**	**Facility**	**Collection date**	**Note**
$Nup153_{1-144}$	6–9	CfMSA	11/2020	Negative control of a cell cycle synchronisation experiment.
Rtn4a	1–4	CfMSA	05/2021	Two biological replicates, each split further into two biochemical replicates.
Sec61β	1–6	CfMSA	03/2021	Three biological replicates, each split further into two biochemical replicates.
End of table				

A.2 Organelle marker proteins

Table A.2: Organelle marker proteins. Table adapted from Chen *et al.*, 2010.

Uniprot ID	Gene name	TMD	Domain
Q8N9A8	CNEP1R1	2	INM
O95476	CTDNEP1	1	INM
P50402	EMD	1	INM
Q14739	LBR	8	INM
Q8NC56	LEMD2	2	INM
P02545	LMNA	-	INM
P20700	LMNB1	-	INM
Q03252	LMNB2	-	INM
Q9Y2U8	LEMD3	2	INM
O14524	NEMP1	5	INM
Q8IXM6	NRM	4	INM
Q9BWN1	PRR14	-	INM
O94901	SUN1	1	INM
Q9UH99	SUN2	1	INM
Q8NF91	SYNE1	1	INM
Q8WXH0	SYNE2	1	INM
Q6ZMZ3	SYNE3	1	INM
Q5SNT2	TMEM201	6	INM
Q96SK2	TMEM209	2	INM
Q9BTV4	TMEM43	4	INM
P42166	TMPOα	-	INM
P42167	TMPOβ/γ	1	INM

Continuation of table A.2			
Uniprot ID	**Gene name**	**TMH**	**Domain**
Q5JTV8	TOR1AIP1	-	INM
P57088	TMEM33	3	INM/ER
Q8NFQ8	TOR1AIP2	1	INM/ER
Q9NRG9	AAAS	-	NPC
Q53GS7	GLE1	-	NPC
O60318	MCM3AP	-	NPC
Q9BTX1	NDC1	6	NPC
P57740	NUP107	-	NPC
Q8WUM0	NUP133	-	NPC
P49790	NUP153	-	NPC
O75694	NUP155	-	NPC
Q12769	NUP160	-	NPC
Q5SRE5	NUP188	-	NPC
P35658	NUP214	-	NPC
Q8NFH5	NUP35	-	NPC
Q8NFH4	NUP37	-	NPC
Q8NFH3	NUP43	-	NPC
Q9UKX7	NUP50	-	NPC
Q7Z3B4	NUP54	-	NPC
Q9BVL2	NUP58	-	NPC
P37198	NUP62	-	NPC
Q9BW27	NUP85	-	NPC
Q99567	NUP88	-	NPC
Q8N1F7	NUP93	-	NPC
P52948	NUP98	-	NPC
O15504	NUP42	-	NPC
Q96HA1	POM121	1	NPC
P78406	RAE1	-	NPC
P49792	RANBP2	-	NPC
P55735	SEC13	-	NPC/ER
Q96EE3	SEH1L	-	NPC
Q9P0U3	SENP1	-	NPC/cytoplasm
Q9HC62	SENP2	-	NPC/nucleus
P12270	TPR	-	NPC
Q8TEM1	NUP210	2	NPC
Q8WYP5	AHCTF1	-	NPC
Q6DD88	ATL3	2	ER
Q9HD20	ATP13A1	7	ER
P16615	ATP2A2	8	ER
P51572	BCAP31	3	ER
P27824	CANX	1	ER
Q07065	CKAP4	1	ER

Continuation of table A.2			
Uniprot ID	**Gene name**	**TMH**	**Domain**
P00167	CYB5A	1	ER
Q86UP2	KNT1	1	ER
P04843	RPN1	1	ER
P04844	RPN2	4	ER
Q9P2E9	RRBP1	1	ER
Q9NQC3	RTN4	2	ER
Q9UGP8	SEC63	3	ER
O95470	SGPL1	-	ER
Q15005	SPCS2	2	ER
O15270	SPTLC2	-	ER
Q9Y5M8	SRPRB	1	ER
P43307	SSR1	1	ER
P51571	SSR4	-	ER
Q969X5	ERGIC1	2	ER/Golgi
Q9Y282	ERGIC3	2	ER/Golgi
Q92896	GLG1	1	Golgi
Q8TBA6	GOLGA5	1	Golgi
O00461	GOLIM4	1	Golgi/Endosome
P49257	LMAN1	1	ER/Golgi
Q12907	LMAN2	1	ER/Golgi
Q16706	MAN2A1	1	Golgi
Q5JRA6	MIA3	-	ER/Golgi
O75396	SEC22B	1	ER/Golgi
Q15363	TMED2	2	ER/Golgi
Q14318	FKBP8	1	ER/Mitochondria
Q16891	IMMT	-	Mitochondria
O95202	LETM1	1	Mitochondria
Q791V5	MTCH2	-	Mitochondria
O43674	NDUFB5	1	Mitochondria
Q13423	NNT	12	Mitochondria
Q02978	SLC25A11	-	Mitochondria
O75746	SLC25A12	-	Mitochondria
Q00325	SLC25A3	2	Mitochondria
P12235	SLC25A4	3	Mitochondria
P05141	SCL25A5	3	Mitochondria
Q3ZCQ8	TIMM50	1	Mitochondria
Q15388	TOMM20	1	Mitochondria
Q9NS69	TOM22	1	Mitochondria
O94826	TOMM70	1	Mitochondria
P45880	VDAC2	-	Mitochondria
P00403	MT-CO2	2	Mitochondria
P33527	ABCC1	16	Plasma membrane

Continuation of table A.2			
Uniprot ID	**Gene name**	**TMH**	**Domain**
Q9HDC9	APMAP	1	Plasma membrane
P29972	AQP1	6	Plasma membrane
P05023	ATP1A1	10	Plasma membrane
Q01814	ATP2B2	8	Plasma membrane
Q93050	ATP6V0A1	7	Plasma membrane
P27449	ATP6V0C	4	Plasma membrane
P19022	CDH2	1	Plasma membrane
O75923	DYSF	1	Plasma membrane
P06756	ITGAV	1	Plasma membrane
P18084	ITGB5	1	Plasma membrane
P11279	LAMP1	2	Plasma membrane
P08473	MME	1	Plasma membrane
Q9UBG0	MRC2	1	Plasma membrane
O15031	PLXNB2	-	Plasma membrane
P08195	SLC3A2	1	Plasma membrane / Lysosome
End of table			

A.3 SUN2 profile plot hits

Table A.3: SUN2 reference profile top 50 hits

Uniprot ID	Gene name	TMD
Q9UH99	SUN2	1
P52948	NUP98	-
P42167	LAP2B	1
Q9Y2W6	TDRKH	1
Q5SNT2	TM201	6
Q5JRX3	PREP	-
Q5JTV8	TOIP1	-
Q96SK2	TMEM209	-
Q9NXE4-4	NSMA3	1
O75694	NU155	-
O60762	DPM1	-
Q9BTV4	TMM43	4
P50402	EMD	1
Q13724	MOGS	1
Q9Y5M8	SRPRB	1
Q96AG4	LRC59	1

Continuation of table A.3		
Uniprot ID	**Gene name**	**TMH**
P04406	G3P	-
A1L0T0	ILVBL	1
Q8N2G8	GHDC	-
Q92604	LGAT1	4
Q07065	CKAP4	1
-	APEX	-
Q8TEY7	UBP33	-
O60725	ICMT	4
P28288	ABCD3	3
O15173	PGRC2	1
Q8NHP6	MSPD2	1
Q03518	TAP1	7
Q96JB5	CK5P3	-
Q96A57	TM230	2
Q9Y673	ALG5	1
Q9NZ01	TECR	4
P49069	CAMLG	-
Q6P1M0	S27A4	2
Q9NP72	RAB18	-
Q8IZ81	ELMD2	-
O14967	CLGN	1
Q86UL3	GPAT4	3
Q9BQE5	APOL2	3
P61619	S61A1	10
Q5BJH7-3	YIF1B	5
Q15392	DHC24	2
Q9BSJ8	ESYT1	2
Q8TCJ2	STT3B	10
O43169	CYB5B	1
P14373	TRI27	-
P17706	PTN2	1
P25398	RS12	-
Q9BVT8	TMUB1	3
Q9HD20	AT131	ER
End of table		

A.4 Emerin profile plot hits

Table A.4: Emerin reference profile top 50 hits

Uniprot ID	Gene name	TMD
P50402	EMD	1
P52948	NUP98	-
Q8NFH5	NUP35	-
P28288	ABCD3	3
Q92841	DDX17	-
Q96A57	TM230	2
Q5JRX3	PREP	-
Q5JTV8	TOIP1	-
Q5BJH7	YIF1B	5
P17706	PTN2	1
O75694	NU155	-
O60762	DPM1	-
Q8NHP6	MSPD2	1
Q15392	DHC24	2
Q5SNT2	TM201	6
P42167	LAP2B	1
P04406	G3P	1
O60725	ICMT	4
P07099	HYEP	-
Q8N2G8	GHDC	-
P18124	RL7	-
Q6P1M0	S27A4	2
Q9UL25	RAB21	-
-	APEX	-
Q8TEY7	UBP33	-
Q03518	TAP1	7
Q9UH99	SUN2	1
Q9NXE4	NSMA3	1
O43169	CYB5B	1
P61619	S61A1	1
Q96JB5	CK5P3	-
Q07065	CKAP4	1
A1L0T0	ILVBL	1
Q86UL3	GPAT4	3
P49069	CAMLG	-
Q9BTV4	TMM43	4
Q9NP72	RAB18	-
Q9HD20	AT131	7
Q9Y5M8	SRPRB	1

Continuation of table A.4		
Uniprot ID	**Gene name**	**TMH**
O15173	PGRC2	1
O15121	DEGS	6
Q96SK2	TM209	2
Q12907	LMAN2	1
O14967	CLGN	1
Q9BSJ8	ESYT1	2
Q9Y673	ALG5	1
Q9NZ01	TECR	4
P14373	TRI27	-
Q13724	MOGS	1
Q9HBH5	RDH14	1
End of table		