

**Characterisation of
Novel Protein
'WDR98/RMC1' and its
Role in Lysosome
Biology**

Kate Densley BSc

MSc by Research
University of York
Biology



August 2019

**“Started a PhD, had a breakdown, Bon appetit.” -
adapted from James Acaster**

Abstract

Following the results from a pilot study indicating the interaction of endosomal maturation proteins MON1B and CCZ1 with the uncharacterised protein 'WDR98/RMC1', this potentially new protein complex was investigated further. Literature research and bioinformatics revealed the highly evolutionarily conserved nature of WDR98 and also its unique C-terminal 'MIC' domain (the N-terminal containing the common protein binding platform domain WD40). Immunoprecipitation methods were used to confirm the, then novel, interaction of WDR98 with MON1B and CCZ1 individually. Furthermore, it was discovered that the WD40 domain is alone capable of eliciting binding between these proteins. Though reported elsewhere that WDR98 was able to bind to RAB5 and RAB7 in a whole-lysate context, we are confident that WDR98 is unable to bind to either RAB in isolation *in vitro*. Immunofluorescence was utilised to localise WDR98 to endosomes both early (RAB5) and late (RAB7) and lysosomes (LGP120), specifically those most active around the nuclear periphery. WDR98 was not seen to localise to autophagosomes (LC3). Both the WD40 and MIC domains also localised to lysosomes individually. This, along with recent literature, indicates an important role for WDR98 in protein trafficking, specifically endosomal maturation and lysosome biology. In addition, its potential role in disease and the emerging disease-contribution of the lysosome, particularly concerning cancer, renders this protein an exciting discovery, with potential as a target for future investigations.

List of Contents

Abstract.....	3
List of Contents.....	4
List of Figures.....	8
List of Tables.....	9
Acknowledgements.....	10
Declaration.....	11
Chapter 1: Introduction.....	12
1.1 Lysosomes.....	12
1.2 Endosomal Maturation.....	14
1.3 WDR98 Identification.....	15
1.4 WDR98 in Literature.....	16
1.5 Lysosomes and WDR98 in Disease.....	16
1.6 Model Systems.....	18
1.7 Aims and Objectives.....	18
Chapter 2: Materials and Methods.....	20
2.1 Materials.....	20
2.1.1 Antibodies.....	20
2.2 Methodology.....	21
2.2.1 General Polymerase Chain Reaction (PCR).....	22
2.2.2 QuickChange PCR.....	22
2.2.3 DNA Electrophoresis and Purification.....	23
2.2.4 DNA Restriction Digests.....	23
2.2.5 Bacterial Transformation and DNA Extractions.....	23
2.2.6 Recombinant Protein Production.....	24
2.2.7 Mammalian Cell Culture.....	25
2.2.8 Mammalian Cell Transfection.....	25
2.2.9 CRISPR/Cas9 Knockouts.....	25
2.2.10 Detergent Soluble Lysates.....	26
2.2.11 Protein Quantification Assay.....	26
2.2.12 Antibody Coupling.....	27
2.2.13 Immunoprecipitation.....	27

2.2.14 Rab Nucleotide Loading.....	27
2.2.15 Pulldowns.....	28
2.2.16 PLATE Transformation of Yeast.....	28
2.2.17 Yeast 2 Hybrid.....	29
2.2.18 SDS PAGE.....	29
2.2.19 Western Blotting.....	30
2.2.20 Mass Spectrometry Sample Preparation.....	30
2.2.21 Liquid Chromatography Mass Spectrometry.....	31
2.2.22 Mass Spectrometry Data Analysis.....	31
2.2.23 Immunofluorescence.....	32
Chapter 3: Identifying and Characterising WDR98.....	34
3.1 Introduction.....	34
3.2 Methods.....	35
3.2.1 Phylogenetic Analysis.....	35
3.2.2 Commercial Antibody Characterisation.....	35
3.2.3 DNA Construct Production.....	36
3.2.4 Antibody Production and Characterisation.....	36
3.2.5 Cell Line Production.....	36
3.2.6 CRISPR/Cas9.....	36
3.3 Results.....	39
3.3.1 Evolutionary Conservation.....	41
3.3.2 Commercial Antibody Testing.....	41
3.3.3 DNA Constructs.....	44
3.3.4 Commissioned Antibody Testing.....	44
3.3.5 Cell Line Characterisation.....	46
3.3.6 CRISPR/Cas9 Knockout Characterisation.....	46
3.4 Discussion.....	47
3.4.1 Assessment of Methods.....	49
3.4.2 Contribution to the Project.....	50
3.4.3 Future Directions.....	50
Chapter 4: WDR98 Protein-Protein Interactions.....	50
4.1 Introduction.....	50
4.2 Results.....	52

4.2.1 Yeast 2 Hybrid Interactions.....	52
4.2.2 WDR98 and WD40 Domain Interacts with CCZ1 and MON1B.....	53
4.2.3 Alternative Immunoprecipitation Interactions.....	55
4.2.4 RAB Pulldowns.....	57
4.2.5 Immunoprecipitation Liquid Chromatography Mass Spectrometry.....	58
4.3 Discussion.....	60
4.3.1 Confirmation of Interactions.....	60
4.3.2 Assessment of Methods.....	61
4.3.3 Future Directions.....	63
Chapter 5: WDR98 Subcellular Localisation.....	64
5.1 Introduction.....	64
5.2 Methods.....	64
5.2.1 WDR98 Overexpression.....	64
5.2.2 Transfection Optimisation.....	65
5.2.3 Image Analysis.....	65
5.3 Results.....	65
5.3.1 WDR98 Subcellular Localisation.....	65
5.3.2 WDR98 Truncated Domains Subcellular Localisation.....	69
5.3.3 WDR98 Overexpression Phenotype.....	71
5.4 Discussion.....	72
5.4.1 Confirmation of Localisation.....	72
5.4.2 Assessment of Methods.....	73
5.4.3 Future Directions.....	74
Chapter 6: Conclusions and Future Directions.....	75
6.1 Characterisation of WDR98.....	75
6.2 Corroboration with Publications.....	75
6.3 WDR98 in Context.....	76
6.3.1 WDR98 in Health.....	76
6.3.2 WDR98 in Disease.....	77
6.4 Future Directions.....	78
Appendices.....	79
A. DNA Construct Coding Sequences.....	79

i. GST-WDR98.....	79
ii. WDR98-HIS.....	80
iii. GFP-WDR98.....	81
iv. GFP-WD40.....	82
v. GFP-MIC.....	83
vi. Δ pMEP4-MYC-GFP-WDR98.....	83
vii. pcDNA5-GFP-humanWDR98.....	84
viii. WDR98 gDNA.....	86
ix. CRISPR/Cas9 Target Sequences.....	94
1. Exon 1.....	94
2. Exon 8.....	94
3. Exon 9.....	94
4. Exon 10.....	94
x. pGAD-WDR98.....	94
xi. pGBDU-WDR98.....	95
xii. GST-RAB5A.....	96
xiii. GST-RAB7A.....	97
List of Abbreviations.....	98
References.....	101

List of Figures

Figure 1: Routes to the Lysosome

Figure 2: The Interactions of MON1 and CCZ1 with Rab5 and Rab7

Figure 3: WDR98 Protein Domains

Figure 4: Evolutionary History of WDR98

Figure 5: Commercial Antibodies Fail to Recognise WDR98 by Western Blotting

Figure 6: Commissioned GST-WDR98 Antibodies Fail to Recognise WDR98 by Western Blotting or Immunofluorescence

Figure 7: WDR98-HIS Exhibits Less Protein Degradation than GST-WDR98

Figure 8: Commissioned WDR98-HIS Antibody may show WDR98 Specificity

Figure 9: WDR98 CRISPR/Cas9 Chromatograms

Figure 10: Yeast 2 Hybrid Method

Figure 11: MON1B and CCZ1 Interact via Yeast 2 Hybrid

Figure 12: WDR98 and WD40 Domain Interacts with CCZ1 and MON1B

Figure 13: WDR98 Potential Protein Interactions Network

Figure 14: WDR98 does not Interact with HDAC6, HSP90, STUB1 or APP

Figure 15: WDR98 does not Directly Interact with RAB5/7 *in vitro*

Figure 16: WDR98 Colocalises with LGP120, CCZ1, MON1B, RAB5 and RAB7

Figure 17: WDR98 and its Domains Localise to the Lysosome

Figure 18: Transfection of GFP-Constructs causes Lysosomal Swelling

List of Tables

Table 1. Details of Antibodies

Table 2. PCR Reaction Parameters

Table 3: MON1B and Structural Proteins are Binding Partners of WDR98

Table 4: Pearson Correlation Coefficients of GFP-WDR98 Subcellular Colocalisation

Table 5: Pearson Correlation Coefficients of GFP-WDR98 Truncated Constructs at the Lysosome

Acknowledgements

Thank you to my main funding body BBSRC and to the University of York for the award of a Facilities Grant which enabled the LC-MS/MS experiments to take place.

Thank you to my supervisors: Dr Paul Pryor, Dr Sean Sweeney and Dr Christoph Baumann for their time and suggestions and to other staff in the Biology department, including members of my TAP panel: Professor Paul Genever, Dr Gonzalo Blanco, Dr Gareth Evans, Dr Sangeeta Chawla, and Dr Adam Dowle for his help with LC-MS/MS, Dr Graeme Park for his help with microscopy and FACS and Professor Gavin Thomas, Professor Richard Waites and Monica Bandeira for their help and advice. Thank you to the Bioscience Technology Facility for the use of their equipment and the advice from staff.

I would like to thank the York Graduate Students' Association, particularly Emma Coward-Gibbs, Susie Gridley, Chris Bovis, Peter Gorbet and Amy Bullard. Your support over the past three years has been invaluable and I feel very lucky to have been at an institution with a postgraduate student union.

The University of York Softball and Baseball Club has made my time at York enjoyable and it was an honour to be your treasurer and president. Thank you to Matt Tomlin for the difference this has made to my life.

Finally, thank you to my friends and family for helping to get me through this incredibly difficult time.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. Exceptions to this are as follows: HHPred analysis, generation of pGAD and pGBDU constructs by the Pryor lab except where stated, generation of Yeast 2 Hybrid data replicates by the Pryor lab, generation of MYC-MON1B NRK cell lines by the Pryor lab, LC-MS and analysis by Dr Adam Dowle, Bioscience Technology Facility. All references to other work are held within the text. This work has not previously been presented for an award at this, or any other, university. All sources are acknowledged as references.

Chapter 1: Introduction

1.1 Lysosomes

Lysosomes use hydrolytic enzymes to break down many biological molecules, including those from the endocytic, phagocytic and autophagic pathways (Luzio *et al.* 2007). They are involved in a variety of cellular processes, not solely degradative, and contribute to cellular homeostasis regulation. As such, lysosomal dysfunction often contributes to disease pathogenesis, including that of lysosomal storage diseases, neurodegeneration and certain cancers (Appelqvist *et al.* 2013). The lysosomal membrane has a uniquely high carbohydrate content owing to the presence of heavily glycosylated proteins, the most abundant being lysosome-associated membrane protein (LAMP)-1 and -2 (Appelqvist *et al.* 2013). Lysosomal membrane proteins mediate many crucial functions, including luminal acidification, transport of small molecules and/or macromolecules, and potentially lysosomal interactions and fusions (Eskelinen 2006).

There are several routes of transport to the lysosome, including autophagy, endocytosis and phagocytosis (Figure 1).

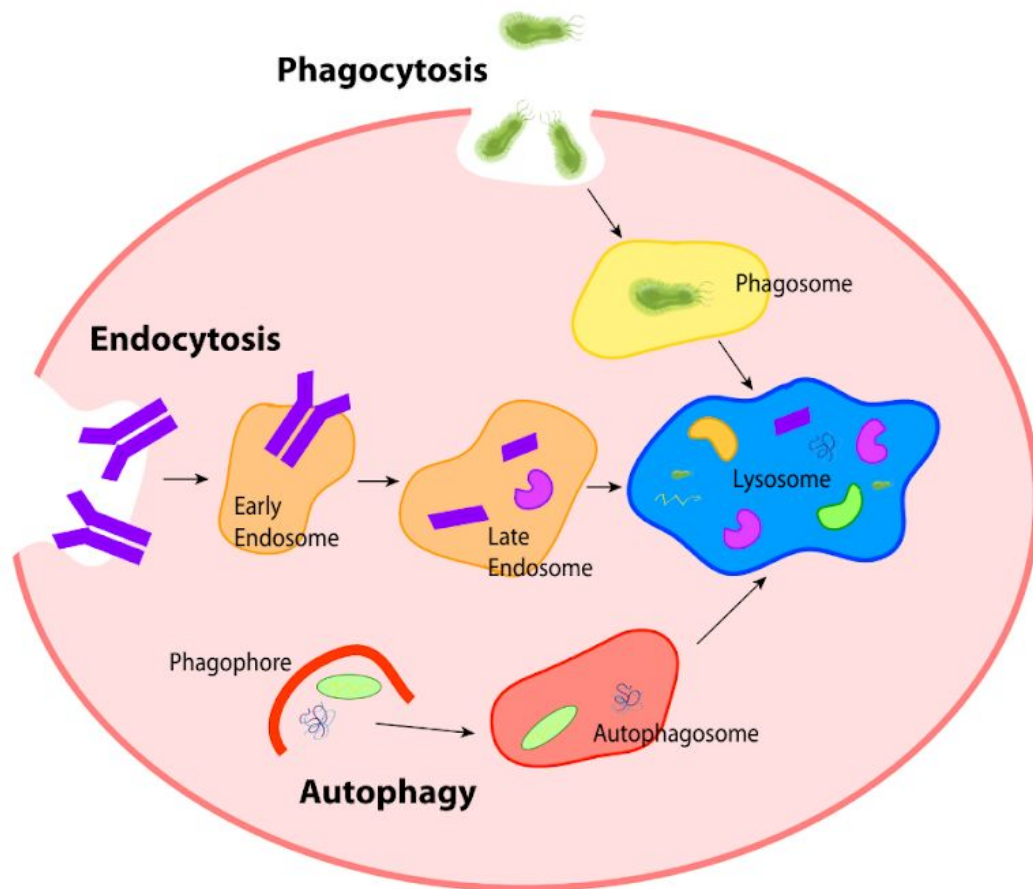


Figure 1: Routes to the Lysosome

Routes of transport to the lysosome: phagocytosis, endocytosis and autophagy. Figure adapted from Xu et al. (2018).

Phagocytosis and endocytosis work to bring molecules external to the cell inside, and traffic these through the generation of vesicles (phagosomes and endosomes) before fusion with the lysosome (Richards and Endres 2017). Autophagy initiates the formation of a partial membrane - a phagophore - which begins to engulf particular intracellular molecules that have been marked for degradation. This matures to an autophagosome before lysosomal fusion (Eskelinen and Saftig 2009).

Autophagy is required to maintain homeostasis and proteomic function by protein degradation and turnover of organelles. There are three types of autophagy, distinguishable by their cargo delivery mechanisms (Kaushik *et al.* 2011). Macroautophagy eradicates damaged organelles or proteins by sequestering regions of the cytosol into double-membraned autophagosomes which fuse with the lysosome (Figure 1) (Kaushik & Cuervo 2012). Microautophagy encases proteins for

degradation inside vesicles via invagination of the lysosomal membrane, which is a more direct engulfment of cytoplasmic material into the lysosome (Cuervo & Wong 2014). Chaperone-mediated autophagy (CMA) is distinct in that it selects individual, soluble proteins from a selective subset, and transports these directly to the lysosomal membrane where they translocate across into the lumen (Kaushik *et al.* 2011).

1.2 Endosomal Maturation

RAB proteins are small GTPases required for membrane fusion. RAB5 is known to function in early endocytic fusion events, and RAB7 in late endosomes, both acting in a coordinated fashion. MON1 has been seen to actively recruit RAB7 to endosomes, and it has been proposed that MON1 mediates the switch between early and late endosomal transition (Figure 2) (Poteryaev *et al.* 2010). Knockdown of MON1 caused early endosomes to accumulate as RAB5 remained trapped on the membranes, unable to exchange for RAB7, which was rectified following MON1 overexpression rescue (Poteryaev *et al.* 2010). MON1 and CCZ1 are proteins identified via gene knockouts to be required for almost all membrane-trafficking pathways, including autophagy, as measured by defective mutants (Wang *et al.* 2002). The MON1-CCZ1 complex has recently been acknowledged as the only known RAB7 guanine nucleotide exchange factor (GEF) and similarly to Poteryaev *et al.* (2010), Yasuda *et al.* (2016) found that CCZ1 knockdown prevented RAB7 activation on late endosomal membranes. This data is also backed up by Hegedus *et al.* (2016), who also identified that the MON1-CCZ1-RAB7 complex is required for autophagosome-lysosome fusion as MON1 and CCZ1 mutants exhibited a reduction in acidic (active) lysosomes. P62 also accumulated, suggesting a loss of autophagic flux.

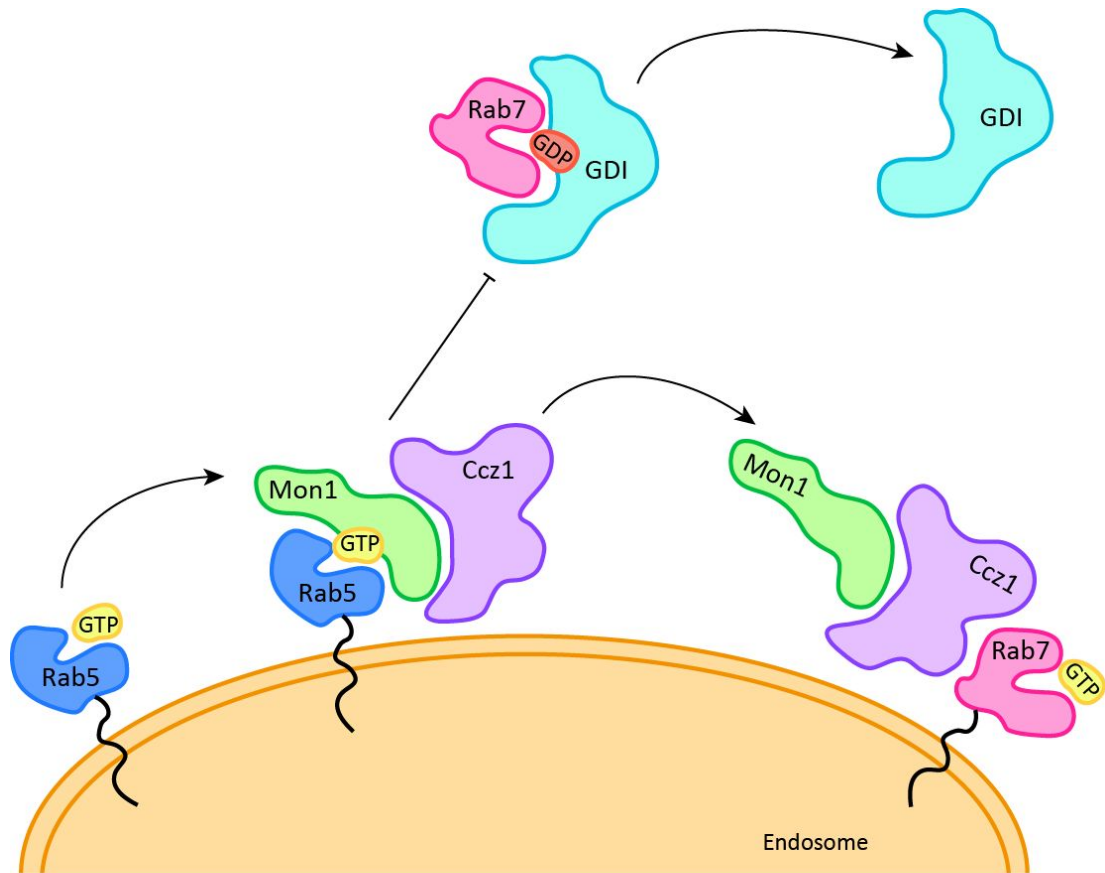


Figure 2: The Interactions of MON1 and CCZ1 with Rab5 and Rab7

RAB5 recruits MON1-CCZ1, which displaces the GDP dissociation inhibitor (GDI) from RAB7, required to initiate endosome maturation. Figure adapted from Kinchen and Ravichandran (2010).

1.3 WDR98 Identification

Data from a large scale immunoprecipitation with mass spectrometry analysis (IP LC-MS/MS) pilot study, conducted by the Pryor lab, found novel protein 'C18orf8/WDR98/RMC1', referred to hereafter as 'WDR98', appearing to bind to MON1 and CCZ1. Following subsequent literature searching, it was determined that this was a protein worth pursuing.

1.4 WDR98 in Literature

WDR98 has previously been identified in liquid chromatography (LC) mass spectrometry (MS) screens of lysosomal homogenate from human placental tissue (Schroder *et al.* 2007) and it was predicted to be membrane-associated. It has also been identified as a RAB5 binding partner through affinity chromatography MS using *Drosophila* Rabs (Gillingham *et al.* 2014) and with a role in autophagy regulation following a genome-wide siRNA screen - knockdown of WDR98 resulted in an increase in autophagic flux (Lipinski *et al.* 2010). WDR98 has also been reported to bind FYCO1, a protein involved in autophagic vesicle transport, which in turn binds MON1B, CCZ1 and RAB7 (Behrends *et al.* 2010).

WDR98 is predicted to have an N-terminal WD40 domain, a common protein domain with a β -propeller structure and a unique C-terminal MIC α -helical domain of unidentified function (Figure 3).

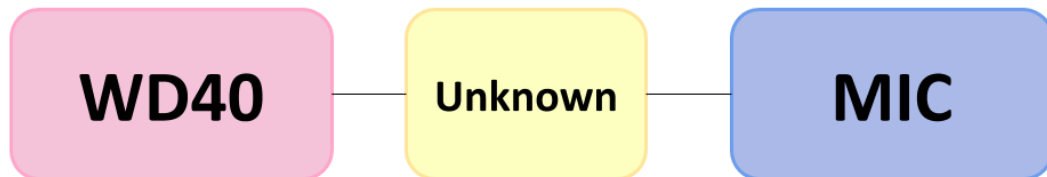


Figure 3: WDR98 Protein Domains

WDR98 is predicted domain structure The area between the WD40 and MIC domains has yet to be defined in terms of its structure.

HHPred (Soding *et al.* 2005) identifies the MIC domain to have a high degree of structural similarity to the N-terminal domain of clathrin heavy chain (92.60%) and vacuolar sorting proteins VPS11 (91.64%), VPS41 (91.11%) and VPS18 (90.86%). These three particular VPS proteins have all been shown to interact with MON1B in *Homo sapiens* (Poteryaev *et al.* 2010).

1.5 Lysosomes and WDR98 in Disease

RAS activation due to missense mutations enables the acquisition of oncogenicity and has been cited as a regulator of RAB5 with the ability to upregulate RAB5-dependent endocytosis (Balaji *et al.* 2012). This can be beneficial to cancer

cells in numerous ways, including the endocytosis and subsequent lysosomal degradation of E-cadherin, occludin and claudin, facilitating the destruction of adherens and tight junctions, which in turn promotes epithelial-mesenchymal transition and eventually metastasis (Mellman and Yarden 2013). Lysosomal proteins have also been implicated in cancer, including the overexpression and presence of LAMP-1 and -2 in the epithelium of metastatic colorectal tissue according to immunohistochemistry analysis (Furuta *et al.* 2001) and increased activation of vacuolar ATPase - promoting an acidic tumour microenvironment and destruction of the extracellular matrix, also enabling tumour progression (Hernandez *et al.* 2012). The lysosome is an emerging cancer therapeutic target and this has been identified as a possible strategy to overcoming cancer drug resistance (Piao and Amaravadi 2016).

WDR98 also has disease relevance, as various research groups have linked it to cancer (Delgado *et al.* 2014, Jones *et al.* 2016) with overexpression in a range of cancers, including thyroid, non-small cell lung, colorectal, melanoma and pancreatic (Kimmelman *et al.* 2008, Doebele *et al.* 2017). Delgado *et al.* (2014) classified WDR98 as an oncogenic open reading frame and found homozygous mutations in a range of tumours. Jones *et al.* (2016) used multiplex targeted mass spectrometry to identify WDR98 as a potential plasma-based protein marker for colorectal cancer detection. Doebele *et al.* (2017) discovered WDR98-NTRK1 (TrkA receptor tyrosine kinase, which has important roles in nervous system development and regulates cell proliferation, differentiation and apoptosis and neuronal survival) fusion genes in a wide range of cancers. They argue this could be used as a biomarker to identify suitable chemotherapeutic treatments and predict response, for example, to RTK inhibitors. Kimmelman *et al.* (2008) observed genetic anomalies in the 18q11 amplicon (a 2Mb span containing WDR98) across multiple tumour types. They knocked down WDR98 in cells cultured from pancreatic ductal adenocarcinomas and observed reduced colony formation *in vitro*. Kimmelman *et al.* argued for the need to characterise this gene given its likelihood as a 'putative PDAC oncogene'.

Considering the context of WDR98 - its likely function in the endosomal-lysosomal system, interaction with essential proteins of this pathway and localisation to the lysosome - it is plausible to explore the possibility of WDR98 as another cancer drug target.

1.6 Model Systems

The primary model systems used throughout this study are normal rat kidney (NRK) cells. These were chosen due to their diploid nature and thin, flat appearance allowing detailed lysosome visualisation via immunofluorescence microscopy. Additional HeLa cell lines were occasionally used, including a previously-produced MYC-MON1B inducible expression line and newly-produced pcDNA5 MYC-GFP-humanWDR98 line in an attempt to achieve a stable cell line expressing GFP-tagged WDR98. Some preliminary work was undertaken in *Drosophila melanogaster* and *Xenopus laevis* which is not reported here due to a lack of biological replicates.

1.7 Aims and Objectives

Given that WDR98 reportedly binds to essential components of the endosomal-lysosomal pathway and has a potential role in lysosomal/autophagosomal biology, there is great cause for investigating its function, magnified by its possible association with cancer. Therefore the aims of this project are to characterise the novel protein 'WDR98' and to probe its function in the wider context of lysosome biology.

In order to achieve this, the following experimental objectives have been defined:

1. To produce a variety of constructs to enable characterisation of WDR98, including GFP-WDR98 and GST-WDR98 (Chapter 3)
2. To produce an antibody to WDR98 (Chapter 3)
3. To assess the protein-protein interactions in which WDR98 participates (Chapter 4)
4. To assess where WDR98 localises in the cell (Chapter 5)

A range of techniques were used to these ends, including yeast 2 hybrid, immunoprecipitations, pulldowns and immunofluorescence. Concurrently, similar work was being carried out by labs in Harvard (Pontano Vaites *et al.* 2017) and Stanford (Tsui *et al.* 2019) universities, and many of our results corroborate with

their data. The novelty of our work comprises that surrounding the WDR98 individual domain data, which has yet to be published by any group.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Antibodies

Details of antibodies used are listed in Table 1.

Table 1. Details of Antibodies

W.B; western blot, IF; immunofluorescence, IgG; immunoglobulin G, AF; Alexa Fluor, HRP; horseradish peroxidase.

Primary Antibodies					
<u>Cat. No.</u>	<u>Source</u>	<u>Target</u>	<u>Host Species</u>	<u>W.B Dilution</u>	<u>IF Dilution</u>
sc-28365	Santa Cruz Biotech	APP	Mouse	1 : 100	-
sc-28365	Santa Cruz Biotech	Beta-Amyloid	Mouse	1 : 100	-
sc-514290	Santa Cruz Biotech	CCZ1	Mouse	1 : 500	1 : 100
ab137276	Abcam	C18orf8	Rabbit	1 : 200	-
ab78596	Abcam	C18orf8	Goat	1 : 1000	-
20111-1-AP	Proteintech	C18orf8	Rabbit	1 : 1000	-
sc-390305	Santa Cruz Biotech	C18orf8	Mouse	1 : 100	-
PA521004	Invitrogen	AFP	Mouse	-	1 : 200
-	Pryor Lab	GFP	Rabbit	1 : 500	1 : 200
-	Pryor Lab	GST	Rabbit	1 : 500	-
sc-133225	Santa Cruz Biotech	HDAC5	Mouse	1 : 100	-
2366	Cell Signaling	HIS	Mouse	1 : 200	-
sc-101494	Santa Cruz Biotech	HSP90	Mouse	1 :100	-
LC3-5F10	Nanotools	LC3	Rabbit	-	1 : 100
Gift	Cambridge	LGP120	Mouse	-	1 : 200
-	Pryor Lab	MYC	Mouse	1 : 200	1 : 100

3547	Cell Signaling	RAB5	Rabbit	-	1 : 100
9367	Cell Signaling	RAB7	Rabbit	-	1 : 100
ab134064	Abcam	STUB1	Rabbit	1 : 10000	-
A	Eurogentec	WDR98	Rabbit	1 : 500	1 : 100
B	Eurogentec	WDR98	Rabbit	1 : 500	1 : 100

<u>Secondary Antibodies</u>					
<u>Cat. No.</u>	<u>Source</u>	<u>Antigen</u>	<u>Label</u>	<u>W.B Dilution</u>	<u>IF Dilution</u>
A11056	Life Technologies	α -Goat IgG	AF 546	-	1 : 300
A150113	Life Technologies	α -Mouse IgG	AF 488	-	1 : 300
A21424	Life Technologies	α -Mouse IgG	AF 555	-	1 : 300
A21429	Life Technologies	α -Rabbit IgG	AF 555	-	1 : 300
A9169	Sigma Aldrich	α -Rabbit IgG	HRP	1 : 8000	-
AP186P	Sigma Aldrich	α -Goat IgG	HRP	1 : 8000	-
AP130P	Sigma Aldrich	α -Mouse IgG	HRP	1 : 8000	-

2.2 Methodology

Note: room temperature (RT) ranged from 17 - 23 °C.

2.2.1 General Polymerase Chain Reaction (PCR)

All PCRs were performed in a Bio-Rad PTC-200 DNA engine cycler. Optimised PCR parameters are displayed in Table 2.

Table 2. PCR Reaction Parameters

Three different enzymes with individual protocols were utilised: HiFi premix (Clontech), Pfu (Promega), Phusion (Thermo Fisher). The reagents required and temperature parameters for the reactions are detailed in the Table below.

	HiFi	Pfu	Phusion
Water	to 20 µl	to 50 µl	to 25 µl
Buffer	-	5 µl	4 µl
dNTPs (10 mM)	1 µl	0.5 µl	1 µl
Primers (10 µM)	1 µl each	2.5 µl each	1 µl each
DMSO	2.5 µl	-	0.6 µl
DNA Template	variable	variable	variable
Polymerase	12.5 µl premix	0.5 µl	0.2 µl
	HiFi	Pfu	Phusion
Initial Denaturation	98°C 10 sec	95°C 2 min	98°C 30 sec
Denaturation	98°C 10 sec	95 °C 45 sec	98°C 10 sec
Annealing	Tm - 5°C 15 sec	Tm - 5°C 30 sec	Tm - 5°C 30 sec
Extension	72°C 1 min/kb	72°C 2 min/kb	72°C 15 sec/kb
Cycles	30	30	30
Final Extension	72°C 10 min	72°C 10 min	72°C 10 min
Final Hold	4°C	4°C	4°C

2.2.2 QuickChange PCR

Primers were designed to insert desired codon into target region of gene, one being the reverse complement of the other. PCR mastermix was prepared as follows: 37.5 µl ddH₂O, 5 µl 10X Pfu buffer + MgCl₂, 0.5 µl 10 mM dNTPs, 2.5 µl 10 µM primers, 0.5 µl Pfu and DNA template at variable amounts. A control PCR was run for each set of primers without the Pfu enzyme present to check for amplification. The reaction parameters were as follows: 95 °C for 2 minutes, 95 °C for 45 seconds, Tm - 5 °C 30 seconds, 72 °C 2 minutes per kb (11 total cycles), 72 °C for 10 minutes, kept at 4 °C.

2.2.3 DNA Electrophoresis and Purification

Agarose gels were made by dissolving 60 g agarose in 75 ml TAE buffer (40 mM Tris, 20 mM acetic acid, 0.5 mM EDTA, pH 8 in 1 L ddH₂O) in a microwave and returning the volume to 75 ml using ddH₂O. SYBR Safe (Invitrogen) was added at 1:10,000 and mixed. DNA was electrophoresed at 60 V for 1 hour and visualised using a blue lightbox. Images captured via Syngene gel imager and UGenius software.

DNA was purified from gel slices using the Qiagen Gel Purification kit as per manufacturer's instructions with the following modifications: 600 µl Buffer QC used to dissolve the gel slice with 200 µl isopropanol added following dissolution.

2.2.4 DNA Restriction Digests

Restriction digests were performed using 4 µl 10X KGB buffer (1 M glutamic acid, 250 mM Tris acetate pH 7.6, 100 mM magnesium acetate, 0.5 mg/ml bovine serum albumin, 5 mM β-mercaptoethanol), 4 µl vector (~ 1 µg), 2 µl appropriate restriction enzyme and 30 µl ddH₂O incubated at 37 °C for 45 minutes. Digested products and undigested controls were analysed by DNA electrophoresis and DNA gel-purified.

2.2.5 Bacterial Transformation and DNA Extractions

DNA was inserted into the vector by combining 2 µl InFusion Enzyme Mix (5X) (Takara), ~ 100 ng vector DNA and ~ 50 ng insert DNA. The mixture was vortexed and incubated at 50 °C for 15 minutes. 2 µl InFusion reaction mixture was added to 50 µl stellar competent cells (Takara) and incubated on ice for 10 minutes, heat shocked at 42 °C for 1 minute and incubated on ice for 2 minutes. 250 µl SOC media (Takara) was added and cells recovered at 37 °C for 1 hour with shaking. Cells were centrifuged at 3500 rpm (Rotor FA-45-18-11) for 2 minutes, resuspended in the media meniscus and plated onto an antibiotic-selective plate.

Plasmid Minipreps from Bacteria:

DNA plasmids were extracted from 5 ml overnight cultures containing the relevant antibiotic using the Qiagen MiniPrep kit as per the manufacturer's instructions.

Plasmid MidiPreps from Bacteria:

DNA plasmids were extracted from 50 ml overnight cultures containing the relevant antibiotic using the Qiagen MidiPrep kit as per the manufacturer's instructions.

Wizard Minipreps from NRK Cells:

1 x T25 flask NRK cells were lysed and DNA extracted using the Promega Wizard Miniprep kit as per the manufacturer's instructions.

DNA samples were sent to either Source Bioscience or Eurofins GATC for sequencing analysis and prepared according to the company's instructions.

2.2.6 Recombinant Protein Production

E. coli (BL21 DE3 pLysS) were transformed with plasmids encoding either GST- or HIS₆-tagged proteins as appropriate. Bacteria were grown in 2 x 600 ml of LB or 2TY broth as follows: 7.5 ml of LB (plus the appropriate selective antibiotics) was inoculated with a single bacterial colony and grown overnight at 37 °C with shaking. The next day, starter cultures were diluted 1:500 into a total volume of 600 ml LB or 2TY broth, and grown at 37 °C until an OD₆₀₀ of 0.6 - 2.0 was obtained. Protein production was induced with IPTG to a final concentration of 0.2 mM, and cultures grown for a further 4 h. The bacteria were pelleted at 2,645 x g for 15 minutes at 4 °C. The cell pellet was resuspended in 10 ml of bacterial lysis buffer (1 % (v/v) Triton X-100, 1 mg/ml lysozyme, 2 mM MgCl₂, 1 U/ml DNase, in PBS) with protease inhibitors (Roche), and frozen at -20 °C overnight. The cell pellet was defrosted on ice and centrifuged at 47,800 x g in a Sorvall SS34 rotor for 20 minutes at 4 °C. For GST fusion proteins, the supernatant was incubated with 1 ml (packed volume) of glutathione sepharose (GE healthcare) for 1 h at 4 °C with gentle rotation. The sepharose beads were then collected and washed with 3 x 20 ml of wash buffer (1 % (v/v) Triton X-100 in PBS). GST/GST-fusion proteins were eluted with 20 ml elution buffer (50 mM Tris, pH 8.0, 10 mM reduced glutathione). These fractions were then analysed by SDS-PAGE. The eluted proteins were then pooled and dialysed extensively into PBS at 4 °C for 2 days and stored at -20 °C until needed.

For the purification of HIS₆-tagged proteins, cell pellets were processed as above. The soluble supernatant was incubated with 134 μ l (packed volume) of HIS-Select resin (Sigma) for 1-2 h at 4 °C with gentle rotation. The beads were washed with 3 x 20 ml wash buffer (10 mM imidazole, 0.3 M NaCl, 50 mM Na₂HPO₄, pH8.0). Proteins were eluted with 20 ml elution buffer (250 mM imidazole, 0.3 M NaCl, 50 mM Na₂HPO₄, pH 8.0) and 1 ml fractions collected. Fractions were analysed by SDS-PAGE. The eluted proteins were then pooled and dialysed extensively into PBS at 4 °C for 2 days and stored at -20 °C until needed.

2.2.7 Mammalian Cell Culture

All cells were cultured in Dulbecco's modified eagle's medium (DMEM) supplemented with 10 % (v/v) FBS, 2 mM glutamine and 1 % (v/v) penicillin-streptomycin (10 U/ml and 10 μ g/ml respectively) in a humidified 5 % CO₂ atmosphere at 37 °C, unless otherwise stated. Flp-In HeLa cells were a kind gift from Prof. M. Lowe (University of Manchester) and isogenic cell lines expressing either GFP or GFP-WDR98 protein were made by cloning genes into pcDNA5/FRT/TO and transfecting cells with this vector along with pOG44. Transfected Flp-In HeLa cells were selected with 50 μ g/ml Hygromycin B (Roche) and protein expression was induced with 1 μ g/ml doxycycline for 24h.

2.2.8 Mammalian Cell Transfection

Transfection reagents were combined with DNA in the ratios 20 μ l X-tremeGENE HP (Sigma Aldrich) to 10 μ g DNA made up to 1 ml total with OptiMEM (Gibco) to 1 μ g DNA and incubated at RT for 20 minutes then added to cells that were split the day prior to obtain confluency of 80-90%.

2.2.9 CRISPR/Cas9 knockouts

An NRK CRISPR/Cas9 knockout protocol was optimised to the following: DharmaFect solution (DFS) was produced with 10 μ l DharmaFect (Dharmacon) and 490 μ l Dulbecco's serum-free media (SFM) (Gibco), incubated at room temperature (RT) for 5 minutes. The transfection mixture contained: 210 μ l SFM, 3 x 5 μ l crRNA with distinct targeting sites (Dharmacon), 15 μ l trRNA (Dharmacon), 3 μ g Cas9

plasmid (Dharmacon) and 300 µl DFS, incubated at RT for 20 minutes before being added to cells containing 2.4 ml DMEM per well of a 6-well plate and cells at ~60% confluency, passaged the day prior.

Cells were incubated overnight at 37 °C in 5% CO₂ before seeding into larger dishes for selection of single colonies. The following day puromycin was added to a final concentration of 50 µg/ml. Once single colonies were established, cells were washed with PBS and colonies isolated with sterile pipette tips following trypsinisation. These were then transferred to a 24-well plate containing DMEM and puromycin and cultured for analyses.

2.2.10 Detergent Soluble Lysates

Tissue culture cells were washed with Dulbecco's phosphate buffered saline (PBS) (Sigma) and scraped into ice-cold cell lysis buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.5% (v/v) NP-40 with EDTA-free protease inhibitors (Roche)). The lysate was incubated at 4°C, 15 minutes and centrifuged at 16,000 x g, 10 minutes at 4 °C.

2.2.11 Protein Quantification Assay

BSA protein stock solution was prepared at 2 mg/ml. 100 µl BSA solution was diluted 1:1 with 0.2 M NaOH and used to set up a standard curve in duplicate with BSA 1 mg/ml ranging from 0 - 10 µl and 0.1 M NaOH ranging from 10 - 0 µl in order to make the total volume 10 µl per well of a 96 well plate. Protein samples to measure were diluted in water, if necessary, at increasing ratios and 5 µl each sample added to 5 µl 0.2 M NaOH in duplicate. BCA reagent and CuSO₄ were combined in the ratio 50:1 and 200 µl this reagent was added to all wells requiring quantification. Plates were incubated at 37 °C for 30 minutes. Absorbance measurements were taken at 560 nm using a Thermo Multiscan GO plate reader to produce a standard curve and subsequently calculate the protein concentration of the sample(s).

2.2.12 Antibody Coupling

Amino-link beads (Pierce) were centrifuged and washed 3 x 500 μ l PBS. 1 μ g GFP nanobody was covalently coupled to 1 μ l packed volume Amino-link beads via addition of 3 μ l 5 M NaCNBH₄ per 200 μ l total volume. The suspension was mixed by rotating at 4 °C for 2 hours, beads were pelleted and washed 3 x 500 μ l 1 M Tris pH 8.0. 500 μ l 1 M Tris and 3 μ l 5 M NaCNBH₄ was added per 200 μ l total volume. The suspension was mixed by rotating at 4 °C for 15 minutes and washed 3 x 500 μ l PBS, followed by 6 x 1 M NaCl and 6 x 500 μ l PBS. Beads were stored at 4 °C as a 50% packed bead slurry in PBS.

2.2.13 Immunoprecipitation

Transfected cells and untransfected controls were washed with ice cold PBS and lysates were produced. Protein quantification assay was used to estimate protein concentration which was normalised across samples. 8 μ l bead slurry was added per 100 μ g lysate and the suspension was rotated at 4 °C for 2 hours. Beads were washed 5 x 500 μ l lysis buffer using Spin-X 0.45 μ m cellulose acetate 8163 filter cups (Corning Costar), and centrifuged at 3000 rpm for 1 minute (Rotor FA-45-18-11). 100 μ l IgG elution buffer (Pierce) was added and beads centrifuged at 3000 rpm for 1 minute (Rotor FA-45-18-11). 100 μ l IgG elution buffer was again added and incubated at RT for 10 minutes before centrifugation at 3000 rpm for 1 minute (Rotor FA-45-18-11). Samples were concentrated for 45 minutes using a vacuum pump.

2.2.14 RAB Nucleotide Loading

RAB proteins were diluted to 1 mg/ml in exchange buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM (NH₄)₂SO₄, 1 mM dithiothreitol, 0.05 mM NP-40) and the solution adjusted to 5 mM EDTA and 250 mM (NH₄)₂SO₄. Each 1 ml solution was incubated with 33 units alkaline phosphatase beads with mixing by rotation for 15 minutes at RT. 250 μ M GMPPNP and 33 units alkaline phosphatase beads were added to each solution and incubated with mixing by rotation for 30 minutes at RT. Beads were washed 2 x 500 μ l exchange buffer using Spin-X 0.45 μ m cellulose acetate 8163 filter cups (Corning Costar), and centrifuged at 3000 rpm for 1 minute

(Rotor FA-45-18-11). Flow throughs were pooled and divided equally into 3 tubes per Rab. Each tube was made up to 2 ml with exchange buffer and GMPPNP / GDP / GTP was added to 500 μ M. Solutions were concentrated using Amicon Ultra Centrifugal Tubes at 3000 rpm for 30 minutes at 4 °C (Rotor FA-45-18-11). 2 ml loading buffer (20 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 1 mM EDTA, 50 mM NaCl, 1 mM dithiothreitol, 0.05 mM NP-40) was added to each tube with 200 μ M nucleotide. Solutions were concentrated using Amicon Ultra Centrifugal Tubes at 3000 rpm for 30 minutes at 4 °C (Rotor FA-45-18-11) and the final step repeated. Each solution was then made up to 1 ml with loading buffer and stored at -80 °C.

2.2.15 Pulldowns

6 μ g bait protein was combined with 6 μ g prey protein and incubated in 500 μ l pulldown buffer for 1 hour with mixing by rotation at 4 °C. Glutathione sepharose was washed 3 x 500 μ l pulldown buffer and 20 μ l unpacked pellet added to solutions and incubated for 1 hour with mixing by rotation at 4 °C. Beads were washed 3 x 500 μ l pulldown buffer using Spin-X 0.45 μ m cellulose acetate 8163 filter cups (Corning Costar), and centrifuged at 3000 rpm for 1 minute (Rotor FA-45-18-11). 40 μ l 95 °C Laemmli sample buffer (3X stock: 188 mM Tris pH 6.8, 6% (w/v) SDS, 30% (v/v) glycerol, 0.03% (w/v) Bromophenol blue, 10% (v/v) β -mercaptoethanol) was added to the beads and incubated at 95 °C for 5 minutes. Samples were analysed using SDS-PAGE and western blotting. Pulldown buffers consisted of loading buffer with slightly different compositions, including 50 mM Tris-HCl pH 7.5 / pH 6, 150-300 mM NaCl, 1-2% (v/v) NP-40, 1-2% (v/v) Tween-20, 1-2% (w/v) PEG 4000.

2.2.16 PLATE Transformation of Yeast

Transformed bacterial stocks were plated onto selective agar plates and cultured overnight at 37 °C. Single colonies were selected to Miniprep. Yeast strains were grown overnight in 5 ml YPD (yeast extract peptone dextrose, 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose) in 50 ml tubes at 30 °C with agitation. Cultures were divided into 500 μ l aliquots, centrifuged at 6,000 rpm (Rotor FA-45-18-11), for 1 minute and the supernatant removed. Pellets were resuspended with 20 μ l ice-cold salmon sperm ssDNA and 20 μ l of miniprep plasmid DNA. Following incubation at RT for 15 minutes, 250 μ l PLATE solution (1 ml 10X TE, 1

ml 1M LiOAc, 8 ml 50% (w/v) PEG 4000) was added to the yeast and the yeast were further incubated overnight at 30 °C without agitation. Cells were heat-shocked at 42 °C, 15 minutes before centrifuging at 6,000 rpm (Rotor FA-45-18-11), for 1 minute and removing PLATE solution. 400 µl YPD was added and incubated at 30 °C for 1-4 hours. The yeast were then pelleted by centrifugation at 6,000 rpm (Rotor FA-45-18-11), for 1 minute then resuspended in 200 µl ddH₂O and 100 µl plated onto selective plates (pGAD vectors onto SD-leu agar and pGBDU onto SD-ura agar) and incubated at 30 °C until transformants could be picked.

2.2.17 Yeast 2 Hybrid

Yeast were plated out onto selective SD-leu/-ura plates and incubated at 30 °C for 48 hours. Yeast were replica plated onto YPD plates using sterile velvets so that pGAD vectors with inserts were mated with each pGBDU. Yeast were allowed to mate for 48 h at 30 °C. YPD plates were replica plated onto SD-leu, -ura plates and incubated at 30°C for 48 hours. Using a blunt-ended sterile cocktail stick, grown yeast from cross-sections were mixed into sterile water and 5 µl pipetted onto new SD-leu -ura plates and incubated at 30 °C for 48 hours. SD-leu -ura plates were replica plated onto SD-leu -ura -ade and incubated at 30 °C until growth observed.

2.2.18 SDS PAGE

Laemmli sample buffer (3X stock: 188 mM Tris pH 6.8, 6% (w/v) SDS, 30% (v/v) glycerol, 0.03% (w/v) Bromophenol blue, 10% (v/v) β-mercaptoethanol) was added to 20 µl lysate samples followed by incubation at 95 °C for 5 minutes. Samples were resolved by gel electrophoresis using discontinuous polyacrylamide gels during which samples move through a stacking gel initially and are separated on a lower resolving gel, using the Bio-Rad minigel apparatus. Resolving gels were made with 7% (v/v) acrylamide (Protogel, National Diagnostics) using 2.5 ml resolving gel buffer (final concentration 0.375 M Tris-HCl pH 8.8, 0.2% (w/v) SDS), 6.1 ml ddH₂O, 50 µl 10% (w/v) ammonium persulfate (APS) and 5 µl TEMED. Gels were sealed with isopropanol. Stacking gels were made with 4% (v/v) acrylamide, 2.5 ml stacking gel buffer (final concentration 0.125 M Tris-HCl Ph 6.8, 0.2% (w/v) SDS), 50 µl 10% (w/v) APS and 10 µl TEMED. A protein ladder (Precision Plus Kaleidoscope ladder, Bio-Rad) was added to the gel along with samples and gel electrophoresis was

performed at 200 V in SDS running buffer (25 mM Tris, 0.1% (w/v) SDS, 192 mM glycine) until bromophenol dye had eluted from the gel.

2.2.19 Western Blotting

Proteins were transferred onto nitrocellulose membranes using the iBlot system (Invitrogen) as per manufacturer's instructions. Membranes were blocked in 5% (w/v) semi-skimmed milk in TBS-Tween (TBST) (100 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.1% (v/v) Tween-20), 30 minutes with agitation. Membranes were incubated with primary antibodies in 5% (w/v) milk TBST, 1 hour with agitation, followed by 3 x 5 minute TBST washes. Membranes were incubated with horseradish peroxidase conjugated secondary antibodies in 5% (w/v) milk TBST, 30 minutes with agitation. TBST washes were repeated and proteins were visualised using ECL reagent (Amersham).

2.2.20 Mass Spectrometry Sample Preparation

Samples were dried in a vacuum concentrator then solubilised in NuPAGE LDS sample buffer (Life Technologies) with heating at 70 °C for 10 minutes before running into a 7 cm NuPAGE Novex 10% (w/v) Bis-Tris gel (Life Technologies) at 200 V for 6 minutes. Gels were stained with SafeBLUE protein stain (NBS Biologicals) for 1 hour before destaining with ultrapure water for 1 hour. In-gel tryptic digestion was performed after reduction with dithioerythritol and S-carbamidomethylation with iodoacetamide. Gel pieces were washed two times with aqueous 50% (v/v) acetonitrile containing 25 mM ammonium bicarbonate, then once with acetonitrile and dried in a vacuum concentrator for 20 minutes. Gel pieces were rehydrated by adding 0.5 µg of sequencing-grade, modified porcine trypsin (Promega) in 200 µl of aqueous 25 mM ammonium bicarbonate. Digests were incubated overnight at 37 °C. Peptides were extracted by washing three times with aqueous 50% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid, before drying in a vacuum concentrator and reconstituting in aqueous 0.1% (v/v) trifluoroacetic acid.

2.2.21 Liquid Chromatography Mass Spectrometry

Samples were loaded onto an UltiMate 3000 RSLCnano HPLC system (Thermo) equipped with a PepMap 100 Å C18, 5 µm trap column (300 µm x 5 mm Thermo) and a PepMap, 2 µm, 100 Å, C18 EasyNano nanocapillary column (75 µm x 150 mm, Thermo). The trap wash solvent was aqueous 0.05% (v/v) trifluoroacetic acid and the trapping flow rate was 15 µl/minute. The trap was washed for 3 minutes before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 1% (v/v) formic acid; solvent B, aqueous 80% (v/v) acetonitrile containing 1% (v/v) formic acid. The flow rate for the capillary column was 300 nl/minute. Column temperature was 40 °C and the gradient profile was: linear 3-10% B over 15 minutes, linear 10-35% B over 108 minutes, linear 35-65% B over 30 minutes, linear 65-99% B over 7 minutes then proceeded to wash with 99% solvent B for 4 minutes. The column was returned to initial conditions and re-equilibrated for 15 minutes before subsequent injections. The nanoLC system was interfaced with an Orbitrap Fusion hybrid mass spectrometer (Thermo) with an EasyNano ionisation source (Thermo). Positive ESI-MS and MS2 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature; 275 °C. MS1 spectra were acquired in the Orbitrap with: 120 K resolution, scan range: m/z 375-1,500; AGC target, 4e5; maximum fill time, 100 ms. Data dependent acquisition was performed in top speed mode using a 1 s cycle, selecting the most intense precursors with charge states +2 to +5. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e3. MS2 spectra were acquired in the linear ion trap with: scan rate, turbo; quadrupole isolation, 1.6 m/z; activation type, HCD; activation energy: 32%; AGC target, 5e3; first mass, 110 m/z; maximum fill time, 100 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time.

2.2.22 Mass Spectrometry Data Analysis

Peak lists were converted from .raw to .mgf format using Mascot Distiller (version 5, Matrix science), stipulating a minimum signal to noise ratio of 2 and correlation (Rho) of 0.6, before submitting to a locally-running copy of the Mascot program using Mascot Daemon (version 2.5.1, Matrix Science). Data were searched against

the *Rattus norvegicus* subset of the UniProt database (21,353 entries, date 20160909) specifying: Enzyme, trypsin; Max missed cleavages, 2; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M), Deamidated (NQ); Peptide tolerance, 5 ppm; MS/MS tolerance, 0.5 Da; Instrument, ESI-TRAP. Peak Lists were also searched against the same database using X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)), with specified variable modifications increased to include: Glu/Gln->pyro-Glu (N-term) and ammonia-loss (N-term). Search results were combined and validated using Scaffold (version 4.8.4, Proteome Software Inc.). Protein identifications were accepted if they could be established at >99.0% probability to achieve a FDR <1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.* 2003). Proteins containing similar peptides that could not be differentiated based on MS2 analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Proteins were annotated with GO terms from goa_uniprot_all.gaf (downloaded 12-Oct-2016) (Ashburner *et al.* 2000). Relative protein quantification was performed by spectral counting analogous to (Dowle, A *et al.* J. Prot. Res. 2016; 15, 3550–3562). Significant differences were calculated using Fisher's exact test with Hochberg and Benjamini multiple test correction applied.

2.2.23 Immunofluorescence

Cells were rinsed with PBS and fixed:

Methanol fixation: cells were incubated in ice-cold methanol at -20 °C for 5 minutes before methanol aspiration.

Formaldehyde fixation: cells were incubated in 4% (v/v) paraformaldehyde in PBS, 20 minutes at RT, followed by paraformaldehyde aspiration and incubation in 50 mM NH₄Cl in PBS, 10 minutes, before aspiration. Cells were permeabilised in 0.2% (w/v) BSA, 0.05% (v/v) saponin in PBS, 10 minutes; all following washes and antibody dilutions were performed in BSA-Saponin-PBS (BSP). For cytosol extraction before fixation, cells were rinsed with cytosol extraction buffer (CEB) (25 mM HEPES-KOH pH 7.4, 25 mM KCl, 2.5 mM magnesium acetate, 5 mM EGTA (sodium free), 150 mM potassium glutamate) and incubated in CEB 0.05% (w/v) saponin, 1 minute, followed by paraformaldehyde fixation.

Coverslips were incubated in primary antibody diluted in BSP, 1 hour at RT followed by 3 x 5 minute BSP washes. Coverslips were incubated in Alexa fluorophore-conjugated secondary antibodies in BSP, 30 minutes at RT followed by 3 x 5 minute BSP washes. Coverslips were rinsed in ddH₂O and mounted onto slides with MOVIOL 4-88.

Slides were imaged using Zeiss LSM 880 Axio Observer.Z1 inverted confocal laser microscope, 63x lens running Zen version 2.3 software. Image processing was completed using FIJI software (Schindelin *et al.* 2012).

Chapter 3: Identifying and Characterising WDR98

3.1 Introduction

To begin this study, all relevant publications and databases were mined for current information on WDR98. Commercially available reagents were assessed and a variety of constructs and cell lines produced to develop the tools necessary to begin establishing the characteristics of WDR98. Pre-existing data was accumulated, including identification of WDR98 as lysosomal in a large scale human placental proteomic analysis (Schroder *et al.* 2007); as a regulator of autophagy whereby WDR98 knockdown increased autophagic flux (Lipinski *et al.* 2010), and reportedly interacting with RAB5 (Gillingham *et al.* 2014) and FYCO1 (Behrends *et al.* 2010), both of which are involved in the lysosomal-autophagy systems. Considering this, it seemed to suggest that the Pryor lab pilot data, which indicated WDR98 interacted with CCZ1 and MON1B, was plausible. Furthermore, HHPred protein structural prediction software (Soding *et al.* 2005) was used to identify proteins with similar structures based on amino acid sequence. The outcomes of this analysis were high degrees of structural similarity between WDR98 and the N-terminal domain of clathrin heavy chain (92.6%) and vacuolar sorting proteins VPS11 (91.64%), VPS41 (91.11%) and VPS18 (90.86%). These three particular VPS proteins have all been shown to interact with MON1B (Poteryaev *et al.* 2010). This information again suggested that looking to characterise WDR98 in the context of lysosome biology was a worthwhile avenue of exploration.

In order to achieve the project aims of characterising WDR98 and to begin to understand its function a variety of constructs were produced. This included GFP-WDR98, for immunoprecipitations and immunofluorescence microscopy, WDR98-HIS for Rab pulldowns and GST-WDR98 for antibody production. The latter being deemed necessary following commercial antibody characterisation which revealed no WDR98-appropriate results, as will be discussed later in this chapter.

Cell lines were also produced in an attempt to generate permanent tools to study the inducible overexpression of WDR98 in a stable line. The generation and characterisation of these cell lines will also be discussed.

CRISPR/Cas9 was also utilised extensively with the aim of obtaining a WDR98-knockout NRK cell line, however, the best result was a heterozygous knockout which displayed no distinct phenotype from wild type NRK cells.

Phylogenetic analysis of the conservation of WDR98 amino acid sequences between a wide range of species, both animal and plant, was carried out to determine the extent to which WDR98 was conserved, and from this to infer whether function was likely similar and how important it was biologically.

Overall, this chapter will discuss the attempts made to generate useful experimental tools and their characterisation, and the compilation of data already available. This will both inform later experiments and begin to build a picture of the role of WDR98 in the cell.

3.2 Methods

3.2.1 Phylogenetic Analysis

NCBI BLAST (protein-protein BLAST algorithm, Altschul *et al.* 1990) searches were performed using the human WDR98 full-length and unique MIC-domain amino acid sequences and subsequently identified sequences from species of decreasing complexity. Cladograms were compiled using Clustal X (Larkin *et al.* 2007) to align the sequences found across varying species and prepare phylogenetic trees. FigTree was used in their manipulation and Figure production.

3.2.2 Commercial Antibody Characterisation

Commercially-available antibodies to WDR98 were characterised using both NRK and HeLa cell lysates, with and without GFP-WDR98 transfections, by western blotting (Methods 2.2.18-19).

3.2.3 DNA Construct Production

After it was decided to use NRK as the main experimental model, all constructs produced (unless otherwise stated) were done so using the *Rattus norvegicus* WDR98 DNA sequence.

3.2.4 Antibody Production and Characterisation

GST-WDR98 protein was produced and purified as described in Methods 2.2.6. This protein was then sent to Eurogentec for antibody production using their 87-Day Classical Custom Polyclonal Antibody programme.

These commissioned antibodies were characterised via western blotting (Methods 2.2.18-19) and immunofluorescence (Methods 2.2.23). In some instances, antibodies (1:1000) were incubated with 10 µg GST-WDR98 protein prior to antibody incubation with the blot, with the aim of saturating the antibody and determining the identity of consistently strong bands.

3.2.5 Cell Line Production

Relevant constructs, pMEP4-MYC-GFP-WDR98 and pcDNA5-GFP-humanWDR98, were transfected into NRK cells and Flp-In-HeLa cells, respectively, and selected using G418 (pMEP4) and Hygromycin B (pcDNA5) to produce cell lines with inducible GFP-WDR98 expression. Expression induction was initiated using cadmium (Δ pMEP4) and doxycycline (pcDNA5). The efficacy of these cell lines was then assessed by immunofluorescence microscopy (Methods 2.2.23-24).

3.2.6 CRISPR/Cas9

An NRK CRISPR/Cas9 knockout protocol was optimised to the following: DharmaFect solution (DFS) was produced with 10 µl DharmaFect (Dharmacon) and 490 µl Dulbecco's serum-free media (SFM) (Gibco), incubated at room temperature (RT) for 5 minutes. The transfection mixture contained: 210 µl SFM, 3 x 5 µl crRNA with distinct targeting sites (Dharmacon), 15 µl trRNA (Dharmacon), 3 µg Cas9 plasmid (Dharmacon) and 300 µl DFS, incubated at room temperature for 20

minutes before being added to cells containing 2.4 ml DMEM per well of a 6-well plate and cells at ~60% confluency, passaged the day prior.

Cells were incubated overnight at 37 °C in 5% CO₂ before seeding into larger dishes for selection of single colonies. The following day puromycin was added to a final concentration of 50 µg/ml. Once single colonies were established, cells were washed with PBS and colonies isolated with sterile pipette tips following trypsinisation. These were then transferred to a 24-well plate containing DMEM and puromycin and cultured for analyses.

Once confluent, cells were transferred to T25 culture flasks and then once confluency was achieved again, these cells were lysed and the DNA extracted using the Promega Wizard Miniprep kit as per the manufacturer's instructions.

DNA was sent to either Source Bioscience or Eurofins GATC for sequencing around the CRISPR-target sites (appendix A ix).

3.3 Results

3.3.1 Evolutionary Conservation

WDR98 appears very well conserved with 90.68 - 100 % identity match among mammals and 24.27 % identity match between *Homo sapiens* and *Arabidopsis thaliana*. The source of this commonality lies in the C-terminal MIC domains, which appears consistently in all species identified. The range of species in which WDR98 orthologs have been found is presented in a non-exhaustive cladogram (Figure 4) intended to represent the level of WDR98 conservation. Efforts were made to identify orthologs in yeast and bacteria, through mining bioinformatics databases, without success.

(i)



Figure 4: Evolutionary History of WDR98

NCBI blast (protein-protein BLAST algorithm) searches were performed using the human WDR98 full-length and C-terminal MIC-domain amino acid sequences and subsequently identified sequences from varying species. This cladogram was compiled using Clustal X to align sequences and prepare phylogenetic trees using data from a small proportion of the sequences found across diverse organisms. FigTree was used to manipulate and organise the trees into cladograms such as that presented here.

3.3.2 Commercial Antibody Testing

Five commercially available antibodies to WDR98 (listed in Table 1) were tested via western blotting to check their specificity, in both wild type NRK lysates (except where stated) and lysates with GFP-WDR98 (100 kDa) transfection (Figure 5). Abcam ab137276 was later tested on HeLa lysates (5E) following a publication which claimed to have found specificity by western blotting in human cells (Pontano Vaites *et al.* 2017) though we were unable to replicate this result.

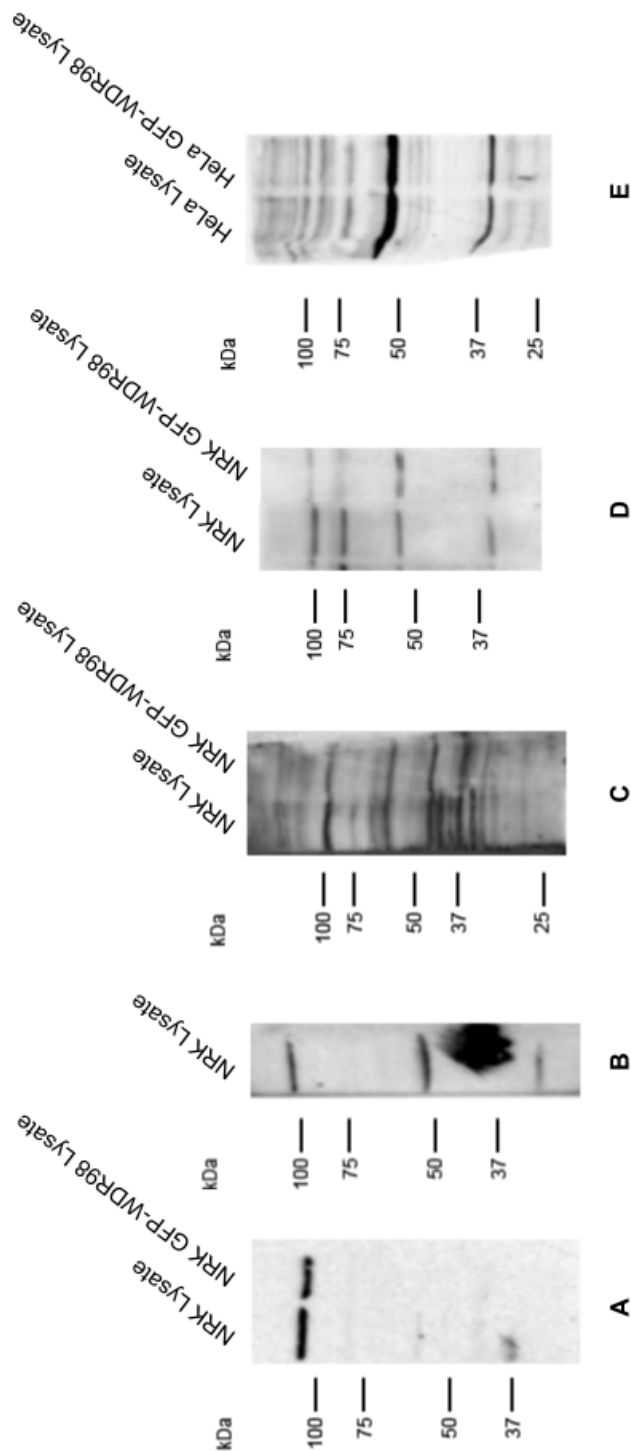


Figure 5: Commercial Antibodies Fail to Recognise WDR98 by Western Blotting

Commercial antibodies to WDR98 (75 kDa) were assessed for specificity via western blotting, using both NRK wild type lysates (or HeLa lysates in E) and lysates from cells which had been transfected with GFP-WDR98 (100 kDa). A - Abcam ab78596, B - SantaCruz sc-390305, C - Proteintech 20111-1-AP, D - Abcam ab137276, E - Abcam ab137276.

3.3.3 DNA Constructs

The following DNA constructs were successfully produced and verified by DNA sequencing (Appendices A i - A xiii):

- GST-WDR98
- WDR98-HIS
- GFP-WDR98
- GFP-WD40
- GFP-MIC
- Δ pMEP4-MYC-GFP-WDR98
- pcDNA5-GFP-humanWDR98
- pGAD-WDR98
- pGBDU-WDR98
- GST-RAB5A
- GST-RAB7A

3.3.4 Commissioned Antibody Testing

In order to assess whether the antibodies from two rabbits raised against GST-WDR98 were indeed specific to WDR98, they were first tested on NRK lysates from cells which had been transfected with GFP-WDR98 (100 kDa) (Figure 6A). This did not produce bands at either the endogenous molecular weight of 75 kDa, nor the weight of the overexpressed protein attached to GFP. There were, however, strong bands at 55 kDa from both antibodies which were consistently present when testing on NRK lysates. This led to the incubation of antibody 'A' with 10 μ g GST-WDR98 pure protein in an attempt to saturate the antibody. The aim of this was to reduce binding to endogenous WDR98 in case it was running much lower than expected at 55 kDa. In Figure 6B, both lanes contain wild type NRK lysate, the first lane was treated with the original antibody without protein incubation, whilst the second lane was treated with the antibody that had been incubated with GST-WDR98. There was no evidence of the 55 kDa band intensity decreasing, nor the intensity of any other bands.

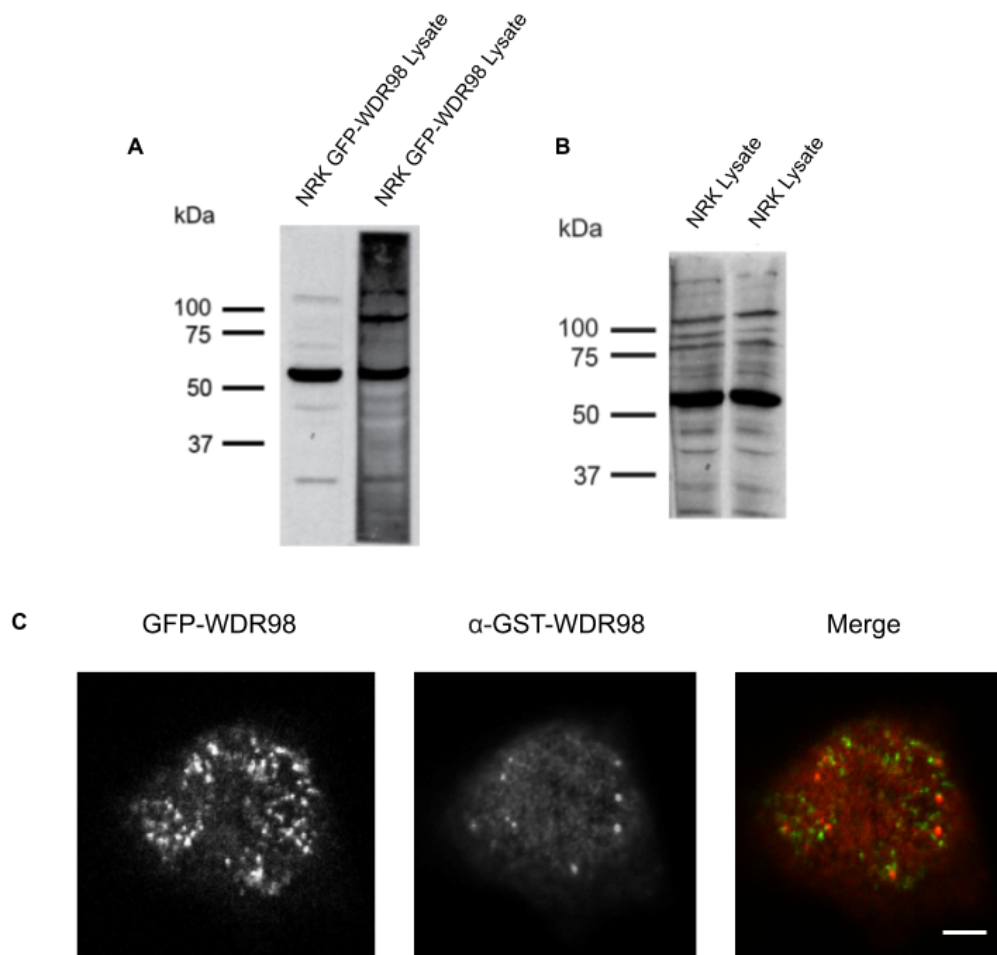


Figure 6: Commissioned GST-WDR98 Antibodies Fail to Recognise WDR98 by Western Blotting or Immunofluorescence

GST-WDR98 protein was produced and sent to Eurogentec for the production of antibodies using two rabbits. Antibodies were characterised by western blotting and immunofluorescence. (A) Both antibodies were tested on NRK lysates from cells transfected with GFP-WDR98, antibody 'A' was used in the first lane, and antibody 'B' in the second. (B) Using NRK lysates, the original antibody 'A' was used in the first lane and antibody incubated with GST-WDR98 was used in the second lane. (C) Antibody 'A' was also tested using immunofluorescence, NRK cells were transfected with GFP-WDR98 and cells were assessed for GFP-WDR98-antibody colocalisation. Pearson correlation coefficient $R = 0.07$. Scale = 20 μm .

Finally, antibody 'A' was also tested via immunofluorescence, again using NRK cells which had been transfected with GFP-WDR98 (Figure 6C). Complete colocalisation would be expected for a WDR98-specific antibody, however the Pearson correlation coefficient produced an R value of 0.07, suggesting a lack of WDR98 specificity.

There were no points of colocalisation and this, combined with the western blotting results, led to the conclusion that these antibodies were not specific to WDR98.

As the first attempt at immunising rabbits against GST-WDR98 was unsuccessful, trial protein productions of GST-WDR98 and WDR98-His were performed, to determine the protein with which we could expect success, with both 4 hour (37 °C) and overnight (21 °C) inductions using IPTG, analysed by western blotting. The GST protein (anti-GST) exhibited large amounts of degradation in both conditions compared to the HIS (anti-HIS) protein (Figure 7), which had the least amount of degradation when induced overnight. WDR98-HIS protein was produced with induction overnight and sent to Eurogentec for new antibody production.

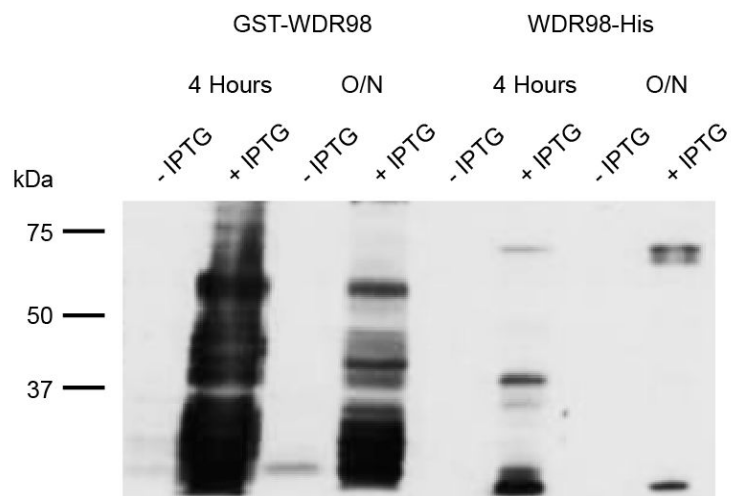


Figure 7: WDR98-HIS Exhibits Less Protein Degradation than GST-WDR98

GST-WDR98 and WDR98-HIS protein was produced.. Tests were performed with both 4 hour (37°C) and overnight (O/N, 21 °C) inductions using IPTG, separated by western blotting, stained with anti-GST and anti-HIS.

Extensive testing has yet to be carried out, but preliminary tests of the new antibodies raised against WDR98-HIS show potential viability with a band matching to that seen with anti-HIS at around 70 kDa, which becomes cleaner after antibody affinity purification (Figure 8).

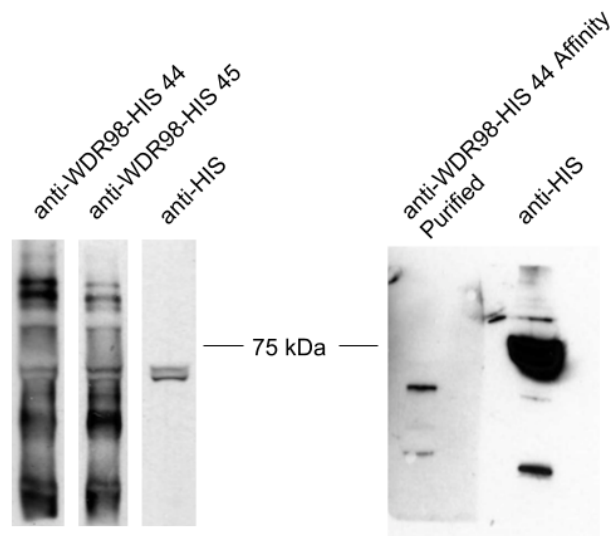


Figure 8: Commissioned WDR98-HIS Antibody may show WDR98 Specificity

Initial tests of the latest WDR98-HIS (75 kDa) antibodies (44 and 45) on WDR98-HIS pure protein compared with anti-HIS on WDR98-HIS pure protein. Blots before (left) and after (right) antibody affinity purification.

The HIS and WDR98 antibody staining does not match up perfectly, so it is possible that there has been some protein degradation. Further optimisation and testing for function in immunofluorescence staining is needed.

3.3.5 Cell Line Characterisation

Cell lines NRK Δ pMEP4-MYC-GFP-WDR98 and HeLa pcDNA5-GFP-humanWDR98 were induced and protein expression assessed by immunofluorescence. The induced expression gave a weaker signal than transiently transfected GFP-WDR98 and over multiple passages, and after resurrection from liquid nitrogen storage, the expression weakened further (data not shown). As such, the decision was made to use transient transfections for all experiments to gain a consistent and higher level of expression.

3.3.6 CRISPR/Cas9 Knockout Characterisation

CRISPR/Cas9 has been employed extensively with the aim of generating a viable WDR98 knockout cell line, however, here we had little success. Due to the lack of a functional WDR98 antibody, cell line characterisation requires gDNA extraction,

PCR and DNA sequencing. Exons 1, 8, 9 and 10 of WDR98's 20 (appendix A viii) have been targeted, with the final trials hitting NRK cells with guide RNAs specific to all four target exons. This yielded one cell line with unusual sequencing in exon 9, including up to three nucleotides identified per base (Figure 9 Panel A). Suspecting a mixed population, cells were further subcloned and exon 9 resequenced, which revealed up to two nucleotides present per base - indicating either a mixed population or a heterozygote. Subcloning was conducted again, with absolute certainty only to clone from single cells, and the results were consistent with previous findings. It is most likely to be a heterozygote cell line, with only one chromosome being affected by the Cas9, as wild type sequence can be identified within the nucleotides reported (Figure 9 Panel B).

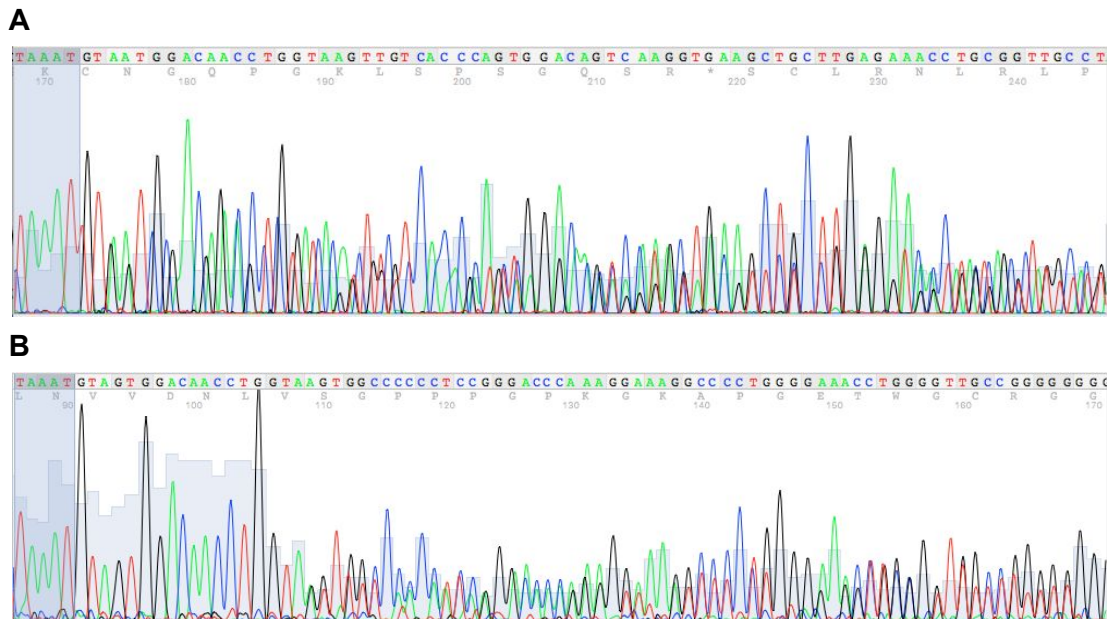


Figure 9: WDR98 CRISPR/Cas9 Chromatograms

Exon 9 sequencing of WDR98 CRISPR/Cas9 knockout attempts in NRK cells, targeting directly after the blue highlighted sequence ‘TAAAT’, up to and including which is WT. Panel A: initial attempt. Panel B: subclone.

The WDR98 Cas9 heterozygote cell line has been characterised, with studies looking at lysosome size and distribution displaying no strong visible difference from wild type (data not shown). Growth rates were also compared, but again there was no notable difference between that of the Cas9 clones or wild type (data not shown).

3.4 Discussion

3.4.1 Assessment of Methods

When characterising antibodies, both commercial and those produced for this project, initial tests on NRK lysates failed to display bands of appropriate molecular weight. The inclusion of GFP-WDR98-transfected cell lysates negated the possibility that this was due to trace amounts of endogenous protein present within the cell, however, this did not improve the results and all antibodies to WDR98 tested failed to show evidence of protein-specificity and therefore useful functionality. Western blotting protocols were adapted to include semi-dry manual as well as iBlot transfers and different membranes were used, including nitrocellulose and PVDF. Unfortunately, none of these methods were successful in improving the antibody

performance and as such, this project has been carried out in the absence of an antibody to WDR98.

All CRISPR/Cas9 resultant cell lines were sequenced as either wild type, with silent mutations or as heterozygous knockouts. Given the numerous and extensive attempts to generate a WDR98 knockout using CRISPR/Cas9; optimisation of the transfection protocol using GFP and the use of a variety and multiple Cas9 guideRNA targets; it was considered that WDR98 complete knockouts were potentially unviable. This was later supported by Pontano Vaites *et al.* (2017) who also struggled to generate a knockout. They instead found success using transient siRNA knockdowns, identifying enlarged endosomal/lysosomal structures using both immunofluorescence and TEM analysis, and accumulation of P62 - indicating a reduction in autophagic flux. RAB7 was also less inclined to localise lysosomally.

A recent publication also supports this, reporting a similar phenotype using a CRISPRi knockdown HeLa cell line. In the absence of WDR98, they observed that early endosomes and lysosomes were enlarged, EEA1 and LAMP1 staining was increased, and EGF and EGFR trafficking to the lysosome was impaired (Tsui *et al.* 2019). Their data supports the theory that WDR98 plays a role in endosomal maturation in concert with MON1B and CCZ1, and its transient nature adds weight to the likelihood that it is not possible to make viable, stable WDR98 knockout cell lines.

3.4.2 Contribution to the Project

The discovery that WDR98 and its unique MIC domain are so well conserved adds weight to the theory that WDR98 performs a fundamental, necessary function and shows that the work carried out during this project to characterise this protein was relevant. The inability to identify an ortholog in yeast using bioinformatics is intriguing, but its complete absence is unlikely. It is more plausible that an ortholog exists which has yet to be identified, perhaps due to quite different amino acid sequences, or the function of a single/group of proteins in yeast may have since evolved to impose distinct functions. Despite this, the phylogenetic findings are interesting, and have yet to be discussed in any other publications.

It was logical to test all available WDR98 antibodies and upon finding none of them suitable, to produce a protein construct with which to attempt to generate our own. It is unfortunate that this too was proven non-specific to WDR98, and this lack of antibody certainly made the overall project more difficult. With hindsight, a GST-tag was perhaps not the best purification tag to use given its large size and inherent immunogenicity. It was considered that the antibody may be specific to GST itself, but western blots containing GST-WDR98 pure protein again did not reveal antigen-specificity (data not shown). If GST had induced a high level of immune response, it is possible the serum received from Eurogentec was a combination of many different antibodies; though these should have been cleared during antibody affinity purification.

Following the unsuccessful antibody trials, the production of several DNA constructs became invaluable. GFP-WDR98 was most commonly used and was used to obtain data from immunoprecipitation and immunofluorescence experiments. Further to this, the generation of GFP-WD40 and GFP-MIC truncated proteins also produced some exciting and novel, results which will be discussed later on in Chapters 4 and 5. GST-RAB constructs and WDR98-HIS were instrumental in conducting the pulldown experiment in the absence of an antibody and the pGAD and pGBDU constructs were used to perform yeast 2 hybrid experiments.

The inducible cell lines did not greatly contribute to the overall project, but did help establish that GFP-WDR98 is not well expressed within cells, through either inducible expression or transient transfection, though the latter yielded higher expression levels and was the method of choice for all experiments discussed.

Similarly, the CRISPR/Cas9 attempts unfortunately did not contribute any data to the project, but the attempts were valid and worthwhile given the breadth of data that stands to be gained from a resource such as a WDR98 knockout. Pontano Vaites *et al.* (2017) and Tsui *et al.* (2019) exemplified this via the data obtained from their knockdowns. WDR98 knockouts could be further exploited to determine the phenotype with regard to expression levels and distribution of MON1B and CCZ1 in the absence of WDR98. This could help elucidate the functionality of this protein complex, how much they rely on each other and whether there is a feedback loop

present or if WDR98 has a role in regulating the expression or function of either of the endosomal maturation proteins it binds to. Finally, given the high level of conservation of WDR98, there are many options for attempting to create mutant organisms to further help understand the function of the protein *in vivo*, as none currently exist, according to publications.

3.4.3 Future Directions

In order to advance research in this field and speed up the acquisition of data, a suitable WDR98 antibody would be invaluable. Further testing of the new WDR98-HIS antibody will hopefully reveal WDR98-specificity.

Increased time spent exploiting bioinformatics techniques and mining sequence databases could eventually lead to the discovery of a yeast WDR98 ortholog, which would also be a highly suitable and cheap research model with which to generate data rapidly.

Finally, siRNA could be a good alternative to generate WDR98 knockdowns and subsequently assess the effect of its reduction in the cell on the endosomal maturation proteins, further elucidating the function of WDR98.

Chapter 4: WDR98 Protein-Protein Interactions

4.1 Introduction

With the aim of uncovering more detail about a possible function for WDR98, a variety of methods were employed to identify its protein binding partners with varying success. Methods selected included yeast 2 hybrid, pulldowns and GFP immunoprecipitations, with further assessment by western blotting and liquid chromatography mass spectrometry in order to maximise the potential to obtain data in the absence of a WDR98 antibody. The two main protein domains of WDR98 were also individually assessed by GFP immunoprecipitation: WD40 and MIC (Figure 3). According to Uniprot databases and JPred predictions, WDR98 has a β -sheet rich propeller-like N terminal 'WD40' domain. Such domains are highly abundant and have well-defined roles in functioning as protein-protein or protein-DNA interaction platforms (Xu and Min 2011). WDR98 also has a highly conserved C terminal α -helical 'MIC' domain of unknown function, which appears to be unique to WDR98 and WDR98-related proteins, of which there are seven isoforms of varying lengths in humans. It has been consistently found that WDR98 interacts with CCZ1 and MON1B (Pontano Vaites *et al.* 2017) and that the WD40 domain is sufficient to elicit binding to these proteins, albeit not as strongly as the full-length protein.

4.2 Methods

4.2.1 Yeast 2 Hybrid Interactions

In addition to existing yeast 2 hybrid pGBDU (bait) and pGAD (prey) vectors containing MON1B, CCZ1, RAB5, and RAB7 that were previously generated by the Pryor lab; pGBDU- and pGAD-WDR98 were created to check for interactions with the aforementioned endosomal proteins. QuickChange PCR was also used to generate RAB constructs locked into the conformation of GDP-bound, GTP-bound and nucleotide free as well as the original wild type construct. SP6 was included as part of a different project and is not relevant to WDR98.

Yeast strains were plated out onto selective synthetic defined (SD)-leucine (pGAD) and SD-uracil (pGBDU) plates (without leucine or uracil in the agar media respectively) and grown individually before mating on YPD (complete amino acids) plates. Plates were later replicated onto SD-leu/-ura plates (no leucine and no uracil combined) and allowed to grow for 48 hours at 30 °C. Yeast from mated cross-sections were mixed with sterile water and pipetted onto new SD-leu/-ura plates and incubated at 30 °C for another 48 hours before replica plating onto SD -leu/-ura/-adenine, the loss of the latter being negated by a successful bait-prey interaction, enabling the production of the missing amino acids and therefore, growth (Figure 10).

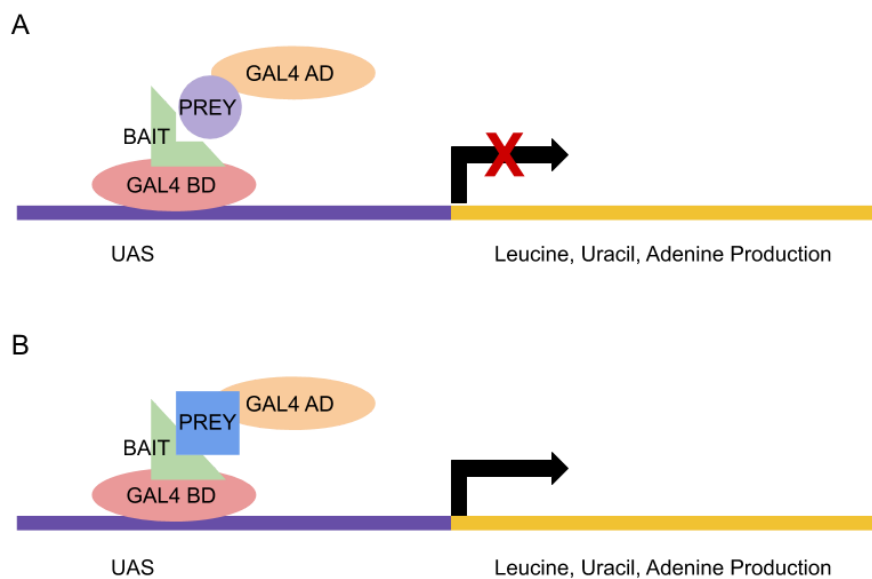


Figure 10: Yeast 2 Hybrid Method

Yeast 2 hybrid utilises the GAL4 transcription factor system which is divided into a binding domain (BD) and activation domain (AD), both of which are required in order to initiate transcription. When this interaction is mediated by the additional interaction of a set of proteins (bait and prey) to which they are fused, transcription is only enabled if there is the ability of the bait and prey proteins to bind. In this instance, transcriptional control was over the genes responsible for producing leucine, uracil and adenine - which were absent from the final yeast plates, and as such, essential for growth. As such, visible yeast growth indicates a positive interaction between the bait and prey proteins. Figure adapted from Lin and Lai (2017).

4.3 Results

4.3.1 Yeast 2 Hybrid Interactions

Yeast 2 hybrid was utilised to assess the potential interactions among proteins involved in the endosomal-lysosomal systems, including WDR98 (Figure 11).

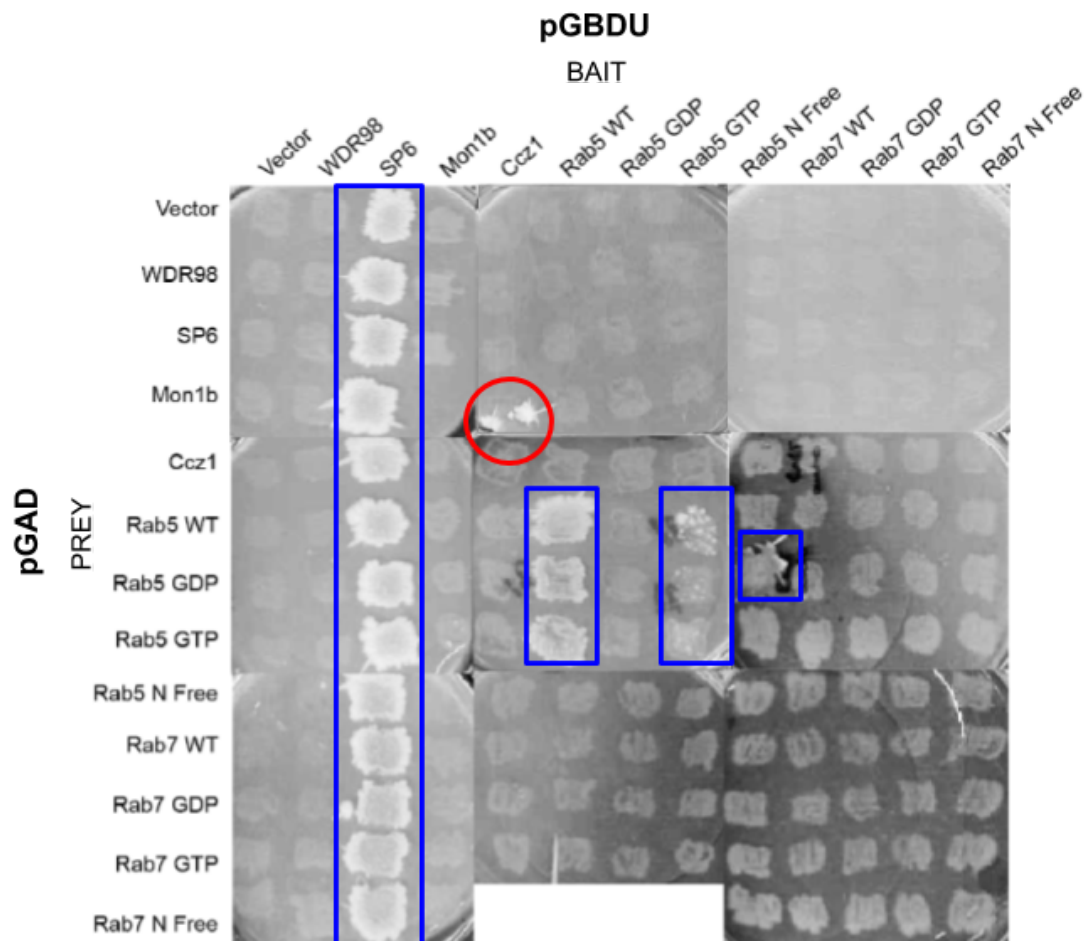


Figure 11: MON1B and CCZ1 Interact via Yeast 2 Hybrid

Yeast 2 hybrid with pGBDU (bait) and pGAD (prey) constructs expressing proteins involved in the endosomal and lysosomal systems, and RABs 5 and 7 locked into different states of nucleotide binding conformations. SP6 was included as part of a different project and is not relevant to WDR98. Yeast were plated out onto selective SD-leucine (pGAD) and SD-uracil (pGBDU) plates and grown individually before mating on YPD plates. Plates were later replicated onto SD-leu/-ura plates and allowed to grow for 48 hours at 30 °C. Yeast from mated cross-sections were pipetted onto new SD-leu/-ura plates and incubated for another

48 hours at 30 °C before replica plating onto SD -leu/-ura/-adenine. Evidence of positive interactions circled red and autoactivation outlined blue. Image taken on day 30.

Following 30 days of incubation for replicate three, the only evidence of positive interactions and resultant growth were between MON1B and CCZ1, with CCZ1 as the bait. Autoactivation also occurred in the pGBDU-SP6 and some of the pGBDU-RAB5 strains. There was no binding evident from any of the WDR98 yeast matings in either direction, though this does not indicate the interactions are not possible, simply that they were not detected via yeast 2 hybrid.

4.3.2 WDR98 and WD40 Domain Interacts with CCZ1 and MON1B

To confirm the interaction between WDR98 and CCZ1 (as indicated by the pilot large scale IP from the Pryor lab), and to examine whether either the WD40 or MIC domain were sufficient for the interaction, an IP was performed using Aminolink beads coupled to an anti-GFP-nanobody and NRK cells transiently transfected with GFP, GFP-WD40, GFP-MIC or GFP-WDR98. The GFP was then immunoprecipitated from lysates. Resultant lysates and immunoprecipitations were separated using SDS PAGE and lysates were blotted for GFP to show levels of GFP-protein expression in lysates pre- and post-IP (Figure 12A, B). Lysates and eluates were also blotted for CCZ1. Data shows consistent CCZ1 bands (SantaCruz) in pre-IP samples (Figure 12C). CCZ1 (55 kDa) was also present in post-IP samples: strongly in full-length GFP-WDR98 and much weaker, but evident, in GFP-WD40 (Figure 12D), suggesting this domain alone is sufficient for a low level of CCZ1-binding.

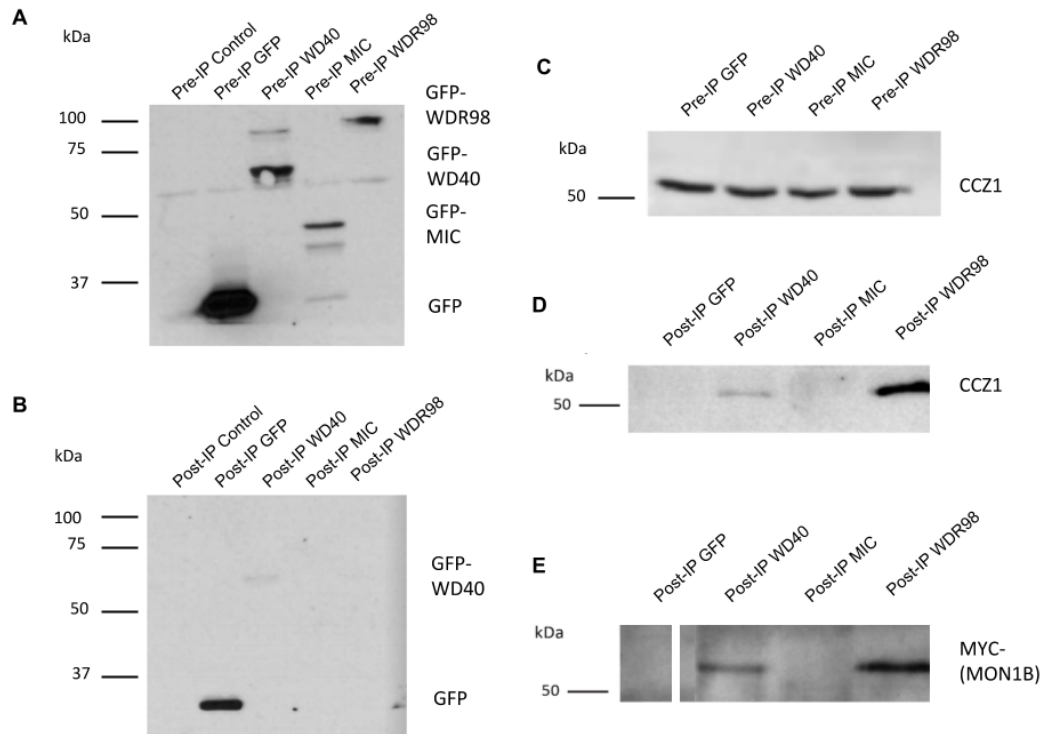


Figure 12: WDR98 and WD40 Domain Interacts with CCZ1 and MON1B

Into NRK cells, GFP, GFP-WD40, GFP-MIC (the two domains of WDR98) and full-length GFP-WDR98 were transfected and GFP immunoprecipitated (IP) and lysates analysed for the presence of CCZ1 and MYC-MON1B (using a MYC-MON1B-inducible cell line). (A) Pre-IP lysates analysed using anti-GFP. (B) Lysates post-IP. (C) Lysates pre-IP following transfections. (D) Post-IP eluate analysis for the presence of CCZ1 (55 kDa). (E) Post-IP eluate analysis for the presence of MYC-MON1B (59 kDa). GFP (27 kDa), GFP-WD40 (70 kDa), GFP-MIC (47 kDa), GFP-WDR98 (100 kDa).

To confirm the interaction between WDR98 and MON1B (again as indicated by the pilot IP from the Pryor lab) Flp-In HeLa cells with inducible MYC-MON1B expression were transfected with GFP, GFP-WD40, GFP-MIC or GFP-WDR98 and the MYC-MON1B expression induced. This was necessary due to the lack of a suitable MON1B or WDR98 antibody. As done previously, the GFP was then immunoprecipitated from lysates, resultant lysates and immunoprecipitations were analysed using SDS PAGE, and lysates were blotted for GFP and MYC. Replicating the data from CCZ1 interaction experiments, MYC-MON1B (59kDa) was present in post-IP samples: strongly in full-length GFP-WDR98 and weaker, but evident, in GFP-WD40 (Figure 12E), again suggesting the WD40 domain is sufficient for binding to MON1B.

4.3.3 Alternative Immunoprecipitation Interactions

Using data compiled from the BIOGRID database (Oughtred *et al.* 2019) and the Pryor lab large scale IP pilot, a protein interaction network was produced to display significant proteins (in terms of function, localisation and disease) which may potentially bind to WDR98, MON1 and CCZ1 (Figure 13). This was used to identify proteins to investigate further in terms of an interaction with WDR98.

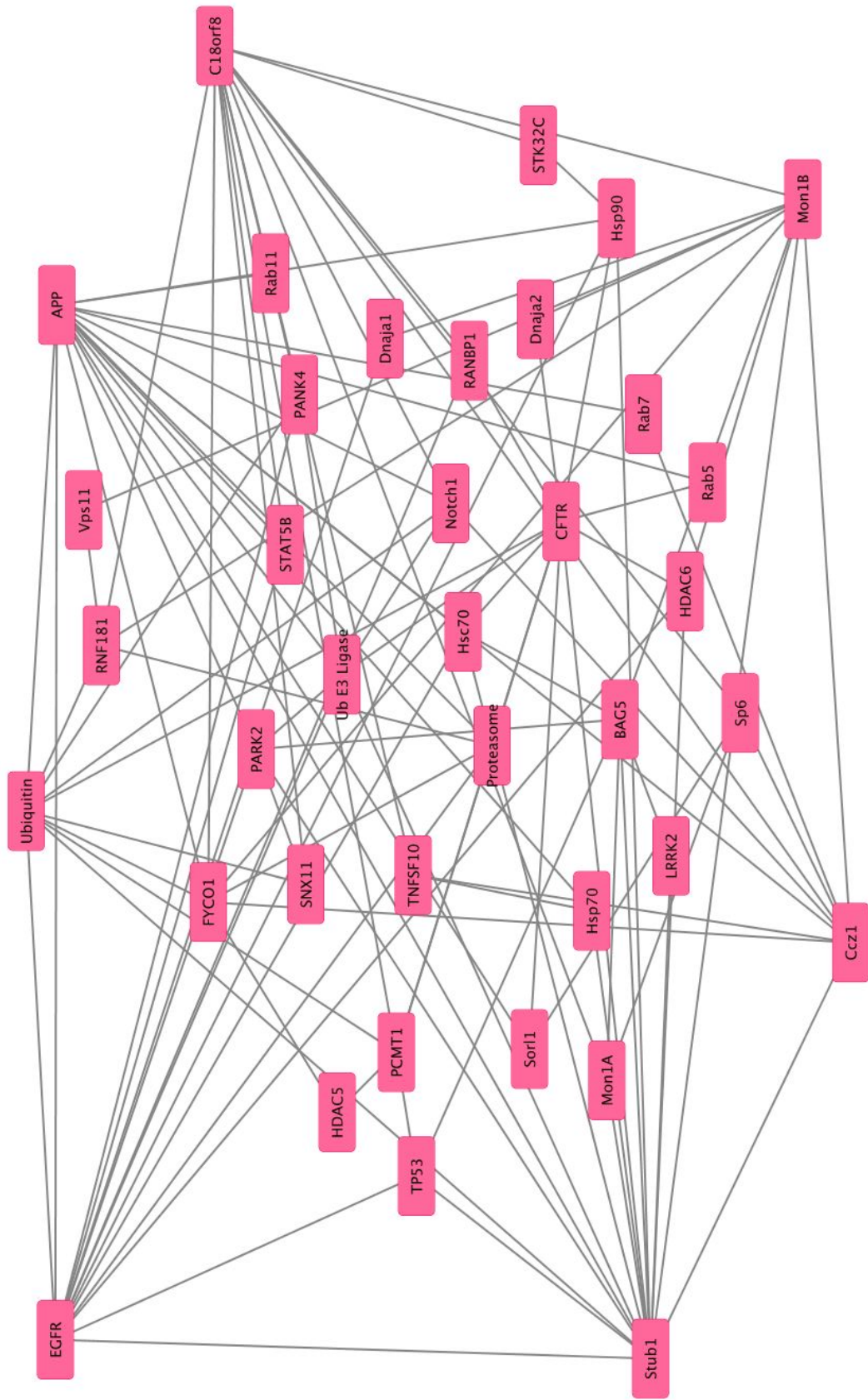


Figure 13: WDR98 Potential Protein Interactions Network

Data compiled from the BIOGRID database and preliminary Pryor lab data was used to produce a potential protein interaction network.

Using specific antibodies, further immunoprecipitations were conducted to look for confirmation of interactions between full-length WDR98 and proteins identified on the potential protein interactions network. Given previous results (figure 12), CCZ1 was used as a positive control. Antibodies against HDAC6, HSP90, STUB1 (Figure 14A), APP and Beta-Amyloid (Figure 14B) were tested on pre-IP lysates and post-IP eluates.

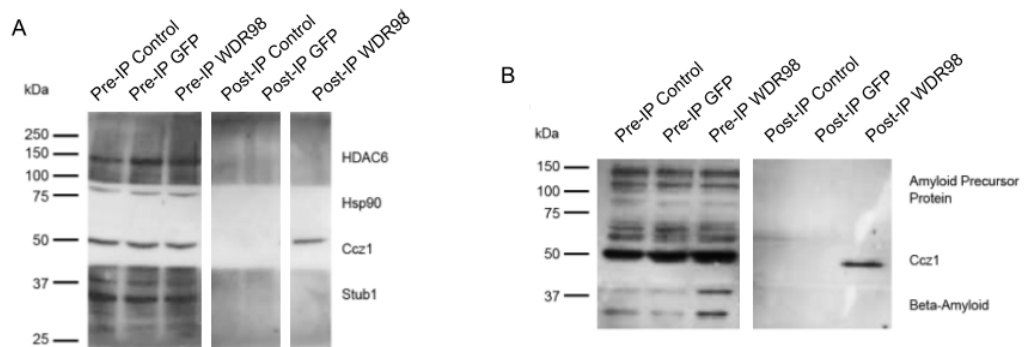


Figure 14: WDR98 does not Interact with HDAC6, HSP90, STUB1 or APP

GFP and GFP-WDR98 were produced in transiently transfected NRK cells and GFP was immunoprecipitated (IP), then lysates analysed for the presence of a variety of proteins involved in trafficking. (A) Lysates analysed using anti-HDAC6, HSP90, CCZ1 and STUB1. (B) Pre-IP lysates analysed using anti-APP, CCZ1 and Beta-Amyloid.

There were no new WDR98 protein interactions identified through these IPs, which can be considered true negatives due to the positive antibody signal in pre-IP lysates and absence of signal in post-IP eluates, in the presence of a positive signal indicating the interaction between GFP-WDR98 and CCZ1.

4.3.4 RAB Pulldowns

To clarify previously reported RAB and WDR98 interactions (Gillingham *et al.* 2014, Pontano Vaites *et al.* 2017), recombinant GST-RAB5 and GST-RAB7 were produced and loaded with nucleotides (GMPPNP, GDP and GTP) before incubation with recombinant WDR98-HIS in the presence of glutathione sepharose beads. GST

bound to glutathione sepharose beads as indicated in Figure 15 and a weak amount of WDR98-HIS was detected in all conditions, including the GST-absent control, suggesting a weak ability to bind directly to the beads themselves. As the level of WDR98-HIS did not increase in any of the GST-RAB conditions compared to the negative control, and smudging in the GST-RAB5 lane was proven to be nonspecific through repetition, this is evidence that there is not a direct interaction between WDR98 or any nucleotide conformation RAB5/7 protein when in isolation *in vitro*.

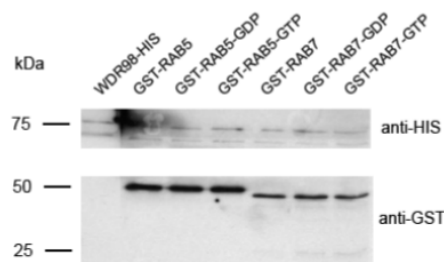


Figure 15: WDR98 does not Directly Interact with RAB5/7 *in vitro*

WDR98-HIS (75 kDa) was incubated with GST-RABs 5 (49 kDa) and 7 (48 kDa) in varying nucleotide-locked conformations (with nucleotides present) to test for protein-protein interactions via pull-down with glutathione beads in a wide range of different buffers (representative example blot shown).

4.3.5 Immunoprecipitation Liquid Chromatography Mass Spectrometry

Immunoprecipitations were conducted again as previously described but on a much larger scale and with liquid chromatography mass spectrometry (LC-MS/MS) analysis of eluates for GFP, GFP-WDR98 and GFP-MON1A, the latter included to aid identification of binding partners specific to WDR98, as trial analysis by LC-MS/MS revealed common interactors for both proteins. Triplicate repeats and Fisher's exact test and Benjamini-Hochberg multiple test correction statistical analysis revealed statistical confidence ($p < 0.00071$) in GFP-WDR98 binding to only a handful of proteins, including MON1B (Table 3).

Table 3: MON1B and Structural Proteins are Binding Partners of WDR98

GFP and GFP-WDR98 were produced in transiently transfected NRK cells and GFP was immunoprecipitated, then eluates analysed via liquid chromatography mass spectrometry. GFP-WDR98 positive results from three replicates were compared to results from the GFP

control analysis, and statistical analysis (Fisher's Exact Test and Benjamini-Hochberg multiple test correction) revealed the proteins displayed in this table were able to bind to GFP-WDR98 with significant statistical confidence ($p < 0.00071$).

<u>GFP-WDR98 Binding Partners</u>	<u>Fisher's Exact Test p =</u>
MON1B	0.0001
PSMC5	0.0001
VIMENTIN CLUSTER	0.0001
MYOSIN CLUSTER	0.0001
FIBRONECTIN	0.0001
CYTOSPIN A	0.0001
PSMC2	0.00014
CLATHRIN HEAVY CHAIN 1	0.00028
HISTONE H2B CLUSTER	0.0003
PLASTIN CLUSTER	0.00036
FLOTILLIN 2	0.00063

The proteins identified mainly included cell-structural and transport proteins, notably clathrin heavy chain 1 which is required for clathrin-coated vesicles in protein trafficking, and two proteins (PSMC5 and PSMC2) of the 26S proteasome regulatory subunit.

MON1B was the only repeat protein that we had previously confirmed interacted with WDR98 to be deemed significant in these experiments. CCZ1 was present in the analysis but not to a high enough level to be considered a significant binding partner following multiple test corrections ($p = 0.0016$).

4.4 Discussion

4.4.1 Confirmation of Interactions

Overall, the main results obtained from these protein-protein interaction studies are confirmation of the interactions between WDR98 and CCZ1 and MON1B of the endosomal maturation complex. Whilst this has since been replicated and published in new literature (Pontano Vaites *et al.* 2017), we have also discovered that the WD40 domain is sufficient to bind to CCZ1 and MON1B, albeit with apparent weaker affinity than the full-length protein, whereas the MIC domain alone does not exhibit any CCZ1 or MON1B binding. The beta-propeller structure predicted to exist in the WD40 domain often acts as a protein-protein or protein-DNA binding platform in other proteins (Xu and Min 2011), which would appear to also be the case in WDR98. This may leave the unique helical MIC domain free to perform an alternative function to the binding of the WD40 domain. These findings localise WDR98 to maturing endosomes and lysosomes within the cell, and also suggest cross-species conservation as the rat-derived WDR98 protein was observed to bind both rat (large-scale IP LC-MS/MS) and human (IP western blotting) MON1B orthologs. Analysis via LC-MS/MS also supports a highly specific role for WDR98 in supporting endosomal maturation and trafficking to the lysosome as very few proteins were found to bind significantly to WDR98 using this method, including MON1B, structural proteins and members of the 26S proteasome regulatory subunit. The significance of the proteasome regulatory subunit-binding again highlights a well-defined role for WDR98 in helping to regulate protein trafficking and degradation. As does the clathrin heavy chain, which is essential for receptor-mediated endocytosis and vesicular structure. However, these two latter interactions have only been identified using this method, and further validation using other approaches would be required to achieve confirmation.

MON1 and CCZ1 have been shown to bind to both RAB5 and RAB7 in *Caenorhabditis elegans* (Poteryaev *et al.* 2010) and *Drosophila melanogaster* (Yousefian *et al.* 2013, Hegedűs *et al.* 2016). *Drosophila melanogaster* WDR98 (CG8270) has previously been shown to bind to RAB5 (Q to L mutant that stabilises GTP bound RABs) (Gillingham *et al.*, 2014), whereas Pontano Vaites *et al.* (2017) detected WDR98 binding to RAB7 (T to N mutant that stabilises GDP bound RABs). However, RAB mutants such as T22N can cause RAB instability, which promotes

non-specific interactions (personal communication with Miguel Seabra, Imperial College, London). Our data suggests there may not be a direct interaction between WDR98 and RAB5 or RAB7. This is plausible as published data from Gillingham *et al.* (2014) and Pontano Vaites *et al.* (2017) cannot rule out indirect interactions via linker proteins. The presence of additional proteins, for example MON1B and/or CCZ1, may be required to enable an interaction between WDR98 and RABs, or they may simply work together as part of the same protein complex, for example during the RAB5 - RAB7 switch in endosomal maturation process. Regardless, this is the first *in vitro* interaction study of RAB5 and 7 with WDR98 in isolation. Following rigorous testing of differing protein concentrations, different types of cross-linking beads and a wide variety of buffers, our conclusion is that they are unable to interact directly *in vitro*.

4.4.2 Assessment of Methods

Yeast 2 hybrid was a relatively quick and economical way to test for interactions between multiple different proteins at the same time, with a visual positive/negative result. However, this method is not without its limitations. Most often, protein-protein interactions are not displayed as positive in both bait-prey directions, as seen here between the well-established interactors MON1B and CCZ1. Their positive interaction was only displayed using CCZ1 as the bait and MON1B as the prey, not the other way around, meaning it is necessary to try both conformations to check for false negatives. There is also the possibility of false positives, as also shown in this study. Autoactivation occurred in the pGBDU-SP6 and some of the pGBDU-RAB5 strains, which is a common limitation for some protein fragments within pGBDU vectors (Galletta and Rusan 2015). Whilst this rarely occurs in both pGBDU and pGAD vectors containing the same protein, it does restrict the data that can be collected, as true positives, only occurring in one bait-prey direction, may be masked by the false positives. For WDR98, there was no autoactivation or evidence of binding from any of the yeast matings in either direction. This does not indicate the interactions are not possible, simply that they are not detected in this yeast 2 hybrid approach.

The use of GFP immunoprecipitation has proven invaluable and can provide data using a variety of different analyses. For known targets, western blotting analysis is

quick and relatively economical once the relevant antibodies are acquired, and there are methods to circumnavigate the lack of a suitable antibody including the generation of inducible-expression cell lines and transfection of tagged DNA constructs, though it is a more lengthy and difficult process to produce the additional molecular constructs required. LC-MS/MS analysis completely removes the need for protein-specific antibodies and also allows proteome-wide, non-targeted forward screening. Here, this was able to illuminate WDR98 interactions with cell structural proteins and proteasome subunits which may otherwise not have been identified. The statistical analysis used, Fisher's Exact Test with Benjamini-Hochberg multiple test corrections are highly stringent, ($p < 0.00071$), however they may be so stringent that we are also dismissing other positive results as not significant, including CCZ1 which had a p value of 0.0016, despite displaying a clear positive result from many IPs with western blotting analysis. Further repeats of these experiments would be useful, however it is costly given the amount of transfection reagent and volume of consumables required to achieve the level of sample necessary for the complex LC-MS/MS analysis, in addition to the costs of the MS itself.

Whilst isolated pulldown experiments can also do away with the requirement for protein-specific antibodies, they too are not without their limitations. Despite trying many different conditions to detect a WDR98-RAB5/7 interaction *in vitro*, it is possible that the optimal condition simply was not found in order to facilitate an interaction, highlighting the inherent difficulties behind *in vitro* studies. Many different replicates were carried out until the decision was made to conclude that we could not find evidence of a direct WDR98-RAB5/7 interaction *in vitro*. A further problem is finding suitable cross-linking beads, because all beads tested (including Ni-NTA agarose) were found to exhibit some level of nonspecific binding to either GST-RAB5/7 or WDR98-HIS. An increase in western blotting band intensity could have been taken as a positive result compared to nonspecific binding in the negative control, however this did not occur within our experiments and the conclusion remains that WDR98 is unable to directly interact with RAB5/7, in any nucleotide conformation, *in vitro*.

4.4.3 Future Directions

It would be very helpful to have an atomic-resolution structure of WDR98 to visualise the distinct domains, thus enabling a model of how the interactions with MON1B and CCZ1 might take place to be created. Atomic-resolution structures of WD40-type domains already exist and this is a domain which is relatively well researched (Zhang *et al.* 2019); the truncated WDR98 WD40 construct could be used for protein crystallisation studies. A crystal structure would also help with the investigation of the secondary structure of the central part of the protein, which is currently unknown. Current understanding is that a complex of both MON1B and CCZ1 dimers are required to form the endosomal maturation complex (Kiontke *et al.* 2017) and a crystal structure may help elucidate how WDR98 fits into that complex. Further studies such as analytical size-exclusion chromatography or analytical ultracentrifugation could determine whether WDR98 too is required in duplicate. Potential obstacles to this include the fact that WDR98 is a rather unstable protein that tends to be insoluble, with attempts to produce high amounts of protein rendering relatively low yields (data not shown).

Repetition of the IPs with LC/MS-MS analysis would be costly but useful to check for any positive results which may have been missed, given that CCZ1 was determined to be a non-statistically significant binding partner.

Finally, as with all studies of this nature, development of a suitable antibody would be invaluable, as would generation of a WDR98 knockout or knockdown model. The model cell line could be used to analyse the effects on the levels and behaviour of MON1B and CCZ1, and the overall effect on endosomal maturation and lysosomal trafficking.

Chapter 5: WDR98 Subcellular Localisation

5.1 Introduction

To further characterise WDR98 and understand its likely function, addressing the final experimental objective, immunofluorescence was utilised to establish where WDR98 localised to within the cell. Due to the lack of an appropriate antibody, this was achieved through transient transfection of a GFP-WDR98 encoding vector into NRK cells (unless otherwise stated) and analysis using confocal microscopy. Given the previous literature claim that WDR98 localised to the lysosome (Schroder *et al.* 2007) we expected to find it colocalising with lysosomal membrane protein LGP120 and checked other cellular locations thereafter. CCZ1 was studied following the results of the pilot immunoprecipitation trial by the Pryor lab which suggested that CCZ1 and WDR98 interacted. Again due to the lack of an antibody to MON1B, a GFP-WDR98 encoding vector was transiently transfected into HeLa cells with inducible expression of MYC-MON1B. Other proteins analysed for colocalisation include RAB5, RAB7 and LC3. The localisation of the distinct protein domains, WD40 and MIC, were also studied using GFP tagged constructs. These were also observed to localise individually to the lysosome. It is important to note that as transfections were required to collect all of this data, thus the WDR98 subcellular localisation and distributions are that of cell lines overexpressing the target protein and may differ from endogenous WDR98 protein.

5.2 Methods

5.2.1 WDR98 Overexpression

Following the production of the DNA constructs, GFP-WDR98, GFP-WD40 and GFP-MIC, these were transformed into bacteria for replication and DNA extracted via Midiprep (Methods 2.2.5). The DNA was then transfected into NRK cells which were incubated for 24 hours prior to fixation and staining (Methods 2.2.23).

5.2.2 Transfection Optimisation

In order to obtain optimal transfections, and as such the highest level possible of protein expression, the transfection protocol was rigorously tested prior to experimentation. Two reagents were assessed: X-tremeGENE HP and Genecellin. Differing ratios of transfection reagent to DNA were trialled, ranging from 1:1 to 4:1 with varying cell confluency levels. The conditions with the brightest GFP signal by microscopy were X-tremeGENE HP 2:1 DNA with cells at 80-90% confluency at the time of transfection.

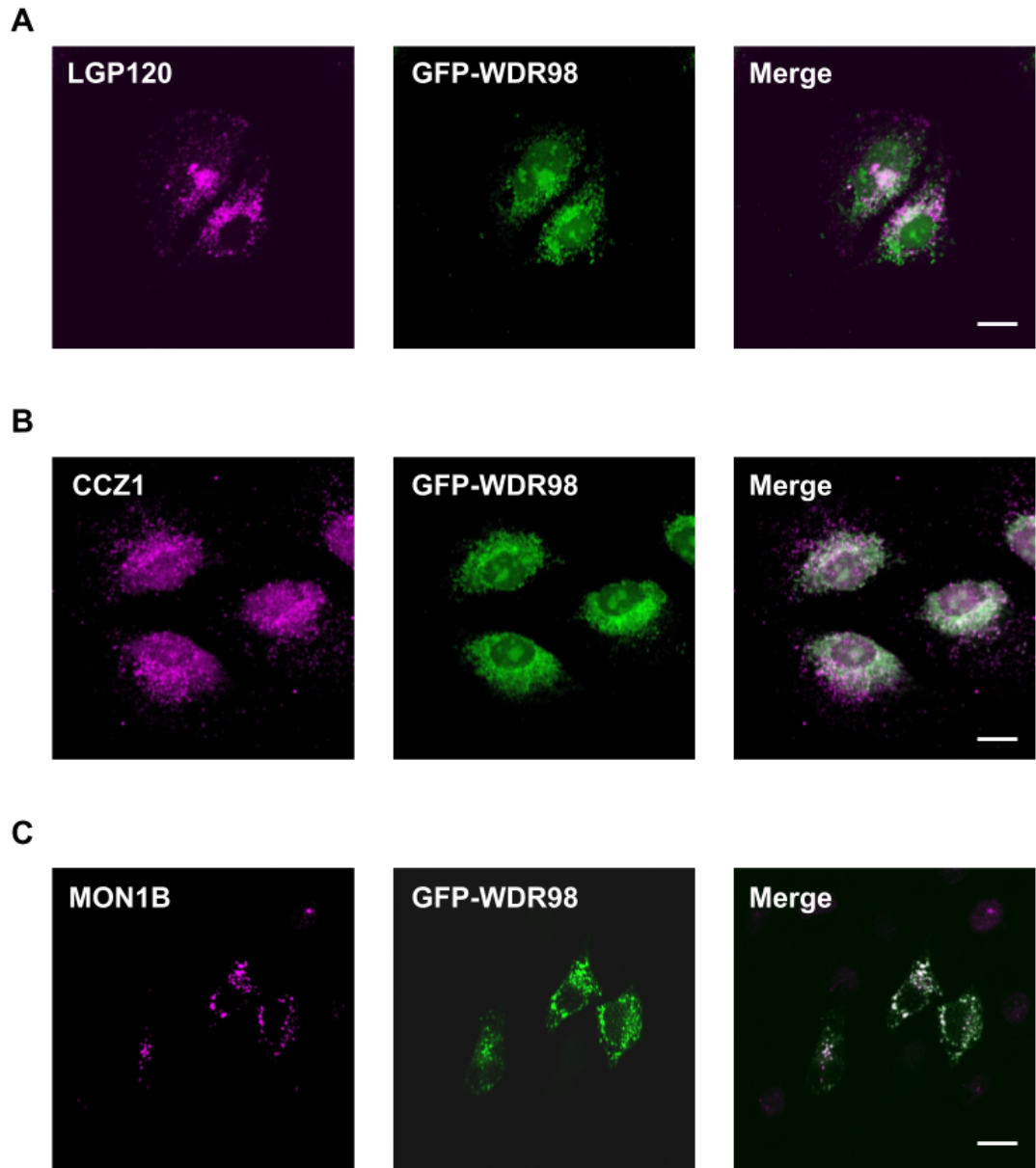
5.2.3 Image Analysis

Images were acquired using a Zeiss LSM 880 Axio Observer.Z1 inverted confocal laser microscope running Zen software version 2.3. Images were collected as Z-stacks and compressed into maximum intensity projections. Further processing was completed using FIJI software (Rueden et al. 2017, Schindelin et al. 2012), and analysis for colocalisation was completed using the Coloc2 plugin. Pearson's correlation coefficient was calculated using images of at least 30 separate cells from at least three replicate experiments and determining the average R value of each. $R = -1$ would indicate localisation at opposite sites, $R = 0$ would indicate no relationship, negative or positive, between signals, and $R = 1$ would indicate perfect colocalisation.

5.3 Results

5.3.1 WDR98 Subcellular Localisation

The subcellular location and distribution of WDR98 was characterised, in the absence of a suitable antibody, by overexpression of GFP-WDR98 using transiently transfected NRK cells. This was assessed against a range of organelle markers, including LGP120 (lysosome), CCZ1 and MON1B (endosomes and lysosomes), RAB5 (early endosomes), RAB7 (late endosomes) and LC3 (autophagosomes). Colocalisation has been qualified both visually (Figure 16) and semi-quantitatively using Pearson's correlation coefficient analysis (Table 4). A GFP antibody was used to supplement the GFP construct transfections due to difficulties in expression and a weak signal.



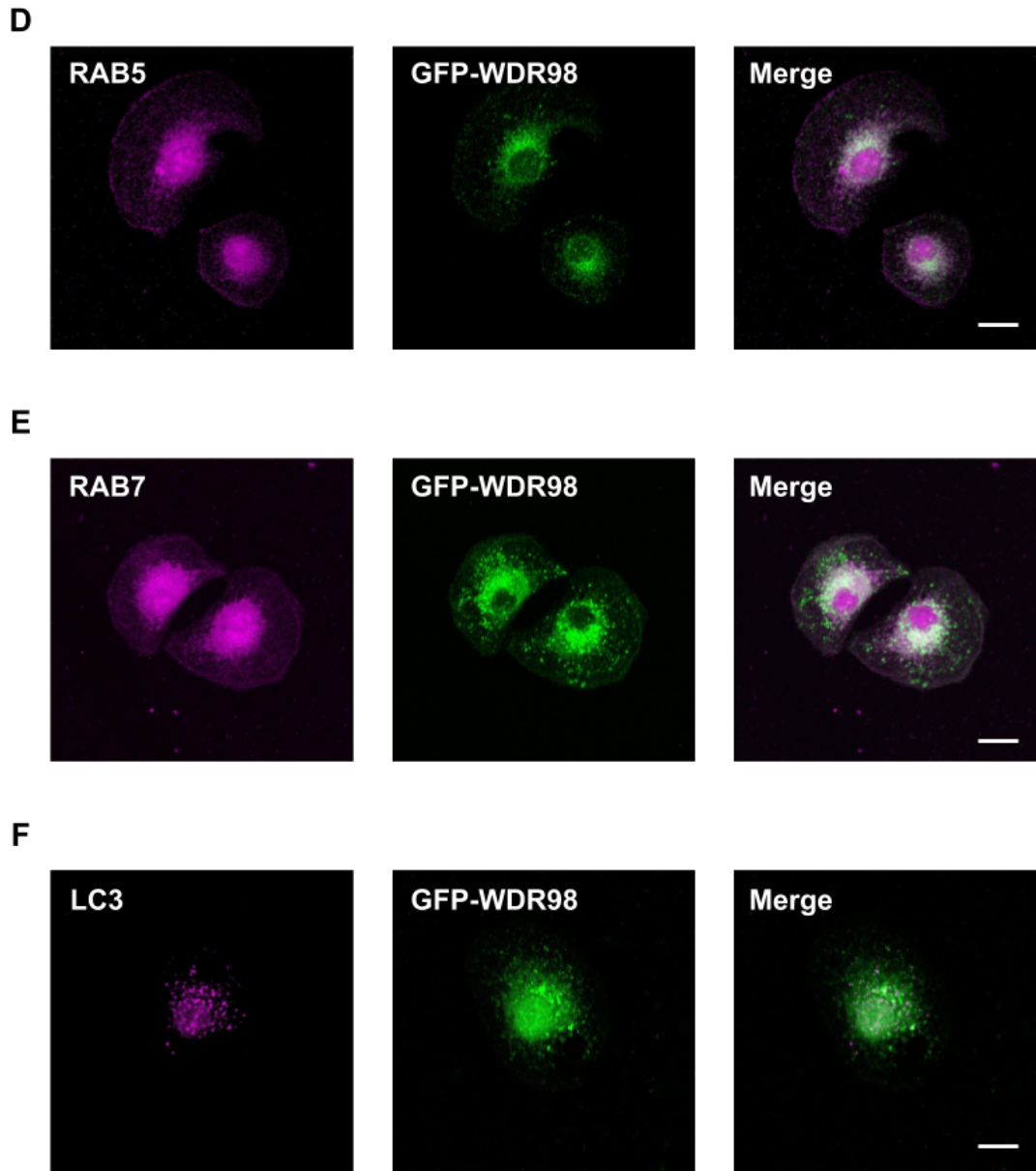


Figure 16: WDR98 Colocalises with LGP120, CCZ1, MON1B, RAB5 and RAB7

GFP-WDR98 was transiently transfected into NRK cells using X-tremeGENE HP. This was assessed against a range of organelle markers, including LGP120 (A, lysosome), CCZ1 (B) and MON1B (C, endosomes and lysosomes), RAB5 (D, early endosomes), RAB7 (E, late endosomes) and LC3 (F, autophagosomes). Images were acquired using a Zeiss LSM 880 Axio Observer.Z1 inverted laser scanning confocal microscope. Images were collected as Z-stacks and compressed into maximum intensity projections. Further processing was completed using FIJI software and analysis for colocalisation was completed using the Coloc2 plugin. Scale bar = 20 μ m.

Table 4: Pearson Correlation Coefficients of GFP-WDR98 Subcellular Colocalisation

Immunofluorescence images of WDR98 subcellular localisation were acquired by transiently transfecting a GFP-WDR98 encoding vector into NRK cells (with the exception of MYC-MON1B inducible expression HeLa cells) and counterstaining with a range of markers for endosomal and lysosomal systems. Images were captured using a Zeiss LSM 880 Axio Observer.Z1 inverted laser scanning confocal microscope. Images were collected as Z-stacks and compressed into maximum intensity projections. Further processing was completed using FIJI software and analysis for colocalisation was completed using the Coloc2 plugin. A minimum of 30 cells were assessed per condition with the average Pearson correlation coefficient calculated as a quantification of colocalisation between markers.

<u>Organelle</u>	<u>Marker</u>	<u>Pearson Coefficient R =</u>
Lysosome	LGP120	0.55
Endosome and Lysosome	CCZ1	0.85
Endosome and Lysosome	MON1B	0.83
Early Endosome	RAB5	0.51
Late Endosome	RAB7	0.78
Autophagosome	LC3	0.16

GFP-WDR98 was able to localise to lysosomes, specifically those surrounding the nuclear periphery (Figure 16A) and reported a moderate association score of 0.55. CCZ1 and MON1B (MYC-MON1B induced expression in HeLa cells) consistently displayed GFP-WDR98 colocalisation (Figure 16B and 16C), with strong correlation coefficients of 0.85 and 0.83 respectively, placing WDR98 at both endosomes and lysosomes. GFP-WDR98 was also observed to colocalise with both RABs 5 and 7 (Figure 16D and 16E), though less so with the former at early endosomes (R = 0.51) and more so with RAB7 at late endosomes (R = 0.78). Finally, there appears to be little to no association of WDR98 with LC3 at autophagosomes (Figure 16F) with a Pearson correlation coefficient score of 0.16.

5.3.2 WDR98 Truncated Domains Subcellular Localisation

The subcellular location and distribution of the individual WDR98 domains, GFP-WD40 and GFP-MIC, were also characterised in a lysosomal context (Figure 17) and Pearson correlation coefficient values were calculated (Table 5).

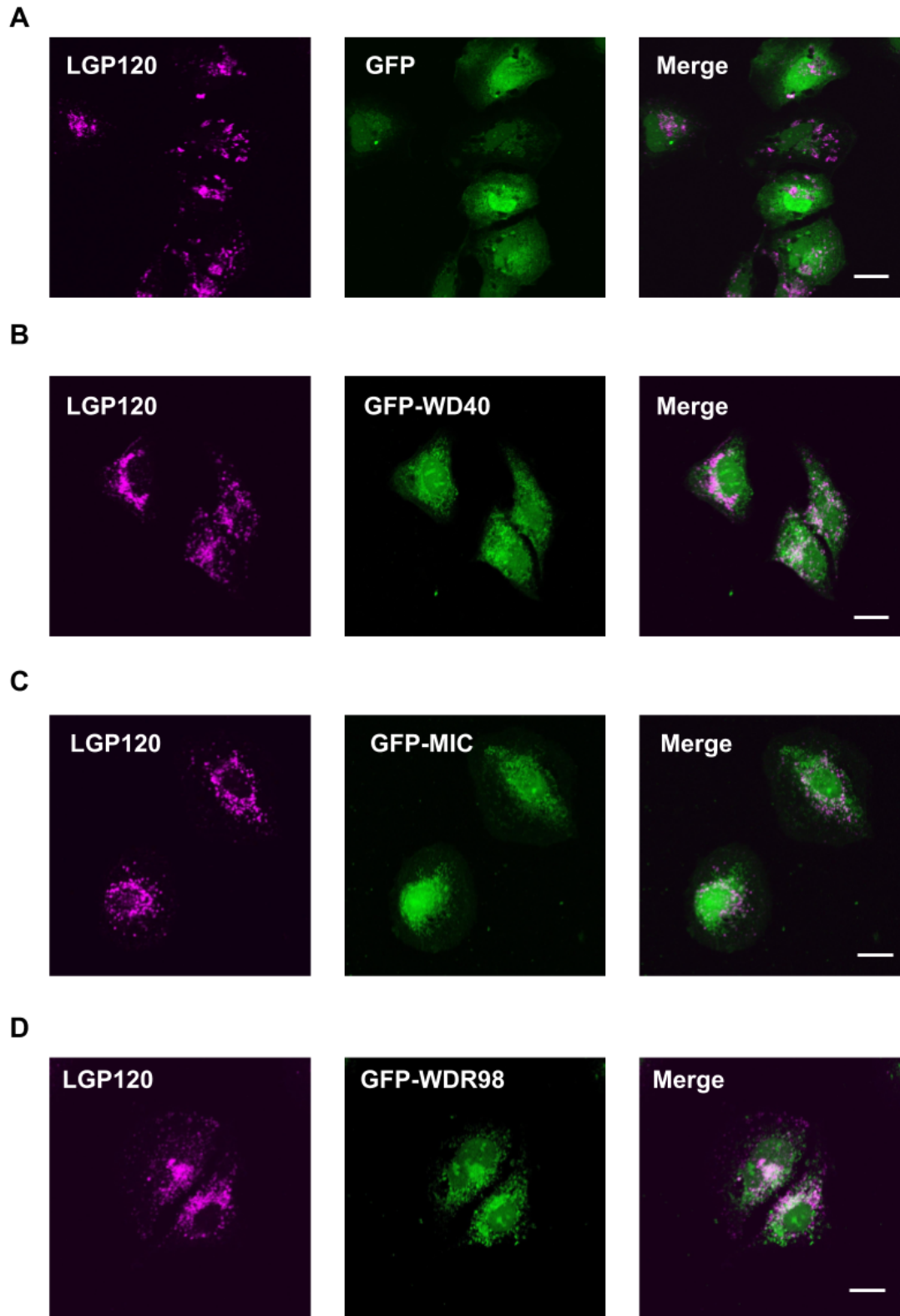


Figure 17: WDR98 and its Domains Localise to the Lysosome

The subcellular location and distribution of GFP, GFP-WD40, GFP-MIC and GFP-WDR98 were characterised by transient transfection of NRK cells with the relevant expression vector using X-tremeGENE HP. This was assessed against LGP120, a lysosomal membrane

protein and marker. Images were acquired using a Zeiss LSM 880 Axio Observer.Z1 inverted laser scanning confocal microscope. Images were collected as Z-stacks and compressed into maximum intensity projections. Further processing was completed using FIJI software and analysis for colocalisation was completed using the Coloc2 plugin. Scale bar = 20 μ m.

Table 5: Pearson Correlation Coefficients of GFP-WDR98 Truncated Constructs at the Lysosome

Immunofluorescence images of GFP, GFP-WD40, GFP-MIC and GFP-WDR98 subcellular localisation were acquired by transient transfection and counterstaining for LGP120 to identify the lysosome. Images were captured using a Zeiss LSM 880 Axio Observer.Z1 inverted laser scanning confocal microscope. Images were collected as Z-stacks and compressed into maximum intensity projections. Further processing was completed using FIJI software and analysis for colocalisation was completed using the Coloc2 plugin. A minimum of 30 cells were assessed per condition with the average Pearson correlation coefficient calculated as a quantification of colocalisation between two markers.

<u>Transfected Construct</u>	<u>LGP120 Colocalisation Pearson Coefficient R =</u>
GFP	0.35
GFP-WD40	0.57
GFP-MIC	0.54
GFP-WDR98	0.55

Both GFP-WD40 and GFP-MIC individually localised to the lysosome (Figure 15B and 17C), with a similar association level to the full length protein (Figure 17D) and slightly more so than GFP alone (Figure 17A). The GFP control transfection obtained a Pearson score of 0.35 in contrast to the WDR98 domain scores of 0.57 for GFP-WD40 and 0.54 for GFP-MIC.

5.3.3 WDR98 Overexpression Phenotype

The approximate number, size and distribution of lysosomes in NRK cells with and without GFP or GFP-WDR98 expression was quantified to check whether transfections or increased expression of this protein affected lysosome morphology. Though there appeared to be no significant perturbations in terms of number of

lysosomes, there were clear differences between untransfected NRK controls and transfected cells, with both GFP- and GFP-WDR98-transfected cells displaying notable lysosomal swelling (Figure 18).

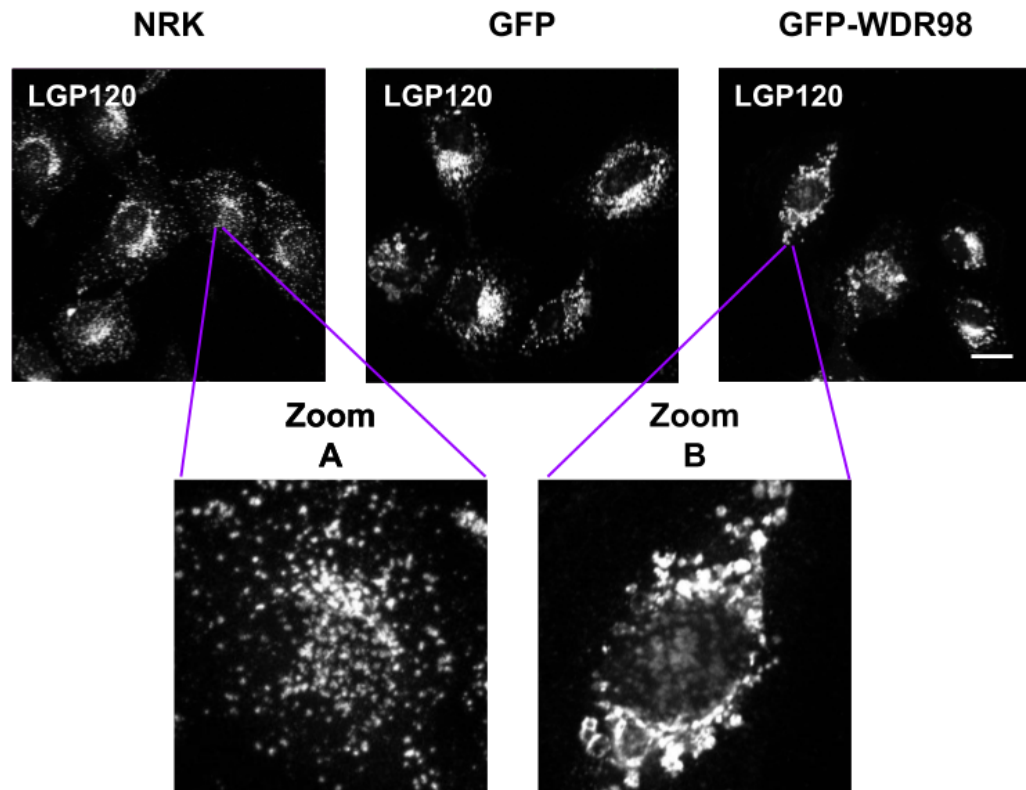


Figure 18: Transfection of GFP-Constructs causes Lysosomal Swelling

The effects of transfection using X-tremeGENE HP to insert GFP-sequences into NRK cells was assessed in terms of lysosomal phenotype using LGP120 as a lysosomal marker. Untransfected controls (Zoom A). Transfected cells (Zoom B). Images were acquired using a Zeiss LSM 880 Axio Observer.Z1 inverted laser scanning confocal microscope. Images were collected as Z-stacks and compressed into maximum intensity projections. Scale bar = 20 μm . 4x Zoom.

5.4 Discussion

5.4.1 Confirmation of Localisation

Data from this study and also from Pontano Vaites *et al.* (2017) and Tsui *et al.* (2019) confirm the localisation of WDR98 to the endosomal and lysosomal systems within the cell. The immunofluorescence images published by Tsui *et al.* (2019) also

indicate a preference for WDR98 to colocalise at lysosomes close to the nuclear periphery, which have been suggested to contain a lower pH and be the more active set of lysosomes, compared to those around the cell membrane (Johnson *et al.* 2016). It is both novel and intriguing that the two separate protein domains, WD40 and MIC, also localised to the lysosome in a manner that was similar to the full length protein. This could mean that either domain is responsible for anchoring the protein to the lysosomal membrane - potentially the MIC domain as the WD40 domain has the capability to bind to MON1B and CCZ1, whereas MIC does not.

There is now an increasing body of evidence that WDR98 functions within the endosomal-lysosomal system, which has been added to by these immunofluorescence studies.

5.4.2 Assessment of Methods

The biggest caveat of the methods used to assess WDR98 localisation is that it is only detectable within a cell line overexpressing the target protein. This requires the use of a transfection reagent, which itself imposes an atypical phenotype on the cell, as suggested by the results in Figure 18. Not only are there a large number of differentially expressed mRNA transcripts, which is induced by transfection alone (Jacobsen *et al.* 2009), but the transfection process itself is likely to result in lysosomal stress as the cell enters survival mode. This is partially backed up by the increased lysosomal swelling observed in all transfected cells.

The use of the large GFP tag may also affect the folding and trafficking of WDR98 as it increases the molecular weight from 75 kDa to 100 kDa. With hindsight, a smaller tag would have been preferable, as due to the difficulty getting NRK cells to express GFP-WDR98 it meant an additional GFP antibody was required, negating the reasons for selecting a GFP-tag initially (although it appears localisation was not affected, as Pontano Vaites *et al.* (2017) found the same localisation using the smaller FLAG-HA tag). Furthermore, there is increasing evidence that GFP itself is harmful to cells, increasing the level of cellular stress in the experimental conditions (Ansari *et al.* 2016). Indeed, our NRK cells when transfected with GFP constructs were only viable for roughly 48 - 72 hours maximum before there was considerable cell death.

5.4.3 Future Directions

Once a suitable antibody has been produced, it would be very worthwhile to repeat all experiments in this chapter in true wild-type conditions, without the need for transfection, overexpression of target proteins, or the introduction of cytotoxic GFP-tagged constructs.

The distribution of endogenous WDR98 in altered metabolic conditions (e.g. starvation to increase autophagic flux and endocytosis inhibition to decrease trafficking to the lysosome) could be analysed to further elucidate its function and potential regulatory role.

Immunofluorescence would also be a great tool for the analysis of a WDR98 knockout/knockdown cell line to probe the effects on the endosomal-lysosomal system, and to determine whether it is possible to rescue some or all wild-type functionality with either the WD40 or MIC domain alone.

Chapter 6: Conclusions and Future Directions

6.1 Characterisation of WDR98

Progress has been made in the characterisation of WDR98, in terms of compiling available data from journal articles and databases, and of conducting new research into its protein binding partners and subcellular location. WDR98 has been shown, via immunoprecipitation and immunofluorescence, to interact, and colocalise, with MON1B and CCZ1. These interactions also occur with the WD40 domain alone. WDR98 and both of its domains in isolation localise to the lysosome, in particular those surrounding the nuclear periphery, which have been reported to be the most active (Johnson *et al.* 2016). We cannot find any evidence of a direct interaction *in vitro* between WDR98 and either RAB5 or RAB7 in any nucleotide-bound conformation.

6.2 Corroboration with Publications

Pontano Vaites *et al.* (2017) were the first to publish a paper with a significant focus on WDR98 (termed RMC1 by the Harvard lab) and their data reflected many, but not all, of my findings. They too identified the ability of WDR98 to interact with CCZ1, MON1A and MON1B as well as its localisation to the lysosome. However, Pontano Vaites *et al.* (2017) predicted the helical structure of the MIC domain to be responsible for binding other proteins, based on structural data alone. In contrast, we have consistently shown that the WD40 domain is sufficient to bind to CCZ1 and MON1B, whereas the MIC domain alone does not exhibit any CCZ1 or MON1B binding. The WD40 domain β -propeller structure in other proteins often acts as a protein-protein or protein-DNA binding platform (Xu and Min 2011), which would appear to also be the case in WDR98.

Drosophila melanogaster WDR98 (CG8270) has been shown to bind to RAB5 (in Q to L mutant that stabilises GTP bound Rabs) (Gillingham *et al.* 2014) and Pontano Vaites *et al.* (2017) found WDR98 binding RAB7 (in a T to N mutant that stabilises GDP bound Rabs). Our data suggests there may not be a direct interaction between WDR98 and RAB5 or RAB7. We argue this is plausible because published data from Gillingham *et al.* (2014) and Pontano Vaites *et al.* (2017) cannot rule out

indirect interactions. Pontano Vaites in particular, based their claim on the mere presence of both WDR98 and RAB7 in the same size exclusion chromatography fractions, which could be coincidence and would require much more explicit demonstration to truly demonstrate an interaction. We conclude that the presence of additional proteins, such as MON1B/CCZ1, may be required to enable an interaction between WDR98 and RABs.

Following unsuccessful attempts to produce a WDR98 knockout cell line using CRISPR/Cas9, we considered whether a WDR98 mutation could be homozygous-lethal. Pontano Vaites *et al.* (2017) also suggested that it was not possible after several failed attempts. Tsui *et al.* (2019) used transient CRISPRi HeLa cells to observe that in the absence of WDR98, early endosomes and lysosomes were enlarged, EEA1 and LAMP1 staining was increased and EGF and EGFR trafficking to the lysosome was impaired. This corroborates findings from Pontano Vaites *et al.*, whose knockdown exhibited swollen endosomes and lysosomes and accumulation of p62, indicating disrupted autophagy; and also supports our theory that WDR98 knockout cell lines are not viable, as there have been no publications of their creation or use. More importantly, their data adds weight to the theory that WDR98 plays a role in endosomal and lysosomal maturation and trafficking.

6.3 WDR98 in Context

6.3.1 WDR98 in Health

As more research groups move to studying WDR98, its full characterisation in terms of its properties and biological function cannot be far away. There is growing interest in this previously uncharacterised protein and given the significance of the proteins it interacts with and the crucial systems it is involved with, it is exciting to think that we could soon have an established, proven function for WDR98. What we know so far is that this function will most likely centre around the regulation of endosomal maturation and lysosomal trafficking. The mechanism by which this is achieved remains unclear.

6.3.2 WDR98 in Disease

Various groups have found WDR98 to be mutated in different cancers, including colorectal (Jones et al. 2016), thyroid, pancreatic (Kimmelman et al. 2008), ovarian, liver, bladder, endometrial, lung, ovarian, prostate, oesophageal, kidney and melanoma (Delgado et al. 2014). Therefore, there is cause to study this protein in disease as well as in health. As discussed in the introduction, Kimmelman et al. (2008) managed to knockdown WDR98 in a pancreatic cancer cell line and intriguingly observed reduced ability to form colonies *in vitro*. Once further understanding of the role of WDR98 has been established, this will more appropriately inform disease-related research. Potential starting points include more knockdowns of WDR98 in different cancer cell lines to see whether they affect cell growth and overexpression in cancer cells, with experiments focusing on growth rates, viability and apoptotic assays.

Recently, a significant body of literature has focused on the roles of endocytosis, autophagy and the lysosome in cancer. Developments in this area included evidence showing that Ras-mutant tumours are able to upregulate RAB5-dependent endocytosis (Balaji *et al.* 2012) and overexpression of LAMP1/2 in metastatic colorectal tissue (Furuta *et al.* 2001). These mutations can benefit cancer cells by facilitating the destruction of adherens and tight junctions by cadherin degradation, which in turn promotes epithelial-mesenchymal transition and eventually metastasis (Mellman and Yarden 2013) and increased activation of vacuolar ATPase promoting an acidic tumour microenvironment and destruction of the extracellular matrix, also enabling tumour progression (Hernandez *et al.* 2012). The lysosome is an emerging cancer therapeutic target and managing the interactions between the lysosome and cancer cells has been identified as a possible strategy to overcoming cancer drug resistance. There are numerous lysosome-targeting cancer drugs at the preclinical stage. Drugs targeting HSP70, a protein which protects the integrity of the lysosomal membrane, block autophagy and lead to apoptosis (Piao and Amaravadi 2016). Chloroquine and hydroxychloroquine also inhibit autophagic flux and are drugs already approved for the treatment of malaria and these drugs could be used in combination with other anti-cancer therapies to circumvent the pro-survival effects of autophagy that can pertain to cancer drug resistance (Verbaanderd *et al.* 2017). Considering the likely integral-lysosomal status of WDR98 and its reported

overexpression in certain cancers, this is a protein worth studying and characterising, both in health and disease.

6.4 Future Directions

It would be interesting to investigate further whether there is a WDR98 ortholog present in yeast. This could be achieved by either a bioinformatics or wet lab approach, to identify proteins with homology and to see if their manipulation matches the phenotype of parallel WDR98 manipulation. Given the level of conservation and the seemingly important role played by WDR98, it would be unlikely for yeast to be missing a protein performing that role if there is no ortholog.

The development of an antibody would be invaluable to assess the subcellular localisation of endogenous WDR98 in a variety of cells and tissues across different organisms if the epitope allowed. It could also be used to study the response of WDR98 to environments affecting lysosomal trafficking, such as starvation and endocytosis inhibition.

It would be worthwhile to further develop and analyse WDR98 knockout or knockdown models to establish its exact function in the cell. Not only would this be valuable to our understanding at a cellular level, but it would be interesting to see if a whole organism would be viable in the absence of WDR98. The use of rescue constructs containing full length WDR98, as well as domains WD40 and MIC individually, could confirm function and help to determine the role each domain plays.

Visualisation of a crystal structure for WDR98 in complex with MON1B and CCZ1 would inform about protein-protein interactions and how the proteins complex together, and whether altered conformations are necessary for different functionality.

Finally, based on the collective results of Delgado *et al.* (2014), Jones *et al.* (2016) and Kimmelman *et al.* (2008), it is worth pursuing further understanding of WDR98 in a disease context. Given the emerging understanding of the role of the lysosome

in cancer, it is also worthwhile studying the organelle as a whole and scrutinising the viability of lysosomal proteins as cancer drug targets.

Appendices

A. DNA Construct Coding Sequences

(All sequences are from rat unless otherwise stated)

i. GST-WDR98

```
ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTT
TTGGAATATCTTGAAGAAAAATATGAAGAGCATTGTGTATGAGCGCGATGAAGGTGATAAA
TGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCTTATTATATTGAT
GGTGTATGTTAAATTAACACAGTCTATGGCCATCATACTGTTATATAGCTGACAAGCACAAC
ATGTTGGGTGGTGTGCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTG
GATATTAGATACGGTGTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTT
GATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTTGAAGATCGTTTATGTCATAAA
ACATATTTAAATGGTGTATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGAT
GTTGTTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTTGTTTTAAA
AAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAATCCAGCAAGTATATAGCA
TGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGAT
CTGGTTCGGGATCCCGGGGCCCCGTACCGCGGGCCGTAGCATGCAGTCGACTAGTATGGG
CAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTCGAGAAGGCGAACCC
GGTTAACTGCGTGTTCCTTCGACGAGGCCAACAAGCAGGTTTTTGTCTGTTCCGGTCAGGTGG
AGCTACAGGAGTGGTTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCTTTTAGAATGGA
CGACAGAGGAGAAGTGAAGTGCATTAAGTTTTCTCTAGAAAATAAGATACTGGCTGTCCA
GAGGACTTCGAAGGCCGTGGACTTTTTGTAATTTTTATTCCAGATAACTCTCAACTGGAGTA
CACACAAGAGTGCAAGACCAAGAATGCCAACATACTAGGATTCTGCTGGACCAGTTCTAC
CGAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTTACCAGGTGTTGCCTGAGAAGCG
AAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAACTGGCACATGTACTGTCCCGA
GAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTTGAGAATGTCCTGCAGCCTTTTTACTT
TAAGGCTGGCACCATGTGCAAGCTTCCCAAGTTTGAGATTGAGTTACCAGCTGCACCGAA
GTCAACTAAACTTAGCCTCTCAGAACGAGACATTGCGATGGCGACCATATACGGGCAGCT
CTACGTCCTCTTCTGCGGCATCATTCCCGGACCTCCAATAGCACAGGAGCGGAGGTGGT
GCTGTACCATCTCCCACGAGAGGGTGCCTGTAAGAAGATGCACATACTGAAGTTGAATAG
GACGGGAAAGTTTGCTCTAAATGTAGTGGACAACCTGGTAGTTGTCCACCATCAGGACAC
AGAGACGTCAGTGATATTTGACATCAGGTTACGGGGAGAGTTTCGACGGCACAGTCACCTT
CCATCATCCTGTGCTTCCAGCGAGATCAATTCAGCCTTACCAGATCCCCCTTGGCAGGTCC
TGCTGCTGTAACCAACCAGTCTCCCGTCCCCTGTAAACTCTATTCCCTCCTCGTGGATCGT
CTTCCAACCAGACATCATCATCAGTGCAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAA
ACTTCAGCCCATAGTGAATCTCTTGCCAGATAAAGGAAGACTGATGGACTTTTCTCTCCA
GAGGAAGGAGTGTAAGCCGTCGTTCTGTCTGTCTGTTTACAGATGCTGAGTGAGTCAGA
CAGAGCCACGCTGCCTGTGATAGCCACTGTTTTCGATAAACTCAACCATGAGTACAAGAA
GTACTTGGATGCTGAACAGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGCCGGAGCAA
CCCACCTCTCAAAGGCCAGTGAGGACCAAGCCGTGGTAGACCAGTCTGATGTGTACAC
ACACGTGCTGTCACCATTTGTGGAAAACAAGGAGACGCCACCCAAGTTTGTGATAGCTGT
GCTGATGGAATACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTGCAGCATTACTTACA
CGAACTTGTGTCATCAAGACACTTGTCCAGCACAACCTATTTTACATGCTCCACCAGTTTCT
```

GCAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGCTTGCTTGTGCTGTCCCTAGAAAAG
CTTTTATCCTCCCGCCCACCAGCTGTCTCTGGACATGTTGAAGAGACTTTCAACAGCAA
TGATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGGTTTCAT
CCGGGGTATTGGCGGCCATGACAACATTTCCGCACGAAAATTCTTAGATGCTGCTAAGCA
GACTGATGATGTCATGCTCTTCTATAACCATATTCGGCTTCTTTGAACAGCGAAACCAGCG
TTTGCGAGGGAACCCTAACTTCACACCAGGAGAACACTGTGAGGAACACGTAGCCTTTTT
CAAGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTCTGA

ii. WDR98-HIS

ATGGGCAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTTCGAGAAGG
CGAACCCGGTTAACTGCGTGTTCTTCGACGAGGCCAACAAGCAGGTTTTTGCTGTTTCG
GTCAGGTGGAGCTACAGGAGTGGTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCT
TTTAGAATGGACGACAGAGGAGAAGTGAAGTGCATTAAGTTTTCTCTAGAAAATAAGA
TACTGGCTGTCCAGAGGACTTCGAAGGCCGTGGACTTTTGTAAATTTTATTCCAGATAA
CTCTCAACTGGAGTACACACAAGAGTGAAGACCAAGAATGCCAACATACTAGGATTC
TGCTGGACCAGTTCTACCGAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTACC
AGGTGTTGCCTGAGAAGCGAAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAA
CTGGCACATGTACTGTCCCAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTGAG
AATGTCTGCAGCCTTTTTACTTTAAGGCTGGCACCATGTCGAAGCTTCCCAAGTTTG
AGATTGAGTTACCAGCTGCACCGAAGTCAACTAAACTTAGCCTCTCAGAACGAGACAT
TGCGATGGCGACCATATACGGGCAGCTCTACGTCCTCTTCCTGCGGCATCATTTCCCGG
ACCTCCAATAGCACAGGAGCGGAGGTGGTGTGTACCATCTCCCACGAGAGGGTGCCT
GTAAGAAGATGCACATACTGAAGTTGAATAGGACGGGAAAGTTTGCTCTAAATGTAGT
GGACAACCTGGTAGTTGTCCACCATCAGGACACAGAGACGTCAGTGATATTTGACATC
AGGTTACGGGGAGAGTTTCGACGGCACAGTCACCTTCCATCATCCTGTGCTTCCAGCGA
GATCAATTCAGCCTTACCAGATCCCCTTGGCAGGTCTGCTGCTGTAACCAACCAGTC
TCCCGTCCCCTGTAAACTCTATTCCTCCTCGTGGATCGTCTTCCAACCAGACATCATC
ATCAGTGCAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAAACTTCAGCCCATAGTGA
ATCTCTTGCCAGATAAAGGAAGACTGATGGACTTTCTCCTCCAGAGGAAGGAGTGTA
AGCCGTGCTTCTGTCTGTCTGTTACAGATGCTGAGTGAGTCAGACAGAGCCACGCTG
CCTGTGATAGCCACTGTTTTCGATAAACTCAACCATGAGTACAAGAAGTACTTGGATG
CTGAACAGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGCCGGAGCAACCCACCTCT
CAAAGGCCAGTGAGGACCCAAGCCGTGGTAGACCAGTCTGATGTGTACACACACGTG
CTGTACCATTTGTGGAAAACAAGGAGACGCCACCCAAGTTTGTGATAGCTGTGCTGA
TGGAATACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTGCAGCATTACTTACACGA
ACTTGTCATCAAGACACTTGTCCAGCACAACTATTTTACATGCTCCACCAGTTTCTG
CAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGCTTGCTTGTGCTGTCCCTAGAAA
GCTTTTATCCTCCCGCCCACCAGCTGTCTCTGGACATGTTGAAGAGACTTTCAACAGC
AAATGATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGG
TTCATCCGGGGTATTGGCGGCCATGACAACATTTCCGCACGAAAATTCTTAGATGCTG
CTAAGCAGACTGATGATGTCATGCTCTTCTATAACCATATTCGGCTTCTTTGAACAGCG
AAACCAGCGTTTTCGAGGGAACCCTAACTTCACACCAGGAGAACACTGTGAGGAACAC

GTAGCCTTTTTCAAGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTCC
ATCATCACCATCACCAC

iii. GFP-WDR98

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCTGG
CCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
ACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG
CACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGAC
GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGGCGACGGCCCCG
TGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAA
CGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTTACTCAGATCTCGAGCTCAAGCTTCGAATTATGGGC
AGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTCGAGAAGGCGAACC
CGGTTAACTGCGTGTCTTTCGACGAGGCCAACAAGCAGGTTTTTGTCTGTTCCGGTCAGG
TGGAGCTACAGGAGTGGTTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCTTTTAGA
ATGGACGACAGAGGAGAAGTGAAGTGCATTAAGTTTTCTCTAGAAAATAAGATACTGG
CTGTCCAGAGGACTTCGAAGGCCGTGGACTTTTGTAATTTTATTCCAGATAACTCTCA
ACTGGAGTACACACAAGAGTGAAGACCAAGAATGCCAACATACTAGGATTCTGCTGG
ACCAGTTCACCGAAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTACCAGGTGT
TGCCTGAGAAGCGAAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAACTGGCA
CATGTACTGTCCCGAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTTGAGAATGTC
CTGCAGCCTTTTTACTTTAAGGCTGGCACCATGTGCAAGCTTCCCAAGTTTGAGATTG
AGTTACCAGCTGCACCGAAGTCAACTAACTTAGCCTCTCAGAACGAGACATTGCGAT
GGCGACCATATACGGGCAGCTCTACGTCTCTCCTGCGGCATCATTCCCGGACCTCC
AATAGCACAGGAGCGGAGGTGGTGTGTACCATCTCCCACGAGAGGGTGCCTGTAAGA
AGATGCACATACTGAAGTTGAATAGGACGGGAAAGTTTGTCTAAATGTAGTGGACAA
CCTGGTAGTTGTCCACCATCAGGACACAGAGACGTCAGTGATATTTGACATCAGGTTA
CGGGGAGAGTTTCGACGGCACAGTACCTTCCATCATCCTGTGCTTCCAGCGAGATCAA
TTCAGCCTTACCAGATCCCCTTGGCAGGTCTGCTGTGTAACCAACCAGTCTCCCCT
CCCCTGTAAACTCTATTCCTCCTCGTGGATCGTCTTCCAACCAGACATCATCATCAGT
GCAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAACTTCCAGCCATAGTGAATCTCT
TGCCAGATAAAGGAAGACTGATGGACTTTCTCCTCCAGAGGAAGGAGTGTAAAGCCGT
CGTTCTGTCTGTCTGTTCACAGATGCTGAGTGAGTCAGACAGAGCCACGCTGCCTGTG
ATAGCCACTGTTTTTCGATAAACTCAACCATGAGTACAAGAAGTACTTGGATGCTGAAC
AGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGCCGGAGCAACCCACCTCTCAAAG
GCCAGTGAGGACCCAAGCCGTGGTAGACCAGTCTGATGTGTACACACACGTGCTGTCA

CCATTTGTGGAAAACAAGGAGACGCCACCCAAGTTTGTGATAGCTGTGCTGATGGAAT
ACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTGCAGCATTACTTACACGAACTTGT
CATCAAGACACTTGTCCAGCACAACTATTTTACATGCTCCACCAGTTTCTGCAGTAC
CATGTGCTCAGCGACTCCAAGCCTCTGGCTTGTGTTGCTGTCCCTAGAAAAGCTTTT
ATCCTCCCGCCCACCAGCTGTCTCTGGACATGTTGAAGAGACTTTCAACAGCAAATGA
TGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGGTTTCATC
CGGGGTATTGGCGGCCATGACAACATTTCCGCACGAAAATTCTTAGATGCTGCTAAGC
AGACTGATGATGTCATGCTCTTCTATACCATATTCCGCTTCTTTGAACAGCGAAACCA
GCGTTTTCGAGGGAAACCCTAACTTCACACCAGGAGAACACTGTGAGGAACACGTAGCC
TTTTTCAAGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTCTGA

iv. GFP-WD40

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
ACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG
CACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGAC
GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCG
TGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAA
CGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTTACTCAGATCTCGAGCTCAAGCTTCGAATTATGGGC
AGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTCGAGAAGGCGAACC
CGGTTAACTGCGTGTCTTTCGACGAGGCCAACAAGCAGGTTTTTGTGTTCCGGTCAGG
TGGAGCTACAGGAGTGGTTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCTTTTAGA
ATGGACGACAGAGGAGAAGTGAAGTGCATTAAGTTTTTCTCTAGAAAATAAGATACTGG
CTGTCCAGAGGACTTCGAAGGCCGTGGACTTTTGTAATTTTATTCCAGATAACTCTCA
ACTGGAGTACACACAAGAGTGAAGACCAAGAATGCCAACATACTAGGATTCTGCTGG
ACCAGTTCTACCGAAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTACCAGGTGT
TGCCCTGAGAAGCGAAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAACTGGCA
CATGTACTGTCCCGAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTTGAGAATGTC
CTGCAGCCTTTTTACTTTAAGGCTGGCACCATGTGCAAGCTTCCCAAGTTTGAGATTG
AGTTACCAGCTGCACCGAAGTCAACTAACTTAGCCTCTCAGAACGAGACATTGCGAT
GGCGACCATATACGGGCAGCTCTACGTCTCTTCCCTGCGGCATCATTCCCGGACCTCC
AATAGCACAGGAGCGGAGGTGGTGTGTACCATCTCCCACGAGAGGGTGCCTGTAAGA
AGATGCACATACTGAAGTTGAATAGGACGGGAAAGTTTGTCTAAATGTAGTGGACAA
CCTGGTAGTTGTCCACCATCAGGACACAGAGACGTGAGTATTTGACATCAGGTTA
CGGGGAGAGTTTCGACGGCACAGTACCTTCCATCATCCTGTGCTTCCAGCGAGATCAA
TTCAGCCTTACCAGATCCCCTTGGCAGGTCTGCTGCTGTAACCAACCAGTCTCCCCT
CCCCTGTAAACTCTATTCTCCTCGTGGATCGTCTTCCAACCAGACATCATCATCAGT

GCAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAAACTTCAGCCATAGTGAATCTCT
TGCCAGATAAAGGAAGACTGATGGACTTTCTCCTCCAGAGGAAGTAA

v. GFP-MIC

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
ACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG
CACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGAC
GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCG
TGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAA
CGAGAAGCGCGATCACATGGTCCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTTACTCAGATCTCGAGCTCAAGCTTCGAATTATGCCC
AAGTTTGTGATAGCTGTGCTGATGGAATACATCCGTTCTCTGAACCAGTTTCAGATCC
CAGTGCAGCATTACTTACACGAACTTGTCAATCAAGACACTTGTCCAGCACAACCTATT
TTACATGCTCCACCAGTTTCTGCAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGCT
TGCTTGTGCTGTCCCTAGAAAGCTTTTATCCTCCCGCCACCAGCTGTCTCTGGACA
TGTTGAAGAGACTTTCAACAGCAAATGATGAGATCGTAGAAGTGCTTCTGTCTAAGCA
CCAGGTGCTGGCTGCCCTAAGGTTCATCCGGGGTATTGGCGGCCATGACAACATTTCC
GCACGAAAATTCTTAGATGCTGCTAAGCAGACTGATGATGTCATGCTCTTCTATACCA
TATTCCGCTTCTTTGAACAGCGAAACCAGCGTTTGGCAGGGAACCCCTAACTTACACC
AGGA

vi. ΔpMEP4-MYC-GFP-WDR98

CGGCCTTGCCGGCCTCGAGCGGCCGCTTACAGAGTATGGTAATTTGTTCTTTAGATTTG
CAGAAATTTATAGAGAAAAGAAAATTACAGCCTGCACACCCAAGACAAGGACGATTCTCCA
ATGAAGCTGGCTGCATCAAAGGTGGACTTCCGCTCAGGCTGTGAGGTTGGAGTCACAGTG
GTATTTGTTGCACCTGCTGATGTGACAACCTGAGGGAGTAGGTGTAGGAGAGGAAGGCCGA
GTTGTAGGCTTAGCTGTAGAATTGGTTGGCACTGGGGTGGTAGGTATTACAGAACAGGAG
TCAGTACTGGTCTTCTGGGTACAATTACTTAATTAATCACTTGAACAATAGGTCTTATTT
GCTTCATGGCAATCTAGCCAGACGCAGGTAATATATTATTAGTCAAGGTGGCGTTGACACAG
GAAACACAGCTGTTGAAGCTCTCACAGGTTTCTGGCGGCTTGGTGGTGGTGAGCGTCGTC
GGGACCACTTTGCTGGTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
CTCTGCGCGGCCGACAGGCAGAGCGCGGCGAGGCAGGTGGCGGCCAGAACAGTCCGCGG
GAGGCGCCGACATCGTGTCTCAGCCCGTCCGTCGACGCTAGCAGCTGGTACGAACAAA
AACTCATCTCAGAAGAGGATCTGGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCG

GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT
CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT
TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACG
TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACA
ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGCG
ACGGCCCCGTGCTGCTGCCGACAACCCTACCTGAGCACCAGTCCGCCCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCCGCCCGGGATCA
CTCTCGGCATGGACGAGCTGTACAAGTTACTCAGATCTCGAGCTCAAGCTTCGAATTATG
GGCAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTCGAGAAGGCGAAC
CCGGTTAACTGCGTGTCTTTCGACGAGGCCAACCAAGCAGGTTTTTGTGTTCCGGTCAGGT
GGAGCTACAGGAGTGGTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCTTTTAGAATG
GACGACAGAGGAGAAGTGAAGTGCATTAAGTTTTCTCTAGAAAATAAGATACTGGCTGTC
CAGAGGACTTCGAAGCCGTGGACTTTTGTAATTTTATTCCAGATAACTCTCAACTGGAG
TACACACAAGAGTGCAAGACCAAGAATGCCAACATACTAGGATTCTGCTGGACCAGTTCT
ACCGAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTACCAGGTGTTGCCTGAGAAG
CGAAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAACGGCACATGTACTGTCCC
GAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTTGAGAATGTCCTGCAGCCTTTTTAC
TTTAAGGCTGGCACCATGTGAAGCTTCCCAAGTTTGAGATTGAGTTACCAGCTGCACCG
AAGTCAACTAACTTAGCCTCTCAGAACGAGACATTGCGATGGCGACCATATACGGGCAG
CTCTACGTCTCTTCCCTGCGGCATCATTCCCGGACCTCCAATAGCACAGGAGCGGAGGTG
GTGCTGTACCATCTCCCACGAGAGGGTGCCTGTAAGAAGATGCACATACTGAAGTTGAAT
AGGACGGGAAAGTTTGTCTAAATGTAGTGGACAACCTGGTAGTTGTCCACCATCAGGAC
ACAGAGACGTCAGTGATATTTGACATCAGGTTACGGGGAGAGTTTCGACGGCACAGTACC
TTCCATCATCCTGTGCTTCCAGCGAGATCAATTCAGCCTTACCAGATCCCCTTGGCAGGT
CCTGCTGCTGTAACCAACCAGTCTCCCGTCCCCTGTAAACTCTATTCCCTCCTCGTGGATC
GTCTTCCAACCAGACATCATCATCAGTGCAAGCCAAGGTTACCTCTGGAACCTGCAGGTG
AAACTTCAGCCCATAGTGAATCTCTTGCCAGATAAAGGAAGACTGATGGACTTTCTCCTC
CAGAGGAAGGAGTGTAAGCCGTCGTTCTGTCTGTCTGTTACAGATGCTGAGTGAGTCA
GACAGAGCCACGCTGCCTGTGATAGCCACTGTTTTCGATAAACTCAACCATGAGTACAAG
AAGTACTTGGATGCTGAACAGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGCCGGAGC
AACCACCTCTCAAAGGCCAGTGAGGACCAAGCCGTGGTAGACCAGTCTGATGTGTAC
ACACACGTGCTGTCACCATTTGTGAAAACAAGGAGACGCCACCAAGTTTGTGATAGCT
GTGCTGATGGAATACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTGCAGCATTACTTA
CACGAACCTTGTCAAGACACTTGTCCAGCACAACTATTTTACATGCTCCACCAGTTT
CTGCAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGCTTGCTTGTGCTGTCCCTAGAA
AGCTTTTTATCCTCCCGCCACCAGCTGTCTCTGGACATGTTGAAGAGACTTTCAACAGCA
AATGATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGGTTT
ATCCGGGGTATTTGGCGGCCATGACAACATTTCCGCACGAAAATTCTTAGATGCTGCTAAG
CAGACTGATGATGTCATGCTCTTCTATACCATATTTCCGCTTCTTTGAACAGCGAAACCAG
CGTTTTCGAGGGAACCTAACTTACACCAGGAGAACACTGTGAGGAACACGTAGCCTTT
TTCAAGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTCTGA

vii. pcDNA5-GFP-humanWDR98

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATG
CCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCC
CGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
TTAGGGTTAGGCGTTTTTTCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAATGGGCG
GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATC
TCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTTAGTGAACCGTCAGATCGCCTGGA
GACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGA
CTCTAGCGTTTTAACTTATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCA
TCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCCAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGG
CCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGGCAGCGCCCCGTGC
TGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGG
ACGAGCTGTACAAGTTACTCAGATCTCGAGCTCAAGCTTCAATTATGGGCGAGGAGGAC
TACTATCTGGAGCTGTGCGAGCGGCCGGTGCAGTTCGAGAAGGCGAACCCCTGTCAACTGC
GTCTTCTTCGATGAGGCCAACAAGCAGTTTTTGTGTTTCGATCTGGTGGAGCTACTGGC
GTGGTAGTTAAAGGCCCAGATGATAGGAATCCCATCTCATTTAGAATGGATGACAAAGGA
GAAGTGAAGTGCATTAAGTTTTTCTTAGAAAATAAGATATTGGCTGTTTCAGAGGACCTCA
AAGACTGTGGATTTTTTGTAAATTTTATCCCTGATAATTTCCAGCTGGAATACACACAGGAG
TGCAAGACTAAGAATGCCAACATTCTAGGATTCTGCTGGACTAGTTCAACTGAAATTGTC
TTCATAACAGATCAAGGAATCGAATTTTACCAGGTATTACCAGAGAAACGGAGTCTGAAA
CTCTTGAAGAGCCACAATCTCAATGTGAATTGGTACATGTACTGCCCCGAGAGCGCCGTG
ATCTTGCTGTCTACCACGGTCCCTGGAGAATGTCTGCAGCCTTTTCACTTTAGGGCTGGC
ACTATGTCGAAGCTGCCCAAATTTGAGATTGAATTACCAGCTGCGCCTAAGTCAACTAAA
CCCAGCCTTTCCGAAAGAGACATCGCAATGGCTACCATATACGGGCAGCTGTATGTTCTC
TTCTTGAGGCATCATTCTCGGACCTCCAACAGCACAGGAGCGGAGGTGGTCCCTCATCAT
CTACCACGAGAAGGTGCCTGTAAAAAGATGCACATATTGAAGTTAAATAGGACGGGAAAG
TTTGCCCTGAACGTGGTGGACAACCTGGTAGTCGTGCATCATCAGGATACAGAGACATCG
GTAATATTCGATATCAAGTTACGGGGAGAGTTTGACGGCTCCGTTACCTTCCACCACCC
GTGCTTCCCGCTCGATCGATCCAGCCCTATCAGATCCCCATCACAGGTCTGCTGCCGTG
ACCAGCCAGTCTCCTGTTCCATGTAAACTCTATTCTTCATCTTGGATTGTCTTTC AACCT

GACATCATTATCAGCGCAAGCCAAGGTTACCTCTGGAACCTCCAAGTGAACTTGAGCCC
ATAGTAAATCTCTTACCAGACAAAGGAAGACTCATGGACTTTCTCCTCCAGAGAAAGGAA
TGCAAGATGGTCATCCTGTCTGTCTGTTACAGATGTTAAGTGAGTCAGACAGAGCATCG
CTGCCCGTGATAGCCACTGTTTTTGTATAAACTCAACCATGAGTATAAAAAGTACCTGGAT
GCCGAGCAGAGTTATGCGATGGCGGTGGAAGCAGGGCAGAGCCGAAGCAGCCCGCTCCTC
AAGAGGCCGGTGCGGACCCAGGCGGTGCTGGACCAGTCAGATGTGTACACCCATGTCCTG
TCAGCCTTTGTGAAAAGAAGGAGATGCCTCATAAATTTGTGATAGCCGTGCTGATGGAA
TACATTCGTTCTCTTAACCAGTTTCAGATTGCAGTACAGCATTACCTACATGAACTTGTT
ATCAAACCCCTGTCCAGCACAACCTCTTTTATATGCTGCATCAGTTCCTGCAGTACCAC
GTCTCAGCGACTCCAAACCTTTGGCTTGTCTGCTGTTATCCCTAGAGAGTTTCTATCCT
CCTGCTCATCAGCTATCTCTGGACATGCTGAAGCGACTTTCAACAGCAAATGATGAAATA
GTAGAAGTTCTCCTTTCCAAACACCAAGTGTAGCTGCCTTAAGGTTTATCCGGGGCATT
GGTGGCCATGACAACATTTCTGCACGAAAATTTTATAGATGCTGCAAAGCAGACTGAAGAC
AACATGCTTTTCTATAACAATATTCCGCTTTTTTGAACAGCGAAACCAGCGTTTGCAGGG
AGCCCCAATTTACACCAGGGGAACACTGTGAAGAACATGTTGCTTTTTTCAAACAGATT
TTTGGAGACCAAGCTCTAATGAGGCCTACAACATTTCTGAAATCACTTGCTGTTTTTTTAT
ATAAAAATGTGTACAAAGTTAATTTATTGCATTAATAAAGCTCTTTAACTATAAAATGT
TATAAAGTGTATCTACAACCTC

viii. WDR98 gDNA

(Exons in bold)

GAAATTTAATGCCCTAATCTCCCTTTATTTCCCTCTTAAATAAAGGCGCCAGTCCCATTGG
CTTAGGGACCTACCTGATGCCTTCAGTTTTCCCTTAATTTCCCTTTGAAGCTTCAAGCACAT
TCTGAGATATTGGGGATTAAACTTCAACACGTAAATTCTAGGGGAGGGGCGAAATCTAG
TCCCTAACATCCTAGCAGATCACAGTACTTGCCAGTATGGCTAAGGCAGACTGCTCAAAA
AATCAGTTTTCTGGGGCTAAACACGGAATT**CCGTTGAGTAATGTAGTTGCCTGG**CGTGAGGG
AGGTCCTGAGGTGAACACAGAAATCAATCCTGGCTCGTAGATCTTTGCTGCACGCTTACA
GTTAGCCTCCTGCCTCATCTCTCGACTAAACAGACAAGAACAGAACCCTGGGCGTAGGCG
ACGCAGAATGTCACCACCCGTAAGCAAATTAGAAACCAGGCTGTCTCCAGCAGCCCCG
AGGCTTCCGGGTCACGTGACCCGACCTCCTCCCTCCTACAAGGGGGCGTGTCTTCGGTGT
TGGGACCCTCCCCGCTCACTGAGGCGCCGCTGGCCCGGTACGTGGAGGGCGGGGCGGG
GCCTAGGGCTGCGGCCAGCGGCACCCGGTTCTTGGACTGCATCTTCTACGGACGAGCTCC
GCCGTTCCAGCCATGGGCAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGT
TCGAGAAGGCGAACC CGGTTAACTGCGTGTCTTTCGACGAGGCCAACAAGCAGGTGAGCGG
CGGCCACGGCGGGGGCCGGGCGCCAGACGCCGGAGTGTCTCTGCACCGGCGCCGAGAGT
GCGGGGTGTATTGTACCAGAGCTGAGACGCGAGTGAAAGTGCATGGGCGGTGTCAGTGTG
CGAGCTGGCTTTTACAGACGGTCAGTCCAGAGTCTGGCCTGCGGTCTTAAAATTCGTCTAA
GCATGTCCTGTTTGTTCGGATGAATCGAGGATGTGGTACGGTAGACATGATTCTTGAAAG
GCTGAGTCTAGGTCCTTATAAAGTCTGACCATTTCATGTTGCCCTTTTTTCCCTCAGGTTTTT
GCTGTTTCGGTCAGGTGGAGCTACAGGAGTGGTCGTTAAAGGTCCAGATGATAGGAATCCCAT
CTCTTTTAGGTAATTTGCACTTTCAGAATAATACCCACATGTAAACACCCGGCGGCTTTCT
AACTCAGAAAGTTGAATCATTTTGTCTCTAAG**CGGCCTGTCGTGTCAGTAAG**AAAAAGCAG

TGGACATTTGCTAGTACAGTTTGGTTACAGCACAACCTGTCAGAGTGAAAATAGGGATTGCT
AGTTCAAGTGGGAACCAAAATCTCCTAGAGCACCCACTTCCTGGTACTTGAAGTTATCGA
GTCCGTGTGTTTTGCACTCATTTCTGCAGTGCAGTGTGGACCTGGGGACTCATTGAC
GTTTATCCCAGCACTTCTAGAGTTTTTGGTGGGTGGAAAATCCAGTGTCTGGTGTA
ATGGAGAACAATCAGGGCTTCCACTGAGGTTGAAGTGACAGACCGGCTTCCAAAGTATG
TAGGTTGACAATGTGATTGACATAGCAGTTCTTTTCATTTTTTACAAGTCCCTCAGCACCT
TGGCAACAGGCCTATAATCCCAGCCCTCTGGGAAGCTAAGTAGCAAATAGCTTGACAGTT
AGTGGCAGTGTAGCAAACCCCGTTTTCAACAACAACGACGACAACAAAAATCTCAGAAA
AATAAATTGCATTTAATATGTGTCAGCTGTTTGAAGTAAAAAAAAAAAAACAGCCAACCT
TGTTTTCCCTTTCTTTTCCTTTGTTACTTCTTCCTTTATTTTTTCATTTTTTTTTGAGACA
GGGTTTCTCAATGTAGCCCTGGCTGTCTGTAACTTAATTCTGTAGACAGGCTGACCTTG
AACTCAGAGATCCTTCTGCTCCTGCCTCCTGAGTGTGGAATCAAGGTGTATTTGACAG
CTTGTCTTGCTTTCTTTAATGGCCCTTGCTCAGCACATCCTGAAAGCAGCTTAACTGG
AGAGAAGAGATTTAGCAAATACTGTTTGCCATGCTAAGCTCTATGCTAAAGCTCTGTTCA
TCGTGTGGGAACACTTAGGGGTGCAGCTTTGACATCAGTAATGGGGTTGAGTAAGGCCAC
AAGGTGGGAAGGGAGTGATTTCTGGTTGGAGTCGAGACCTAGATAACACCAGGAGCAAAG
GTGTGATGTTCTGGGGAGTCTCCTCCTGTGTTGGGACACTGTCTCACTGGCACATTAC
CATGGCAAGGGCTCCTTAAGGGCTTGCTTGCCCTCTGTCTCCACAGCCCTGGAATTATAAG
GGATGGCACAGTGTGAGATTTTTTTTTTTTAAAACAGCTTTATTGAGTTAAAATTGGCTC
ATAAATTGCACATTTACTGTGCATTGATATATCTATACACCCAGGAACTGTCACTACAA
CTAAGGTAATGAATGCAGCTTTATCCACAAGATCTCCGGTAGTCTTCAGGTGACGTGTG
TGACCCCAAGTCTCCAACCACTTGCTGTGTGTCATCTGTCCACCTCACTAGCAAGTCGTTG
CCTGTAGATTCTGGTACAGCTTCAGAAATGCAATTCCTAACATTTGGCTCAGGGCTGCT
GCTTCCCAGAGTTCTGGCTTTTTTGTGTCAGTGGGAGTAAGGACTAGAGTTGAGAGCGGCTTC
TCCCCTAACCGTGCTTTCTTCTTCAG**AATGGACGACAGAGGAGAAGTGAAGTGCATTAAGT**
TTTCTCTAGAAAATAAGATACTGGCTGTCCAGAGGACTTCGAAGGCCGTGGTAAGACATAG
CTTAGAACAGCATCGGAAAGAGAGTGGGTGCTTGGAAGAGACAGAGTCTGAGTTCTGGCT
GTTTGTCTCGTCCCTCAGAGTCAGTGAAGTGGGGGCTCCTGAGAGCGTTTGCAGATAAGAGG
ACGAACGAGCTTACTTATCATTTGTCTGTCCCTAAAGGCAGTGTGTATTCATGATGCGAA
ATGCTGAGCCTCATCATGCCATTGTCATACGTGTTTATCGTGTCTGTGATAATGTTCTG
CCCCATTGCCTTCTCCTGGCCCCCTCCTGGACCCCTTCTGGCCTCCCCCTCCTCTAAG
TGGCCCTTTTTCTACTCTGATCATTTGTATGTGTGTGGATGTGTGTACCATGTGTGTGCAG
TACCTGAGGAGGCCAGAAGAGGGCACAGGATCCCCTGCAACAGGAACTACAGGCGGTTGG
GTGCTGACAATCGAAGCCGGCTGTTGAGCCAGCTCTCTATCCTCTATTTTTGTTTTAAGTAA
TTTGGATTATCATGCATCTTTATATTGCTTGTTAACTGGCAACATGGATTATGGGAGGTA
TATTTTGATTACTTGGTAAATGAATCGTATTTCAAGTCACTTCAGCATCTGGTTCTTTCT
CCACCAGATATGCTGGAGAATATTGGATTGGGATATCCAATGCTGGATTGGGTAGGAACA
GGCAGATGAACACAGTTGGGGAGTTTGTGGTGCTTCATCTCTTGGACATCCTCACTATC
GCTTGCCCTGGGTTGGGATGCAGTGTGATTAGCCTAGCATGAGGAAATCACAGGTTTGCT
CCTAGTGTCTCATGTGTCAGTCAATTATTCATCCCGTTCCTGCTGTCATCAGTAAAGTAGGA
AGGAGGTGGTGGGAAGACAGGCAGACAGCTGTGTTTGTGCTGCTTTATTTCTCTGGATT
TCCAGGAAGCCTGGGGGAAGTGATGTCACCCATTTGAGGAATATATTTCTCCTTTCCCTCC
CTCCTCCTTTCTTTCTGCTCCTTCTCTCTCCTTTCTCTCTATTTCCCGCTCCTTTTATGG
TCAAGGTTGACCTGGAATTTACTACATAGAACAAGTTGGTGTGAACTTAAAGCAATCCT
CCAGTCTCAGCTTCTTGGGTGCTGAGATCCTAGGTATGAGCTGCCCACTTGGCTTGGGA
ATGAGTTTTAGAAGTCTGTTCATGAGTTCCAACCTGAGTGTTTAGGTGGACTCCTTTTAGT

TGACTGGTAAATCTGGTTTCATTTGATCCTAATTGTCTTTATATCTTTTCTCTGTGGGTG
TTTGTCCAGCTGTTTTTTAATCACTTTAAAAATCTGCTTGTGAGGCATAAGGGGACACAC
TCTTAATACCCCCTAGCACTCAGAAGGCAGAGGTAGGCAGCTCTCTCTGCCAGCACTTCC
TACTGAGACCCTGCCTCAGAACAAAAACAAAAAAGTTGTCCATTTTATTTTCGTAGTAGGC
TATTTAATCTGATAATTTCCCTCTGTGTATTTTTTCTGTCTTTTCAG**GACTTTTGTAAATTT**
ATTCCAGATAACTCTCAACTGGAGTACACACAAGAGTGCAAGGTACACGGCTGGTGTGGGG
GAGGGGGCTTCATTCCCTCACTGCAGTGTCCACTGTTGTTTGTGTGTGGAGAACAGGTTGG
GTAGTAGACACACATTGGAATCTCTGTCTGGGGAGGTGTCTGTGTTGACGGTGAGCTGTA
GAGGTTGGTTGGGGGTTTACAGCTATGCCAGACAGTTTACCCAGTTTTGTCTCTTAGCTAT
AAAGTGGAGGACTGATGGTAAAGGGTCACCGTCAAGGCTCCCTGGGGCTTCCTCACCGTG
CTTGCAGGCCTCTGTGTCTCAGGCTTGGATCCCACCTGCCCATCTTGTAAGGTTCTCTG
AAGTGTGTTTCCCTCAATGAGTCTCCCTCCTGTATCCCTGCACCCTATGTAGGTATAGCC
CTAGTCTCTCTTAATTAGAATAGCTGGGTGTAGAAAATGAAGCTGGTGGGAGGGAGCTAT
GCCATGGTCAGCCAGGGGCAGCTGGGAGAGTAACCACAGTTGTGAGTCGCAGGAGTAAC
TTTGGTGTACCTCATTGTTCTTGGTAAGTGGCACTCAAGATGGGGTCCCTTATCTGTGAGA
GGAGACTCAGGACTTAACCAAACCTCCTTCCAGCCCTACAGGTGCTCTAAAGGAACCGCTC
CTTGTCTGAGCCCTGCTTTGCCCTTCTGTCTTTGAGCCACCATTCTGTGGTCTCAGGC
CCTGGAACCCATCTTGGTTTGTGTTAACTGAAGAACTTTAGCCCTTGATATTATTGTAATG
ACTGTGAGAATGTTGCAGTTTGGAGGGCAGAAGTTGGAATTTCTGGCCTCTGGGTATTAT
CTTTCTATAAAAATTTAGGCTTCTTTCTCTGTTTCAGCAGGGTACAAGGATTCAGAATCTTT
GGTAGAAATGTTAGGACAAAATCGGGAGAGCTGTGTTTAAAGAGCGTTTGAATCTGGCAT
GGAAATGCTCCCCTGTAGCCAGCATCTGGGCGGGTGCATCAGGAGAGTCCTGAGTCCAG
GCTGCCTGGGCTGTAGGGCTCCAGGCAGCCTTTGTGAGAGTGTGGGATGGAGCCTAATGA
GTGAGTGACCGTAACAGCCGAGCATTGAGGTTTTGAAGGGAGCTGCACAGCCTTCAAACC
AGAGCATTACACTTGTAGAGCTACAGTGTGTCTTGATTTGTGTGTGGTGTAGGTACG
CCTGTCTCACAGGTCAGTGGAGGATGTGGAGTGTCTTGCAGTGTGACATTTTATCCCT
TGAGGCAGGTCTCTTACTGAACCTGGAGATAGGCTAGTGGCCAGCAAGCCCAAGTCCCTC
CTCTCTGCTCCTCATGTCCCACACCAGCACTGTGATTATGGCCACTTGTGGCCATTTATG
TGGTACTGGGGATTTGAACTCAAGTCTCCTGCTGTGAGCAGGGGCTCTTACCCACAGG
GGCATCTCTCTAGCCCTTGCTTTTTTGTAAATTTTCACTTATTTATTTTTTTGGAGATAG
TGTTTCATGTCCAGACTGGCCTGCATCTTGAAATGTACCCATGGATGGCCTTGAATTCCC
AAATGCCCTGCCCTAGCCTCCTGGGTGCTGGGATTATAGACATCACAATCACAGCCAGCT
AAAAACCACCTTTTTTTTGTAAATCTTCCAG**ACCAAGAATGCCAACATACTAGGATTCTGTCT**
GGACCAGTTCTACCGAAATTGTCTTTCATAACGGATCAAGGCATCGAATTTTACCAGGTATTA
TATCAGTTATAGATTTGCTAGAGAAAGCAAAGGCAGTTCAGCAGCCTGTGTTGTATCAAA
AAGTCTGAAACCAACCTCTTAGTTGAGGAAACATTATACAAAAGTAAAAGTTCCATAA
ATTGACTGTCAGCCCCAGTTTGTACATAGTTCAGTTTTCTCTGGACACGTGCATTTATC
ATTGTTATTTTTATTTTTATTTTTTTTGCAGTGGGGCTCCTGCACATAAACTTGTGTGGT
GTAGTTTTTTCTTGCATAGCAGTATGCTTTGGGGTGTCCCAGATTTCTGCTCTGAGAA
TAGTCTGCTCTGTGGACAATCTGCTGTTTGGCATGTCACTTCCCTGTTACAAGATGTCTG
GCTTGATTTCACTCTTTATGGTTGGTACTCCAGAAGTTCACGAGCCGTCTTACATGTTT
CTTCATGTTTCATGGGCAAGCAGTTCTTGTAAATTGCATAGATACAGACAGACACCATTACA
TTTTTCACAGATGTCCATGCTTTGAGAAATAGTGTGGGACTCTCTTGTCCACATTGCTGC
TGGGCCTGAATCTCATTGATCTGTAGACAGTTGCCTTCTCAGACGGGGAAAGATTCTCAC
TGTCTCTTAGGACAGTGGGTACAATCCCTATCACAAGGGTAACCTAAGACCATCAGAA
AACACAAATATTTGTAACCATGGGCTTTTTTTTTTTTTTACCTGCACCTCAGTTCCTAAATA

ATGACTCTGAGACTGAATTATTTATGAGTAAATGCCTAGGTCTTAGACTTGATCTTATTC
CCTGACCAGCCTTTAACTTAATTAACCCATTTATTCTATTCTGTGAGTGCCACAAGGCTG
GTTATCTCTCCTCAGATTCATGTCTGTCTCCTCAGAGTTCAGGGCAAATCTCTTCTGAG
CCTGACTTTATCCCAGAGTCTCCTTTTCTGACGGGAACTCCCATATCCTATTTTCTGACT
CAGCATCATGGAGGAGCCTTGACAGCAGTGTAAAATCAAATAATGAGCATGAAGAGCACT
GTAAGTTTTAAATCCATACACGGAGAAGGGATAGAGGACCCCGGAGGATAGAGGCCTCCCT
ACAGTCATTTTAAAAATATCTGTGCTGGTGTCTAAGAGCACTTTCAGGATCATAGTGGGT
CACAGACATCCGAACCTTCAGTGCCAAGGTATTCAAACCTCTTCTGACTTTCACAGGTGT
AAGACCTGCGCATAGTTCATATCATATTTTCCAACACACACACACAAAACATTTATAT
ACATAAAACAAAATAACACTGGAAAAACTCGGTTCTCTCTAGTGAGTAAGTATGTCAG
ATCCCCAAGACATGTTAGGGCTGGTGAGATGGGTCAGCACTTGTACCCAGTCCTTGTGA
TCTGAGTTCCTTCTGTTCATGAGAGAACTGTCACCCCGAAGACGTCTTCTGATTTCCATGT
GCACATTATGACATACTCGTGTCTCCATATACATAATAACAACAAGAAATAGATTA
AGAAAGAGAACTCCAGCATGCTTCTTTATGGCTCTCAGTGTGTGTGTCTGTGCCACAGTG
CACATGTGGAGGTCAGAGGAGAACTCTGTGGAGTCCCTTATTTCCCTCCGTCTTTTTGTG
GGTCTGGGATCAGGTTAGGGCAGGCAGTCTTTACTTGCTGAGCCATCTTTCTGGCCGTG
TACGTTCTCATTTTGTGAGGCTTGGGTGAGGATGTTTCATGCTTGGCCTTGATTTCCCTCG
TGGGGTGAGTGTGGTATCTTGAGTCTGCAGTGTCTTAGTGTGCCAGAATATGGAGAAT
AGATGTGCCTAGATTCTGAAGGGCTGCACTTGCAAGCGGTCTTACGAATAGTGGGGTTAG
GAAGAAGAAGCCTGGCCTGATCAGCCAGGTGGAGTGTAGAGAGAGCTGGTGGCTCTCATG
TTGTGCTGTCTCTCAGAGTGTAGGAAAGAGTAAATTAATGAGGGGTGAAAGCGCCAGGC
ACACAGCGCCATAATAAATACAAGTAGCAGTTAGCTGTACGACCCTGGGCTGCATTGTTT
GGGTTTTGTGGTGGTGGTGGTGAAGTGGCTGGGCCTCCTGTTTCTTATGCAAGCTC
TTGTCCACTGGGTTTTATCCCGACCTTGGTGGGCTTTCTGTGCTCTTTATTTACCTCACTT
GTCGTAGCAGTCTTTTAAACGTGAAGATCCCTCTTCTTGACCTTCGTCCAG**GTGTTGCCT**
GAGAAGCGAAGCCTCAAACCTGAAGAGCCACAACATCAATGTGAAGTGGCACATGTACTG
TCCCGAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCTTGAGAATGTCTGCAGCCTTTTT
ACTTTAAGGTGAGAGTGAGCCTCAGCCCCTTAAAAATAACCTCTCTACCCAGCTCTAAAGC
TGGTCCGCTCCTATGAATGCCGCCTCTCAGACCACTAGAGGGTGCCCTTACCATCCAGC
AGTCCAAGAATGTCTGAGGGGGAGCATTGAAGACCTGTCTCCATTGCTTTCATGCCGGG
CAGGCTAAAACATGGAAGGAGTGCCCTAAATGTTCCCTAAACTTCCCCCTAAACAG**GCTGG**
CACCATGTGCGAAGCTTCCCAAGTTTGAGATTGAGTTACCAGCTGCACCCGAAGTCAACTAAAC
TTAGCCTCTCAGAACGAGACATTGCGATGGCGACCATGTATGTCTGTCTTTCAGAGGGTGCT
GTGTGTAACATGACACTTTTTTTAAAAAGTTTTTATTTTTCTGGGTATGCTTTATCTGTATG
TGTGTGCACCTCAGAGGCCATAGAGGGTGTCAAGTCCCCTGGAAATGGAATTACAGCTGC
CATGTGGGTGCTGGGAACCTGAACCCGGGAGCTGAGGTGTGACTTATCACAGTAGAATGCA
CTTTCCTTCTTCTGTCTGTCTGCTTGGCCACCCAGTTTAGGGAGTTGAAAAGTGAAGTGC
TTTTTTGGCTCAATTCTGGGTGGGGACTGGTAGGCGGGCCAAGCTCATTGCTGGCACCC
AGTAGGTGACATCTGCTGTGGAGAGTGCCTGGCCATCCTTCTGTTACGGCACATGTCCT
CGTTCGTGAAAGGAAGCAGCTAGAAGGGAAATAACCAGTGGGTGTTGTGACTGTTGACC
TGTATTTCTTCCCTGCCTTGAATTGTGTGCTTGGCTGTAGTCAAACCTGTATGTAGAGGAC
TCTTACTATCCCTGTAAGGCTGGGTGGAACTTGTCTGTAGTCAAGATGACTTGAACCTT
CTGATCCTCCTGAGTGTGAAATTCTGAACATAAGCCACCATACTCAGTTTATGGCATGA
AGGGGCTCAAGCCAGGCCTTTGTCCAGCTAGGTAAGCCTTACTCTAGCAGCTCAGTTACG
TACCCAGCGGAGAGAAAGGACTCTGGAATGGCCACGATGACTTCTCTCTCTCCGACTAGA
TACGGGfCAGCTCTACGTCCTCTTCCCTGCGGCATCATTCCCGACCTCCAATAGCACAGGAG

CGGAGGTGGTGTGCTGTACCATCTCCCACGGTACGTCTGCCTCATGCCCTCTTCTCTCTTTGC
TTTTAAGTTGAACTGAGGTGTGCTGACCTCTTCCTCTGACTTCTGTGCTCTCTTACAGAG
AGGGTGCCTGTAAGAAGATGCACATACTGAAGTTGAATAGGACGGGAAAGTTTGCTCTAAAT
GTAGTGGACAACCTGGTAGTTGTCCACCATCAGGACACAGAGGTAAAGTCTCCTGGGGAAG
CTGCGGTTGTCTGTGTGGGGCCCTTTGATAGGCTTTGTGAGAACCGGCAAACCTGACAAG
GGTGGGAAAGAGCGTAAGCTCAAAGGAGTTAAGTTTAACTGATTCTCTAACAGTTTATTA
AATAAAAATGATTCCAGTTTGGGGATGCAGTTTGTCTAGGCAGAGGGCTTGCTTCATTCGT
CCCATCCCAGCTTTGCTTAAAGATCAGCTGTGATGGTGCACACCTGTTATCCCAGCACGC
GGGAGGAAGAGGCAGGAGGATGAGAAGTTCAAACAGTCGTTAGCTCTACAGTGGGTGGA
GGGCAGCTTGGGCTACATGAGCTGCTGTCTCAAAAAACAAAACAAAATAAAAAAAA
AACAAACCAAAACAAAACAAAAAATAAGGAAATTA AAAACTTTAAGAAGACATCATTTGA
TGTGATCTTTGTACTCTTGCTCAATCTTGGATTGATCCAGCTGAGCTGGGGCCTGCTGCA
GACCTGGGTGTGGGTGTGGATAGCAGTATCTGACCAGGGTGTGGACACTGCTCTGCCCA
TCAGTCTGCAATAGGTCTTGCTTTTCCAAGTGGTTAGCTGAGCGCCTGGTGGTAGGCCAG
CCACACCTTTCTCTACTTTGCTGAAGGGTCTGCTGCTACGTCTCTAACTAATGGAGAT
GTTGTTTCAAGTTC AAGTGAAGTTTATATGCCTGTGTCTCTTTGAGATCTTACAAGGAAG
CAGGAAGCATTTGTGAAAGCAATAGCAGTGGGAAGGAGAGTGTGATGTGCATAGCCGTCC
TCTGAGGGAGAGCTGGGCTGGGGTGGCTGGGCACAGGCTTCCAGACCGGCAACAGGTGAT
GTGTTTGCTCCTACCGTAG**ACGTCAGTGATATTTGACATCAGGTTACGGGGAGAGTTCGAC**
GGCACAGTCACCTTCCATCATCTGTGCTTCCAGCGAGATCAATTCAGCCTTACCAGATCCC
CTTGGCAGGTAAAGAGAAGTTCTGCGAAATGGCAGTGCTGTCTTTTCCCCACAGGGCCATA
TTCTTGTTAGAATCTTCTGGGGTCAGGCTCAGAATGGGACAGGGAACGAGGAGAATCCCC
AATCTGCTTTATTTTGCAGAGGAACATTTGTTTTCTTGTTTTGGGATGCTGGGGGATTAA
TTAACATTTTCTGTTTTCTGTTTCTGTAACCTGGGATTAATTCAGCTGACTATGGCT
AAATAGTTTTATTTGGGCTTACATGCCCTGCATTTTCATACTCTGTGTTGGTTTTTTGT
TTTTTTTTTAGATTTGCTTTTTTACGTATGTGTTTGTATGTGTCTGCATGTATGTATGAGCG
TGTGTGTTTTGGGTGCTGTCAGAAGGCAGAAGAGGGTGTGGAGTCCCTGGAGCTGGAGTT
ACATGTGGTTGTATCTTTGAGGATGAGGATGCTGGAGACTGAATTCAGGTCCTGTGTTAG
AGCAACAAGGACTTTTAAACCACTGAGCCAGCTTGCCAAACCTTGTTTTGGTTATAGCAGA
TGTGGCCAATGGTTAGTCACAGTAACCCACAGTCGGGGAGGAGGTCAGGAGAGTAGTAG
GTCAGTAGTTCAGAAAAGAAATAGTCCCATTCATAACAGGCTCTAATAATACTGTTACT
GTAATTTTATAGGAGACTACAGTCCAGTTAGCTGGTTGTCTCTGTTTTGTTAGGAGACAT
ACATGTCAGCTGGCCTGTGGCAGTTCTCAGCCTCTGATTTCTTGCTTTTTGTGGAGAC
AGTTGAGAGCCACGCTCTAGGTGGTAGATTGGTCAGCAGGCGCTCAGCCTCGAGTAGCTG
GCTCATCTTACAGGCTCAGAAAATGTCCAGTGGCATCTGGCTTTTCTGGATAACATTAG
CAGTAAGCAACGGATGAGAGCAAATTTTTGACATGGGTAAAACCACTGTAAACATGTAAA
ACCCTCAATTTTGATCCAGTAATAAAGGAGCAATAAACAGGCTGTTTTGCTTGAGCAAGGT
GCAAATTGAAGTGGATCAGATTAGTCTAGGAAGGACTGTCCAACATTTCTTTTATTCTGGC
AGTAATTACGTAGCAAGCCTAGTTGTCTAGAAAGGTCACCTCTGACTTACGAAAATAATG
AAAATAAATAATTGCCGAGAGGGCAGGAATACATGGGTTGCTGGGGTTTTAGGCAGTCTG
GGCAGGTAGAGCATGCCTTCCCTTTGTGTTTTCTGCTGCACATGGGCCCTGGGATATCT
GAGACTAATGTGCCACCTGGCACCGGTTGGGTGGGCATGTACCATAAGCTGATTGTGA
ACGCATTGGTTTTGTCTTGCTGAACAGGTAGGTGGCTTGTGTTTACTGTGTTAGGAACA
CCAGGGGTCTCTGGAGATTGAAATGCCCATGCTCTCCTGAGGCCACGATAGGCAATGGA
ATTTTCTCTTCAGTGAGCCAGTGCTTCGCTGAAGCTTGAGTGACTGCACGGCCATCATT
TGGATGTGGACTGTGTTAGATGGCTGTCTTACCTATGCAGCTAGGCAACACGGCCTTCTG

ATTGTAGTTGGCCAGGTGGCACAGAAATGGGAACCCTGCTTAGCTTGAGGAGTGCTCTCAT
TCCCCAGGCTGCAGGACAGCCAGCAGAACTGGTTGAATGAGGCTCTGCCATGAGAGGGTG
TGCTGCTTAGGTTGGCCCTCTTAGGCCTAGATCTGTTTCAGTGCAGCACAGAGGGTAGGGA
AGCTCACTGTTTCAGCAATTGCCACAAGCCGGCTTCATTTCGACCCATCAGTGAATCTGGG
TAGTATTTTTGAGATATCTGGGTCAATTTCTCAGGGGTTAATGTGCACATTTAAACGGT
GATTCTTCATTTCTGGTTAAAAACAAGCTCAGGGCAGAGACGTGGCTTATGGTGAAGAGC
ACTTGCCGAGCCTGCAGAGGACCTGGGTTTGGTCTCTGGTACCCGTGTGGTGGTGGACAA
CTATCTGTAACCTCCTCTCTTGGCCTCTTCAGGCACCCGGACTTGGATGTGGAACGAGTGCA
CACACTCGTGCGCGCACGCACACACACACACATAACACACACACACACACAAACACACA
CAC
ACCAACTTCTGTGTGGGAAGCTTTGACAAATTTCTCTTCTTGGAGACACATTTTCATGC
CAGTGCTGACACCCTCCCTCTTGTGCTCAGAGGTGTCATTAATGAGGAGGGTCTCCCC
CTCCCCCTCCCCCTCCCTCCCCCGCCCCTTCTCTCCCCCACCTCGCCCTAACTACA
GGTCAAGTCCCTGCTGCTGGAGCCCATTATAAAGTCTGTCTTTACTGTCATCAAGACAAAA
CGGCAACCAACAGTTGGGTAAAAGATCTTTACCAATCCTACAACAAGATAGAGGCTACAA
TATCCAAATATACCCAAGAACTCAAGAAGTGAGACTCCAGAGAGCCAAATCACCCCTATTA
AAAAAGGTACAAAGATAAACAAAACATTTCTCAGCTGAGGATTATCGAATGGCAAAAAGC
ACCTAAAGAAATGTTCAACGATCTTAGTCATCATGGGAAATGCAAATCCAAAACAACCCT
GAGATTGTCACCCTCACACCAGTCAGAATGGGCTAAGATCAAAAACCTCAGGTGACAGCAG
ATGTGGCGAGGATGTGGAGAAAGAGGGAAACACTCCTCCATTGTTTTGGTGGGATTGCAGAC
TGGGTAACAACCCTCTGAAAATCAGTCTGGAGGTTCTCAGAAAATTTGGACATTACATT
ACCTGAGGATCCAGCTAATACCTCTCTTGGGCATATACCCAAAAAGATGCTCCACCTACA
AACAAAGACACGTGCTCCACTATGTTTCATAGCAGCCTTATTTATAATAGCCAGAAGCTG
GAAAGAACCCAGATGCCCTTCAACAGAGGAATGGATTAAAAAAATGTGGTACATCTACAC
AATGGAATACTACTCAGCTATCAAAAACAATGACTTCGTTTTTGATATTCATATCAAAC
ATGAAATTCATAGGCAAATGGAATGAACTAGAAAATATCATCCTGAGTGAGGTTACCCAA
TCACAGAAAAACACACTTGGTATGCACTCCTTGATAAGTGGATATTAGCCAAAAAGCTCA
AATTACCCAAGATGCAATCCACAGACCACAGGAAGCTCAAGAAGAAGGATGACCAAAATG
CAGATGCTCCACTCCTTCTTAAAAGGGGGAAAAATATCATAGAAGGGGATATGGAAGCAA
AGTTTAGAGCAGCGACTCAAGGAATGGCCATTCAGAGCCTGACACACATGTGGCCCATAT
ATATACAGCCACCAAACTAGATAAGATTGATGAAGCTAAAAAAATGCATGCAGAAAGGG
ACTCTTCTGAGAAACACATCCAGATTATGTCCAATACAGAGGTCAATGCTAGCAGCAAAC
CACTGAACCTGAGAATAGGACCCCTTGGGGTAATTAGAGGAAGAGTTGAAGGAGCTTGCA
ACCCATAAGAACAACAATGCCAACCAACCAGAGCTTCCAGGGACTAAACCACTGCTGAA
AGACTATACATGGACTGACCCAGGGCTCCAACCTGCATATGTAGCAGAGAATAGCCTTGTT
GGGGCACCAGTGGAAGGGGAAGCCCTTGGTCTTGCCAAGATTGGACCCCCAGTGTAGGGG
AATATGGGGGAGGGCAATAAGGGGGATTTATAGGGGGAACACCTGTATGGGGGAAGGGAG
GGAATGGGGGCTTGACAGGAAACCAGAAAGGGGAATAACTTTTGAAATGTAAGTAAAGA
AATATATAATAAAAATTAAGAAAAAGAAAAATAATAATAAAAAAATACCTAAAAACAAA
CAAAATCTGTCTTTGCTCGGCTCGGTTCTCCTTCTAGCAGGCTCCCTGACTGCCTCCTCA
CACTTTGGTTTTCTGTTTTGCAG**GTCTGCTGCTGTAACCAACCAGTCTCCCGTCCCCTGTA**
AACTCTGTATCCTCACGAAGTGGTGGCGCTGTGCACACAGGGCAGTCTCCCTGCCCCAC
TGCTGGAGCTGCTCTGGGAGGCAGGCCTTCTGTGACCTCAAACGTGGGGTTGGGCAAG
TGAATGGGGCTTCCCTTTTATACCCAATGTTCCAGTGTGGACTTTCTTTAACGTACA
GATTCCTCCTCGTGGATCGTCTTCCAACCAGACATCATCATCAGTGCAAGCCAAGGTAGGT
CTGGGGGCCATTGCCTACTTTAGTCTGCCAGAGTGACTTCAGTATGAGTCTTTGAGGTG

ACTGACCCTCCCCTCAGCCTTGAAAGTGGGACTGCGTTTCATCCTATAACCCTCCTCTGCA
AGGGTTTTTCCATGAGCATGTATGATTCTTATAGCTAAAGTGACTAAGAATGGCCAATCC
CAGTTTGATTCCCACGGCCATTTAGAAAAGCTGCCTATGGTGAGTGGTCTGCATTCTCAT
TCTCTAGCCCCACTCCCTGTGTTCCCTGCCAAGACCCCTGCCCTTACCTTTCCAGCAGA
CAGCTCCTCAGCCTGCTGTCTCTCCCTGGGTAACGGCCTGCGGCATTGCACCTGAGA
TCCCACCAGTTTCCATGGTTTTTGAGTGTCCAGCCTCACCCCTCTGGGGTTCTTAGGTAGAC
ACCCACTGTATATAGTGAGTGCTTCTCTTCTGTATACTGACAGTAGGGCCAGCTCTTGGA
AACTGAGAAAGGCGTCACTTTTTTTTTTTTTTTTTTGTATGTGACCTCTGGTGGCACCAGCT
TGCTGTTAGCACTCAAATAGGCCGATGAGTGGGTGGGTATTTCACTGATAGCCTATGGG
GTTACCTCCTGCACCCATGAGCTAGTACAGCAAAAACACAAGTACCTGGTACACAGCCAA
TGAGGGGGCTAGCACGTTAGAGGGGAAGGAGTCTCTGGATTGAGCTGCTGTCTTGGAG
ACAACCGTGAGAGTGACTCTTAGTGCAATTCTTCACGCATTAGAAATCAGCTAAGTTTTTC
TGATAGTGTGACTGTGAGGCTTCCACAGTCAGCCTCCTTCCCTGGACATCCTAGCTCCTT
GTTGGTCGGGAAGGAAGGGCTGTTAACTGTATTTGCAGTCCTTAACCTGGGGCAGTTGGG
ACATTTCTTTGTTGGGTTTGGTCTGGCTAGCGCACACTGATGAACATGCTACCTCTTAAC
ATCCAG**GTTACCTCTGGAACCTGCAGGTGAACTTCAGCCCATAGTGAATCTCTTGCCAGAT**
AAAGGAAGACTGATGGACTTTCTCCTCCAGAGGAAGGAGGTAAAGCCGTCGTTCTGTCTGT
CTGTTACAGAGTAAGTGCACCTGACCCTTTCCCCTGTCTGTGGAAAAGCACCATCTCTT
TATTTAATTATGCCTTGCGGCCTGTGCTTAGTACCTGCTGAGCCTCCTTCCAGGGTGCCC
AACCCACCTTACAACCTGGGAATTGAACATGCCTTTCACACTCAGGCCGGCACACTTGAA
GGATGGATTCTTAACCTCATAGGAAAGACTGGGAGATAAACACAGGACTTGTAGACTGAC
AGAAACCCACCAGTAGGCAGAAAGCTGATTACATCCTCTGGTGCCACTGGTCTGACTGA
TTTTAACATTTCTTGAAGTCCTCAAGGCAGGGTTTGCATGGGTACCCCTCTGAGACTG
CCATGTCACCAGCACGTGGGGATCCTGATTACAGTGGGTCTGACCGTGTGCATTGTACTT
TCCAG**TGCTGAGTGAGTCAGACAGAGCCACGCTGCCTGTGATAGCCACTGTTTTCGATAAAC**
TCAACCATGAGTACAAGAAGTACTTGGATGCTGAACAGAGCTACACAATGGTAAGACAGAC
TGTATTGGAGAGTCCCTGCAGTGTGATGTGGTCATACAGCGTAACAGCTGAGGGCACCGTG
ACAGTGGCAAGCCAGAGTGCATCTGGCTCAAGGCTTCACTGTTCTGCTTTTTAAAAATCTG
TTTGTGCTTTTTGTTTGGTTATGTTGTAGAAAGACAATCAGCCGTGGCTTTTTAACAGAGCA
AATTAGAGATGTGATTTCTTTCCACTGGTGTCTGCTTTCTCCTTTGCAAATATTTAGAA
ATATCAGGTTATCTAGAGGGACATTTGGGCAGGGCGTGTCCAACATGCAACACGTACTTG
TATGAAACCTAAAAGTAAATTTCTAAAAGATGGTTATTTCAAGTAAGAATAGACGAATGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATCCCCACCCCCACCCCAG
CTGAGAACCAAACCCAGGGCCTTGTGCTTGTAGGCAAGCACTCTACCCTGAGCTAAAT
CCCCAACCCCTTGTGGTTTTTTTTTTTTTTTTTTTTGTTTTGTTTTGTTTTGTTTTAAAGACAGGG
TTTCTCTATGTTGCTGTCCAGGTTGGTCTTGAACCTCAGAGGTCCACCTGTCTCCTGAGTG
CTGGGATTAAGCCTTGTTTTTTTTTTAGAGTCTCACTGTGTACCTCTACCTCTGGCTGGCC
TGGAACCTGGTTATAGAGAGTCCCTCCTAAGTATTAGGCATGTGACACGACACCTGGCCAA
AATTTTATTGGTGAATGTGTCCTTTGACAATTTCTGTACGTCTGCAGCGTTGTTTCCT
CACTTCCCAGTCACTCCCTCCTACCTGTGTCTCCCTCCTGCCCTTACTCCTCTTTCTCTG
TTACGATGGCTTTGTTTTATGACCCACTGATTTAATGAGGACTGTCTAATAAGCACAGA
TTTGCAAAGTCCCTTGGAGCCTGGTGGGATCACTATTGGGTACTCAACCGGAAGGCCA
TGGCTGTCTCTTCCCAGACAGTTCAGCAGGGAAATTTAGGGTCCCATGAGCCCTCCAGA
TCCACGGCTGACGGCCAGTGCAGTGCAGTTGTGTGAGATGTCTGTGTCAAGCTTATGGGT
GGTATTTTGTAGCCCTTTCTTGGTCTTCTGGCGTTTACATTTCTTGTGATGTTCTCTGAA
TCTCAGAAGGGGCTGAATATTTCAACATTTATCCTCAGCATCTGGGCAGACACACAGATAT

CCTAGGACAAGTACCCTGTGAGCCACTGGTTAAGTACCAGCACCATGCATTTTCAGTTAAT
AGACAGCACGTGTTTCAGTGAAAGGTAGGGAAAAAGTTACTGTGGCTTATTACTAGGTACG
AAACTTAGTTCTTCCACTTTTCTAAAGCGTGAAAGAACTTGGGACAAGTGGGAGAAATT
ATTTCAAGGACATGCTTAGCAACAGTGGACACTGTTTTAAGCAACAACTGAGACCCCCA
CTTTGCATGGCTATTTGAGGTTAGTGAGTCCTGGGCCACTGAAGGGCAGGCTGCCTACAT
TACCTCTCCTTGTTCGTGTCAGCCGTCTGGCTGTGCCTCCTTCCTCCTCTTGGTCCTTC
TGGTGGTGTCTGGGCTGAGGGCTTGAGCCCAGTGGCAGACTCTGGGCTATGCCAGACTG
CTTATCAGCTGTCAGGAAGCAGAGGTCTGCACTGAGCTGCTGCTCACTAGAGATCAAACA
GCTGGCACAGCTCACAGCTCTGTCTGTAATAATTCCCATTACACTGGCCTGTGGCTGT
AACTTGGTGGTAGAAATGTAGATTGGTGAAGCTCTGGATCCATTTCAAGGACTACAGAAT
GAATTCTAAAATGACAGCATAGAGCTTTTACTGTCTGGCAGAACTTCATGAATGAGGGCG
CAGCTGGCCAGGAGAGTCAGGTCAGTCAATCTTCAGAAGTTGGGAAAAACATGACTGGC
TTTATTTGATTGCAGTCTCCACAAACACTGGAGTGTCTGGCTTGTGTCTAGAAACTTAA
GTCTTGAGCTAAGTATAGTGGGTACCATTTCTGTGGAAGATGACTAAGTCCACAAGTTGT
CCCTCTTTAG**GCAGTAGAAGCAGGGCAAAGCCGGAGCAACCCACCTCTCAAAGGCCAGTG**
AGGACCCAAGCCGTGGTAGACCAGTCTGATGTGTACACACACGTGCTGTCACCATTTGTGGA
AAACAAGGTATGTAATGGTTTTCTGGTCTCCCAAGTGGGTACTGAGAGGGCAGCAACAGA
ATCCACACTATGGTTAAATTAAGTGGTATTGCCCTGGGAGAGGGTCTTGGGCCATGAGGC
TGGCTTCTAGTAGACTTATACTTGGTCACTAGGACGTACGGTCCATAAGGATTTGTCTAC
ACCAGCAGACAGGATCAGTGGGCTGCCTGTAACCTCTTTTGCAACTCACATTACAAGACC
CCCTTCTTCTCCACAG**GAGACGCCACCCAAGTTTTGTGATAGCTGTGCTGATGGAATACAT**
CCGTTCTCTGAACCAGTTTTCAGATCCAGTGCAGGTACCTTCAGAGGGCTGCTTGCCATCC
TGAGTTAGAGGAAGGAGGTCAGAAGAGATGGCTGAAAAATCCGCAGACCCCGGGGTGATG
TTAACATGTCAGTCATCAGTGGGGATGGGGAATCTGGTTGAGAATGCTTTGTTACAGCTA
AATTATGGGAATGTGGCCAGTCACATAGTCCGGCCACTCAGCGATGTTATTTACCTTCTT
CCTGTGTCCCCAG**CATTACTTACACGAACTTGTTCATCAAGACACTTGTCCAGCACAACTA**
TTTTACATGCTCCACCAGTTTTCTGCAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGTCTG
TATCATTGGATGTCCATTCTGTAGTCTCTAAACCTCCCCGTGGGAACCTTCCAAGTGGG
CCGGCTGAAACCCTGTCTTTCAG**GCTTGCTTGGTGGCTGTCCCTAGAAAGCTTTTATCCTCC**
CGCCCACCAGCTGTCTCTGGACATGTTGAAGTACCCGCCCTGCCAGCTTCAGGTCGCCA
GCGTTCTCTCCTGTGCTTGCCCTGACTGCTGTCTTCCCTCACTTT**CAGAGACTTTCAACAGC**
AAATGATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGGTTCA
TCCGGGGTATTGGCGCCATGACAACATTTCCGCACGAAAATCTTAGATGCTGCTAAGCAG
ACTGATGATGTCATGCTCTTCTATACCATATTCCGCTTCTTTGAACAGCGAAACCAGCGTTT
GCGAGGGAACCCTAAC TTCACACCAGGTAAGGGAAGGCCAGGGTGACCTCAGGCAACGGG
GGGGGGGGGGGGGGCTTTCAAGAAGTTCAGGGTGGAAATTTAGGCTTCTGGGGCATATTC
TATATCTGGCCTTGGAAATCTCTATATGGACGAATCAGGGACAGGAGTGCCTCAGTTTTAA
ATGGTGATAAAACAGGATGCAGGAGAGGAGCTTATGCTGTGTGTACACAGTACAGGTGGC
CTGAGTGACAGGTCAGGGATGCCCTGGTCTGACAAGGATGTAAGCCGTGACTGCTGAAGG
TACGAGTTCTGCAACAGACCTCAAAAATGTAGGAGGGATTTTATTAAGAAAAATAAGTTA
AAACCAGCATAACAACCATGAGCATTGTCTGCACTCCGTCTTCCCTGTGCCACAACCAGC
TGGCACCTAAGCTACATTTTTTCTG**CAGGAGAACACTGTGAGGAACACGTAGCCTTTTTTCA**
AGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTTCTGAAGTCCCTCAACGGTT
TTAATATAAAAATGTACAAAGTTAATTTATTGCATTAATAAAGCTCTTTAAACTACAAAAGT
GTTGTAATAAAGTACCTACAGCATCCTCAAATGTTGCTGACTCCCAAACAGGAGAGAGCAT
GTGGGCTGGCTTTTTCAGAATTGGGGATCAGGTCACAGTTGTAGCCACCACCTGTGAAGTGGA

CAGTCAGTTGTGTTGGTGTTTTGCTGGGGCAGACACAGGTGAAAGGCTAAGGCAGACTTG
TGAAGGACTGTTTCCCTAAAGCAAGCACGTGAAAGGACATGAATGACTGGCATGTATTGC
TTGGTCTGCCTTACATTGCGTAGCTGAGCTCTGTTTGTGGGACTCCACAGAGAAGGAAA
TAGAGAGCTAAGCTGTGGAACATTTGTTGGGTCTTGAGTCTTCGCTGTGAGAGGCTCAGC
AGAGAGCATCGAGAGTTCCTGTTGGCCCCGCCCTCCTGCTGATGGCCATTTCAGCCGAGG
CCTGGCCACTGCTGCTGCCATCCCGATACTATTGAACTGGACAGCTGATGTATCCGTGAG
GTGTTTGAGTGATCGAGTTGCTGACCTATGAACTGAACTGCTGATTTCCCTGACAACACAG
ATAGGATATGCTGCAAAGAACCTTTCTTAACACATCCACTTCCACTCTCTGGTGGGTGT
TGGACTAGAAGGGGTTAAAGTTTTAATGATGGTTTGATAGCCATCATTTAAAGTAGGCAT
TGAAAGAATGAAAGTTGTAGAACTGCCCTTCTTGTGCCCAAGATGAGGTTTCTTAG

ix. CRISPR/Cas9 Target Sequences

1. Exon 1

GTAACTGCGTGTTCTTCGA

2. Exon 8

CTCTCTCTCCGACTAGATAC

3. Exon 9

GGTAGTTGTCCACCATCAGG

4. Exon 10

GTGCTTCCAGCGAGATCAAT

x. pGAD-WDR98

TGCATGCCTGCAGGTCGAGATCCGGGATCGAAGAAATGATGGTAAATGAAATAGGAAA
TCAAGGAGCATGAAGGCAAAGACAAATATAAGGGTCGAACGAAAAATAAAGTGAAAA
GTGTTGATATGATGTATTTGGCTTTGCGGGCCGAAAAACGAGTTTACGCAATTGCA
CAATCATGCTGACTCTGTGGCGGACCCGCGCTCTTGCCGGCCCGGCGATAACGCTGGG
CGTGAGGCTGTGCCCGGCGAGTTTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCG
CTAAGGGGCGAGATTGGAGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGA
CCACGACAACCTGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAACA
TGAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGTGGCA
ACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAACCGCTAGAGTACTTTGA
AGAGGAAACAGCAATAGGGTTGCTACCAGTATAAATAGACAGGTACATACAACACTGG
AAATGGTTGTCTGTTTGAGTACGCTTTCAATTCATTTGGGTGTGCACTTTATTATGTT
ACAATATGGAAGGGAACTTTACACTTCTCCTATGCACATATATTAATTAAGTCCAAT
GCTAGTAGAGAAGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTTCTAAACC
GTGGAATATTTCCGATATCCTTTTGTGTTTCCGGGTGTACAATATGGACTTCCTCTT
TTCTGGCAACCAAACCCATACATCGGGATTCTATAATACCTTCGTTGGTCTCCCTAA
CATGTAGGTGGCGGAGGGGAGATATAACAATAGAACAGATACCAGACAAGACATAATGG
GCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAACTAA

TACTGTAGCCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACTACCCTTTTTCC
ATTTGCCATCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTTCTTTTTTTTTCTT
TTCTCTCTCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAATGATGGAAG
ACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAG
CTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCTTCCTTCATTACGCACACT
ACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCAGTTACTTGAATTTGAAATAAA
AAAAAGTTTGCTGTCTTGCTATCAAGTATAAATAGACCTGCAATTATTAATCTTTTGT
TTCCCTCGTCATTGTTCTCGTTCCTTTCTTCCTTGTTCCTTTTTTCTGCACAATATTT
AAGCTATAACCAAGCATAACAATCAACTCCAAGCTTTGCAAAGATGGATAAAGCGGAATT
AATCCCGAGCCTCCAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCGCCAATTTT
AATCAAAGTGGGAATATTGCTGATAGCTCATGTGCCTTCACTTTCACTAACAGTAGCA
ACGGTCCGAACCTCATAACAACCTCAAACAAATTTCTCAAGCGCTTTCACAACCAATTGC
CTCCTCTAACGTTTTCATGATAACTTCATGAATAATGAAATCACGGCTAGTAAAAATTGAT
GATGGTAATAATTCAAAACCACTGTCACCTGGTTGGACGGACCAACTGCGTATAACG
CGTTTGGAACTACTACAGGGATGTTTAATACCACTACAATGGATGATGTATATAACTA
TCTATTTCGATGATGAAGATACCCACCAACCAAAAAAGAGATCTTTAATACGACT
CACTATAGGGCGAGCGCCGATGGAGTACCATAACGACGTACCAGATTACGTCATA
TGGCCATGGAGGCCAGTATGGGCAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCC
AGTGCAGTTCGAGAAGGCGAACCCGGTTAACTGCGTGTCTTCGACGAGGCCAACAAAG
CAGGTTTTTGTGTTTCGGTCAGGTGGAGCTACAGGAGTGGTTCGTTAAAGGTCCAGATG
ATAGGAATCCCATCTCTTTTAGAATGGACGACAGAGGAGAAGTGAAGTGCATTAAGTT
TTCTCTAGAAAATAAGATACTGGCTGTCCAGAGGACTTCGAAGGCCGTGGACTTTTTGT
AATTTTATTCAGATAACTCTCAACTGGAGTACACACAAGAGTGAAGACCAAGAATG
CCAACATACTAGGATTCTGCTGGACCAGTTCTACCGAAATTTGTCTTCATAACGGATCA
AGGCATCGAATTTTACCAGGTGTGCCTGAGAAGCGAAGCCTCAAACCTCCTGAAGAGC
CACAACATCAATGTGAACTGGCACATGTACTGTCCCGAGAGCTCGGTGATCCTGCTGT
CCACGACAGTCCTTGAGAATGTCTGCAGCCTTTTTACTTTAAGGCTGGCACCATGTC
GAAGCTTCCCAAGTTTGAGATTGAGTTACCAGCTGCACCGAAGTCAACTAAACTTAGC
CTCTCAGAACGAGACATTGCGATGGCGACCATATACGGGCAGCTCTACGTCCTCTTCC
TGCGGCATCATTCCCGACCTCCAATAGCACAGGAGCGGAGGTGGTGTGCTGTACCATCT
CCCACGAGAGGGTGCCTGTAAGAAGATGCACATACTGAAGTTGAATAGGACGGGAAAG
TTTGCTCTAAATGTAGTGGACAACCTGGTAGTTGTCCACCATCAGGACACAGAGACGT
CAGTGATATTTGACATCAGGTTACGGGGAGAGTTTCGACGGCACAGTACCTTCCATCA
TCCTGTGCTTCCAGCGAGATCAATTCAGCCTTACCAGATCCCCTTGGCAGGTCCCTGCT
GCTGTAACCAACCAGTCTCCCGTCCCCTGTAAACTCTATTCCCTCCTCGTGGATCGTCT
TCCAACCAGACATCATCATCAGTGAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAA
ACTTCAGCCATAGTGAATCTCTTGCCAGATAAAGGAAGACTGATGGACTTTCTCCTC
CAGAGGAAGGAGTGTAAGCCGTCGTTCTGTCTGTCTGTTACAGATGCTGAGTGAGT
CAGACAGAGCCACGCTGCCTGTGATAGCCACTGTTTTCGATAAACTCAACCATGAGTA
CAAGAAGTACTTGGATGCTGAACAGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGC
CGGAGCAACCCACCTCTCAAAGGCCAGTGAGGACCCAAGCCGTGGTAGACCAGTCTG
ATGTGTACACACACGTGCTGTCAACATTTGTGGAAAACAAGGAGACGCCACCCAAGTT
TGTGATAGCTGTGCTGATGGAATACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTG
CAGCATTACTTACACGAACTTGTTCATCAAGACACTTGTCCAGCACAACTATTTTACA
TGCTCCACCAGTTTCTGCAGTACCATGTGCTCAGCGACTCCAAGCCTCTGGCTTGCTT
GTTGCTGTCCCTAGAAAGCTTTTATCCTCCCGCCCACCAGCTGTCTCTGGACATGTTG

AAGAGACTTTC AACAGCAAATGATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGG
TGCTGGCTGCCCTAAGGTT CATCCGGGGTATTGGCGGCCATGACAACATTTCCGCACG
AAAATTCTTAGATGCTGCTAAGCAGACTGATGATGTCATGCTCTTCTATACCATATTC
CGCTTCTTTGAACAGCGAAACCAGCGTTTTCGAGGGGAACCCTAACTTCACACCAGGAG
AACACTGTGAGGAACACGTAGCCTTTTTTCAAGCAGGCTTTTGGAGAACAAGCTCTAAT
GAGGCCTACAACATTCTGA

xi. pGBDU-WDR98

TGATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTT
GTTCCAGAGCTGATGAGGGGTATCTCGAAGCACACGAACTTTTTTCCTTCCTTCATTC
ACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCAGTTACTTGAATT
TGAAATAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAATAGACCTGCAATTATTA
ATCTTTTGTTCCTCGTCATTGTTCTCGTTCCTTTCTTCCTTGTTTCTTTTTCTGCA
CAATATTTCAAGCTATACCAAGCATAACAATCAACTCCAAGCTTGAAGCAAGCCTCCTG
AAAGATGAAGCTACTGTCTTCTATCGAACAAAGCATGCGATATTTGCCGACTTAAAAAG
CTCAAGTGCTCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACCTGGGAGT
GTCGCTACTCTCCCAAACCAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGT
GGAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATTTTTCTTCGAGAAGAC
CTTGACATGATTTTTGAAAATGGATTCTTTACAGGATATAAAAGCATTGTTAACAGGAT
TATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGATAGATTGGCTTCAGTGGA
GACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCAGACATCATCATCGGAA
GAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGCCGGAATTCCCCGGGATGG
GCAGCGAGGACTACTACCTGGAGCTGTGCGAGCGGCCAGTGCAGTTCGAGAAGGCGAA
CCCCGTTAACTGCGTGTTCCTTCGACGAGGCCAACAAGCAGGTTTTTGCTGTTCGGTCA
GGTGGAGCTACAGGAGTGGTCGTTAAAGGTCCAGATGATAGGAATCCCATCTCTTTTA
GAATGGACGACAGAGGAGAAGTGAAGTGCATTAAGTTTTCTCTAGAAAATAAGATACT
GGCTGTCCAGAGGACTTCGAAGGCCGTGGACTTTTTGTAATTTTATTCCAGATAACTCT
CAACTGGAGTACACACAAGAGTGCAAGACCAAGAATGCCAACATACTAGGATTCTGCT
GGACCAGTTCTACCGAAATTGTCTTCATAACGGATCAAGGCATCGAATTTTACCAGGT
GTTGCCTGAGAAGCGAAGCCTCAAACCTCCTGAAGAGCCACAACATCAATGTGAACTGG
CACATGTACTGTCCCGAGAGCTCGGTGATCCTGCTGTCCACGACAGTCCCTTGAGAATG
TCCTGCAGCCTTTTTACTTTAAGGCTGGCACCATGTGCAAGCTTCCCAAGTTTGAGAT
TGAGTTACCAGCTGCACCGAAGTCAACTAAACTTAGCCTCTCAGAACGAGACATTGCG
ATGGCGACCATATAACGGCAGCTCTACGTCTCTTCTGCGGCATCATCCCCGGACCT
CCAATAGCACAGGAGCGGAGGTGGTGTGCTGTACCATCTCCACGAGAGGGTGCCTGTAA
GAAGATGCACATACTGAAGTTGAATAGGACGGGAAAGTTTGCTCTAAATGTAGTGGAC
AACCTGGTAGTTGTCCACCATCAGGACACAGAGACGTCAGTGATATTTGACATCAGGT
TACGGGGAGAGTTTCGACGGCACAGTCACCTCCATCATCCTGTGCTTCCAGCGAGATC
AATTCAGCCTTACCAGATCCCCTTGGCAGGTCCTGCTGCTGTAACCAACCAGTCTCCC
GTCCCCTGTAAACTCTATTCCTCCTCGTGGATCGTCTTCCAACCAGACATCATCATCA
GTGCAAGCCAAGGTTACCTCTGGAACCTGCAGGTGAAACTTCAGCCCATAGTGAATCT
CTTGCCAGATAAAGGAAGACTGATGGACTTTCTCCTCCAGAGGAAGGAGTGTAAAGCC
GTCGTTCTGTCTGTCTGTTACAGATGCTGAGTGAGTCAGACAGAGCCACGCTGCCTG

TGATAGCCACTGTTTTTCGATAAACTCAACCATGAGTACAAGAAGTACTTGGATGCTGA
ACAGAGCTACACAATGGCAGTAGAAGCAGGGCAAAGCCGGAGCAACCCACCTCTCAA
AGGCCAGTGAGGACCCAAGCCGTGGTAGACCAGTCTGATGTGTACACACACGTGCTGT
CACCATTTGTGGAAAACAAGGAGACGCCACCCAAGTTTGTGATAGCTGTGCTGATGGA
ATACATCCGTTCTCTGAACCAGTTTCAGATCCCAGTGCAGCATTACTTACACGAACTT
GTCATCAAGACACTTGTCCAGCACAACCTATTTTACATGCTCCACCAGTTTCTGCAGT
ACCATGTGCTCAGCGACTCCAAGCCTCTGGCTTGCTTGTGCTGTCCCTAGAAAAGCTT
TTATCCTCCCGCCCACCAGCTGTCTCTGGACATGTTGAAGAGACTTCAACAGCAAAT
GATGAGATCGTAGAAGTGCTTCTGTCTAAGCACCAGGTGCTGGCTGCCCTAAGGTTCA
TCCGGGGTATTGGCGGCCATGACAACATTTCCGCACGAAAATTCTTAGATGCTGCTAA
GCAGACTGATGATGTCATGCTCTTCTATAACCATATTCCGCTTCTTTGAACAGCGAAAC
CAGCGTTTGGCAGGGAAACCCTAACTTCACACCAGGAGAACACTGTGAGGAACACGTAG
CCTTTTTCAAGCAGGCTTTTGGAGAACAAGCTCTAATGAGGCCTACAACATTCTGA

xii. GST-RAB5A

ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTC
TTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTGTATGAGCGGATGAAGGTGA
TAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTAT
ATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACTTATATAGCTGACA
AGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGG
AGCGGTTTTGGATATTAGATACGGTGTTCGAGAATTGCATATAGTAAAGACTTTGAA
ACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTTCAAGATC
GTTTATGTCAATAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTT
GTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCA
AAATTAGTTTGTAAAAAACGTATTGAAGCTATCCACAAATTGATAAGTACTTGA
AATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGG
CGACCATCCTCCAAAATCGGATCTGGTTCGGATCCCGGGGCCCGGTACCGCGGCCGC
TAGCATGCAGTCGACTAGTATGGCTAGTCGAGGCGCAACAAGACCCAACGGCCCAAAT
ACTGAAAATAAAATATGCCAGTTCAAACTAGTACTTCTGGGAGAGTCCGCTGTTGGCA
AATCAAGCCTAGTGCTTCGTTTTGTGAAAGGCCAATTTTCATGAATTTCAAGAGAGTAC
CATTGGGGCTGCTTTTCTAACCCAAACTGTATGTCTTGATGACACTACAGTAAAGTTT
GAAATCTGGGATACAGCTGGTCAAGAAGGATACCATAGCCTAGCACCAATGTACTACA
GAGGAGCACAAGCAGCCATAGTTGTATATGATATCACAAATGAGGAGTCCTTTGCAAG
AGCAAAAAATTGGGTAAAGAACTTCAGAGGCAAGCAAGTCCTAACATTTGTAATAGCT
TTATCGGGAAAACAAGGCCGACCTAGCAAATAAAAAGAGCAGTAGATTTCCAGGAAGCAC
AGTCCATGCAGATGACAATAGTTTATTATTTCATGGAGACATCCGCTAAAACATCAAT
GAATGTAAATGAAATATTCATGGCAATAGCTAAAAAATTGCCAAAGAATGAACCACAA
AATCCAGGAGCAAATTCGCCAGAGGAGGAGTAGACCTTACCGAACCCACACAAC
CAACCAGGAATCAGTGTTGTAGTAACTAAACCTCTAGTTTGAAC

xiii. GST-RAB7A

ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCT
TTTGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATA
AATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATT
GATGGTGTATGTTAAATTAACACAGTCTATGGCCATCATACTGTTATATAGCTGACAAGCA
CAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGG
TTTTGGATATTAGATACGGTGTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTC
AAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTTCGAAGATCGTTTATG
TCATAAAACATATTTAAATGGTGTATCATGTAACCCATCCTGACTTCATGTTGTATGACG
CTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCAAAATTAGTT
TGTTTTAAAAACGTATTGAAGCTATCCACAAATTGATAAGTACTTGAAATCCAGCAA
GTATATAGCATGGCCTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTC
CAAATCGGATCTGGTTCGGATCCCGGGGCCCGGTACCGCGGCCGCTAGCATGCAGTC
GACTAGTATGACCTCTAGGAAGAAAGTGTGCTGAAGGTTATCATCCTGGGAGATTCTG
GAGTCGGGAAGACATCACTCATGAACCAGTATGTGAATAAGAAATTCAGCAATCAGTAC
AAAGCCACAATAGGAGCTGACTTTCTGACCAAGGAGGTGATGGTGGATGACAGGCTAGT
CACAATGCAGATATGGGACACAGCAGGACAGGAACGGTTCAGTCTCTCGGTGTGGCCT
TCTACAGAGGTGCAGACTGCTGCGTTCGGTATTTGATGTGACTGCCCCCAACACATTC
AAAACCCTAGATAGCTGGAGAGATGAGTTTCTCATCCAGGCCAGTCCCCGAGATCCTGA
AAACTTCCCATTTGTTGTGTTGGGAAACAAGATTGACCTCGAAAACAGACAAGTGGCCA
CAAAGCGGGCACAGGCCTGGTGTCTACAGCAAAAACAACATTCCCTACTTTGAGACCAGT
GCCAAGGAGGCCATCAACGTGGAGCAGGCGTTCAGACGATTGCACGGAATGCACTTAA
GCAGGAAACGGAGGTGGAGCTGTACAACGAATTTCTGAACTATCAAACCTGGACAAGA
ATGACCGGGCCAAGGCCTCGGCAGAAAGCTGCAGTTGCTGA

List of Abbreviations

ADE - adenine

AF - Alexa Fluor

AGC - automatic gain control

APP - amyloid precursor protein

APS - ammonium persulphate

BCA - bicinchoninic acid assay

BLAST - Basic Local Alignment Search Tool

BSA - bovine serum albumin

BSP - BSA-saponin-PBS

CEB - cytosol extraction buffer

CMA - chaperone mediated autophagy

CRISPR - clustered regularly interspaced short palindromic repeats

ddH₂O - double distilled water

DFS - DharmaFect solution

DMEM - Dulbecco's modified eagle medium
DNase - deoxyribonuclease
EDTA - ethylenediaminetetraacetic acid
EEA1 - early endosome antigen 1
EGF - epidermal growth factor
EGFR - epidermal growth factor receptor
ESI - electrospray ionisation
FDR - false discovery rate
GDI - guanine dissociation inhibitor
GDP - guanosine diphosphate
GEF - guanine nucleotide exchange factor
GFP - green fluorescent protein
GLU - glutamic acid
GLN - glutamine
GMPPNP - 5'-guanylyl imidodiphosphate
GST - glutathione s-transferase
GTP - guanosine triphosphate
HCD - higher-energy collisional dissociation
HDAC6 - histone deacetylase 6
HeLa - Henrietta Lacks
HIS - histidine
HRP - horseradish peroxidase
HSP90 - heat shock protein 90
IF - immunofluorescence
IgG - immunoglobulin G
IP - immunoprecipitation
KDA - kilodaltons
LAMP - lysosomal associated membrane protein
LC-MS - liquid chromatography mass spectrometry
LDS - lithium dodecyl sulfate
LEU - leucine
LGP120 - lysosomal membrane glycoprotein 120
LSM - laser scanning microscope
NCBI - National Center for Biotechnology Information
NRK - normal rat kidney

OD - optical density
PAGE - polyacrylamide gel electrophoresis
PBS - phosphate buffered saline
PCR - polymerase chain reaction
PEG - polyethylene glycol
PVDF - polyvinylidene fluoride
RMC1 - regulator of MON1-CCZ1 complex
RT - room temperature
SD - synthetic defined
SDS - Sodium dodecyl sulfate
SFM - serum free media
SOC - super optimal broth with catabolite repression
STUB1 - STIP1 homology and U-Box containing protein 1
TAE - tris-acetate-EDTA
TBS - tris buffered saline
TBST - tris buffered saline TWEEN 20
TEMED - tetramethylethylenediamine
URA - uracil
V/V - volume/volume
WB - western blot
WDR98 - WD40 repeat 98
WT - wild type
W/V - weight/volume
YPD - yeast extract-peptone-dextrose
Y2H - yeast two hybrid

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Ansari, A. M., Ahmed, A. K., Matsangos, A. E., Lay, F., Born, L. J., Marti, G., Harmon, J. W., Sun, Z. (2016). Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Rev.* **12**:553-559.

Appelqvist, H., Waster, P., Kagedal, K., Ollinger, K. (2013). The lysosome: from waste bag to potential therapeutic target. *J Mol Cell Biol.* **5**:214-226.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* **25**:25-29.

Bagley, D. C., Paradkar, P. N., Kaplan, J., Ward, D. M. (2012). Mon1a protein acts in trafficking through the secretory apparatus. *J Biol Chem.* **287**:25577-25588.

Balaji, K., Mooser, C., Janson, C. M., Bliss, J. M., Hojjat, H., Colicelli, J. (2012). RIN1 orchestrates the activation of RAB5 GTPases and ABL tyrosine kinases to determine the fate of EGFR. *J Cell Sci.* **125**:5887–5896.

Behrends, C., Sowa, M. E., Gygi, S. P., Harper, J. W. (2010). Network organization of the human autophagy system. *Nature.* **466**:68-76.

Cuervo, A. M. & Wong, E. (2013). Chaperone-mediated autophagy: roles in disease and aging. *Cell Res.* **24**:92-104.

Delgado, A. P., Brandao, P., Chapado, M. J., Hamid, S., Narayanan, R. (2014). Open reading frames associated with cancer in the dark matter of the human genome. *Cancer Genomics Proteomics.* **11**:201-214.

Eskelinen, E. L. (2006). Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. *Mol Aspects Med.* **27**:495-502.

Eskelinen, E. L., and Saftig, P. (2009). Autophagy: A lysosomal degradation pathway with a central role in health and disease. *BBA Mol Cell Res.* **1793**:664-673.

Furuta, K., Ikeda, M., Nakayama, Y., Nakamura, K., Tanaka, M., Hamasaki, N., Himeno, M. *et al.* (2001). Expression of lysosome-associated membrane proteins in human colorectal neoplasms and inflammatory diseases. *Am J Pathol.* **159**:449-455.

Galletta, B. J. and Rusan, M. N. (2015) A Yeast Two-Hybrid approach for probing protein-protein interactions at the centrosome. *Methods Cell Biol.* **129**:251-277.

Gillingham, A. K., Sinka, R., Torres, I. L., Lilley, K. S., Munro, S. (2014). Toward a comprehensive map of the effectors of rab GTPases. *Dev Cell.* **31**:358-373.

Johnson, D. E., Ostrowski, P., Jaumouille, V., Grinstein, S. (2016). The position of lysosomes within the cell determines their luminal pH. *J Cell Biol.* **212**:677.

Hegedus, K., Takats, S., Boda, A., Jipa, A., Nagy, P., Varga, K., Kovacs, A. L., Juhasz, G. (2016). The Ccz1-Mon1-Rab7 module and Rab5 control distinct steps of autophagy. *Mol Biol Cell.* **27**:3132-3142.

Hernandez, A., Serrano-Bueno, G., Perez-Castineira, J. R., Serrano, A. (2012). Intracellular proton pumps as targets in chemotherapy: V-ATPases and cancer. *Curr Pharm Des.* **18**:1383-1394.

Jacobsen, L., Calvin, S., Lobenhofer, E. (2009). Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells. *Biotechniques.* **47**:617-624.

Jones, J. J., Wilcox, B. E., Benz, R. W., Babbar, N., Boragine, G., Burrell, T., Christie, E. B. *et al.* (2016). A plasma-based protein marker panel for colorectal cancer detection identified by multiplex targeted mass spectrometry. *Clin Colorectal Cancer.* **15**:186-194.

Kaushik, S., Bandyopadhyay, U., Sridhar, S., Kiffin, R., Martinez-Vincente, M., Orenstein, S. J., Wong, E., Cuervo, A. M. (2011). Chaperone-mediated autophagy at a glance. *J Cell Sci.* **124**:495-499.

Kaushik, S. & Cuervo, A. M. (2012). Chaperone-mediated autophagy: a unique way to enter the lysosome world. *Trends Cell Biol.* **22**:407-417.

Kimmelman, A. C., Hezel, A. F., Aguirre, A. J., Zheng, H., Paik, J., Ying, H., Chu, G. C. *et al.* (2008). Genomic alterations link Rho family of GTPases to the highly invasive phenotype of pancreas cancer. *Proc Natl Acad Sci USA.* **105**:19372-19377.

Kinchen, J. M. and Ravichandran, K. S. (2010). Identification of two evolutionarily conserved genes regulating processing of engulfed apoptotic cells. *Nature.* **464**:778-782.

Kiontke, S., Langemeyer, L., Kuhlee, A., Schuback, S., Raunser, S., Ungermann, C., Kummel, D. (2017). Architecture and mechanism of the late endosomal Rab7-like Ypt7 guanine nucleotide exchange factor complex Mon1-Ccz1. *Nature Communications*. **8**:14034.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, F., Valentin, I. M. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**:2947-2948.

Lipinski, M. M., Hoffman, G., Ng, A., Zhou, W., Py, B. F., Hsu, E., Liu, X. *et al.* (2010). A genome-wide siRNA screen reveals multiple mTORC1 independent signaling pathways regulating autophagy under normal nutritional conditions. *Dev Cell*. **18**:1041-1052.

Luzio, J. P., Pryor, P. R., Bright, N. A. (2007). Lysosomes: fusion and function. *Nat Rev Mol Cell Biol*. **8**:622-632.

Mellman, I. and Yarden, Y. (2013). Endocytosis and cancer. *Cold Spring Harb Perspect Biol*. **5**:a016949.

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. **46**:D8-D13

Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. **75**:4646-4658.

Nordmann, M., Cabrera, M., Perz, A., Brocker, C., Ostrowicz, C., Engelbrecht-Vandre, S., Ungermann, C. (2010). The Mon1-Ccz1 complex is the GEF of the late endosomal Rab7 homolog Ypt7. *Curr Biol*. **20**:1654-1659.

Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., Kolas, N. *et al.* (2018). The BioGRID interaction database: 2019 update. *Nucleic Acids Research*. **47**:D529-D541.

Piao, S. and Amaravadi, R. K. (2016). Targeting the lysosome in cancer. *Ann N Y Acad Sci.* **1371**:45-54.

Pontano Vaiteș, L., Paulo, J. A., Huttlin, E. L., Harper, J. W. (2017). Systematic analysis of human cells lacking ATG8 proteins uncovers roles for GABARAPs and the /MON1 regulator C18orf8/RMC1 in macro and selective autophagic flux. *Mol Cell Biol.* doi:10.1128/MCB.00392-17.

Poteryaev, D., Datta, S., Ackema, K., Zerial, M., Spang, A. (2010). Identification of the switch in early-to-late endosome transition. *Cell.* **141**:497-508.

Richards, D. M., and Endres, R. G. (2017). How cells engulf: a review of theoretical approaches to phagocytosis. *Reports on Progress in Physics.* **80**:126601-126621.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics.* **18**:529.

Schindelin, J. Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods.* **9**:676-682.

Schroder, B., Wrocklage, C., Pan, C., Jager, R., Kusters, B., Schafer, H., Elsasser, H. P. *et al.* (2007). Integral and associated lysosomal membrane proteins. *Traffic.* **8**:1676-1686.

Settembre, C., Di Malta, C., Polito, V. A., Garcia Arencibia, M., Ventrini, F., Erdin, S., Edrin, S. U. *et al.* (2011). TFEB links autophagy to lysosomal biogenesis. *Science.* **332**:1429-1433.

Soding, J., Biegert, A., Lupas, A. N. (2005). The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**:W244-W248.

Tsui, K. C., Barfield, R. M., Fischer, C. R., Morgens, D. W., Li, A., Smith, B. A., Bertozzi, C. R. *et al.* (2019). Systematic identification of regulators of antibody-drug conjugate toxicity using CRISPR-Cas9 screens. *Antibiotics (Basel)*. **8**:18.

Verbaanderd, C., Maes, H., Schaaf, M. B., Sukhatme, V. P., Pantziarka, P., Sukhatme, V., Agostinis, P., and Bouche, G. (2017). Repurposing Drugs in Oncology (ReDO)—chloroquine and hydroxychloroquine as anti-cancer agents. *Ecancermedicalscience*. **11**:781.

Wang, C. W., Stromhaug, P. E., Shima, J., Klionsky, D. J. (2002). The Ccz1-Mon1 protein complex is required for the late step of multiple vacuole delivery pathways. *J Biol Chem*. **277**:47917-47927.

Xu, C. and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein Cell*. **2**:202-214.

Xu, J., Camfield, R., Gorski, S. M. (2018). The interplay between exosomes and autophagy - partners in crime. *J Cell Sci*. **131**:jcs215210.

Yasuda, S., Morishita, S., Fujita, A., Nanao, T., Wada, N., Waguri, S., Schiavo, G. *et al.* (2016). Mon1-Ccz1 activates Rab7 only on late endosomes and dissociates from the lysosome in mammalian cells. *J Cell Sci*. **129**:329-340.

Yousefian, J., Troost, T., Grawe, F., Sasamura, T., Fortini, M., Klein, T. (2013). Dmon1 controls recruitment of Rab7 to maturing endosomes in *Drosophila*. *J Cell Sci*. **126**:1583-1594.

Zhang, P., Fan, Y., Ru, H., Wang, L., Magupalli, V. G., Taylor, S. S., Alessi, D. R., Wu, H. (2019). Crystal structure of the WD40 domain dimer of LRRK2. *PNAS*, **116**:1579-1584.