Quantitative proteomic studies of fast endophilin-mediated endocytosis and motor neuron degeneration

Quantitative Proteomic studies of Fast Endophilin-mediated Endocytosis and Motor Neuron Degeneration

Shaan Subramaniam, MSci

Birkbeck, University of London

Institute of Structural and Molecular Biology

Thesis submitted for the degree of Doctor of Philosophy

Declaration

I hereby declare that all work presented here is my own. Where information has been derived from other sources, or is of a collaborative nature, I confirm that this has been explicitly indicated in the text.

Acknowledgements

I would like to thank my two amazing supervisors, Emmanuel Boucrot and Kostas Thalassinos, who have always pushed me and supported me. For their patience and generosity, I owe them more than I can say.

I also wish to extend a heartfelt thanks to my collaborators, Prof Michael Cheetham and Dr David Parfitt, whose support and samples made a huge difference.

To the wonderful crew of Darwin Office 106—ma cherie Thomas, Abu, Suniya, Charlie—you guys taught me so much and I will treasure our times together. To One, Michael and the various supervisors of the UCL Anatomy building stores, thank you so much for always lifting my spirits, I was lucky to be assigned stores duties and see you guys every day. To the 106-ers now long gone, who made me feel at home when I first got here, Claudi and Kieran, thank you! To Nathanael, for all the snacks. The juicy, juicy snacks. To Katelyn, thank my lucky stars I have a friend in you! I was also fortunate to have the advice and support of Sebastien Lane during the long coronavirus-induced wait for my viva.

Sid, Mahesh, Alex, Sophie and Sansa, words cannot express how much I love you guys and will fondly remember our Fairthorn times together. To the V-gang, I will always remain true. To JP, I owe a great debt. I would also like to thank my UCL undergrad support group: Zammy Fairhurst-Hunter, Lucy Nevard, Charlotte Flanagan, Magda Meier and Jack Munns (I guess).

And of course, to Ale, I wouldn't have got through any of it without you. You walked the hardest road of all of us, you are an inspiration.

Lastly, to my wonderful parents, for their example, their love - and especially their financial aid - I am deeply indebted.

Abstract

Quantitative proteomics is a powerful and versatile tool for the study of biological systems. In this work, quantitative proteomics was applied to the investigation of a novel endocytic pathway and to the characterisation of a murine model of motor neuron disease.

Fast endophilin-mediated endocytosis (FEME) is a recently discovered clathrin-independent endocytic route notable for its speed and its reliance on receptor-ligand interaction for cargo uptake. Active in a range of cell types, the pathway is known to be hijacked by bacterial toxins and is important in signalling and cell motility.

Incomplete knowledge of the receptors internalised by FEME limits understanding of the pathway's functional roles. Label-free quantitative mass spectrometry (MS) workflows were optimised to identify candidate FEME receptors. Plasma membrane extracts from mammalian cells were profiled by MS following genetic perturbation of different endocytic routes. This data was used to identify Junction Plakoglobin, Dysferlin and others as candidate FEME-associated proteins for future study. Progress was also made towards establishing an affinity-purification mass spectrometry (AP-MS) pipeline for probing plasma membrane extracts with FEME bait proteins.

Growth factor availability in cell media was shown to modulate FEME in mammalian cell culture. AP-MS experiments aimed at investigating how the endophilin interactome changes upon manipulation of growth factor supply identified several candidate FEME regulators, including GSK3β. Further studies showed that GSK3β, but not GSK3α, interacts with the endophilin SH3 domain under conditions of low growth factor signalling, shedding light on how kinases regulate FEME.

Finally, quantitative MS experiments were conducted to help elucidate the causes of neuron death in a murine model of motor neuropathy. Lumbar spinal cord extracts from mice lacking the co-chaperone HSJ1 were compared to their wild-type counterparts using in-gel tryptic digestion

followed by quantitative MS. These studies failed to identify differentially expressed proteins at the depth of coverage achieved, but network analysis of our qualitative data showed potential disruption of mTORC1 signalling and lipid homeostasis and identified several proteins as candidate HSJ1 clients including Gdap1, Ataxin-10 and Ubiquilin-1.

The datasets presented here serve as a resource for future studies of FEME and HSJ1-associated neurodegeneration.

Table of contents

List of Figures

List of Tables

List of Equations

Abbreviations

ANOVA	Analysis of variance
AP2M KD	Adaptor protein complex 2 μ2-subunit knockdown
APEX2	Ascorbic acid peroxidase-mediated proximity labelling (generation 2)
AP-MS	Affinity purification-mass spectrometry
ATP	Adenosine tri-phosphate
BAR	Bin1/Amphiphysin/Rvs
BioID	Proximity-dependent Biotin identification
CID	Collision-induced dissociation
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CME	Clathrin-mediated endocytosis
CMT2	Charcot-Marie-Tooth syndrome type 2
CSC	Cell surface capture
CV	Coefficient of variation
Da	Dalton
DAVID	Database for Annotation, Visualization and Integrated Discovery
dHMN	Distal Hereditary Motor Neuropathy
DDA	Data Dependent Acquisition
DDM	n-dodecyl- b-D-maltopyranoside
DIA	Data Independent Acquisition
DTT	Dithiothreitol
eGFP	enhanced green fluorescent protein
EndoA2	Endophilin A2
EndoTKD	Endophilin triple knockdown
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
FASP	Filter-aided sample preparation
FDR	False discovery rate
FEME	Fast endophilin-mediated endocytosis

FWHM	Full width at half maximum
GPCR	G-protein coupled receptor
GDP	Guanosine di-phosphate
GO	Gene Ontology
GST	Glutathione S-transferase
GTP	Guanosine tri-phosphate
HPLC	High performance liquid chromatography
IAA	Iodoacetamide
iBAQ	Intensity-based absolute quantification
IMS	Ion mobility separation
IP	Immunoprecipitation
iPSC	induced pluripotent stem cell
KD	Knockdown
KO	Knockout
LC-MS	Liquid chromatography-mass spectrometry
MALDI	Matrix-assisted laser desorption ionisation
MN	Motor neuron
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	mass-to-charge ratio
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCA	Principle component analysis
PLGS	Protein Lynx Global Server
PPI	Protein-protein interaction
PRM	Proline-rich motif
QTOF	Quadrupole Time of Flight
RPE	Retinal pigment epithelium
RLE	Relative log expression
RNAi	RNA interference

RTK	Receptor tyrosine kinase
ROS	Reactive oxygen species
SCA10	Spinocerebellar ataxia type 10
SD	Standard deviation
SDS	Sodium dodecyl sulphate
s.e.m	standard error of the mean
siRNA	small interfering RNA
SH3	Src homology 3
SMA	Spinal muscular atrophy
TAP	Tandem affinity purification
TBS	Tris-buffered saline
TIL	Third intracellular loop
THSD	Tukey's honest significant difference
TWIMS	Travelling wave ion mobility separation
UBA	Ubiquitin-associated domain
UBL	Ubiquitin-like domain
UPR	Unfolded protein response
UIM	Ubiquitin-interacting motif
WT	Wild-type

1. Introduction

1.1. Quantitative proteomics by Mass spectrometry

1.1.1. Mass spectrometry in the biological sciences

Mass spectrometry is an analytical technique which enables the characterisation of ionised molecules based on their mass-to-charge (m/z) ratio. A variety of mass spectrometer configurations exist, but all possess the following 3 features: a source where analytes are ionised and volatised, a mass analyser where analytes are separated on the basis of their m/z and a detector where analyte m/z values are recorded.

Due to their thermal instability and non-volatile nature, proteins were historically unsuited for analysis by MS. The development of two 'soft' ionisation techniques in the late 1980s and early 1990s – electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI) – enabled the production of ionised biomolecules in the gas phase without excessive damage to the sample, thus marking the advent of MS-based proteomics (Karas and Hillenkamp, 1988; Fenn *et al.*, 1989).

1.1.2. Electrospray Ionisation

In electrospray ionisation analytes in solution are infused into a narrow capillary under atmospheric pressure. An electric potential (3 – 6 kV) is applied between the tip of the capillary and a counter electrode. This causes the emission of a fine spray of solvent droplets, with accompanying analytes, from a Taylor cone at the probe tip (Taylor, 1964). Droplets (initially ~10 µm) rapidly shrink due to solvent loss by evaporation but retain their charge, resulting in droplet fission by "coulombic explosion" as the accumulated repulsive forces overcome the cohesive surface tension of the solvent (Figure 1.1 and Equation 1.1). This process produces smaller progeny droplets which are susceptible to undergoing further fission events. There is still debate

as to the exact mechanism of ion formation (Figure 1.2). The ion evaporation model (IEM) proposes that the build-up of charge in a Rayleigh-charged nanodroplet (< 10 nm radius) would be sufficient to eject solvated analyte ions of low molecular weight from the droplet surface (Iribarne, 1976). Under the charged residue model (CRM), the analyte remains stably hydrated within a charged droplet that evaporates to dryness. As the solvent shell is removed, its charge is transferred to the analyte. This has been shown experimentally for intact globular proteins between 5-500kDa in size (Iavarone and Williams, 2003; Kaltashov and Mohimen, 2005). Unfolded proteins appear to undergo ionisation by the chain ejection model (CEM). The hydrophobic nature of unfolded protein chains causes their migration to the surface in a Rayleigh-charged nanodroplet. A single chain terminus is expelled and the polymer is ejected in a stepwise fashion. Mobile protons are deposited on the analyte through charge equilibration (Ahadi and Konermann, 2012; Konermann, Rodriguez and Liu, 2012).

Whatever the mechanism, this process culminates in the formation of desolvated, multiply-charged analyte ions in gas phase. The tendency to form multiply-charged ions is particularly significant, as this improves instrument sensitivity and allows the use of mass analysers with a relatively poor mass range.

ESI performance can be improved by using narrow capillary tips (<10 µm) and very low flow rates (tens of nl min^{-1}). It is believed that the smaller initial droplet sizes produced by nanoESI require fewer fission/evaporation cycles before analytes are fully desolvated, resulting in increased sensitivity and ionisation efficiency (Wilm and Mann, 1996; Covey, Thomson and Schneider, 2009).

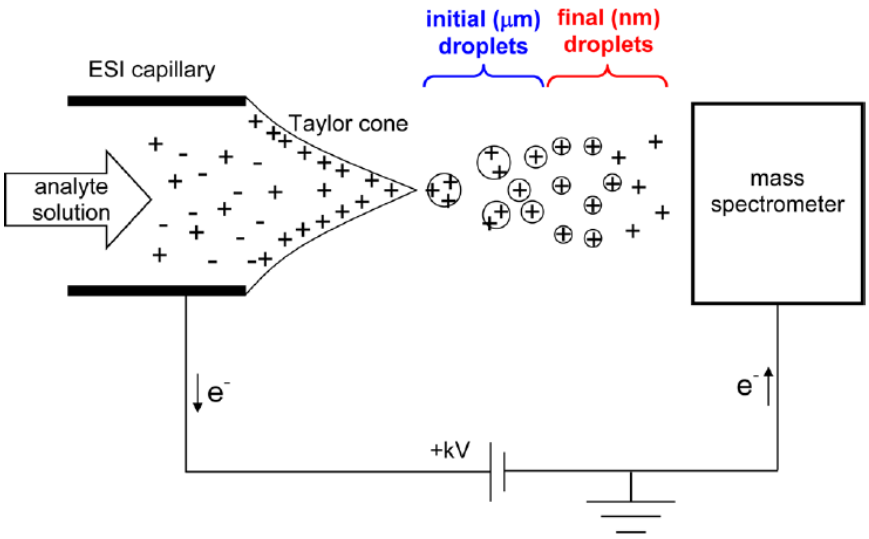

Figure 1.1 Schematic representation of an ESI source for generating positive ions. Oxidation of solvent at the capillary tip produces free protons. Under the influence of the electric field charge accumulates at the surface of the analyte solution and is deformed into a Taylor cone upon the application of sufficient voltage. Emmitted droplets reduce in size due to the combined forces of solvent evaporation and droplet fission. Analytes are eventually fully desolvated and enter the gas phase as multiply charged ions by a number of proposed mechanisms based on analyte properties.

$$z_R = \frac{8\pi}{e}\sqrt{\varepsilon_0 \gamma R^3}$$

Equation 1.1 The Rayleigh limit z_R describes the maximum number of charges *e* that a liquid droplet of given radius (*R)* and surface tension (γ) can hold. ε_0 describes the vacuum permittivity of the system. If charge accumulates above this limit, droplets undergo fission.

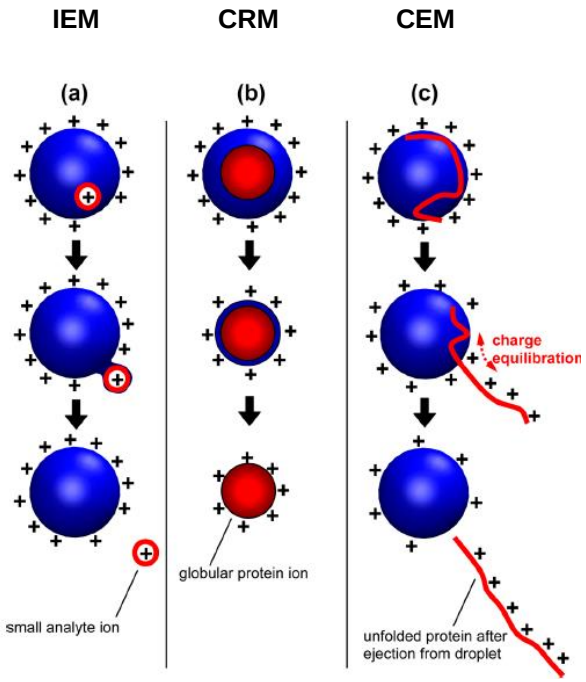

Figure 1.2 Experimental evidence and in-silico modelling support a number of ESI ionisation mechanisms, shown here in positive mode.

(adapted from Konermann et al., 2013)

1.1.3. Measuring mass-to-charge using a Time-of-flight instrument

In the field of proteomics, the Orbitrap and the Quadrupole-Time of Flight (QTOF) are the two most commonly used instrument configurations. In this report, MS data was collected using a NanoESI-QTOF instrument (Waters Synapt G2-Si). Ions entering the instrument first pass through a quadrupole analyser which uses electric fields emanating from a set of 4 electrodes to filter the sample, excluding ions outside of a given m/z range (see Figure 1.3). The ions pass through an ion mobility separation (IMS) cell between two collision cells, before entering the TOF analyser. Here, they are accelerated toward and bounced off a reflectron, which increases the distance travelled and helps coordinate the arrival of ions with the same m/z at the detector.

Time of flight analysers operate on the principle of measuring the time taken for an ion to travel from an ion source to a detector. Ions from an ion source are accelerated towards a detector by a pulse V, following which they enter a field free region (maintained at high vacuum). All ions receive the same kinetic energy in the acceleration region, which is expressed by:

Equation 1.2

$$qV = zeV = E = \frac{1}{2mv^2}$$

Where q = total ion charge, V = push voltage, z = ion charge, e = electron charge, E = total energy and v = ion velocity.

The time taken to traverse the field free region is given by:

Equation 1.3

$$t = \frac{s}{v}$$

Where t = time, s = distance and v = ion velocity. Substitution of v with Equation 1.2 gives:

Equation 1.4

$$t = \frac{s}{\sqrt{\frac{2ezV}{m_i}}}$$

Rearrangement of Equation 1.4 reveals the relationship between time taken for an ion to reach the detector and its m/z:

Equation 1.5

$$t = \frac{s}{\sqrt{2eV}}\sqrt{\frac{m_i}{z}}$$

Given that $\frac{s}{\sqrt{2eV}}$ is a constant, we see that *t* and m/z are proportional. The mass analyser can be calibrated with reference to a compound of known m/z to calculate the constant in Equation 1.5.

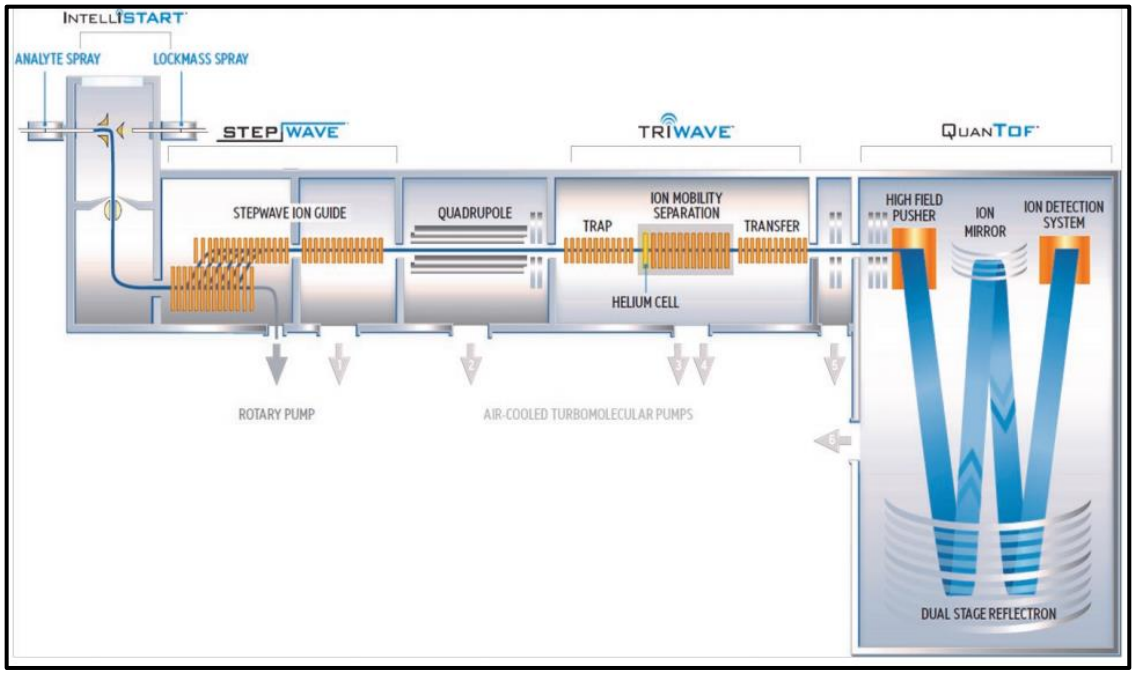

Figure 1.3 Schematic of a Waters Synapt G2-Si QTOF instrument used in this work (Image from Waters Corp.).

1.1.4. Tandem mass spectrometry and Collision induced dissociation

Tandem MS refers to the collection of mass spectra for a single analyte before and after a dissociation event, yielding information about the composition of the analyte under study. CID is a fragmentation process during which ions analytes are accelerated through a chamber that contains inert gas molecules (e.g. H_2, Ar or N_2). Collisions between the analyte and the gas molecules can cause fragmentation within the analyte. Because CID tends to cause fragmentation of the weakest bonds within an ion between analysed, peptides undergoing CID

tend to fragment at the amide bond. This produces a convenient series of so-called b- and y-ions, from which the original sequence of a peptide can be assembled, given that a sufficient range of peptide fragments are measured.

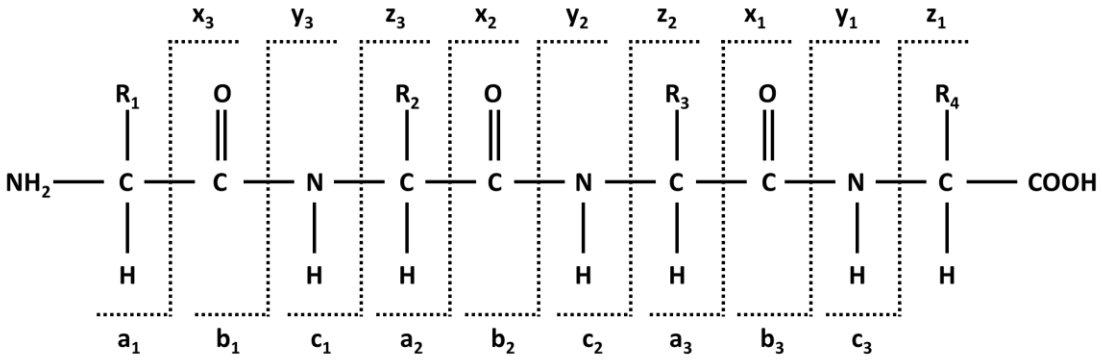

Figure 1.4 Fragmentation ion nomenclature for peptides.

The Roepstorff-Fohlmann-Biemann (Roepstorff and Fohlman, 1984; Biemann, 1992) product ion nomenclature is based on which side of the fragmented analyte retains charge. Retention of charge on the N-terminal fragment results in b-ion formation, whilst retention of charge on the C-terminal fragment results in y-ion formation. The ions are numbered in sequence from the corresponding N- or C-terminus as shown in Figure 1.4.

The use of a protease with well-defined specificity, such as trypsin (which only cuts at lysine and arginine residues C-terminally, but not before proline) to generate peptides for analysis by tandem MS simplifies the search parameters considerably.

1.1.5. Proteomic workflows

The goal of bottom-up proteomics is the identification and quantification of proteins in a complex sample based on their mass spectra. A typical bottom-up workflow involves sample clean-up followed by enzymatic digestion. The resultant peptides are then analysed by tandem mass spectrometry and their m/z spectra recorded. Data processing software then compare these recordings to a database of theoretical peptide spectra, generated *in silico* from genomic data. Matches are assigned based on correlation between the experimental dataset and the theoretical dataset, resulting in a 'hit list' of statistically significant protein groups. Features of the collected mass spectra such as spectral counts or ion intensity can be used to estimate protein quantity in the original sample.

1.1.6. Peptide separation in proteomics experiments by liquid chromatography

A typical biological sample exhibits differences in protein abundance in the region of 6 orders of magnitude, whilst very complex samples such as blood plasma can exhibit protein abundances spanning 9 orders of magnitude. This level of complexity currently exceeds the dynamic range of LC-MS instrumentation (Zubarev, 2013). A solution to this problem is to partition the sample following enzymatic digest on the basis of peptide physicochemical properties.

ESI is particularly suitable for the analysis of complex mixtures as it can be readily coupled to liquid chromatography (LC) (Hunt *et al.*, 1992). The most common form of LC used in the context of MS proteomics is reversed-phase LC. In this form of chromatography, peptides produced by enzymatic digestion are loaded onto an analytical column containing C18 stationary phase. The proportion of non-polar solvent in the mobile phase is gradually increased, with polar peptides eluting faster than non-polar ones. This allows for pre-separation of the sample of the basis of hydrophobicity, improving peak capacity, peptide resolution and sensitivity.

An important innovation which facilitated highly reproducible analysis of complex biological samples was the commercial availability of chromatography systems capable of operating at backpressure limits > 15,000 psi. This allowed the use of chromatographic columns packed with sub-2µm particles, dramatically increasing peak capacity and resolution (Nováková, Matysová and Solich, 2006).

1.1.7. Ion-mobility-assisted MS^E data acquisition

Data-dependent acquisition (DDA) is the traditional mode of data acquisition in bottom-up proteomics experiments. A survey scan is conducted using a wide m/z range to identify prominent m/z peaks eluting at a given time from the chromatographic column. A selection of these (usually the Top 10 – 30 most intense) is selected for tandem MS/MS. The ions are isolated with a quadrupole or an ion trap mass analyser and fragmented to obtain simplified precursor and product ions. This cycle of survey, selection and tandem MS is repeated throughout the chromatography gradient to acquire tandem MS/MS data for all peptides from the sample. Limitations of this approach include the bias towards abundant peptides and the reliance on a survey scan to select ions to fragment. Variability in chromatographic performance can shift survey scan timings away from a peptides chromatographic apex in repeated runs, resulting in the loss or gain of peptides with intensities near the threshold for selection, even when analysing the same sample.

Data-independent modes of acquisition were developed as an unbiased alternative to DDA methods by removing the requirement for peptide pre-selection. The MS^E data acquisition mode developed by Silva and colleagues (Silva *et al.*, 2005, 2006) is a form of data-independent tandem mass spectrometry (MS/MS) suitable for peptide identification from a complex sample. Under MS^E, the instrument rapidly and continuously switches between a high and low energy mode of operation. In the low energy mode, ions are accelerated through the Trap collision cell at low kinetic energy (4 eV), resulting in transmission of ions through the cell with no fragmentation. In the high energy mode, an energy ramp (15-35 eV) is used to facilitate collision induced dissociation (CID) of the analyte through high energy collisions with the Argon atoms in

the collision cell. Mass spectra from the low energy mode represent intact precursor ions (MS1), with spectra acquired in high energy mode representing their fragmented counterparts (MS2). Cycling between both modes is usually completed in just over a second. This high scan rate ensures peptides species are not lost between scans.

Ion mobility separation is an analytical technique whereby ionised analytes are separated in a drift cell on the basis of their shape and size. The drift cell is filled with inert gas molecules which impede the transit of the analyte. As larger/ more extended molecules experience more collisions, they take a longer period of time to traverse the drift cell. A variant of this method known as travelling wave ion mobility separation (TWIMS) can be implemented in the IMS cell of the Waters Synapt G2-Si. Ions are accelerated through the IMS cell by a series of stacked ring ion guides (SRIG). Collisions with a neutral gas (typically N_2 maintained at 0.2 mBar) facilitate separation of analytes over a millisecond timescale.

Peptides undergoing UPLC separation elute over a chromatographic width in the order of seconds, whilst TOF measurements are completed in a µs timescale. IMS and can therefore be easily incorporated into bottom-up proteomics workflows as an additional chromatographic dimension. IMS-assisted MS^E data acquisition is termed High-definition MS^E ($HDMS^E$) and conveys a variety of advantages when analysing complex samples. $HDMS^E$ workflows display up to 60% better proteome coverage in comparison to MS^E alone. This effect is attributed to several factors including measurement of IMS drift times enabling better matching between precursor and fragment ions during database searching, improved fragmentation of peptides, and an increased peptide detection dynamic range, particularly for lower abundance peptides. An additional factor which boosted identifications is the ability to separate isobaric species in the IMS cell prior to fragmentation on the basis of their shape (Geromanos *et al.*, 2012; Shliaha *et al.*, 2013).

1.1.8. Database searching

Making sense tandem mass spectra requires significant computational investment. A number of MS search algorithms automate data processing to obtain monoisotopic peaks for each peptide/ peptide fragment, which are then aligned by retention time to assign hits with reference to a database of protein sequences in the FASTA file format (Perkins *et al.*, 1999; Craig and Beavis, 2004; Geer *et al.*, 2004; Cox *et al.*, 2011). For MS^E data produced by Waters instruments, this is performed by the Protein Lynx Global Server (PLGS) (G. Z. Li *et al.*, 2009; Geromanos *et al.*, 2009) or Progenesis software packages.

Because MS^E data is acquired continuously, precursor and fragment ions must be matched following data acquisition. PLGS is a database search tool which aligns precursor and product ions by retention time to reduce the complexity of the data. Tentative precursor/product associations are then matched to peptides and proteins in a 3-stage process using an iterative ion depletion strategy. Briefly, PLGS implements 2 algorithms to pre-process data after lockmass correction. Apex3D performs noise subtraction and integrates ion current across the chromatographic profile of all putative ions. The algorithm also identifies the chromatographic peak apex, which serves as the ion's retention time. The Pep3D algorithm then de-isotopes and performs charge-state reduction on all features. This process associates the combined ion current from all isotopic peaks and all instances of the same peptide at a higher charge state with a single $[M+H]^{1+}$ monoisotopic mass. This generates an exact mass and retention time (EMRT) for all precursor and product ions.

After ion detection, retention time alignment of precursors and products is performed. Precursor and product ions with closely matching retention times (product RT ±10% of the FWHM chromatographic peak width of precursor RT) are putatively associated. Ion mobility drift times can also be used to align precursor and product ions if data were acquired in $HDMS^E$ mode. Product ions can be associated to multiple precursors at this stage.

EMRT clusters are now submitted to the database search algorithm. The user provides a database of protein FASTA sequences which is digested *in-silico* to generate a list theoretical peptide masses to be queried. Database searching is performed in three stages ("passes").

During Pass 1, EMRT clusters are matched to fully tryptic peptides (with fixed modifications) on the basis of exact mass. Peptides matches are putatively associated with proteins, and these matches are scored using physicochemical models which account for peptide retention time, preferred fragmentation sites and number of product ion spectra for a protein of given mass and intensity among other parameters (G. Z. Li *et al.*, 2009). Proteins are ranked by score and the highest ranked protein is considered identified. This protein and its top-ranked peptide/product EMRT clusters are removed from consideration from subsequent iterations of the search (ion depletion). The database search process is then repeated until one of two conditions is met: none of the proteins identified have scores higher than the minimum threshold or, the user-set false discovery rate (FDR) is reached. FDR is controlled empirically by using a randomised/ reversed (decoy) database. When the number of hits from the decoy database exceeds the user-set limit, the iterative search is stopped.

During Pass 2, the search space is expanded to identify modified forms of the proteins identified in Pass 1. These modifications include missed cleavages, post-translational modifications (PTMs), oxidation of methionine, neutral losses, acetylation, deamidation and in-source fragmentation. Finally, a third search (Pass 3) is implemented which can identify highly labile peptides which are susceptible to in-source fragmentation. The search parameters are expanded further to include peptides with multiple modifications. The search iterates until the user-specified FDR is reached.

Progenesis is a database search software which uses the PLGS search engine to perform EMRT matching to a FASTA database but incorporates novel steps in the ion detection (peak picking) part of the algorithm. Progenesis converts data from MS^E or $HDMS^E$ experiments into a 3D map with axes representing m/z, retention time and ion intensity. The software selects a reference run

on the basis of similarity to all other runs. The retention times of ions in all runs are adjusted in a non-linear manner to align them to the reference run. This is achieved by placing a series of landmarks called vectors on selected ions. Vectors are representations of the distance between ion retention times in the reference run and the run being aligned.

During automatic alignment, the software first performs adjustments to vector retention times to bring the two runs into general alignment. Next, vectors are adjusted in smaller increments to finely optimise the alignment.

Following retention time alignment, Progenesis creates an aggregate dataset of all runs. Ions detected in any given run are searched for in all other runs in a process analogous to the "match between runs" feature in the MaxQuant software package (Cox *et al.*, 2014). This process confers greater sensitivity to the workflow by using information from all runs during peak picking. A further feature is the elimination of missing data as an ion current is measured even where peaks are not present, facilitating downstream statistical analysis.

Selected ions are converted into EMRTs using the Apex3D and Pep3D algorithms as detailed above for PLGS. EMRTs are matched to peptides and proteins from a FASTA database using the PLGS database search algorithm.

1.1.9. Quantitation by mass spectrometry

The relationship between amount of analyte in a sample and signal intensity in MS is not straightforward and mass spectrometers are thus inherently poor quantitative instruments. Numerous innovations have enabled quantitative measurements using MS which can be divided into label-based and label-free methods.

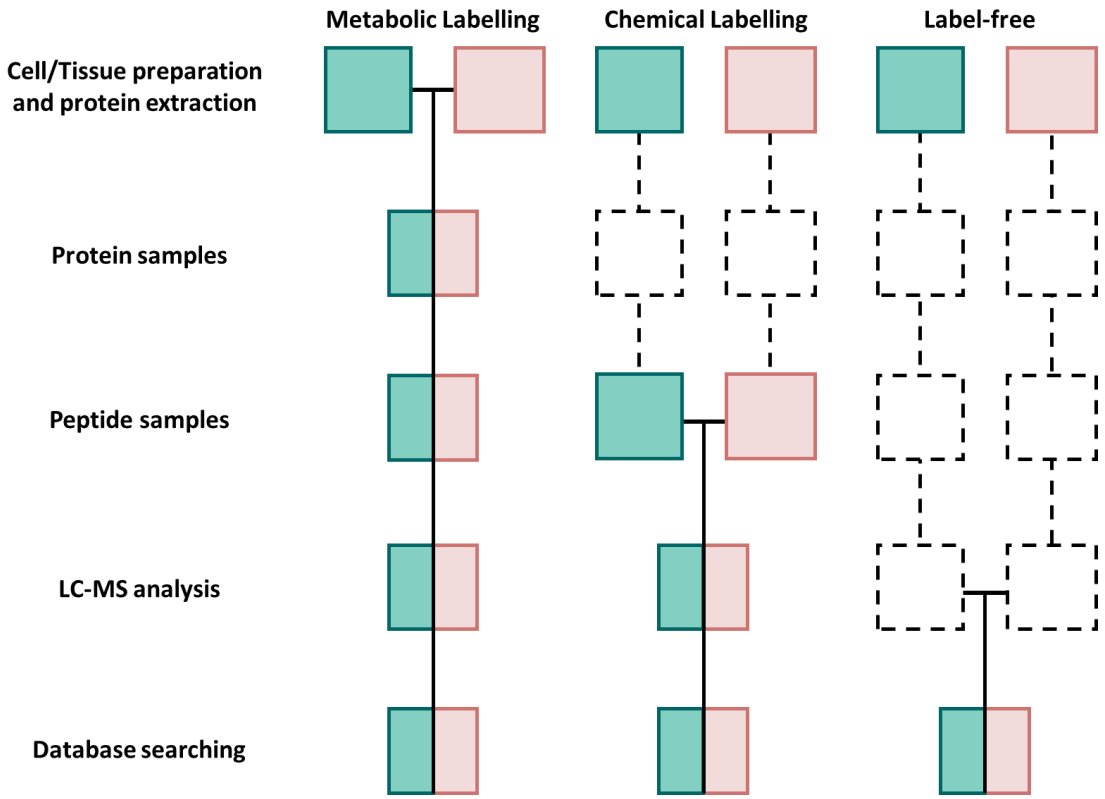

Figure 1.5 Three common quantitative proteomics workflows. Coloured boxes indicate two experimental conditions. Horizontal lines indicate when samples are combined. Empty boxes indicate where samples are kept separate during the workflow.

Label-based methods involve the addition or incorporation of a chemical label to distinguish proteins/peptides from different experimental conditions, allowing relative or absolute quantitation of proteins on the basis of differences in MS1 or MS2 spectra. Stable isotope labelling with amino acids in cell culture (SILAC) was an early development in this field and gained widespread popularity (Ong *et al.*, 2002). In this technique, cells from one experimental condition are grown in media containing heavy isotope-labelled amino acids (e.g. C^{13} Lysine or

C^{13} Arginine), such that virtually all the proteins in cells from this condition have incorporated heavy amino acids. Following cell lysis, protein extracts are mixed and undergo enzymation digestion and MS analysis together (Figure 1.5). The distinctive difference in m/z introduced by stable isotope labelling allows confident identification of peptides from the same protein and quantitation of their relative abundance in MS1 scans. Because almost all experimental steps are performed under the same conditions, the technique displays excellent analytical precision through minimisation of handling errors (e.g. pipetting errors) and run-to-run variability in instrument conditions. Initially restricted to experiments performed on mammalian cell culture, the technique has now been successfully expanded to include in vivo labelling of many animal species (Sowell *et al.*, 2007; Larance *et al.*, 2011; Konzer *et al.*, 2013). Although a version of the technique which enabled comparisons of 5 different conditions simultaneous was quickly developed (Blagoev *et al.*, 2004), in practice the complexity introduced in MS1 spectra upon multiplexing, and the high cost of reagents often limits comparisons to duplex or triplex experiment designs.

Isobaric labelling techniques are alternative label-based methods which feature MS2-based quantitation. The most popular versions of this method are Tandem Mass Tag (TMT) and isobaric tag for relative and absolute quantitation (iTRAQ) (Ross *et al.*, 2004; Dayon *et al.*, 2008). These techniques feature labelling of samples from different conditions with different isobaric chemical tags, generally conducted post digestion. These tags are composed of a reactive group which facilitates covalent attachment to peptides (e.g. via reactions with primary amines on the N-terminus or on Lysine residues), a reporter group of varying mass, and a balancer group which ensures that all tags are of equivalent total mass. Following labelling, peptides are mixed and co-analysed by MS (Figure 1.5). Because all tags are isobaric, peptides from different conditions are detected at the same m/z during MS1 and are selected for fragmentation by the mass analyser. However, because the reporter group is attached to the tag by a labile bond, these are released upon fragmentation by CID during MS2 acquisition. The relative intensities of the reporters in the MS2 scan reflect the relative intensities of each peptide in the different conditions under investigation enabling quantitation. Isobaric labelling techniques boast two major advantages. Firstly, these techniques display very high technical precision, often outperforming

SILAC-based workflows (Sonnett, Yeung and Wühr, 2018). Secondly, they offer the highest degree of multiplexing currently possible with label-based methods, with simultaneous analysis of up to 11 different conditions the same time with TMT tags (Pappireddi, Martin and Wühr, 2019). This high degree of multiplexing represents a major saving in instrument time, especially in comparison to label-free methods. The main weakness of isobaric labelling techniques is their reliance on isolation of single peptides for MS2 analysis. Current instrumentation is unable to isolate peptides in a mass window narrower than ~0.5 m/z. Consequently, MS2 scans often isolate more than one peptide for fragmentation. Often, this has been shown to cause a "ratio distortion" effect, bringing the measured ratios of reporter tags in a given MS2 scan close to 1:1. This problem has been addressed in newer instrumentation through implementation of an MS3 scan which provides greater selectivity for peptides of interest, but remains a major technical hurdle (McAlister *et al.*, 2014).

Label-free quantitative proteomics is enjoying a surge in popularity. These techniques do not rely on prior labelling steps, but instead use intrinsic features of MS spectra to provide relative and even absolute quantitation of protein abundances between samples analysed separately (Figure 1.5). The earliest attempts to glean quantitative data from non-labelled samples utilised spectral counts (Liu, Sadygov and Yates, 2004; Old *et al.*, 2005). This technique was based on the observation that the total number of spectra featuring peptides assigned to a given protein was roughly proportional to its abundance. A popular iteration of this method known as Exponentially Modified Protein Abundance Index (emPAI) takes the number of tryptic peptides produced by a protein into account (Ishihama *et al.*, 2005). Whilst easy to implement, the method never performed comparably to label-based techniques.

By contrast, the quality of quantitative data obtained with label-free methods which rely on comparisons of ion intensity has improved enormously in recent years. This is due in no small part due to technological factors. These techniques typically rely on quantifying ion intensity on the basis of integration of the area under a chromatographic peak, or by finding the peak apex. Improvements to chromatographic instrumentation and mass analyser resolving power and especially the maturation of LC-nanoESI has greatly improved the comparability of analyte ion

intensity in separate analytical runs. The smaller inital size of droplets generated by nanoESI is believed to lower the contribution of ionic matrix components (e.g. salts) to variability in ESI efficiency. This has greatly reduced the imperative to perform quantitative MS experiments on peptides in the same run.

Intensity-based absolute-quantification (iBAQ) is a technique for absolute quantitation of proteins by label-free MS which uses the summed extracted ion chromatograms (XICs) of all peptides for a protein and normalises this to the number of peptides theoretically generated by enzymatic digestion to generate an abundance score (Schwanhüusser *et al.*, 2011). This technique was used to generate protein copy numbers/cell where the original cell number was known. The MaxLFQ algorithm developed by the Mann group also features the use of summed XICs from all peptides assigned to a protein to calculate protein abundances suitable for comparison with samples analysed in separate runs (Cox *et al.*, 2014).

An alternative method of label-free method of protein quantification known as Top3/ Hi3 was described by Silva and colleagues (2006). The authors showed that the average intensity of the 3 most intense peptides for a given protein correlates with its concentration in the source sample (Silva *et al.*, 2006). This property can be used to assign absolute concentrations for all proteins in a sample with reference to a spiked-in standard of known concentration and displays an average CV of 10-15% (Silva *et al.*, 2006). Importantly, Hi3 can only be used for proteins with at least 3 unique peptides, and measurement error increases for proteins with < 5 peptides (Silva *et al.*, 2005; Dator, Gaston and Limbach, 2014). The performance of Top3 and iBAQ were assessed using a dataset of 27 different human tissues and bodily fluids, representing an MS-derived map of the human proteome. The methods performed comparably, but Top3 displayed a slightly worse quantitative error in protein abundance among proteins with < 5 peptides, in agreement with the other reports (Wilhelm *et al.*, 2014).

Overall, label-free methods still suffer lower quantitative precision and place heavier demands on instrument time than label-based methods owing to the fact that samples are analysed separately. However, they avoid costly and laborious labelling steps and do not have intrinsic

limits on the number of conditions which can be compared. In this work, the ability to identify and quantify proteins from small amounts of complex samples through quantitative label-free LC-MS has been applied to the study of endocytosis and neurodegeneration.

1.2. Endocytosis and fast endophilin-mediated endocytosis

1.2.1. Endocytosis in eukaryotic cells

Endocytosis is the process by which cells internalise extracellular molecules and surface plasma membrane components. Internalisation is achieved through the invagination of portions of the plasma membrane, along with receptors and their ligands, which are then sealed off in intracellular compartments called vesicles and released for onward trafficking. Endocytosis is a key component of many cellular processes including nutrient uptake, signalling, growth, mitosis and motility (Doherty and McMahon, 2009).

Several pathways of endocytosis exist, delineated by vesicle size, target ligands and the protein and lipid factors required for vesicle maturation. Of these, clathrin-mediated endocytosis (CME) is the best-studied, while the so-called clathrin-independent endocytosis (CIE) pathways have proven more difficult to characterise.

CME is constitutive in all known eukaryotic cell types and does not strictly require receptor activation for internalisation. In mammals, CME cargoes are assembled into clathrin-coated vesicles 60-200nm in diameter over the course of ~1 minute (Gaidarov *et al.*, 1999; Kaksonen and Roux, 2018). Separation of the maturing vesicle from the plasma membrane (termed “scission”) is dependent on the action of the mechanochemical enzyme dynamin (Kosaka and Ikeda, 1983), which is recruited to the neck of the forming vesicle by Bin/Amphiphysin/Rvs (BAR) domain containing proteins (Sundborger *et al.*, 2011). Uptake of the largely CME-specific transferrin receptor is often used as a proxy for CME efficiency (Motley *et al.*, 2003).

A number of distinct CIE pathways have been described but the predominance of CME in most cells has complicated efforts to study them. Experiments involving deletion or knockdown of CME components tend to cause disruption to cellular function, resulting in dramatic compensatory changes in clathrin-independent pathways. A lack of pathway-specific cargoes (such as

transferrin for CME) further obscures the picture, resulting in a much poorer understanding of the functional role of CIE.

1.2.2. A novel clathrin-independent endocytic pathway

Perturbation of the BAR domain protein Endophilin A causes defective endocytosis in flies, worms and mice (Guichet *et al.*, 2002; Verstreken *et al.*, 2002; Schuske *et al.*, 2003; Milosevic *et al.*, 2011). Endophilin was also found to recruit the endocytic proteins dynamin and synaptojanin (Ringstad, Nemoto and De Camilli, 1997). The protein was initially assigned as a component of the CME pathway, after it was observed to colocalise with clathrin-coated pits in the early stages of the CME pathway (Taylor, Perrais and Merrifield, 2011) and was also found to play a role in uncoating of clathrin-coated vesicles during the latter stages of CME (Milosevic *et al.*, 2011). Further studies, however, argued against a central role for endophilin in CME. For instance, endophilin was only detected in ¼ of early-stage clathrin-coated pits (Taylor, Perrais and Merrifield, 2011), and knockdown of all three Endophilin A genes (A1, A2 and A3) did not have any effect on the internalisation of transferrin (Meinecke *et al.*, 2013). Interestingly, endophilin was found to bind directly or indirectly to several membrane receptor proteins (Tang *et al.*, 1999; Soubeyran *et al.*, 2002).

Investigating these interactions led Boucrot *et al*. (2015) to the discovery of a novel clathrin-independent pathway of endocytosis, marked and controlled by endophilin. This endocytic route is characterised by the rapid formation of endophilin-dependent, clathrin-independent tubular vesicles upon receptor activation, and was dubbed fast endophilin-mediated endocytosis (FEME). Activation of a variety of receptors was tested for the ability to stimulate FEME, including the β1-adrenergic receptor (β1-AR). Stimulation with the β_1-AR agonist denopamine caused accumulation of endophilin at clathrin-free zones of the leading edge (front of a migrating cell) of BSC1 cells and subsequent formation of β1-AR-positive endocytic tubules ~1µm in size. These tubules then rapidly (<7 secs) proceeded to the interior of the cell. For all tested FEME receptors, disruption of clathrin and/or the central CME adaptor protein AP-2 had no effect on cargo uptake, demonstrating the clathrin-independent nature of this pathway. Furthermore, receptor activation

(through agonist stimulation) was found to be an absolute requirement for FEME activity. Small molecule and genetic manipulation established that the budding of endophilin-positive tubules from the plasma membrane was dependent on the activity of dynamin, Rac, phosphatidylinositol-3-OH kinase (PI3K), PAK1 and actin polymerization.

A molecular mechanism by which FEME events occur has been elucidated. The inhibition of FEME in class I PI3K-deficient cells, coupled with the prominence of FEME at the cell's leading edge suggested a role for phosphatidylinositol (PtdIns) patterning in FEME initiation (Boucrot *et al.*, 2015). The leading edge of a migrating cell is the area of the plasma membrane which is being extended outwards in the direction of migration, and thus features active cytoskeletal reorganisation (Ridley, 2011). In lammellipodia of growing cells, the inner leaflet of the plasma membrane is enriched in PtdIns(3,4,5)P_3 through the action of class I PI3Ks downstream of receptor activation (Servant *et al.*, 2000; Srinivasan *et al.*, 2003; Nalbant *et al.*, 2004; MacHacek *et al.*, 2009). This results in local activation of the RhoGTPase Cell division control protein 42 homolog (Cdc42), which recruits the F-BAR domain proteins Cdc42-interacting protein 4 (CIP4) and Formin-binding protein 1 (FBP17) via their Ras exchanger motif (REM) domains (Chan Wah Hak *et al.*, 2018). These, in turn, recruit the 5' phosphatases Src homology 2 (SH2) domain containing inositol polyphosphate 5-phosphatase 1 (SHIP1) and Src homology 2 (SH2) domain containing inositol polyphosphate 5-phosphatase 2 (SHIP2) to the plasma membrane where local activity of SHIP1/2 converts PtdIns(3,4,5)P_3 into PtdIns(3,4)P_2 (Chan Wah Hak *et al.*, 2018). This promotes the CIP4/FBP17-assisted recruitment of lamellipodin, a scaffolding protein which contains at least 10 proline rich motifs (PRMs) suitable for targeting by Endophilin A SH3 domains (Vehlow *et al.*, 2013). This serves to cluster multiple Endophilin A proteins at the leading edge, ready to initiate FEME upon receptor activation. In the absence of receptor activation these endophilin clusters disassemble quickly (within a few seconds) through the action of GTPase activator proteins (GAPs) which stimulate Cdc42's GTPase activity and thus exert tight control over FEME events (Chan Wah Hak *et al.*, 2018). This priming cycle is summarised in Figure 1.6.

In the presence of activated FEME cargo receptors, Endophilin A is stabilised and can oligomerise on the plasma membrane, bending it to form tubular invaginations (Boucrot *et al.*,

2015). FEME cargoes are sorted into these tubules through direct or indirect contacts between the Endophilin A SH3 domain and receptor cytoplasmic tails. Endophilin recruits the GTPase dynamin to constrict the vesicle neck and separate it from the plasma membrane. Endophilin-coated FEME carriers are then transported into the cell interior through the action of Dynein as summarised in Figure 1.7 (*Ferreira et al, submitted*).

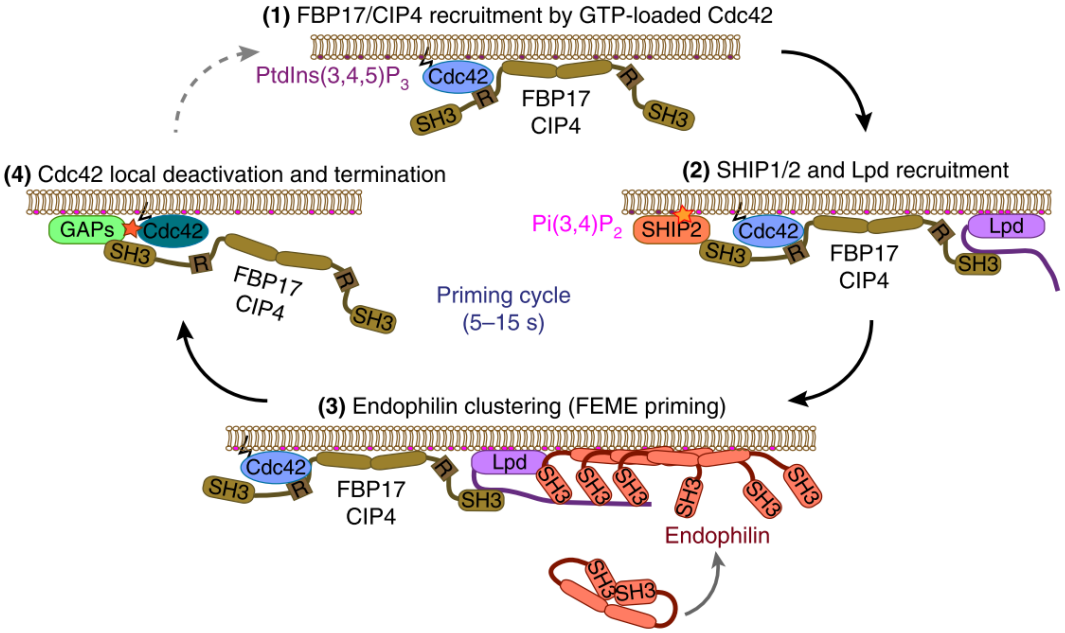

Figure 1.6 The FEME priming cycle. A model summarizing the priming cycle of FEME in resting cells: active GTP-loaded Cdc42 recruits FBP17 and CIP4 through their REM domains (step 1). FBP17 and CIP4 cluster the 5' -phosphatases SHIP1 and SHIP2 as well as lamellipodin through their SH3 domains (step 2). Lamellipodin is further stabilized by PtdIns(3,4)P_2 , which is locally produced by SHIP1/2. Endophilin is recruited and concentrated by lamellipodin (step 3). Pre-enriched endophilin mediates prompt FEME carrier formation upon cargo activation. In the absence of cargo activation, the FEME priming complex disassembles (step 4) upon local Cdc42 deactivation by GTPase-activating proteins (GAPs).

(adapted from Chan Wah Hak et al., 2018)

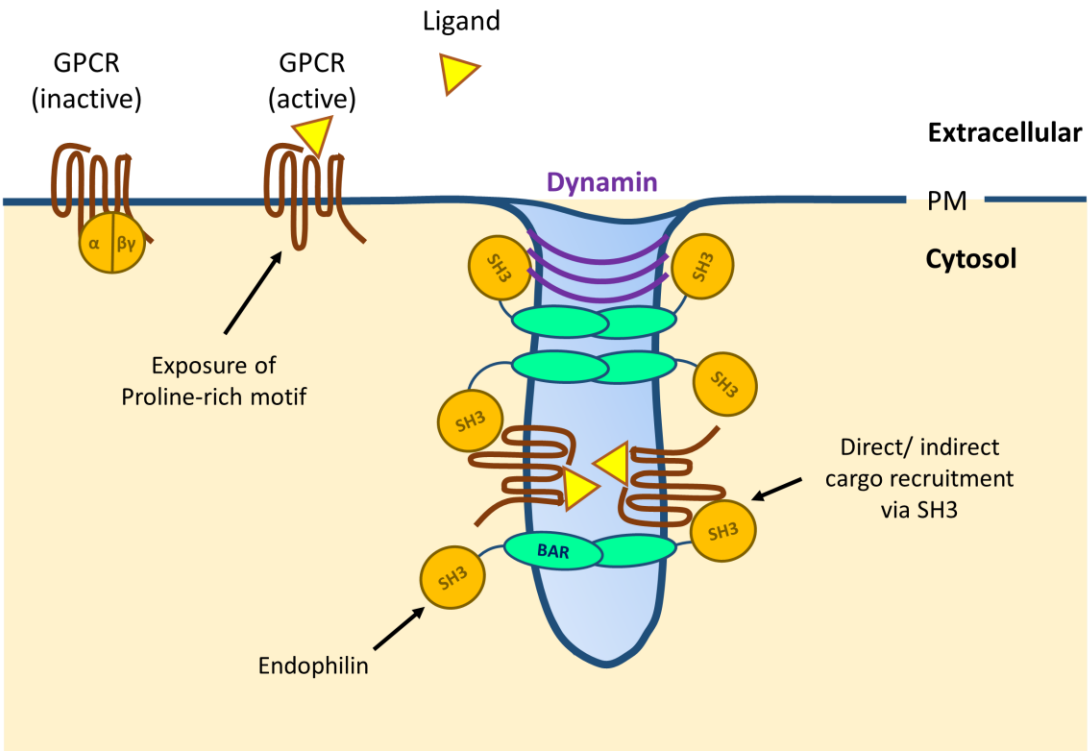

Figure 1.7 FEME carrier budding after receptor activation. FEME cargo activation by extracellular ligands, here illustrated in the case of a GPCR, facilitates direct engagement by the endophilin SH3 domain. In the case of RTK cargoes, this engagement is indirect. In the presence of active cargo, Endophilin dimers can oligomerise on the inner face of the plasma membrane (PM) and bend it to form tubular invaginations. Cargoes are sorted into the nascent FEME carriers. Endophilin recruits dynamin to the neck of mature tubules to trigger scission.

1.2.3. What is the functional role of FEME?

At the time of writing, the FEME pathway is known to internalise 15 receptors including a range of G-protein coupled receptors (including α_{2a}- and β_1-adrenergic, muscarinic acetylcholine receptor 4 and dopmaninergic D3 and D4 receptors), receptor tyrosine kinases (including Hepatocyte growth factor receptor and Epidermal growth factor receptor) and the Interleukin 2 receptor (Table 3.1). With such a broad range of cargoes, FEME is likely to play a role in many cellular processes.

Renard *et al*. (2015) have demonstrated that the pathway is subverted by the clathrin-independent cargoes Shiga and cholera toxin to gain entry into cells. Notably, the membrane deformation caused by Shiga toxin appears to be sufficient to induce FEME without receptor activation. It is probable that a host of pathogens and toxins gain entry to cells by exploiting this pathway.

Examining the dynamics of this novel endocytic route in a range of cell types will greatly inform our understanding of cell function. For example, it was reported that an ultrafast endophilin-dependent, clathrin-independent endocytic pathway is favoured over CME during the bulk uptake of synaptic vesicles in highly stimulated hippocampal neurons (Kononenko *et al.*, 2014). Excitingly, this phenomenon may be the neuronal manifestation of FEME, where the pathway might play a role in neurite outgrowth and function. In the broader field of endocytosis, the relative contribution of CME and CIE pathways to overall endocytic flux has been a subject of some contention (Howes, Mayor and Parton, 2010; Bitsikas, Corrêa and Nichols, 2014). It will be interesting to see how the discovery of FEME factors into this debate, ultimately shedding light on the reason why distinct endocytic pathways exist at all.

A study of the role of FEME in disease is also warranted. Internalised receptors are often recycled back to the cell surface in clathrin-mediated endocytosis. This tendency becomes problematic if cycling receptors are being expressed at oncogenic levels. If FEME cargoes are found to be target cargoes to a degradative fate, forceful downregulation of oncogenic signalling by

upregulating FEME activity could be a viable cancer therapy. In combatting pathogens which hijack the pathway like Shiga and cholera toxins, inhibition of FEME may be a worthwhile intervention. Another promising aspect of FEME is its dependence on ligand stimulation for activity. This feature could prove advantageous for specifically targeting drugs to tumours or infected cells.

1.2.4. Research questions and strategy

There are many questions to be addressed in the newly discovered field of FEME. A major hindrance in the analysis of CIE is a dearth of marker cargoes. The identification of a FEME-specific cargo receptor will greatly benefit research in this field. The normal human hTERT-immortalised cell line RPE1 was chosen as a model system for this work, as it displays a normal diploid karyotype and adopts a stretched morphology ideal for recording endocytic events via microscopy. Furthermore, FEME has been well characterised in this human cell line from past work.

Owing to the disparate nature of the analyses undertaken in the course of this work, additional introductory information is provided at the start of each results chapter.

2. Materials and methods

2.1. Cell culture and siRNA knockdown

2.1.1. Mammalian cell culture

Normal human epithelial h-TERT-RPE1 cells (ATCC, CRL-4000) were maintained at 37°C, 5% CO2 in DMEM:F12 HAM (1:1 v/v; Sigma), 0.25% sodium bicarbonate (w/v; Sigma), 1mM GlutaMax-I (Gibco), 10% foetal bovine serum (FBS, Life Technologies), 1X antibiotic-antimycotic and 20 µg/ml hygromycin. HEK293 cells were grown in DMEM supplemented with 10% FBS, 1X antibiotic -antimycotic (Gibco) and 1 mM Glutamax (Gibco). Cells were passaged by 5-10 min incubation in enzyme-free cell dissociation solution (S-014-B, Millipore) at 37 °C, resuspension in full growth medium and re-seeding at a 1:5 to 1:10 ratio, 3 times per week. Cells were regularly tested for mycoplasma contamination by polymerase chain reaction (PCR) with mycoplasma DNA-specific primers.

2.1.2. siRNA-mediated knockdown of gene expression

Human RPE1 cells were counted using a CASY Cell Counter and Analyser Model TT (Roche Applied Science) and seeded at the densities between 3.3 and 13.3 x10^3 cells/cm^2, as indicated in main text 6 hours prior to transfection. Oligofectamine (Thermo Fisher) or Lipofectamine RNAiMAX (referred to as RNAiMAX hereafter; Thermo Fisher) were used as transfection reagents in this work as indicated (Table 2.1). siRNA oligonucleotides and indicated transfection reagent were diluted separately in 500 µl OptiMEM (Thermo Fisher), gently mixed and incubated at room temperature for 20 min (proportions shown in Table 2.1) before pipetting dropwise onto cells in culture media (Day 1). A second transfection was performed 24 hours after the first transfection, preceded by exchange of cell media 1 hour in advance (Day 2). Cell media was changed 24 hours after the second transfection (Day 3), and cells were harvested on Day 4 or 5. siRNA oligonucleotide sequences are shown in Table 2.2.

Reagent	35 mm Single KD	35 mm Triple KD	100 mm Single KD	100 mm Triple KD
siRNA (total in 2 transfections using 20 µM stock solutions)	80 pmol (2 x 2 µl)	80 pmol (2 x 2 µl)*	624 pmol (2 x 15 µl)	624pmol (2 x 15 µl)*
Oligofectamine/µL siRNA	0.4 µl	0.27 µl	0.4 µl	0.27 µl
RNAiMAX/ µL siRNA	0.4 µl	0.27 µl	0.4 µl	0.27 µl

Table 2.1 Reagents used for RNA interference. *siRNA amounts for triple knockdown experiments are reported per duplex.

siRNA	Sequence	Manufacturer
µ2-adaptin (Human *AP2M1*) HSS101955	5-'CCGCCAGAUGGAGAGUUUGAGCUUA-3' 3'-UAAGCUCAAACUCUCCAUCUGGCGG-5'	Thermo Fisher
Endophilin A1 (Human *SH3GL2*) HSS109709	5'-GCUGAGUCAAGCAUGUUCAAUCUCU-3' 3'-AGAGAUUGAACAUGCUUGACUCAGC-5'	Thermo Fisher
Endophilin A2 (Human *SH3GL1*) HSS109707	5'-CCCAAGAUCGCAGCUUCAUCGUCUU-5' 3'-AAGACGAUGAAGCUGCGAUCUUGGG-5'	Thermo Fisher
Endophilin A3 (Human *SH3GL3*) HSS109712	5'-CAAUGGAGUUUCCACCACCUCUGUA-3' 3'-UACAGAGGUGGUGGAAACUCCAUUG-5'	Thermo Fisher

Table 2.2. SiRNA sequences and manufacturers. Sense and antisense oligonucleotides are shown.

2.2. Sample preparation for liquid chromatography mass spectrometry

2.2.1. In-gel digestion

For In-gel digestion, proteins were separated by SDS-PAGE and stained using Coomassie Brilliant Blue G250 (Bio Basic Canada Inc.), InstantBlue protein stain (Expedeon) or Pierce™ Silver Stain for Mass Spectrometry (Thermo Scientific) as indicated. All staining and destaining steps were performed in clean covered containers on a rotating platform (75 rpm) in a fume hood. All gel cutting steps were performed in a laminar flow hood.

For Coomassie staining, gels were incubated on a rocking platform in Coomassie staining solution (0.3% w/v Brilliant Blue G250; 45% methanol, 10% glacial acetic acid in ddH_2O) for ≥ 1 hour. Gel pieces were washed twice in ddH_2O (10 min) and fixed by incubation in 30% methanol for 5 mins. Background was reduced by treatment with Coomassie destain solution (40% ethanol, 10% glacial acetic acid in ddH_2O) (2 x 30 min washes, followed by overnight incubation in 1:1 Coomassie destain: ddH_2O). For InstantBlue Protein staining, gels were incubated on a rocking platform in InstantBlue solution for 15 min – 1 hour. Destaining was performed in water to reduce background staining.

For Silver staining using Pierce™ Silver Stain for Mass Spectrometry (Thermo Fisher), the following procedure was followed. The gel was washed twice in LC-MS grade water for 5 min, then fixed in Fixing solution (Fixing Solution: 30% ethanol, 10% acetic acid) for 2 x 15 min incubations with agitation (20 rpm). The gel was washed twice in 10% Ethanol for 5 min, then washed twice in LC-MS grade water for 5 min. A sensitiser working solution was prepared by mixing 1 part Silver Stain Sensitizer (Thermo Fisher) with 500 parts LC-MS grade water and incubated with the gel for 1 min. The gel was then quickly washed in LC-MS grade water for 2 x 1 min washes. A silver stain enhancer solution was prepared by mixing 1 part Silver Stain

Enhancer (Thermo Fisher) with 100 parts Silver Stain (Thermo Fisher) and immediately added to the gel, and incubated for 5 min. A developer working solution was prepared by mixing 1 part Silver Stain Enhancer with 100 parts Silver Stain Developer. Gels were washed twice briefly in LC-MS grade water and incubated in developer solution for 2-3 min to allow bands to appear. When staining was optimal, developer solution was exchanged for a 5% acetic acid "Stop solution" and gel bands were immediately excised.

Gel lanes were cut into sections as indicated in the text using a clean scalpel, and sections were cut into ~1 mm^3 pieces. Pieces from each section were collected into separate Lobind 1.5 ml tubes (Eppendorf) for destaining. Coomassie stained gel pieces were destained in a mixture of 40% acetonitrile, 60% 50 mM ammonium bicarbonate (2 x 15 min washes on sample tube rotator (750rpm; Fisher Scientific). Silver stained gel pieces were destained by 2 x 15 min in a Destaining working solution (74 µl of Silver Destain Reagent A, 245 µl of Silver Destain Reagent B and 4 ml of LC-MS grade water). Gel pieces were then washed 3 times in a silver staining wash solution (50% acetonitrile in 25 mM ammonium bicarbonate).

Following destaining, neat acetonitrile was used to dehydrate the pieces and removed. Proteins were reduced by incubation in a solution of 10 mM dithiothreitol in 50 mM ammonium bicarbonate for 30 min at 80°C. Exposed cysteine residues were then capped by alkylation with 55 mM iodoacetamide in 50 mM ammonium bicarbonate, at room temperature, for 20 min in the dark. The pieces were washed in 50 mM ammonium bicarbonate and subjected to 2 "shrink and swell" steps: pieces were shrunk by dehydration in neat acetonitrile before rehydration in 50 mM ammonium bicarbonate. Pieces were then dehydrated in acetonitrile and dried by incubation with tubes open at 50 °C for 5 min.

Only LC-MS grade reagents were used beyond this point unless otherwise stated. Gel pieces were immersed in 10 ng/µl trypsin (Promega) in 50 mM ammonium bicarbonate (using a volume sufficient to cover pieces) overnight at 37°C unless otherwise stated. An equal volume of Extraction buffer (1% formic acid; J. T. Baker, 2% acetonitrile; Biosolve, in LC-MS grade water;

Thermo Scientific) was added following digest and samples were agitated in a sample tube rotator (750 rpm; Fisher Scientific) for 15min. Supernatant was collected and the extraction step was repeated. A final extraction step was conducted in a small volume of neat acetonitrile. The peptides were dried down in a vacuum concentrator (SpeedVac DNA 120, Thermo Scientific) and re-solubilised in 10/20 µl of mobile phase solution A (3% LC-MS grade acetonitrile, 0.1% formic acid in H_2O) for processing by LC-MS or frozen at -20°C for storage.

For samples prepared from mouse spinal cord, a C18 StageTip clean-up step was performed following elution of peptides from gel pieces. A 200 µl pipette tip was plugged with a frit cut from an Empore C18 disc (3M). Following extraction from gel pieces, peptides were dried in a SpeedVac and resuspended in 20 µl 0.5% acetic acid. Using Table 2.3 as a reference, StageTips

Solution	Composition
1	100 % Methanol
2	0.5 % Acetic acid in water
3	0.5 % Acetic acid in 80% Acetonitrile
4	0.5% Acetic acid, 20mM Ammonium Acetate, 20% Acetonitrile
5	0.5% Acetic acid, 50mM Ammonium Acetate, 20% Acetonitrile
6	0.5% Acetic acid, 80mM Ammonium Acetate, 20% Acetonitrile
7	0.5% Acetic acid, 150mM Ammonium Acetate, 20% Acetonitrile
8	0.5% Acetic acid, 250mM Ammonium Acetate, 20% Acetonitrile
9	0.5% Acetic acid, 500mM Ammonium Acetate, 20% Acetonitrile
10	0.5% Acetic acid, 1000mM Ammonium Acetate, 20% Acetonitrile

Table 2.3 Solutions used during 'In-gel-SCX' protocol and C18 SPE cleanup.

were washed in methanol, then equilibrated in Solutions 3 then Solution 2. Peptides were added and washed once in 50µl Solution 2, then eluted in 20 µl Solution 3. Peptides were then dried down in a vacuum concentrator in preparation for resuspension in LC-MS running buffer.

For profiling of plasma membrane extracts from RPE1 cells with RNAi-induced endocytic inhibition, an SCX fractionation step was performed using C18-SCX Stop-and-go extraction tips

(StageTips). Solutions used for In-gel-SCX are shown in Table 2.3. C18-SCX StageTips were assembled with two plugs each of Empore C18 discs (3M) and Empore Strong Cation exchange discs (SCX; 3M) taking care to avoid tight packing and ensuring correct orientation as described in Rappsilber, Mann and Ishihama, (2007). Following extraction from gel pieces, peptides were dried in a SpeedVac and resuspended in 20 µl 0.5% acetic acid. Using Table 2.3 as a reference, StageTips were washed in methanol, then equilibrated sequentially in 50 µl of Solutions 3, 2 then 10 and finally 2 again. Peptides were added and washed once in 50 µl Solution 2, then eluted from the C18 bed in solution 50 µl Solution 3. Peptides were then eluted from the SCX matrix sequentially into separate sample tubes using 20 µl of Solutions 6, then 8 and finally 10. Peptides were dried in a vacuum concentrator and resuspended in LC-MS running buffer prior to LC-MS analysis.

For preparation of lumbar spinal cord extracts using in-gel digestion, sections of lumbar spinal cord (L1-L5) weighing between 20-25 mg were prepared from 3 *Hsj1*$^{+/+}$ wild-type (WT) and 3 *Hsj1*$^{-/-}$ knockout (KO) mice 10 days after birth (P10), coincident with a detectable unfolded protein response (UPR), but before neuronal death. HSJ1 KO mice were generated by genOway (Lyon, France) in a C57BL/6 mouse strain by deletion of exons 2 and 3 in Hsj1. Neither HSJ1 isoform could be detected by immunoblotting of HSJ1 KO brain extracts (Cheetham Lab). Lumbar tissue was lysed in a Tris-SDS lysis buffer (0.1 M Tris-HCl pH 7.8, 0.05 M DTT, 2% SDS with protease and phosphatase inhibitors). Samples were boiled for 5 minutes and centrifuged at 16,000 x *g* for 10 min at 4°C. The cleared lysate was mixed with SDS-PAGE sample buffer and 30ul of each sample separated on a 1D SDS-PAGE gel (10%) and prepared for LC-MS analysis by in-gel digestion (6 bands) followed by C18-StageTip clean-up as described in above. All the above steps were performed by D. Parfitt (Cheetham Lab, UCL), save the SPE clean-up.

2.2.2. On-bead digestion and On-bead digestion coupled to SCX (Plasma membrane fraction)

On-bead digestion and on-bead digestion coupled to strong cation exchange ('On-bead-SCX') were performed on affinity purified plasma membrane fractions from 6 x 100 mm dishes of 80 - 90% confluent RPE1 cells. Solutions used in this protocol are listed in Table 2.4 and were all MS-grade. Following biotinylation, cell surface proteins were captured on 500 µl neutravidin beads and washed 4 times in proprietary kit wash buffer (Thermo Fisher,), then washed again 3 times in 800 µl Tris-buffered saline (TBS) to reduce detergent concentration. Beads were resuspended in 400 µl Tris-Urea buffer (TU; Table 2.4) containing 115/300/600/1800 ng of trypsin as indicated and incubated at 27°C for 30 min. Sample was centrifuged (1000 x *g*, 1min) and supernatant was transferred to a fresh sample tube. Beads were washed twice in 200 µl DTT buffer and supernatants from wash steps were combined with the supernatant from the trypsin step. The reaction mixture was incubated at room temperature overnight for a combined reduction/ digestion step. Alkylation was then performed by adding 138 µl of IAA Buffer (Table 2.4) and incubating mixture for 20 min in the dark, before adding formic acid to 1% total volume to stop the trypsin reaction.

Solution	Composition
TU	2 M Urea, 50 mM Tris (pH 7.5) in ddH_2O
Trypsin Buffer	2 M Urea, 50 mM Tris pH 7.5, 1.8 µg Trypsin/digest
DTT buffer	2 M Urea, 50mM Tris (pH 7.5), 2mM DTT
IAA buffer	0.2 M Urea, 50mM Tris (pH 7.5), 10mg/ml IAA
TBS	50mM Tris (pH 7.5), 150mM NaCl in ddH_2O

Table 2.4 Solutions used in on-bead digestion protocols.

Samples were desalted using Sep-Pak tC18 cartridges (Waters). All solutions utilised beyond this point were MS-grade. Cartridges were washed with 500 µl acetonitrile and equilibrated with 2 x 500 µl washes in SPE wash solution (5% acetonitrile, 0.1% formic acid in water). Samples were then loaded onto the cartridge and washed with 2 x 500 µl SPE wash solution. Samples

were then eluted using 2 x 250 µl of SPE elution solution (50% acetonitrile, 0.1% formic acid in water). Samples were dried in a vacuum centrifuge and resuspended in LC-MS running buffer or coupled to C18-SCX StageTip fractionation.

Solution	Composition
1	100% Methanol
2	0.5% acetic acid
3	80% acetonitrile, 0.5% acetic acid
4	20mM ammonium acetate, 20% acetonitrile, 0.5% acetic acid
5	50mM ammonium acetate, 20% acetonitrile, 0.5% acetic acid
6	100mM ammonium acetate, 20% acetonitrile, 0.5% acetic acid
7	500mM ammonium acetate, 20% acetonitrile, 0.5% acetic acid
8	1000mM ammonium acetate, 20% acetonitrile, 0.5% acetic acid

Table 2.5 Solutions used in On-bead-SCX protocol.

For On-bead-SCX samples, StageTip cleanup was coupled to 5-fraction SCX on SCX-StageTips. Solutions used for On-bead-SCX are shown in Table 2.5. SCX StageTips were assembled in 200 µl pipette tips with two plugs of Empore Strong Cation exchange discs (SCX; 3M), taking care to avoid tight packing and ensuring correct orientation as described in Rappsilber, Mann and Ishihama (2007). Using Table 2.5 as a reference, StageTips were washed in methanol, then equilibrated sequentially in 50 µl of Solutions 3, 2 then 7 and finally 2 again. Peptides were added and washed once in 50 µl Solution 2, then sequentially eluted into separate sample tubes from the SCX bed in 20 µl using of Solutions 2, 4, 5, 6 and finally 7. Peptides were dried in a vacuum concentrator and resuspended in LC-MS running buffer (3% acetonitrile. 0.1% formic acid in water) prior to analysis.

2.2.3. On-bead digestion coupled to SCX (eGFP-EndophilinA2 AP-MS)

Following immunoprecipitation, GFP-Trap® beads were washed 3 times in 500 µl ice-cold Tris-buffered saline (TBS) solution using a magnetic rack. Solutions used in this protocol are listed in Table 2.4 and were all MS-grade. Beads were resuspended in 120 µl Tris-Urea buffer (TU; Table 2.4) containing 400 ng of trypsin and incubated at 27°C for 30 min. Beads were sedimented on a magnetic rack and the supernatant was transferred to a fresh "Digest" sample tube. Beads were washed twice in 50 µl DTT buffer (Table 2.4) and supernatants from wash steps were combined with the supernatant from the trypsin step in the "Digest" tube. The reaction mixture was incubated at room temperature overnight for a combined reduction/ digestion step. Alkylation was then performed by adding 40 µl of IAA Buffer (Table 2.4) and incubating the mixture for 20 min in the dark, before adding acetic acid to 0.5% total volume to stop the trypsin reaction.

Samples were desalted using C18-SCX StageTips in an effort to reduce the amount of NP-40 detergent in the final sample for injection onto the LC column. C18-SCX StageTips were assembled in 200 µl pipette tips with two plugs each of Empore C18 SPE and Empore Strong Cation exchange discs (SCX; 3M), stacking the C18 plugs above the SCX plugs. Care was taken to avoid tight packing and ensuring correct orientation as described in Rappsilber, Mann and Ishihama (2007).

Using Table 2.5 as a reference, C19-SCX StageTips were washed in methanol, then equilibrated sequentially in 50 µl of Solutions 8, 3 and finally 2. Peptides were added and washed once in 50 µl Solution 2, then eluted into the SCX bed using 50 µl of Solution 3. Peptides were eluted from the SCX bed using 20µl of Solution 8. Peptides were dried in a vacuum concentrator and resuspended in LC-MS running buffer (3% acetonitrile. 0.1% formic acid in water) prior to analysis.

2.2.4. Filter-aided sample preparation (FASP)

Samples were mixed with denaturing and reducing solution (100 mM dithiothreitol; Fisher Scientific, 8 M Urea; Sigma, in 100 mM Tris-HCl pH 8.5) to a final volume of 400 µl and incubated for 20 min at room temperature in a Vivacon 500 30K molecular weight cutoff (MWCO) column (Sartorius). Samples were centrifuged at 14,000 x *g* for 20 min. DTT was exchanged out by washing in 200 µl Urea solution (8 M Urea in 100 mM Tris-HCl pH8.5) (3 x 20 min washes, 14,000 x *g*). Cysteine residues were alkylated by resuspending protein in 100 µl Iodoacetamide solution (50 mM iodoacetamide; Sigma, 8 M Urea in 100 mM Tris-HCl pH 8.5) and incubating for 20 min in the dark. Iodoacetamide was exchanged out (3 x 20 washes in Urea solution, 14,000 x *g*) and Urea was then removed through 3 x 200µl washes in 50mM ammonium bicarbonate (10 min, 14,000 x *g*).

Trypsin digest was conducted using a 1:100 enzyme:protein ratio (w/w). An appropriate amount of trypsin (Promega) based on sample quantitation was added to a final volume of 40 µl 50 mM ammonium bicarbonate and incubated at 37°C overnight.

Peptides were recovered by centrifugation (14,000 x *g*, 10min) and dried in a vacuum dryer. Peptides were then subjected to C18 StageTip desalting or C18-SCX fractionation.

For unfractionated FASP samples, a C18 StageTip clean-up step was performed following elution of peptides from gel pieces. A 200 µl pipette tip was plugged with a frit cut from an Empore C18 disc (3M). Following extraction from gel pieces, peptides were dried in a SpeedVac and resuspended in 20 µl 0.5% acetic acid. Using Table 2.3 as a reference, StageTips were washed in methanol, then equilibrated in Solutions 3 then Solution 2. Peptides were added and washed once in 50µl Solution 2, then eluted in 20 µl Solution 3. Peptides were then dried down in a vacuum concentrator in preparation for resuspension in LC-MS running buffer.

For FASP-SCX, an SCX fractionation step was performed using C18-SCX Stop-and-go extraction tips (StageTips). Solutions used for FASP-SCX are shown in Table 2.3. C18-SCX

StageTips were assembled with two plugs each of Empore C18 discs (3M) and Empore Strong Cation exchange discs (SCX; 3M) taking care to avoid tight packing and ensuring correct orientation as described in Rappsilber, Mann and Ishihama (2007). Following extraction from gel pieces, peptides were dried in a SpeedVac and resuspended in 20 µl 0.5% acetic acid. Using Table 2.3 as a reference, StageTips were washed in methanol, then equilibrated sequentially in 50µl of Solutions 3, 2 then 10 and finally 2 again. Peptides were added and flow through was retained as 0 mM fraction and washed once in 50µl Solution 2, then eluted from the C18 bed in solution 50 µl Solution 3. Peptides were then eluted from the SCX matrix sequentially into separate sample tubes using 20 µl of Solutions 4 through 10. Peptides were dried in a vacuum concentrator and resuspended in LC-MS running buffer prior to LC-MS analysis.

2.3. SDS-PAGE and Western Blotting

Samples were suspended in 2X SDS-PAGE sample buffer (125 mM Tris/HCl (pH 6.8), 20% glycerol, 4% SDS and 0.02 % bromophenol blue), heated at 95 °C for 5 min and centrifuged at 16,000 x *g* for 1.5 min prior to loading on Bolt® Bis-Tris Plus precast gels (Thermo Scientific). Samples were separated by SDS-PAGE in MOPS buffer (45 min at constant 150 V, or 60 min at 130 V for quantitation) then transferred to a PVDF membrane using a Lightning Blot Semi-Dry Transfer System (Perkin Elmer) (20-24 min at constant 24V).

Membranes were blocked for 1 hour in Blocking Buffer (5% BSA in 0.2% TBS-Tween) at room temperature then incubated overnight with primary antibody at 4°C as indicated in Table 2.6. Membranes were washed in 0.2% TBS-Tween (3 x 10 min washes) and incubated with HRP-conjugated secondary antibody for 30 min at room temperature. Membranes were then washed in 0.2% TBS-tween (3 x 10 min washes) and immersed in ECL Plus Western Blotting Substrate (Thermo Scientific) for 5 min then imaged using an X-ray film developer.

Antibody target	Company	Species	Dilution factor or final concentration (µg/ml)
EGFR	CST	Rabbit	1:1000
Calnexin	Santa Cruz	Rabbit	1:500
Transferrin Receptor	Millipore	Rabbit	1:2000
Adaptin-α clone 8/Adaptin-a	BD Biosciences	Mouse	1:250
GAPDH	Santa Cruz	Mouse	1:20000
Histone H3	Upstate	Mouse	0.5
Goat anti-Mouse IgG Secondary Antibody, HRP	BioRad	Goat	1:10000
Goat anti-Rabbit IgG Secondary Antibody, HRP	BioRad	Goat	1:10000
β_1-AR	Abcam	Rabbit	1:500
HGFR	R&D systems	Rabbit	1:1000
Integrin-αV	CST	Rabbit	1:1000
Integrin β1	CST	Rabbit	1:1000
Endophilin A2	Santa Cruz	Mouse	1:500
eGFP	Roche	Mouse	1:1000
GSK3α/β	CST	Rabbit	1:500

Table 2.6 Antibodies used in the study with dilutions.

2.4. Silver staining

Following SDS-PAGE, gels were stained using the Pierce™ Silver Stain for Mass Spectrometry (Thermo Fisher). The gel was washed twice in LC-MS grade water for 5 min, then fixed in Fixing solution (30% ethanol, 10% acetic acid) for 2 x 15 min incubations with agitation (20 rpm). The gel was washed twice in 10% Ethanol for 5 min, then washed twice in LC-MS grade water for 5 min. A sensitiser working solution was prepared by mixing 1 part Silver Stain Sensitizer (Thermo Fisher) with 500 parts LC-MS grade water and incubated with the gel for 1 min. The gel was then quickly washed in LC-MS grade water for 2 x 1 min washes. A silver stain enhancer solution was prepared by mixing 1 part Silver Stain Enhancer (Thermo Fisher) with 100 parts Silver Stain (Thermo Fisher) and immediately added to the gel, and incubated for 5 min. A developer working solution was prepared by mixing 1 part Silver Stain Enhancer with 100 parts Silver Stain Developer. Gels were washed twice briefly in LC-MS grade water and incubated in developer solution for 2-3 min to allow bands to appear. When staining was optimal, developer solution was exchanged for a 5% acetic acid “Stop solution”. Gels were washed in water and stored in the cold.

2.5. Densitometric analysis

Images were analysed using ImageJ v1.49. Raw images were converted to 8-bit grayscale and adjusted to obtain optimum signal. Profile plots for each lane were obtained using ImageJ gel analysis tools. The rightmost signal (representing the bottom of each lane) was assumed to represent background noise and used to draw a horizontal line across the plot. The area above this line was used for calculation of relative protein levels. Alternatively, densitometric analysis was performed using ImageStudio Lite v 5.2. Signals for each band were corrected using relative GAPDH signal, with GAPDH acting as a proxy for differences in total cell number.

2.6. LC-MS analysis of samples

1 µl – 4 µl of each sample was loaded onto a reversed-phase UPLC Symmetry C18 Trap column (180 µm internal diameter, 20 mm length, 5 µm particle size, Waters Corp.). Samples were desalted (99.5% Buffer A) at a flow rate of 8 µl/min for 2 min. Peptides were then separated using a linear gradient (0.3 µl/min, 35°C column temperature; 3% - 40% mobile phase B) by a BEH130 C18 analytical column (75 µm internal diameter, 400 mm length, 1.7 µm particle size, Waters Corp.) over the course of 40 min, 60 min or 90 min as indicated.

Mobile phase A: LC-MS grade H_2O, 0.1% formic acid
Mobile phase B: LC-MS grade Acetonitrile, 0.1% formic acid

The nanoUPLC was coupled through a nanoflow sprayer to a quadrupole time-of-flight (QToF) mass spectrometer (HDMS Synapt G2-Si; Waters Corp.) operating in Resolution mode. The ToF analyser was externally calibrated from m/z 175.11 to 1285.54 using the fragment ions of a 320 fmol/µl solution of [Glu1]-fibrinopeptide B (GFP; Sigma Aldrich). Data were lockmass-corrected following acquisition using the monoisotopic mass of the doubly charged precursor of GFP (785.8426 m/z), delivered to the mass spectrometer via a LockSpray interface. This reference sprayer was sampled every 30 s. Mass measurements were made using a data independent mode (MS^E) of acquisition. Briefly, energy in the collision cell was alternated from low energy (4 eV) to high energy (energy ramp from 15-35 eV) whilst continuously acquiring MS data. Measurements were made over an m/z range of 50-2000 Da with a scan time of 0.6s One cycle of MS and MS^E data were acquired every 1.2s.

In $HDMS^E$ acquisition mode, IMS was performed by applying a constant wave height of 40 V whilst a constant wave velocity of 650 m/s was maintained. Wave heights within the trap and transfer were both set at 4 V whilst the wave velocities were 311 and 175 m/s, respectively. MS data were acquired over 50-2000 m/z for each mode. Spectral acquisition time for each mode was 0.5 s with a 0.015 interscan delay. During the low energy MS mode data were acquired whilst applying a constant collision energy of 4 eV within the transfer travelling wave region. High

energy MS/MS data were acquired by ramping the collision energy within the transfer cell, post IMS, between 25 V and 55 V. One cycle of low and elevated energy data was acquired every 1.1 s. To ensure that ions with m/z less than 350 observed in the LC-MS data were derived exclusively from peptide fragmentation within the transfer region the radio frequency applied to the quadrupole mass analyser was adjusted to optimise transmission within the region of 350 – 2000 Da.

2.6.1. Database searching

Data collected by LC-MS was processed using Protein Lynx Global Server (PLGS) v3.0.2 (Waters Corp.) or Progenesis v4.0 (Waters Corp.). For experiments presented in chapters 3 and 4, data were queried against a *Homo sapiens* FASTA protein database (UniProt proteome: UP000005640) concatenated with a list of common contaminants obtained from the Global Proteome Machine (The Global Proteome Machine, 2016) and the Hi3 standard, *E. coli* ClpB. For experiments presented in chapter 5, data were queried against a *Mus musculus* protein database (UniProt proteome: UP000000589), concatenated with a list of common contaminants obtained from the Global Proteome Machine (The Global Proteome Machine, 2016) and the Hi3 standard, *E. coli* ClpB. In both cases, a 1-fold randomised decoy database was concatenated onto the target database to facilitate false discovery rate (FDR) estimation. Carbamidomethyl-C was specified as a fixed modification and Oxidation (M) was specified as a variable modification. For experiments described in Section 3.2.4, CAMthio-propanoylation of lysines and N-termini were selected as variable modifications. For experiments described in chapter 5, phosphorylation of serine, threonine and/or tyrosine were specified as variable modifications. For all experiments, a maximum of 2 missed cleavages were tolerated in the analysis to account for incomplete digestion. For peptide identification 3 corresponding fragment ions were set as a minimum criterion whereas for protein identification a minimum of 7 fragment ions were required. Protein-level FDR was set at 1% as estimated by the number of proteins identified from the randomised decoy database. Where indicated, samples were spiked post-digestion with *E. coli* ClpB peptide (50 fmol/injection). Absolute or relative quantitation was performed by PLGS or Progenesis v4.0 (Waters Corp.) using the Hi3 approach (Silva *et al.*, 2005, 2006).

2.6.2. Data normalisation

Data was normalised to global intensity or to median Intensity as described. Normalisation was performed using the NormalyzerDE platform where indicated (Willforss, Chawade and Levander, 2019). In the following paragraph, “protein intensity” refers to Top3 intensity (the summed intensity of the top3 best ionising peptides for a given protein) or Hi3 protein abundance (the absolute amount of a given protein as calculated via the Hi3 method with reference to a spiked-in peptide standard of known concentration).

Normalisation to global intensity (GI) was performed by dividing the measured protein intensity for a given protein in a given replicate by the sum of protein intensity of all proteins in that replicate, then multiplying this by the median of summed protein intensities among all replicates. Data was then $\log_2$ transformed. Normalisation to median intensity was performed by dividing the measured protein intensity for a given protein in a given replicate by the median protein intensity for all proteins in that replicate, then multiplying this by the median of protein intensities among all replicates. Data was then $\log_2$ transformed.

2.7. Statistical analysis and data visualisation

2.7.1. Analysis of Western blots

Measurements of western blot signal were performed as described in Section 2.5. Normalised protein amounts were compared pairwise between conditions using an unpaired, 2-way Student's *t*-test assuming equal variances.

2.7.2. Volcano plots, PCA analysis and multi-scatterplots

P-value calculations for volcano plots were performed in Perseus, using a 2-sided *t*-test with equal variances and correcting for multiple comparisons using a permutation-based FDR correction (FDR cut-off of 0.05). Proteins were denoted as differentially expressed upon fulfilment of two criteria: *p*-value < 0.05 and a fold change (ratio of average protein intensity) greater than 1.3-fold in either direction. Quantitative measurements of peptides using MS^E have been shown to display an error in signal intensity of 10-15% (Silva *et al.*, 2006; Patel *et al.*, 2009). Thus, a fold change cut-off of 1.3 represents a value 2-3 times larger than that expected given normal technical variation. Plots were rendered using Rstudio. PCA analysis and multi scatter plots with Pearson correlation calculations were created using the tools in the Perseus software platform (Tyanova *et al.*, 2016).

2.7.3. Network analysis in STRING

Following annotation as differentially expressed, proteins were processed using the STRING database search tool to identify interaction networks. Interaction sources were limited to Experiments, Databases and Gene Fusion to increase confidence. The minimum interaction score required was 0.7 and clusters were demarcated using k-means clustering assuming 5 clusters.

2.7.4. DAVID functional annotation

Gene Ontology enrichment analysis was performed using the DAVID functional annotation suite assessing biological process and cellular compartment terms for enrichment (Huang *et al.*, 2009; Huang, Sherman and Lempicki, 2009). The software calculates enrichment based on the number of gene ontology terms present in a chosen background. A *p*-value is calculated by comparing the number proteins with a given gene ontology (GO) term present in a query list to the number present in the background, in relation to the total number of proteins tested. Enrichment analysis in DAVID was performed using all 1099 well-replicated proteins as a background and enrichment was defined as a p value ≤ 0.05, corrected for multiple testing using a FDR cut-off of 0.05 (Benjamini and Hochberg, 1995).

2.8. Preparation of plasma membrane fractions

2.8.1. Surface biotinylation of plasma membranes

Surface Plasma Membrane proteins were isolated using the Pierce Cell Surface Protein Isolation Kit (Thermo Scientific). RPE1 or HEK293 cells in 100 mm dishes were washed four times in 2 ml ice-cold Phosphate Buffered Saline (PBS), then incubated in 8mL of ice-cold sulfo-NHS-SS-Biotin (Thermo Scientific) for 30 min at 4°C with gentle agitation. Cells were covered in a 1 ml of a 1:1 mixture of TBS and proprietary Quenching Solution (Thermo Scientific) and gently scraped into solution and transferred to a 50 ml sample tube. Plates were rinsed with 5 ml Tris Buffered Saline (TBS) and the rinse volume was added to the 50 ml sample tube. Where indicated, cells from 6 plates were pooled. Cells were pelleted at 500 x *g* for 3 min and washed once in ice-cold TBS, then resupended in 600 µl Lysis Buffer (Thermo Scientific) with EDTA-free Protease Inhibitor (Cell Signalling). Cells were sonicated at low power (1.5) on ice using 5 x 1-second bursts (Soniprep 150, MSE). Samples were incubated on ice for 30 min, vortexing briefly every 5min. The lysate was sonicated again at low power (1.5; 5 x 1-second bursts). The lysate was

then centrifuged at 10,000 x *g* for 2 min at 4°C. Clarified supernatant was collected in a fresh 1.5 ml sample tube. The pellet was resuspended in SDS-PAGE sample buffer (Thermo Scientific) and used as “Pellet” control.

The clarified supernatant was then incubated with 400 µl pre-washed 50% NeutrAvidin Agarose slurry (Thermo Scientific) for 1 hour at room temperature. Beads were centrifuged at 1000 x *g* and supernatant was retained as “Flow Through” control. Beads were then washed (1000 x *g*) in 500 µl Wash Buffer (Thermo Scientific) with protease inhibitor 4 times. Supernatants from the first and fourth washes were retained as “Wash 1” and “Wash 4” controls. Affinity purified proteins were then eluted from beads in SDS-PAGE sample buffer by boiling for 5 min (for western blotting) or 1 hour incubation at room temperature (for in-gel digestion) or prepared for LC-MS directly by on-bead digestion.

Alternatively, captured proteins were subjected to mild elution in a range of buffers and elution conditions as indicated in Section 3.3.2.

2.8.2. Preparation of crude membrane extracts

All steps were performed on ice. RPE1 cells were subjected to mechanical lysis in hypotonic buffer. RPE1 cells grown in 100 mm dishes were grown to 80-90% confluence, washed 2 times in hypotonic buffer (20 mM HEPES, 10 mM KCL, 2 mM MgCl2, 1 mM EDTA, 1mM DTT with phosphatase and protease inhibitors) then scraped into a sample tube and allowed to swell for 15 min in 1 ml hypotonic buffer. Cells were then lysed by 20/40 passes through a 23-gauge/ 30-gauge needle or 100 strokes of a Dounce homogeniser as indicated. Intact cells and nuclei were collected by centrifugation of the lysate at 1000 x *g* for 5 min and the supernatant was centrifuged for 1 hour at 100,000 x *g* to collect a membrane pellet and cytosolic fraction (supernatant). Where indicated, the membrane pellet was washed 3 times by resuspension in 400 µl hypotonic buffer and centrifugation at 100,000 x *g* for 1 hour.

2.9. Secondary GST pulldown from plasma membrane fraction

2.9.1. Overexpression of eGFP constructs

HEK cells were grown for 24 hours to reach a confluence of 70% in 100 mm dishes. For each 100 mm dish of HEK293 cells to be transfected, 300 µl of OPTIMEM (Thermo Fisher) was mixed with 18 µl GeneJuice (Millipore) and 6 µg of plasmid containing β_1-AR-eGFP (EB1019) or Dyn1-eGFP construct (B148) was added, and the solution was mixed and allowed to incubate at room temperature for 20 min. The transfection mixture was added in a drop-wise manner to cells in full growth medium. Cells were placed in an incubator and used for further analysis 16-24 hours post transfection.

2.9.2. Making GST and GST-EndoA2-SH3 coated magnetic beads

A single colony of pre-selected BL21 transformed with a glutathione S transferase (GST) vector plasmid (GV38) or a GST-EndoA2-SH3 domain construct (EB5007) was used to inoculate a 20ml starter culture of autoclaved LB media with ampicillin (additional GST construct details are listed in Table 2.7). Starter cultures were grown overnight at 37°C to reach stationary phase. 10ml of this culture was used to inoculate 0.5L-1L of autoclaved LB media, which was grown at 37°C on a rotating platform. IPTG was added as to a final concentration of 1mM upon reaching OD600 0.5-0.6 and cultures were grown overnight at 25°C.

Pellets were obtained by centrifugation (4500 x *g*, 20min, 4°C) (SX4750A rotor, Beckman Coulter). Pellets were frozen at -20°C overnight prior to resuspending in 5 volumes of Lysis buffer (137 mM NaCl, 2.7 mM KCl, 12.5 mM Na_2HPO_4, 2mM KH_2PO_4, 0.5% Triton X-100 with DNaseI, Lysozyme and protease inhibitor) at 4°C on a rotating platform (100 rpm). The lysate was sonicated if required then centrifuged at 11,000 rpm in (rotor JA 25.50, Beckman Coulter) for 1 hour at 4°C.

The supernatant was concentrated using a Centricon Plus-70 assembly (10K MWCO; Millipore) by centrifugation at 3,500 x *g*, 4°C for 40 min to 2 hours as necessary, to a final volume of 5-10 ml. Concentrated extract was incubated with a minimal volume of Glutathione agarose magnetic beads (Thermo Fisher) on a wheel rotor for 1 hour at 4°C. Beads were then washed 10 times in PBS (5 min washes on a wheel rotor, 4°C), and stored at 4°C, or stored in 50% glycerol at -20°C.

2.9.3. Pulldown from plasma membrane eluate

Plasma membrane fractions were obtained from RPE1 cells or from HEK293 cells overexpressing β1-AR-eGFP as indicated. Following release in low DTT CHAPS buffer, the plasma membrane eluate was incubated with 2/5/10 µl decorated GST beads (pre-washed 3 x in 500 µl low DTT CHAPS) overnight at 4°C with end over end rotation. Following incubation beads were collected at the side of the tube using a magnetic rack and the supernatant was retained as “Flow Through” control. Beads were then washed 3 times in 500µl Low DTT CHAPS buffer and bound proteins were eluted in 50 µl SDS-PAGE sample buffer for 15 min at 65°C.

2.10. Immunoprecipitation of endogenous endophilin A2 with modulation of serum in cell media

2.10.1. Stimulation of growing RPE1 cells with serum and immunoprecipitation of endogenous endophilin A2

Cell culture and immunoprecipitation and subsequent in-gel digestion steps were performed by Mr J. Panambalana. RPE1 cells were grown on 100 mm dishes to a confluence of 80% before harvest. For Stimulated condition, cell media was supplemented with an extra 10% FBS (20% final) for 5 min at 37°C. For Resting condition, no additional treatment was performed prior to cell lysis. Cells were washed 2 times in ice-cold PBS before being gently scraped into lysis buffer (10 mM Tris HCL pH7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP40 with protease and phosphatase inhibitor cocktail (Thermo Scientific)) and incubated on ice for 30 min, then centrifuged at 17,000 x *g* for 10 min. Anti-endophilin antibodies (Endophilin II A-11) were coupled in-house to hydrazide-terminated magnetic beads (H-beads). 10 µl of H-beads pre-washed in lysis buffer was incubated with 470 µl of cell lysate overnight at 4°C with end-over-end rotation. The beads were washed 3 times in lysis buffer then boiled in 50 µl SDS sample buffer for 10 min to release bound proteins.

2.11. GST pulldown of GSK3 from RPE1 cell lysate

2.11.1. Serum stimulation/ withdrawal, cell lysis and GST Pulldown

For the experiments described here, A. Casamento prepared RPE1 cell extracts and performed pulldowns, S. Subramaniam performed Western blotting and analysis. RPE1 cells grown to a confluence of 80% in 100 mm dishes and serum levels in cell media were modified prior to cell lysis as follows. For serum starvation, cell media was aspirated and cells were washed 3 times in 2 ml of pre-warmed (37°C) RPE1 cell media without FBS (serum-free; SF). 12 ml of SF media was then added to each dish and cells were incubated in SF media for 2 hours. For serum stimulation, pre-warmed (37°C) FBS was added to cell media to a final concentration of 20% v/v (resting RPE1 cell media contains 10% FBS v/v) and incubated at 37°C for 5 min in a tissue culture incubator. For Resting conditions, no treatment was performed.

Following treatment, cells were washed 3 times with 1 ml ice-cold PBS and harvested by gently scraping into 500 µl Lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100 with protease and phosphatase inhibitors added at 1X working concentration). Samples were incubated on ice for 30 min then sonicated (10 sec On, 30 sec Off, 3 cycles, 50% amplitude) using a probe sonicator (Soniprep 150, MSE). Samples were then centrifuged at 17,000 x *g* for 10 min to remove the nuclear fraction and insoluble material. The resulting supernatant was mixed with 10 µl of GST or GST-EndoSH3-coated magnetic agarose slurry (beads were prepared as described in Section 2.9.2) and incubated for 16-18 hours at 4°C with end-over-end mixing. Using a magnetic rack, beads were washed 3 times in 500 µl Lysis buffer prior to elution in 100 µl SDS sample buffer at 95°C for 10 min. Samples were analysed by western blotting as described in Section 2.3.

2.12. Expression and immunoprecipitation of GFP-tagged constructs in RPE1 cells

2.12.1. Transient overexpression of eGFP constructs

RPE1 cells were seeded into 100 mm dishes (10^6 cells/ dish) and grown for 24 hours. Four dishes were seeded for each sample to generate enough material for MS sample preparation. For each 100 mm dish of RPE1 cells to be transfected, 500 µl of OPTIMEM (Thermo Fisher) was mixed with 15 µl GeneJuice (Millipore) and 3 µg of plasmid was added (V04 or EB1004; see Table 2.7) to the solution, with mixing. Following 20 min incubation at room temperature the transfection mixture was added in a drop-wise manner to cells in full growth medium. Cells were placed in an incubator and used for further analysis ≥24 hours post transfection.

2.12.2. Kinase inhibitor treatment, cell lysis and immunoprecipitation

Cells were grown in an incubator for 24 hours following transfection. For kinase inhibition, cells were treated with the CDK1/2/5/9 inhibitor Dinaciclib (MedChemExpress) and GSK3α/β inhibitor CHIR-9902 (Cayman Chemical Corp.) at a working concentration of 1 µM for 10 min at 37°C in an incubator prior to cell lysis. Following treatment, cells were washed 3 times with 1 ml cold PBS and harvested by gently scraping into 500 µl Lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40 with protease and phosphatase inhibitors added at 1X working concentration) and combining the lysates from all 4 dishes into a single 2 ml sample tube. Samples were incubated on ice for 30 min then centrifuged at 17,000 x *g* for 10 min to remove the nuclear fraction and insoluble material. The resulting supernatant was mixed with 20 µl of GFP-Trap® magnetic agarose slurry and incubated for 3 hours at 4°C with end-over-end mixing to facilitate capture of eGFP-tagged proteins in complex with cytosolic binding partners. Using a magnetic rack, beads were washed 3 times in 500 µl Lysis buffer prior to starting On-bead-SCX digestion as described in section 2.2.3.

2.13. Plasmid constructs used in this study

Name	Vector	Insert	Tag	Promoter	Origin	Insert Species
V04	pEGFP-C2	-	eGFP	CMV	Boucrot Lab	-
EB1004	pCi N-EGFP DEST	EndophilinA2 FL	N-term eGFP	CMV	McMahon Lab	*Homo sapiens*
GV38	pGex 6P2 GST-PrS-DEST	-	GST	tac	McMahon Lab	-
EB5007	pGex 6P2 GST-PrS-DEST	EndophilinA2 SH3 domain (311-end)	GST	tac	McMahon Lab	*Rattus norvegicus*
B148	pEGFP-C1	Dynamin1 FL	N-term eGFP	CMV	McMahon Lab	*Bos taurus*
EB1019	pCi C-EGFP DEST	β_1- Adrenergic receptor FL	C-term eGFP	CMV	McMahon Lab	*Mus musculus*

Table 2.7 Plasmids used in this study. CMV – Cytomegalovirus promotor, FL – full length.

3. Identifying candidate cargoes for fast endophilin-mediated endocytosis (FEME) using label-free quantitative mass spectrometry

3.1. Introduction

3.1.1. Research aims and strategy

The FEME pathway rapidly internalises a range of receptors upon receptor-ligand engagement. At the time of writing, 15 receptors have been identified (Table 3.1), including receptor tyrosine kinases (RTKs), G-protein coupled receptors (GPCRs) and the Interleukin 2 receptor (IL2R) (Boucrot *et al.*, 2015). Given the wide range of functions represented by this pool, it can be assumed that FEME plays a role in diverse cellular processes. Expanding the set of known FEME cargoes will help shed light on the physiological roles of FEME and assist in pathway characterisation, with a view to define therapeutic targets for future studies.

The central role of endophilin in FEME can be exploited for the discovery of novel cargoes. Endophilin acts as an adaptor for cargo recruitment into FEME carriers through its SH3 domain. Direct interactions between the endophilin SH3 domain and proline-rich regions in the third intracellular loop (TIL) of a number of GPCRs were shown to mediate their uptake via FEME (Boucrot *et al.*, 2015). Similarly, for the RTKs EGFR and HGFR, recruitment into FEME carriers was mediated by the endophilin SH3 domain indirectly via the CIN85-Cbl complex. Two approaches to cargo discovery were developed based on the central role of endophilin in FEME.

Receptor	Abbreviation	Type	Publication
β_1-adrenergic receptor	β_1-AR	GPCR	Boucrot *et al.*, 2015
α_{2a}-adrenergic receptor	α_{2a}-AR	GPCR	Boucrot *et al.*, 2015
Dopaminergic D3 receptor	DR3	GPCR	Boucrot *et al.*, 2015
Dopaminergic D4 receptor	DR4	GPCR	Boucrot *et al.*, 2015
Muscarinic acetylcholine receptor 4	mAchR4	GPCR	Boucrot *et al.*, 2015
Epidermal growth factor receptor	EGFR	RTK	Boucrot *et al.*, 2015
Hepatocyte growth factor receptor	HGFR	RTK	Boucrot *et al.*, 2015
Platelet-derived growth factor receptor	PDGFR	RTK	Boucrot *et al.*, 2015
Neuronal growth factor receptor	NGFR	RTK	Boucrot *et al.*, 2015
Insulin-like growth factor 1 receptor	IGF1R	RTK	Boucrot *et al.*, 2015
Interleukin-2 receptor	IL2R	Type I cytokine receptor	Boucrot *et al*. (2015)
Vascular endothelial growth factor receptor	VEGFR	RTK	Boucrot *et al.*, 2015; Genet *et al.*, 2019
Tropomyosin receptor kinase B	TrkB	RTK	Fu *et al*., 2011
Plexin A1	PLXNA1	Semaphorin complex co-receptor	Ferreira *et al*., *submitted*
Roundabout homolog 1	ROBO1	Roundabout receptor	Ferreira *et al*., *submitted*

Table 3.1 Receptors known to internalise via FEME. GPCR – G-protein coupled receptor, RTK – receptor tyrosine kinase.

3.1.2. Part A: Identifying candidate FEME cargoes by comparing plasma membrane extracts from FEME-competent and FEME-blocked cells using quantitative mass spectrometry

The pool of receptors at the plasma membrane is maintained by a homeostatic balance between secretory/ recycling pathways which localise receptors to the plasma membrane, and endocytic pathways which remove them. Blockade of an endocytic pathway's activity can result in accumulation of receptors which are normally internalised via that pathway at the plasma membrane. Endocytic pathways can be blocked by depletion of key adaptor proteins from the cell by RNA interference (RNAi) For instance, Boucrot and colleagues (2015) showed that RNA interference (RNAi)-based knockdown of all three endophilin A proteins blocked formation of FEME carriers and caused steady-state accumulation of β_1-AR and EGFR at the plasma membrane, in cells grown in full serum and without stimulation of the receptors in Table 3.1. It has also been previously shown that RNAi-induced blockade of Clathrin-mediated endocytosis (CME) via knockdown of components of adaptor protein complex 2 causes plasma membrane accumulation of ~40% of plasma membrane proteins as defined by surface biotinylation (Bitsikas, Corrêa and Nichols, 2014), consistent with the notion that a large proportion of receptors at the plasma membrane are internalised via CME. Figure 3.1 shows the hypothetical impact of endocytic blockade on the populations of different receptors at the plasma membrane. Receptors that internalised exclusively via FEME can be expected to accumulate at the plasma membrane when FEME is blocked relative to untreated cells but will not accumulate at the plasma membrane of CME-blocked cells (Figure 3.1), and vice-versa. In reality, receptors often enter cells via multiple endocytic routes so the behaviour of receptors which are internalised by multiple pathways is difficult to predict. In addition, some receptors might only use an endocytic route as a compensatory mechanism upon inhibition of their usual portal of entry into cells. By comparing the levels of receptors at the cell surface under different conditions of endocytic blockade, it should be possible to define a pool of candidate FEME receptors for further validation.

Changes in protein levels of candidate FEME cargoes at the plasma membrane could be assessed by antibody-based methods such as surface labelling followed by fluorescent

microscopy; or by quantitative western blotting of a plasma membrane extract. However, these methods would only target a small number of candidates at a time, would be limited by cost and availability of antibodies against candidates. They would also be biased by candidate preselection.

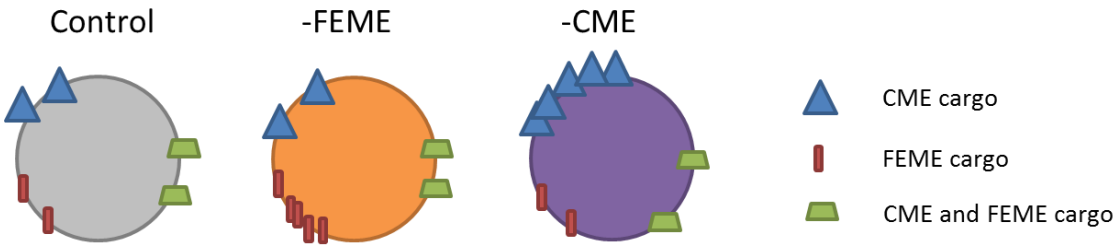

Figure 3.1 Expected effect of endocytic blockade on cell surface receptor populations at the plasma membrane. Circles represent mammalian cells. Other shapes represent cargo receptors which enter cells exclusively via FEME/CME or via both pathways.

Quantitative proteomics provides a powerful alternative for unbiased profiling of subcellular protein populations. Workflows have steadily increased in quality and, given appropriate instrumentation and sample preparation, it is now possible to identify and quantify thousands of proteins from low amounts of complex sample (<20µg) (Wiśniewski *et al.*, 2009; Hughes *et al.*, 2014; Kulak *et al.*, 2014; Sielaff *et al.*, 2017). This opens the possibility of performing quantitative proteomics experiments on plasma membrane isolates and the lower requirements for starting sample amount also help compensate for the cost of experimental procedures upstream of subcellular fractionation.

A proposed quantitative proteomics workflow for comparing surface receptor levels in plasma membrane fractions from cultured mammal cells following RNAi-induced endocytic blockade is described in Figure 3.2. As shown, a number of steps must be optimised. Specific inhibition of a given endocytic pathway can be achieved by disrupting the expression or action of a key pathway component. For example, overexpression of the carboxy terminus of the neuron-enriched CME cargo adaptor AP180 causes a dominant negative phenotype. The COOH-terminus of AP180 has many AP2 and clathrin binding sites, so overexpression titrates away free AP2 and clathrin from the cytosol, reducing the number of copies of available for recruitment into clathrin-coated

pits at the plasma membrane below functional levels (Ford *et al.*, 2001). Despite the efficacy of this method, it relies upon inhibition of clathrin activity which may be undesirable due to disruption of clathrin's other roles in the cell (Raiborg *et al.*, 2001; Royle, Bright and Lagnado, 2005; Deborde *et al.*, 2008). CME inhibition can also be achieved using small molecule inhibitors which target Clathrin such as Pitstop 2 (von Kleist *et al.*, 2011) or Ikarugamycin (Elkin *et al.*, 2016). It can also be achieved by using drugs targeting dynamin such as dynasore (Macia *et al.*, 2006), Dyngo 4a (Mccluskey *et al.*, 2013), quinone, Indole or Dynole series (Hill *et al.*, 2009; Gordon *et al.*, 2013; Macgregor *et al.*, 2014) but these would also block other endocytic pathways, including FEME. Furthermore, the toxicity of these fast-acting inhibitors means they cannot be used for the extended periods of time (multiple days) required to induce significant accumulation of receptors on the cell surface.

An alternative technique for inhibition of CME is through RNAi-induced knockdown of central CME components. Knockdown of specific proteins can be achieved through transfection of small interfering RNA (siRNA) oligomers which target complementary sequences on target mRNAs for cleavage via the RNA induced silencing complex (RISC) and eventual degradation (reviewed in Wilson and Doudna, 2013). The resultant post-transcriptional gene silencing is almost always incomplete (especially for long-lived proteins whose levels drop mostly by dilution upon cell division) and require longer treatment to observe maximal effects but can effectively reduce the concentration of a target protein below functional levels. Effective CME inhibition has been demonstrated by transfecting mammal cells with siRNA oligomers targeting the μ- and α-subunits of the AP2 complex (Motley *et al.*, 2003; Boucrot *et al.*, 2010), as depletion of either subunit is sufficient to destabilise AP2 complexes and induce degradation of the other subunits. Because AP2 is the main adaptor recruiting Clathrin to the plasma membrane, its depletion blocks CME without affecting other cellular functions relying on Clathrin, such as in endosomal sorting (Raiborg *et al.*, 2001), trans-Golgi network export (Deborde *et al.*, 2008) and mitosis (Royle, Bright and Lagnado, 2005)). Similarly, simultaneous targeting of all three Endophilin A proteins using the same technique has been shown to inhibit FEME in an effective manner (Boucrot *et al.*, 2015). Endocytic blockade will be achieved by optimisation of a siRNA knockdown procedure (Figure 3.2).

The second step in the proposed workflow is the isolation of plasma membrane proteins for LC-MS analysis. Due to the hydrophobic nature and low abundance of transmembrane receptor proteins, it is not generally possible to profile cell surface proteomes (surfaceomes) without an enrichment step. Many techniques have been developed for the enrichment of plasma membranes, some of which are explicitly designed for LC-MS coupling.

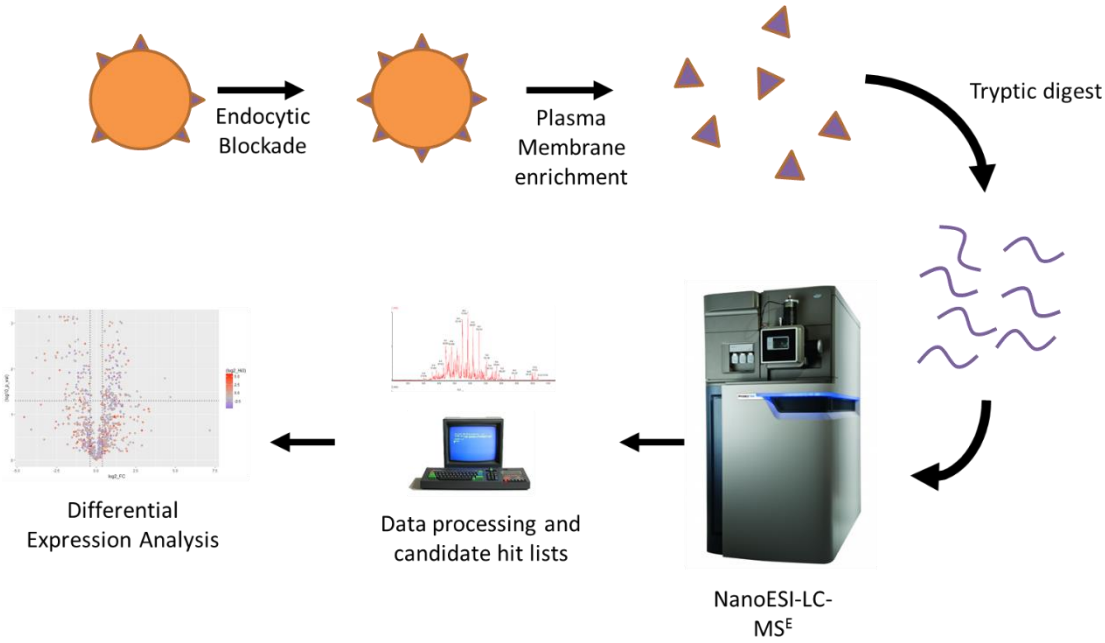

Figure 3.2 Proposed workflow for quantitation of cell surface receptor populations by LC-MS.

Historically, ultracentrifugation of cells following gentle lysis (hypotonic shock, Dounce homogenisation, or needle shearing) was used to separate membranes from soluble proteins. Separation during centrifugation is dependent on differences in buoyant density, size and shape between intracellular proteins and organelles; and is divided into two categories: differential centrifugation and gradient density centrifugation (Brakke, 1953; Duve, 1975; Dormeyer *et al.*, 2008; Lewandrowski *et al.*, 2009). Differential centrifugation separates cellular components primarily on the basis of size, with larger components tending to sediment faster. Gradient-density centrifugation is performed in a medium containing a discrete or continuous density gradient, usually with the use of sucrose or Percoll. Following layering of the sample at the top or bottom of a gradient, centrifugation separates components of the sample on the basis of size (rate zonal centrifugation) or buoyancy (equilibrium centrifugation). An effective technique for

obtaining a plasma membrane fraction is to perform a differential centrifugation to pellet cell membranes following lysis (crude membrane fraction), and then separate this membrane fraction by gradient centrifugation to obtain organellar subfractions (Golgi, endoplasmic reticulum, lysosomes, CCVs and plasma membrane among others). Due to heterogeneity in organelle size, and tight associations between some cytosolic proteins with integral membrane proteins, this technique displays lower purity relative to other plasma membrane isolation approaches.

Biotinylation of cell surface proteins is an additional method for isolation of plasma membrane proteins. A range of chemical linkers have been developed which conjugate a biotin moiety, often with a linker region, to various functional groups on the cell surface (Table 3.2). If a membrane-impermeable biotinylation reagent is used, only proteins exposed at the extracellular face of the plasma membrane will be labelled. This is followed by cell lysis and affinity purification of biotin-conjugated proteins with an avidin resin. The strong biotin-avidin interaction (K_d~10^{-15}M; Green, 1975) allows for harsh washing steps, but presents difficulties during elution from the resin. Indeed, an early study using this technique reported just 2% of protein bound to avidin beads could be eluted through competition using free biotin (deBlaquiere and Burgess, 1999). This problem has been addressed by incorporation of a cleavable linker region in the biotinylation reagent, such as in sulfo-NHS-SS-Biotin (Figure 3.3). This molecule features a disulphide linker which can be cleaved under reducing conditions. Alternatively, enzymatic digestion can be performed on the avidin resin following capture of biotinylated protein, but risks overwhelming the eluate with avidin peptides. Eluted proteins can then be analysed by LC-MS.

The use of sulfo-NHS-SS-Biotin was attractive due to its applicability to both approaches pursued in this work to facilitate the discovery of novel FEME cargo receptors. In addition to profiling quantitative changes in plasma membrane protein amounts between conditions, a complementary experimental workflow was optimised featuring secondary GST-pulldown from a plasma membrane fraction (described in Section 3.1.3). Ultracentrifugation techniques allow for purification of functional receptors but are laborious and produce relatively impure plasma membrane isolates. We reasoned that captured plasma membrane proteins labelled with sulfo-

NHS-SS-Biotin could be eluted under gentle conditions, making this technique suitable for use in both discovery approaches (see also Section 3.1.3). Finally, the chosen reagent is membrane impermeable and reacts with accessible primary amines at the N-terminus or at Lysine side chains. This feature was expected to assist in labelling of GPCRs and RTKs (virtually all of which have extracellular N-termini), both classes of receptors with great relevance to FEME. Thus,

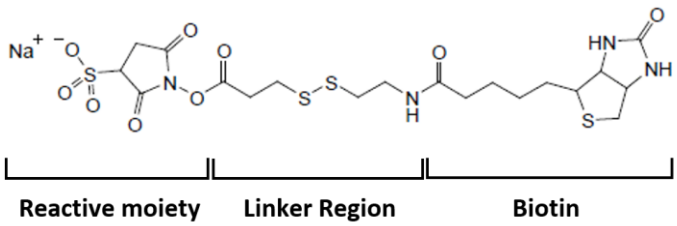

Figure 3.3 Chemical structure of sulfo-NHS-SS-Biotin.

Reactive group	Target
NHS	Primary Amine
Sulfo-NHS	Primary Amine
Pentafluorophenyl	Primary and Secondary Amine
Tetrafluorophenyl	Primary and Secondary Amine
Iodoacetyl	Sulfhydryl
Maleimido	Sulfhydryl
Pyridythiol	Sulfhydryl
Hydrazide	Aldehyde
Amine	Aldehyde

Table 3.2 Biotinylation reagents and their targets.

biotinylation of surface proteins using sulfo-NHS-SS-Biotin was selected for plasma membrane isolation.

A quantitative proteomics workflow was optimised to compare surface receptor levels in plasma membrane fractions from RPE1 cells following RNAi-induced endocytic blockade with the aim of identifying candidate FEME receptors and is described in Part A (Section 3.2).

3.1.3. Part B: Identifying candidate FEME cargoes by GST-Endophilin-SH3 pulldown from a plasma membrane extract

Endophilin proteins play multiple roles in FEME, including cargo recruitment at the plasma membrane. This role is likely to be mediated by direct and indirect interactions between the endophilin SH3 domain and proline rich motifs in target cargoes and adaptor proteins. A workflow was developed to investigate the interaction of endophilin SH3 with integral membrane or membrane-associated proteins.

Due to a number of biological and biochemical factors, the discovery and characterisation of membrane protein-protein interactions has proven challenging. Firstly, most membrane proteins are expressed at relatively low abundances in the cell (~1-100,000 copies/cell), posing a challenge for mass spectrometry detection. This can be circumvented though the use of large amounts of starting material (cultured cells or tissues), but this may be prohibitively expensive if treatment conditions are costly to induce and maintain. Overexpression of receptor proteins can be performed but is risky due to the exponential nature of signalling cascades whereby activation of a single receptor at the cell surface can trigger changes in thousands of downstream effector molecules. Further, integral membrane proteins are anchored in the lipid bilayer by transmembrane (TM) domains made up of hydrophobic amino acids. These TM domains are poorly soluble in aqueous solution, necessitating the use of detergents in most biochemical handling steps, which can interfere with downstream applications such as MS. Removal of detergents requires additional sample preparation steps, which reduce peptide yield and increase cost. Additionally, the wide range of non-denaturing detergents display heterogeneity in their capacity to functionally solubilise membrane protein. Solubilisation conditions must therefore be determined empirically on a case-by-case basis.

These difficulties have prompted researchers to utilise indirect methods to screen for membrane protein interactions. One successful approach which does not necessitate the solubilisation of membrane proteins is the split ubiquitin yeast two-hybrid assay (first described in Johnsson and

Varshavsky, 1994). Pairs of candidate integral membrane protein binding partners are tagged with complementary fragments of the ubiquitin protein, one of which also features a transcription factor for a reporter gene. Close association of the partners *in vivo* causes functional reconstitution of a pseudo-ubiquitin protein, and its consequent cleavage by deubiquitinating enzymes. The transcription factor is released upon cleavage and can translocate to the nucleus to activate reporter gene expression. A recent adaptation of this system for use in mammalian cells shows great promise (Petschnigg *et al.*, 2014; Saraon *et al.*, 2017), but still suffers from several limitations: all tests are conducted pairwise between a set of pre-selected, manually tagged candidate proteins, and overexpression of candidate fusion proteins is a highly artificial situation which may not reflect protein behaviour *in vivo*.

A powerful alternative for the discovery of novel protein-protein interactions is affinity purification-mass spectrometry (AP-MS). In a general AP-MS workflow, the protein or protein domain of interest serves as a molecular bait and is exposed to a lysate or subcellular fraction containing potential prey proteins. The bait is purified by affinity chromatography under conditions which preserve protein-protein interactions, thus copurifying prey proteins. The resulting bait-prey complexes are then subjected to enzymatic digestion for identification by tandem mass spectrometry. AP-MS analyses are therefore not biased by pre-selection of candidates, and do not necessarily rely on overexpression of candidate proteins.

Numerous factors contribute to the success of an AP-MS workflow. The affinity purification must be optimised for specificity and must provide enough starting material (generally a minimum of 20 µg) for mass spectrometry sample preparation. The choice of MS sample preparation method is also reliant on the final elution conditions from the affinity purification step. Therefore, AP-MS experiments targeting membrane proteins must feature optimisation of the bait-prey binding step following detergent solubilisation and must feature appropriate MS sample preparation strategies for detergent removal post-elution.

The endophilin SH3 domain is an ideal bait for use in an AP-MS workflow to discover novel FEME cargoes. SH3 domains are relatively short (~60 amino acids) soluble domains which recognise proline-rich motifs (Kaneko, 2008; Kurochkina and Guha, 2013) and often serves as a dedicated scaffolding domain to promote protein-protein oligomerisation. Boucrot *et al.* (2015) described two modalities by which the endophilin SH3 domain recruits cargo into FEME carriers. GPCR cargoes were recruited by directly contact with targets whereas RTK cargoes were captured indirectly by association with cytosolic adaptors. Boucrot *et al.* (2015) showed that the endophilin A1 and A2 SH3 domains were able to interact directly with proline-rich motifs in a number of GPCR third-intracellular loops (TILs). Consistent with this observation, only GPCRs which could interact with the SH3 domain could be internalised by FEME upon agonist stimulation. The recruitment of EGFR and HGFR into FEME carriers was shown to be dependent on bridging of endophilin to the RTKs by the CIN85 adaptor and Cbl ubiquitin ligase following receptor activation (Petrelli *et al.*, 2002; Soubeyran *et al.*, 2002; Boucrot *et al.*, 2015). The bridging of endophilin to tropomyosin receptor kinase B (TrkB) via retrolinkin is a further example of this indirect recruitment (Fu *et al.*, 2011).

The use of full length endophilin constructs as bait was avoided. Endophilin proteins share a common domain architecture, with an N-terminal N-BAR domain and a C-terminal SH3 domain connected by and a variable linker region. The N-BAR domain features an amphipathic helix (H0) which inserts into lipid bilayers to sense, stabilise or induce membrane curvature (Gallop *et al.*, 2006; Masuda *et al.*, 2006; Chen *et al.*, 2016; Simunovic *et al.*, 2016). Endophilin proteins dimerize through interactions between their BAR domains adopting a rigid crescent-shaped structure (Gallop *et al.*, 2006). In solution, endophilin A1 homodimers appear to display an autoinhibitory mechanism whereby the H0 helix and the SH3 domain make intermonomer, intradimer contacts, preventing both membrane sensing by the H0 helix, and protein-protein interactions by the SH3 domain (Wang *et al.*, 2008; Meinecke *et al.*, 2013; Vázquez, Unger and Voth, 2013; Chen *et al.*, 2014). Indeed a similar autoinhibitory mechanism has been observed in the F-BAR protein Pacsin, where tubulation of liposomes was suppressed by the presence of full-length protein (Wang *et al.*, 2009). Using full-length endophilin as a bait could retrieve mostly autoinhibited endophilin dimer complexes and thus reduce workflow sensitivity. Furthermore,

studies have reported protein-protein interactions between the endophilin family (Ralser *et al.*, 2005; Kristensen, Gsponer and Foster, 2012), raising the possibility that endophilins can heterodimerise with other BAR domain proteins. Thus, use of a full-length endophilin bait could purify complexes with no relevance to FEME, and may be inhibited through interactions with endogenous endophilin. Finally, Boucrot *et al*. (2015) compared the importance of all endophilin A proteins in FEME-mediated GPCR uptake, showing that endophilin A2 plays the most important role in non-neuronal cells, and established that all proteins bound by endophilins A1 and A3 are also bound by endophilin A2. Bait-prey binding conditions were therefore optimised using a GST-EndoA2-SH3 construct.

AP-MS efficacy can also be enhanced through subcellular fractionation. Affinity purification from a total cell lysate exposes the bait molecule to all possible binding partners in the cell, including ones which are ordinarily localised in separate compartments. This heightens the risk of detecting false positive and nonspecific interactions. Furthermore, low abundance true interactors (such as membrane proteins) may be unable to compete with highly abundant soluble proteins for occupancy in bait-prey complexes. During MS analysis, peptides with similar physicochemical properties co-elute during liquid chromatography and compete for charge during ionisation. Peptides from highly abundant proteins in an unfractionated sample can elute over an extended period and partially or completely mask the signal from lower abundance peptides. Pre-enrichment of a relevant cellular compartment reduces interference from nonspecific and/or highly abundant proteins during affinity chromatography and aids mass spectrometric identification of proteins by reducing sample complexity and thus the frequency of co-eluting peptides.

FEME cargo recruitment occurs at the plasma membrane, where transmembrane receptors are localised by secretory pathways. Well-established techniques for the enrichment of membrane proteins rely on separation of a plasma membrane fraction through differential or equilibrium density-gradient centrifugation (Brakke, 1953; Duve, 1975; Castle, 2003; Dormeyer *et al.*, 2008). An alternative methodology relies on biotinylation of surface proteins with subsequent affinity purification (Deblaquiere and Burgess, 1999; Weekes *et al.*, 2010; Bitsikas, Corrêa and Nichols,

2014). A workflow was developed employing recombinant GST-EndophilinA2-SH3 (GST-EndoA2-SH3) to screen plasma membrane fractions from human RPE1 cells for potential FEME cargoes by AP-MS (Figure 3.2). Progress to optimise this workflow is presented in Part B (Section 3.3).

3.2. Part A: Identifying candidate FEME cargoes by comparing plasma membrane extracts from FEME-competent and FEME-blocked cells using quantitative Mass spectrometry

3.2.1. Plasma membrane enrichment via biotinylation of cell surface proteins

As explained in section 3.1.2, we elected to perform plasma-membrane enrichment through biotinylation of plasma membranes using the sulfo-NHS-SS-Biotin reagent. A commercially available plasma membrane isolation kit utilising the cell surface biotinylation paradigm was tested following the manufacturer's instructions (Pierce Cell Surface Protein Isolation Kit) (Figure 3.4). Briefly, surface proteins on RPE1 cells grown to 80-90% confluence in 100 mm dishes were labelled with sulfo-NHS-SS-Biotin, a membrane impermeable reagent reactive against primary amines. The reaction was quenched and cells were lysed in a proprietary lysis buffer. Solubilised, biotinylated proteins were affinity purified from the lysate by exposure to agarose beads coated with NeutrAvidin monomers, which associate strongly with biotin. The bead resin was washed 4

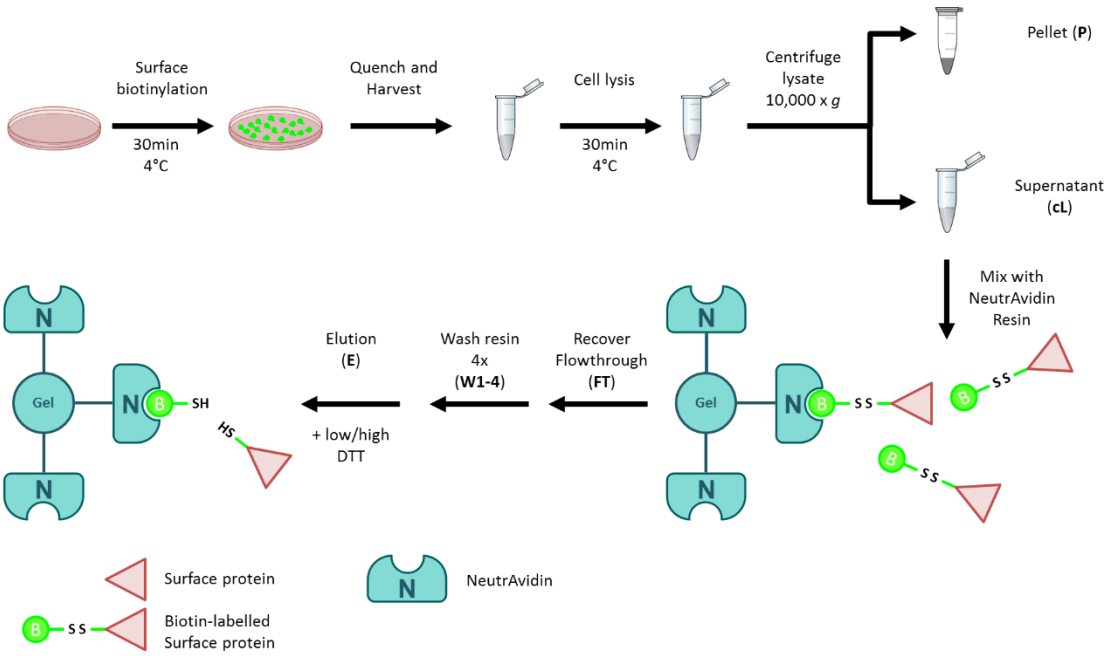

Figure 3.4 Schematic of surface biotinylation followed by affinity purification of plasma membrane protocol. Note the reduction of the disulphide linker in the elution step to release captured surface proteins.

times prior to elution. Cleavage of the disulphide linker by reduction in SDS-PAGE sample buffer supplemented with 50 mM dithiothreitol (DTT) released captured surface proteins in the elution step.

Western blotting was used to characterise plasma membrane eluates obtained using this method. In a subcellular fractionation protocol, marker proteins with well validated localisation to specific cellular compartments can be used as a proxy for the purity of a given fraction. As shown, a number of plasma membrane proteins (EGFR, HGFR, Integrins β1 and αV and the Transferrin receptor) are well solubilised and retained in the eluate (Figure 3.5). Marker proteins for the cytosol (Glyceraldehye 3-phosphate dehydrogenase; GAPDH), endoplasmic reticulum (Calnexin) and nucleus (Histone H3; HH3) were significantly depleted in the eluate. The restriction of the majority of HH3 signal to the Pellet fraction indicates that nuclei remained largely intact under these lysis conditions, in agreement with the manufacturer's claim that a "mild detergent" is used in the proprietary Kit lysis buffer. Notably, some GAPDH signal is visible in the eluate and may be explained by reports that sulfo-NHS-SS-biotin is not strictly membrane-impermeable (Luo, McDonald and Hanrahan, 2008). In addition, GAPDH is a highly abundant housekeeping protein and may simply adhere non-specifically to the surface of the agarose beads used in the affinity purification step.

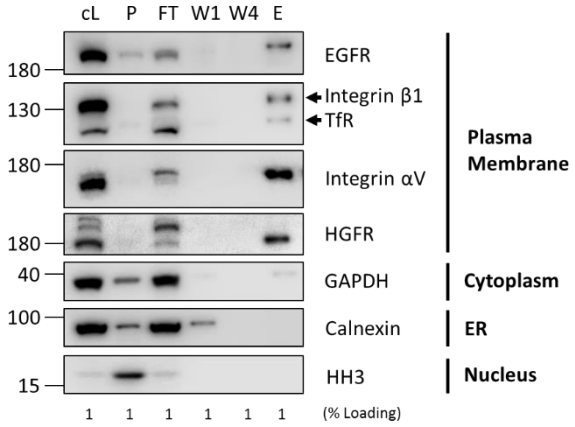

Figure 3.5 Western blotting of affinity purified plasma membranes demonstrates high purity in eluates. The presence of proteins from various cellular compartments in aliquots from different stages of the plasma membrane isolation workflow was assessed by western blotting. Elution was conducted in SDS-PAGE sample buffer with 50mM DTT. cL – Clarified Lysate, P – Pellet, FT – Flowthrough, W1-4 – Wash 1-4, E – Elution, WCL – Whole Cell Lysate.

Comparison of the Pellet (P) and Clarified lysate (cL) fractions gives a measure of the efficacy of the lysis step. The presence of signal in the Pellet fraction for EGFR, GAPDH and Calnexin indicated incomplete lysis, so the ratio of lysis buffer to wet cell pellet volume was increased in future experiments. Interestingly, the capture of surface proteins was non-uniform (Figure 3.5). Comparison of the Clarified lysate, Flowthrough (FT) and Elution signal provides an estimate of the amount of available receptor which is successfully captured. A large proportion of available Integrin αV and Hepatocyte growth factor receptor (HGFR) were captured, as indicated by the near depletion of these species in the FT fraction. In contrast, a much smaller proportion of available Integrin β1, Transferrin receptor (TfR) and especially EGFR are retained in the eluate. This difference could be a function of the availability of extracellular primary amines for biotinylation or could represent an intracellular pool of receptors which are inaccessible to the Sulfo-NHS-SS-Biotin reagent.

Quantitation of protein yield in plasma membrane fractions was important to determine the required quantity of starting material for tryptic digestion. Aliquots of affinity purified plasma membrane fractions from single 80-90% confluent 100 mm dishes were poorly visualised by Coomassie staining. Quantitation was then attempted on an increased scale (6 x 100 mm dishes, 90% confluent RPE1 cells) using a detergent-compatible Lowry assay (Bio-Rad RCDC protein assay), but aliquots of the 400 µl Elution solution were measured below the lower limit of the assay dynamic range (0.2 mg/ml). This set an upper bound on protein amount in affinity purified plasma membrane to 80 µg/ 6 dishes RPE1 (assuming 0.2 mg/ml concentration in 400 µl) To obtain a rough estimate for the lower bound of protein amount in affinity purified plasma membranes, a 360 µl eluate from 6 x 100 mm dishes (90% confluent RPE1) was concentrated 12-fold using a 30 kDa molecular weight cutoff (MWCO) spin filter. The concentrated eluate was assayed with the detergent-compatible Lowry assay at 2 dilution levels (0.7x and 0.4x) with success (Appendix 8.1) to obtain a lower boundary estimate of total protein yield from 6 x 100 mm dishes of 31.5 µg ± 3.15µg (see Appendix Table 8.1, Table 8.2 and Figure 8.1). Given the expected loss of sample through use of the spin filter, and further possible losses of proteins larger than 30 kDa due to denaturing in SDS, this figure represents a lower boundary and the

yield of plasma membrane protein. Therefore, the yield of biotinylated surface protein from 6 x 100 mm dishes of RPE1 cells lies between 30 µg – 80 µg.

3.2.2. Optimisation of siRNA knockdown experiments

RNAi-induced endocytic blockade of FEME and CME was optimised in RPE1 cells in order to generate biologically informative alterations in receptor populations at the cell surface for subsequent LC-MS analysis (see Figure 3.2). Endocytic blockade was first attempted on RPE1 cells growing in 100 mm dishes. For CME suppression, 625 pmol of Invitrogen Stealth siRNA oligonucleotides targeting the µ2 subunit of adaptor protein complex 2 (HSS101955: 1 oligo targeting Human *AP2M1*) were used (AP2M KD). For FEME suppression, (Endophilin triple knockdown: EndoTKD) 625 pmol each of Invitrogen Stealth siRNA oligonucleotides targeting all three endophilin A proteins (HSS109709: 1 oligo against human *SH3GL2*; HSS109707: 1 oligo against *SH3GL1*; HSS109712: 1 oligo against human *SH3GL3*) were co-transfected into RPE1 cells. Control cells were treated with 625 pmol of Invitrogen Stealth control (scrambled) oligo 138782.

Anticipating reduced cell growth in FEME- and CME-blocked cells, RPE1 cells were seeded onto 100 mm dishes at two different densities (2.1×10^5 cells/dish for Control dishes, 2.3×10^5 cells/dish for AP2M KD and EndoTKD dishes), 6 hours prior to transfection. siRNA oligonucleotides and Oligofectamine transfection reagent were diluted separately in 500 µl OptiMEM (Thermo Fisher), gently mixed and incubated at room temperature for 20 min before pipetting dropwise onto cells in culture media (Day 1). A second transfection was performed 24 hours after the first transfection, preceded by exchange of cell media 1 hour in advance (Day 2). Cell media was changed 24 hours after the second transfection (Day 3), and cells were harvested on Day 4 or 5. The use of 2 transfections increases the percentage of cells transfected, improves the longevity of RNAi and the extended period of growth (4-5 days) prior to harvesting allows for multiple cell divisions to dilute the pool of target protein with a low turnover rate to achieve maximal knockdown efficiency. These optimisations were shown to be essential for achieving >90% reductions µ2-Adaptin levels (Motley *et al.*, 2003) and >80% reduction in endophilin A protein levels (Meinecke *et al.*, 2013; Boucrot *et al.*, 2015) through RNAi.

Finally, it has been shown previously that knockdown of µ2-Adaptin causes instability and degradation of other AP2 subunits including α-Adaptin (Motley *et al.*, 2003; Boucrot *et al.*, 2010). α-Adaptin is therefore an excellent proxy for AP2M knockdown and is used as such hereafter.

Total cells extracts were obtained from siRNA-treated cells by harvesting them in 500 µl SDS sample buffer on Day 4 of the siRNA transfection procedure. Lysates were mildly sonicated (three

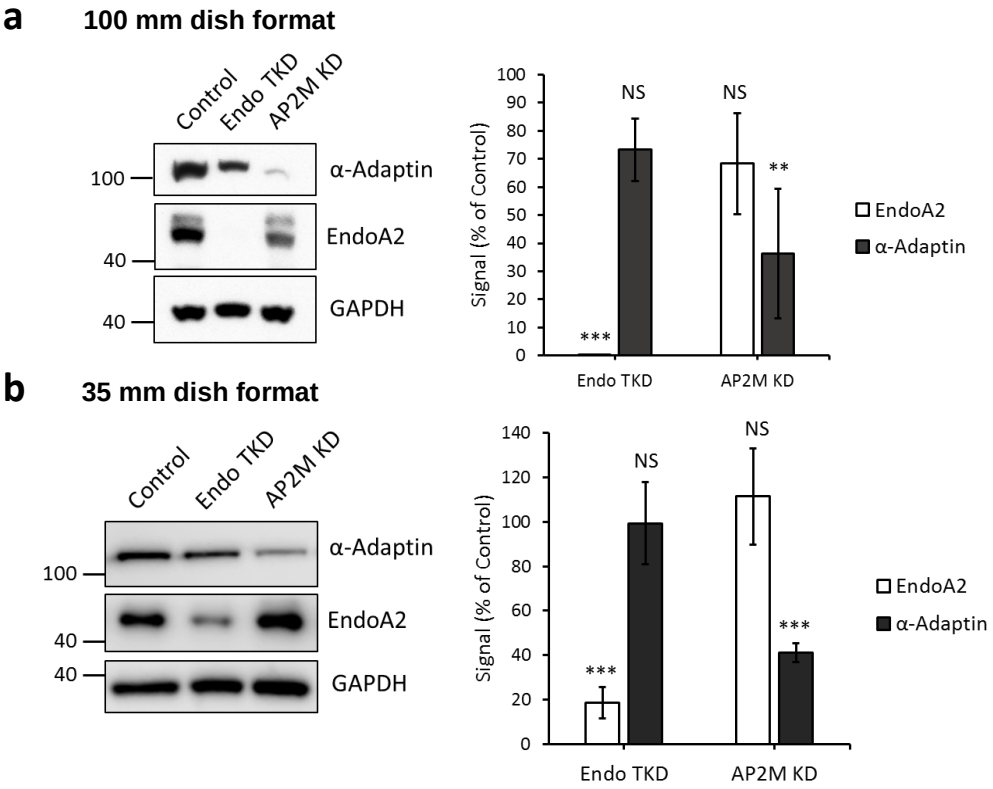

Figure 3.6 Optimisation of siRNA knockdown using Oligofectamine transfection reagent. **a**, Left, representative western blot showing α-Adaptin and EndoA2 HRP signal upon different siRNA treatments. Right, Densitometric quantitation of siRNA efficacy. Bar graphs show GAPDH-normalised α-Adaptin/EndoA2 signal expressed as a % of Control levels within each treatment condition (means ± s.e.m. from three independent experiments; *P < 0.05, **P < 0.01, ***P < 0.001; 2-sided, unpaired Student's *t*-test versus Control). **b,** Left, representative western blot showing α-Adaptin and EndoA2 HRP signal from siRNA experiments in 35 mm dishes. Right, Densitometric quantitation of siRNA efficacy. Bar graphs show GAPDH-normalised α-Adaptin/EndoA2 signal expressed as a % of Control levels within each treatment condition; means ± s.e.m. from four independent experiments; *P < 0.05, **P < 0.01, ***P < 0.001; 2-sided, unpaired Student's t-test versus Control.

10 second pulses, 10 µm) to shear DNA, then centrifuged (16,000 x*g*, for 5 min) to remove insoluble material. Clarified lysates were then analysed by western blotting. Densitometric analysis of western blot images indicated highly significant and almost complete knockdown of EndoA2 levels (0.2% ± 0.1% of Control; P < 0.0001) in the EndoTKD condition (Figure 3.6a).

However, knockdown of µ2-adaptin appeared to be incomplete in the AP2M KD condition (36.3% ± 23.1% of Control) with a high degree of variability in RNAi.

Following RNAi induced knockdown of µ2-adaptin to levels undetectable by western blotting, it is still possible for clathrin-coated pits to form (though at a greatly reduced density and rate) (Motley *et al.*, 2003; Boucrot *et al.*, 2010). It is probable that a population of cells with ~36% remaining µ2-adaptin may retain enough CME function to reduce the magnitude of receptor accumulation at the plasma membrane and thus reduce workflow sensitivity.

Efforts to improve AP2M KD were undertaken in a smaller format (35 mm dishes) to conserve reagents and test multiple conditions. Inconsistent cell growth was observed whilst performing the initial 100 mm dish experiments and likely contributed to the large variability in knockdown performance observed in AP2M KD samples (Table 3.3, Original). To ameliorate this, small scale trials were conducted to ascertain a suitable seeding density for a 4-day RNAi protocol in 35 mm dishes (Table 3.3). These trials indicated that seeding 48,000 cells for Control and AP2M KD single knockdown conditions, and 96,000 cells for the EndoTKD triple knockdown condition would provide optimal growth during a 4-day RNAi experiment.

Trial	Dish size	No. of cells seeded in Control	Seeding ratio (Control:EndoTKD:AP2M)	Control (Day 4)	EndoTKD (Day 4)	AP2M KD (Day 4)
Original	100	210,000	1:1.1:1.1	68%±23%	35%±21%	53%±28%
1	35	48,000	1:3:1	80%	95%	70%
2	35	96,000	1:3:1	95%	100%	100%
3	35	48,000	1:2:1	85%	90%	80%

Table 3.3 Optimisation of initial seeding conditions and seeding ratios for siRNA experiments. Original Trial refers to the data shown in Fig 3.3. Confluence is expressed as mean ± s.e.m. where repeated measures are available.

Triplicate analysis of siRNA treated RPE1 cells in 35 mm dish format using optimised seeding was performed. Densitometric analysis of western blot images show that these technical adjustments had mixed effects (Figure 3.6b). Firstly, the quality of EndoA2 knockdown in the EndoTKD condition deteriorated relative to initial experiments in the 100mm dish format (Figure 3.3a), with a mean of 18.6% ± 7% of Control levels. The optimisations appeared to greatly reduce variability in α-Adaptin levels following AP2M KD (almost 5.5-fold decrease in s.e.m. relative to original trial in Figure 3.6a), although the levels of α-Adpatin still indicated incomplete knockdown (41% ± 4.2% of Control).

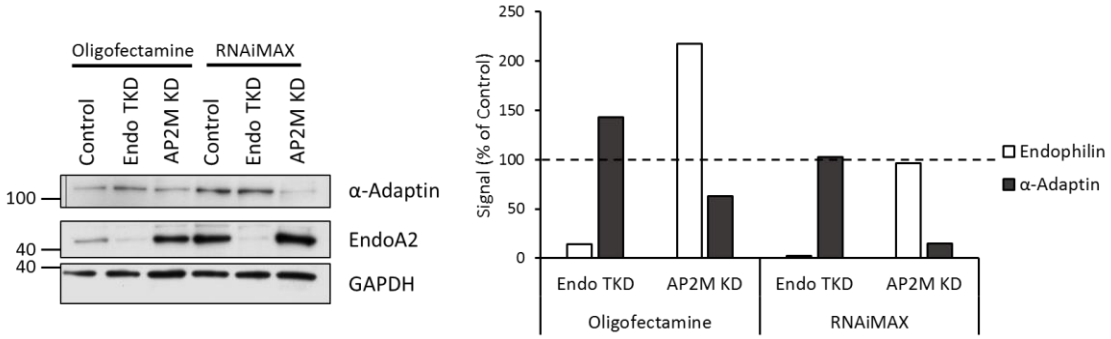

Figure 3.7 Comparison of siRNA knockdown efficiency using different transfection reagents. Left, western blot to show α-Adaptin and EndoA2 HRP signal upon siRNA transfection with Oligofectamine or RNAiMAX reagent. Experiments were performed in 35 mm dishes. Right, Densitometric quantitation of siRNA efficacy. Bar graphs show GAPDH-normalised α-Adaptin/EndoA2 signal expressed as a % of Control levels within each treatment condition (n=1).

In an effort to improve AP2M KD performance, siRNA was transfected into RPE1 cells using an alternative transfection reagent: Lipofectamine RNAiMAX (Thermo Fisher). Knockdown was performed exactly as previously described, but with exchange of the transfection reagent in 35 mm dishes. A set of experiments was performed using Oligofectamine in parallel. When treated with RNAiMAX-siRNA complexes, Endophilin levels in EndoTKD were reduced to 2% of Control levels, indicating excellent knockdown (Fig 3.7). Interestingly, the reduction in α-Adaptin signal observed in RNAiMAX-treated AP2M KD cells (14% of Control) was the best knockdown observed for α-Adaptin so far, representing a 4.5-fold improvement in knockdown efficacy over the corresponding Oligofectamine-treated cells (Figure 3.7). Despite the use of a single biological replicate, these results were promising enough to warrant further trials using RNAiMAX.

Seeding trials were performed to assess the optimal seeding density for 4-day siRNA treated using RNAiMAX in 100 mm dishes. These trials established optimal Day 1 seeding densities for Control (200,000 cells), EndoTKD (750,000 cells) and AP2M KD (400,00 cells) representing a seeding ratio of 1 : 3.75 : 2 (Control: EndoTKD: AP2M KD). Using these optimised conditions, a >95% reduction in target proteins was achieved following siRNA treatment of RPE1 cells in 100 mm dishes (Figure 3.8). These conditions were taken forward and used for the discovery of FEME cargoes as described in section 3.2.4.

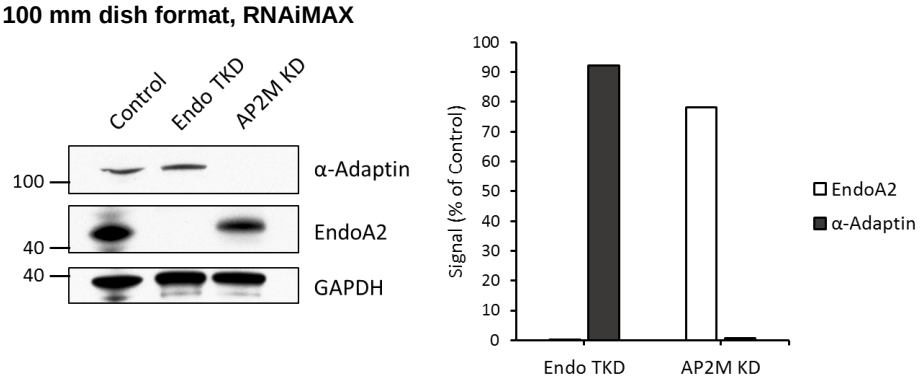

Figure 3.8 The use of RNAiMAX transfection reagent causes large reductions in target protein levels following siRNA transfection in 100 mm dishes. Left, western blot to show α-Adaptin and EndoA2 HRP signal upon siRNA transfection with RNAiMAX reagent. Experiments were performed in 100 mm dishes. Right, Densitometric quantitation of siRNA efficacy. Bar graphs show GAPDH-normalised α-Adaptin/EndoA2 signal expressed as a % of Control levels within each treatment condition (n=1).

3.2.3. Optimisation of a liquid chromatography-mass spectrometry workflow for quantitation of membrane proteins

Traditionally, sample preparation for mass spectrometry involves preparation of a protein sample for enzymatic digestion to generate peptides (by denaturing then alkylation and reduction of cysteines to prevent refolding during digestion), followed by peptide recovery and clean-up. Many reagents commonly used in the preparation of protein lysates are detrimental to MS performance or enzymatic digestion. For example, chelating agents such as EDTA, salts and detergents can suppress ionisation during ESI, either by competing with peptides for charge or by ion-pairing and neutralising analytes (Xu *et al.*, 2005). Detergents such as sodium dodecyl sulphate (SDS) or chaotropes such as 8M Urea will denature and inactivate enzymes at concentrations commonly used in lysis buffers. In addition to these requirements, the target proteins of this analysis – membrane proteins – pose a number of challenges for conventional sample preparation in bottom-up proteomics. They possess dedicated hydrophobic regions which make them difficult to solubilise, are relatively low in abundance compared to cytosolic proteins and tend to produce large hydrophobic peptides upon tryptic digest, due to the low availability of lysine and arginine residues for cleavage in hydrophobic transmembrane regions. Techniques developed for the preparation of membrane protein samples for LC-MS often feature pre-enrichment of membrane proteins followed by “in-gel” or “in-solution” digestion.

Pre-enrichment for plasma membrane proteins has been achieved using biotinylation and affinity purification of surface proteins (Section 3.2.1). Following capture and washing of surface proteins, an in-gel or in-solution paradigm can be followed.

In-gel digestion involves resuspension or direct elution of a sample in SDS sample buffer. The sample is boiled and subjected to 1 or 2-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE) to denature proteins and remove charged contaminants such as salts, with the added advantage of fractionating the sample by molecular weight (Shevchenko *et al.*, 2007). Proteins in a gel lane are then manually excised, and the gel pieces are destained and washed.

Advantages of this method for the digestion of membrane proteins include elution of membrane proteins in SDS, which reduces the risk of sample precipitation prior to digest, and the opportunity to apply enzymes to fully unfolded proteins, avoiding potential steric hindrance from detergent molecules in in-solution techniques. Drawbacks of this method include labour-intensive sample processing and relatively poor peptide yield due to enzyme accessibility issues. Of particular concern for quantitative workflows, in-gel techniques feature more handling steps than in-solution techniques and thus introduce greater experimental error, especially during gel excision steps.

Using an in-solution paradigm, the captured proteins could be eluted NeutrAvidin beads using a 8M Urea/ DTT solution, then alkylated. As trypsin retains activity in low concentrations of Urea (<2 M), the sample can then be diluted and subjected to trypsin digest. This option is unattractive due to the danger of sample loss when diluting mass-limited samples. When handling samples with very low amounts of protein (<100 µg), loss of protein through adsorption to the walls of sample tubes is a constant danger. This can be ameliorated by reducing the number of handling steps, maintaining peptides and proteins at high concentration (i.e. low volumes) and the use of Lo-bind sample tubes (Feist and Hummon, 2015). For these reasons, a traditional in-solution digest was not considered.

Alternatively, filter-aided sample preparation (FASP) is an attractive in solution/ on-filter digest method for membrane proteins. The sample can be eluted in SDS, and is then desalted and detergents removed through centrifugation and buffer exchange in a spin-filter prior to on-column tryptic digest (Wiśniewski *et al.*, 2009). Importantly, by minimising handling steps and permitting the use of SDS during upstream steps, the technique has been shown to be effective in situations where sample starting material is low (<20 µg) (Sharma *et al.*, 2012), although sample losses during preparation can be as high as 50%.

Both traditional in-solution digest and FASP rely on an initial elution step, followed by sample alkylation and digestion. A number of “on-bead” digestion strategies have been reported which skip this elution step, featuring addition of trypsin to captured proteins without elution (Hubner *et*

al., 2010; Fukuyama *et al.*, 2012; Turriziani *et al.*, 2014). This has the advantage of minimising handling steps, labour time and quantitative reproducibility of the workflow. In the context of membrane proteomics, this method is limited by high dependency on the accessibility of extramembrane regions to enzymatic digest and on the effective removal of detergents during washing steps prior to digestion.

The efficacy of in-gel digestion, FASP and on-bead digestion were tested using samples obtained from plasma membrane isolation (without siRNA knockdown) from RPE1 cells. To improve sensitivity, peptide level fractionation using strong cation exchange (SCX) was also trialled in conjunction with an on-bead digest ('On-bead-SCX'). A FASP protocol was also attempted, but owing to poor initial results, was not revisited.

In-gel digestion was performed on a plasma membrane fraction obtained from a single 100 mm dish of RPE1 cells (90% confluent). Briefly, biotinylated cell surface proteins were captured on neutravidin beads and washed. Beads were incubated in SDS-PAGE sample buffer (with 50mM DTT) for 1 hour at room temperature to release bound protein. 50 µl of the eluate was subjected to SDS-PAGE then stained with a mass spectrometry-compatible silver stain. Gel lanes were cut into 6 bands, and bands were further diced into 1 mm^3 pieces. Gel pieces were destained, reduced and alkylated and finally washed using alternating incubations in ammonium bicarbonate and acetonitrile. Gel pieces were incubated in 10 ng/µl trypsin overnight at 37°C, and peptides were recovered by incubation in 1% formic acid, 2% acetonitrile extraction buffer and a final elution in acetonitrile. Peptides were then dried in a vacuum concentrator (SpeedVac DNA 120; Thermo Scientific) and re-solubilised in the desired volume of mobile phase solution (3% LC-MS grade acetonitrile, 0.1% formic acid in H_2O) and processed by LC-MS.

On-bead digestion and On-bead-SCX was performed on six 100 mm dishes of 90% confluent RPE1 cells. Briefly, biotinylated cell surface proteins were captured on neutravidin beads and washed then was again in TBS to reduce detergent concentration. Beads were resuspended in

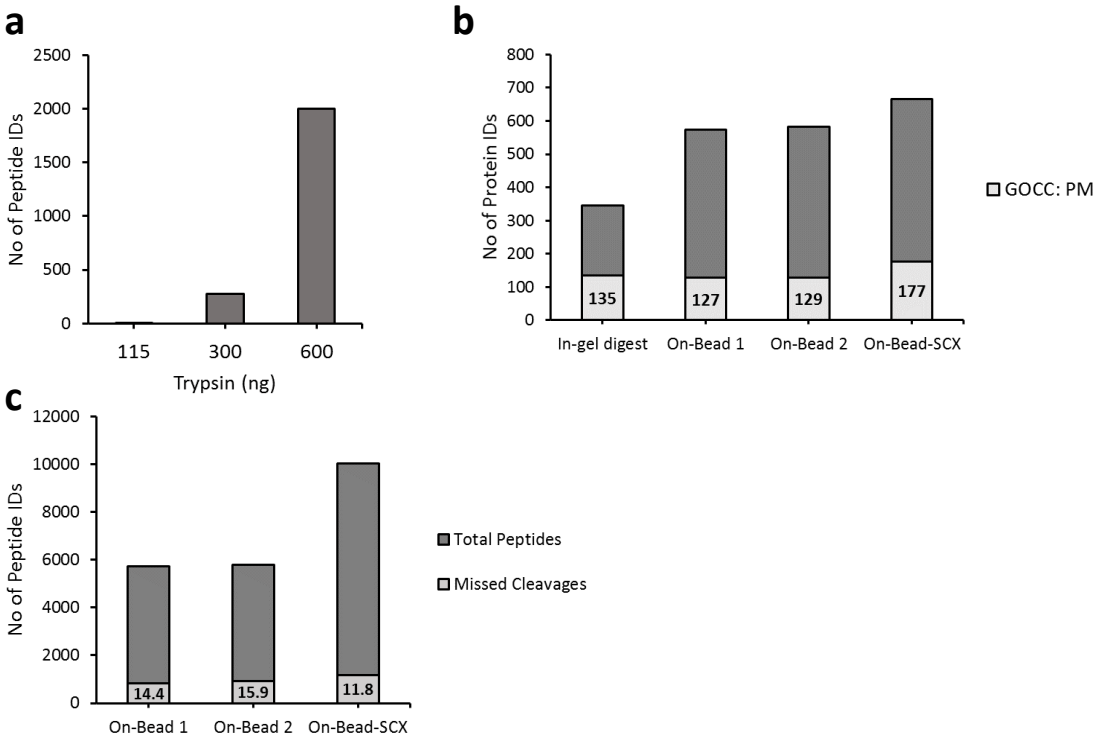

Figure 3.9 Comparison of sample preparation techniques for quantitative membrane proteome profiling on the basis of protein and peptide identifications. **a**, Bar chart showing number of peptides detected upon trypsin digest of equimolar amounts of plasma membrane fraction. **b**, Stacked bar chart to show total proteins ID/ condition. Numbers inset represent absolute number of IDs with GOCC:plasma membrane annotation. **c**, Stacked bar chart to compare number of peptide IDs in each On-bead/ On-bead-SCX workflow. Inset numbers represent % peptides with missed cleavages.

a Tris-Urea buffer containing 115/300/600/1800 ng of trypsin as indicated and incubated at 27°C for 30 min. Trypsin solution was transferred to a fresh sample tube and incubated at room temperature overnight for a combined reduction/ digestion step. Alkylation was performed on the following day, and peptides were subjected to solid-phase extraction (SPE) StageTip desalting. For On-bead-SCX samples, StageTip cleanup was coupled to 5-fraction SCX. Peptides were then dried in a vacuum concentrator and prepared for LC-MS analysis.

During in-gel digestion, trypsin diffuses through the polyacrylamide matrix to access and cleave the denatured proteins within. Due to the low efficiency of this process, a relatively high amount

of trypsin (10 ng/µl) is usually employed to ensure complete digestion of the sample. For on-bead digestion workflows, enzymatic cleavage is performed in solution, so lower amounts of trypsin can be used. This is desirable as it reduces the level of contaminating protease peptides in the final sample, improving quantitation of co-eluting peptides which otherwise may have been poorly ionised.

A plasma membrane fraction from 6 x 100 mm dishes of RPE1 (90% confluence) was split into three equal parts and subjected to on-bead digestion with 115 ng, 300 ng or 600 ng trypsin (Figure 3.9a). Equal volumes of the resulting peptide samples were subjected to LC-MS analysis using a 90 min gradient, and peptide identifications were used to compare digest efficacy. The least efficient digest was with 115 ng (just 1 peptide detected) and the most effective digest was achieved using 600 ng trypsin (275 peptide IDs). Scaling this ratio up, sample preparation of plasma membrane isolations from a larger amount of starting material (six 100 mm dishes of RPE1 cells) was conducted using 1.8 µg trypsin.

The results of each workflow were compared on the basis the following parameters: overall sensitivity, plasma membrane protein sensitivity, plasma membrane purity, digest quality and practicality (Figure 3.9 and Table 3.4). On-bead-SCX outperformed on-bead digestion without fractionation, identifying ~15% more protein groups (667 vs 579). Despite the use of 6-fold less starting material, in-gel digestion identified 345 proteins (Figure 3.9b). Gene ontology cellular compartment (GOCC) annotations are manually curated or computationally generated assertions of a protein's subcellular localisation. In this study, proteins with a GOCC "plasma membrane" annotation were deemed to be plasma membrane proteins. In absolute terms, On-bead-SCX identified the highest number of proteins with a plasma membrane GO annotation (177) (Figure 3.9b). Interestingly, in-gel digestion displayed comparable plasma membrane sensitivity to on-bead digestion (135 and 128 proteins, respectively), and the highest plasma membrane purity (39.1%) of all datasets tested. The deciding factor between these digestion strategies is the proportion of plasma membrane identifications of sufficient quality to provide quantitative data (≥ 3 peptides). Again, On-bead-SCX performed best with 149 quantifiable plasma membrane

	In-gel digestion	On-bead digest	On-bead digest-SCX
No. of Protein IDs	345	579 ± 6	667
No. of Peptide IDs	*	5757 ± 56	10,033
GOCC: Plasma Membrane	135	128 ± 1	177
% PM	39.1	22.1	26.5
Quantifiable GOCC: Plasma Membrane	131	110 ± 1	149
% Missed Cleavages	*	14.4 ± 1	11.8
Starting material (RPE1 cell culture)	1 x 80-90% confluent 100 mm Dish	6 x 80-90% confluent 100 mm Dish	6 x 80-90% confluent 100 mm Dish
Average sequence coverage (%)	23.3	24.6 ± 0.1	19.9
No. of Fractions	6	1	5
Protocol length	4 hours + overnight digest**	1 hour + overnight digest	2 hours + overnight digest
Total instrument time	9 hours (90 min/ fraction)	120 min	7.5 hours (90 min/ fraction)

Table 3.4 Comparison of sample preparation techniques for profiling affinity purified plasma membrane fractions by LC-MS. Values in On-bead digestion column represent means ± standard deviation (n=2). * Peptide data unavailable. ** Increases rapidly with higher numbers of samples.

proteins, with in-gel digestion performing well (131 proteins) and on-bead digestion without fractionation identifying ~110 quantifiable plasma membrane proteins. Interestingly, overall quality of digests as assessed by average sequence coverage was comparable across techniques ranging from a low of 19.9% (On-bead-SCX) to a high of 24.6% ±0.1% (On-bead digest) (Table 3.4). Similarly, quantifying the number of missed cleavages in each on-bead dataset did not reveal striking differences (Figure 3.9c).

Comparison of isoelectric point (pI) and molecular weight (MW) at the protein level were performed to assess if the sample preparation methods displayed any physicochemical biases (Figure 3.10). All techniques exhibited a strong bias towards identification of proteins with an isoelectric point below physiological cell pH (pH 7.4). In-gel digestion exhibited the largest spread in pI and a bias towards larger proteins (Figure 3.10a). It was hypothesised that the bias towards larger proteins could be caused by the more effective fractionation in the in-gel digestion protocol. Consistent with this, in-gel sample fractions contain ~4-fold higher percentage of unique proteins per fraction than On-bead-SCX fractions (Figure 3.10b).

Practical aspects of each technique were also compared. In-gel digestion was the most laborious, and the relatively high number of fractions (usually 8-20 in the literature) necessitate very long sample preparation and analysis times. On-bead digestion without further fractionation provided the best balance of labour to sample coverage, with the addition of a 5-fraction SCX step only boosting sensitivity by 15%.

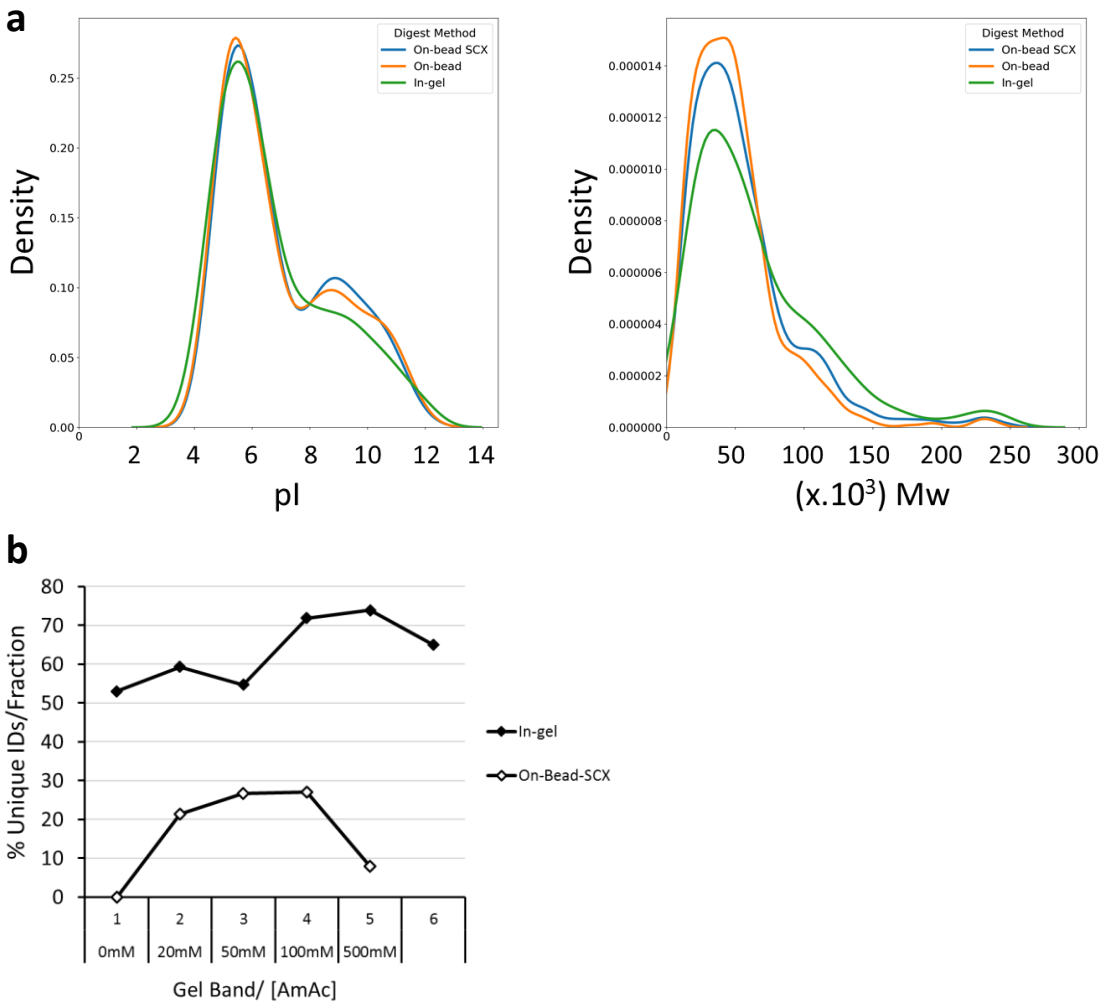

Figure 3.10 Fractionation efficiency assessment and physicochemical properties of identified proteins using different sample preparation strategies. **a**, Left, density plot to show spread of protein pI scores. Right, density plot to show distribution of protein MW. **b**, Line chart to show % of Unique protein IDs/ fraction. The x-axis features numbered gel bands or Ammonium acetate concentration in SCX fractionation solution.

In this analysis, in-gel digestion displayed superior purity of plasma membrane protein despite much lower starting amounts. However, the inevitable introduction of experimental error from this technique, and its burgeoning cost in labour and instrument time upon sample multiplexing,

represented significant drawbacks. A compromise approach “In-gel-SCX” was proposed. Proteins would be eluted in SDS-PAGE sample buffer and run on a gel for a very short time period, allowing excision and in-gel digestion of the whole sample in a single gel band. This would be followed by fractionation at the peptide level by SCX. The compromise protocol was applied to the analysis of plasma membrane proteins, as described in section 3.2.4.

3.2.4. Identifying candidate FEME cargoes by comparing plasma membrane extracts from FEME-competent and FEME-blocked cells using quantitative Mass spectrometry

A complete optimised workflow was performed to assess quantitative differences in plasma-membrane receptor populations under conditions of CME or FEME endocytic blockade (Figure 3.2). Six 100 mm dishes were seeded at optimised densities for each RNAiMAX treatment: Control, EndoTKD and AP2M KD (See section 3.2.2). Experiments were performed in biological triplicate. Following treatment with 625 pmol of respective siRNA oligos, plasma membrane

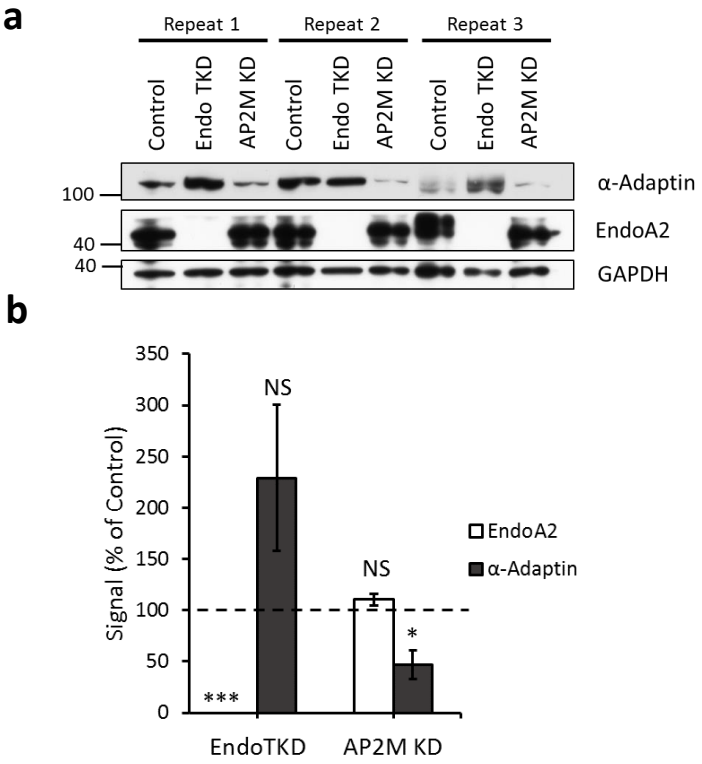

Figure 3.11 Assessment of siRNA knockdown for full discovery workflow. **a**, Western blotting to assess knockdown efficacy of EndoTKD and AP2M KD treatments. **b**, Bar charts to show densitometric quantification of western blot shown in **a**. EndoA2 and α-Adaptin signals are shown following normalisation to GAPDH loading control and relative to within-biological repeat Control signal (means ± s.e.m. from 3 independent experiments; *P < 0.05, **P < 0.01, ***P < 0.001; 2-sided, unpaired Student's *t*-test versus Control)

fractions were harvested on Day 4 using by affinity purification following cell surface biotinylation. An aliquot of clarified lysate was retained for assessment of knockdown efficacy. Following capture of biotinylated protein using and washing of the plasma membrane fraction, proteins were prepared for LC-MS analysis using an 'In-gel-SCX' protocol.

Captured proteins were released through cleavage of the Sulfo-NHS-SS-Biotin linker through incubation in SDS-PAGE sample buffer containing >50 mM DTT at room temp for 1 hour. Samples were not boiled as this was found to greatly increased the incidence of Avidin monomers in eluates (data not shown). Following elution in 300 µl SDS-sample buffer, samples were normalised using quantitation of a silver stain, and up to a 50 µl of a sample was loaded into a gel for 1-dimensional SDS-PAGE. After running the gel for a short time (< 3min), and before the lagging electrophoretic front had exited the stacking gel, a second aliquot of protein (up to 50 µl, in accordance with normalisation), was loaded. Electrophoresis was resumed and proteins were allowed to proceed ~1 cm into the gel before stopping the run and excising the whole sample as a single gel band. Gel bands containing whole samples were subjected to in-gel digestion as per the protocol described in Section 3.2.3, but with the use of 300 ng of trypsin for overnight digest in accordance with the lower estimate of ~30 µg of capturable material using the surface biotinylation protocol optimised in section 3.2.1. Following in-gel digestion, the resultant peptides were resuspended in 0.5% acetic acid and subjected to SCX fractionation using 80 mM, 250 mM and 1000 mM Ammonium acetate as elution buffers. Peptides from each of the 3 fractions were dried in a vacuum concentrator and prepared for LC-MS analysis separately.

Western blot analysis of the clarified lysate from plasma membrane harvests was performed to assess the efficacy of RNAi. RNAi-induced knockdown of Endophilin was highly effective with with almost no Endophilin A2 signal detectable in the EndoTKD condition in any of the biological repeats (Figure 3.11). Endophilin A2 levels in the AP2M KD condition did not significantly different from control levels and did not display wide variation (110% ± 6% of Control) However, despite optimisation for seeding density and RNAiMAX-mediated transfection in 100 mm dishes, RNAi-induced knockdown of α-Adaptin was incomplete (46% ± 14% of Control). This indicated that CME was not fully inactivated in AP2M KD samples.

Peptides from fractionated samples were analysed by nano-ESI-LC-HDMS^E using a 40 min linear gradient in technical triplicate. Proteins were quantified using the Hi3 method, which uses the sum of intensity for the top 3 best ionising peptides for each protein identification as a proxy for

the protein's concentration in the sample (Silva *et al.*, 2006). A minimum of 3 peptides per protein were required, therefore, for consideration in quantitative analysis. A total of 1130 proteins were identified in all conditions. The Hi3 protein intensity for all technical repeats was normalised to total intensity protein intensity within each replicate. The median protein intensity from normalised technical replicates was then summed across all fractions from the same sample to generate a Hi3 intensity score for each protein within a sample. Due to the stochastic nature of peptide ionisation during ESI, missing values are often found in MS datasets. Only proteins that were observed in at least 2 of the 3 biological replicates were carried forward for quantitative analysis, leaving a subset of 560 proteins.

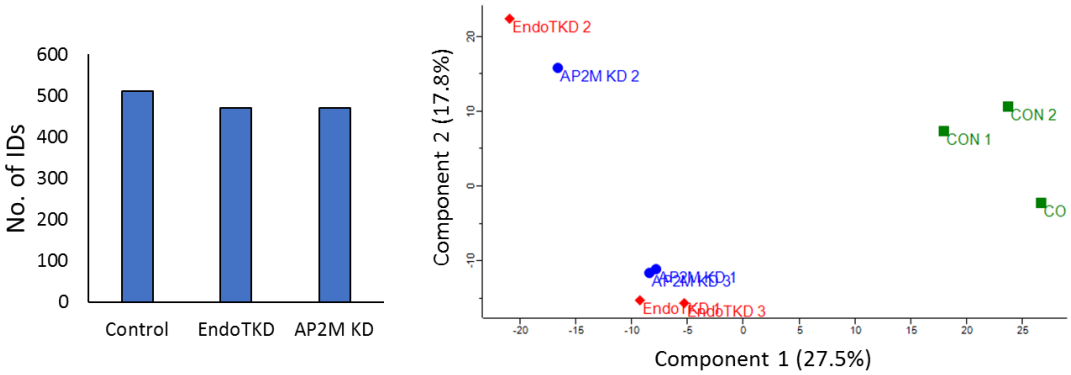

Figure 3.12 Global assessment of Plasma membrane dataset. Left**,** Bar chart to show total number of IDs with a minimum of 2 valid values in all 3 datasets. Right, Principle component Analysis (PCA) plot performed on relative Hi3 protein abundance to represent data similarity in 2 dimensions. CON – Control, EndoTKD – Endophilin triple KD, AP2M KD – µ-Adaptin knockdown. Numbers indicate biological replicate identifier.

Comparing the dataset on a global level, the total number of proteins identified with a minimum of 2 valid measurements was consistent across each condition. Principle component analysis on relative Hi3 protein abundance indicated that samples separated into 3 distinct groups. All Control samples were clustered and siRNA-treated samples from biological repeats 1 and 3 were separated from siRNA-treated samples from biological repeat 2.

Further investigation revealed that a lower quantity of protein has been measured in these runs, resulting in a much lower no. of protein IDs (~250 vs 400+) and lower correlation in protein abundance when compared with other samples in the same condition. These 2 repeats were

therefore excluded from further analysis. The final dataset for ANOVA consisted of 283 proteins, 129 of which annotated with the GOCC:plasma membrane term.

Variation in this final dataset was analysed. Repeated measurements of Hi3 intensity within technical replicates displayed a median Pearson correlation coefficient of 0.85. Median Pearson correlation among biological replicates was also found to be 0.85, indicating high technical and biological reproducibility (Figure 3.13). PCA analysis was also performed on the minimal dataset and shows that the data separate into 3 clusters in agreement with experimental treatments (Figure 13.4).

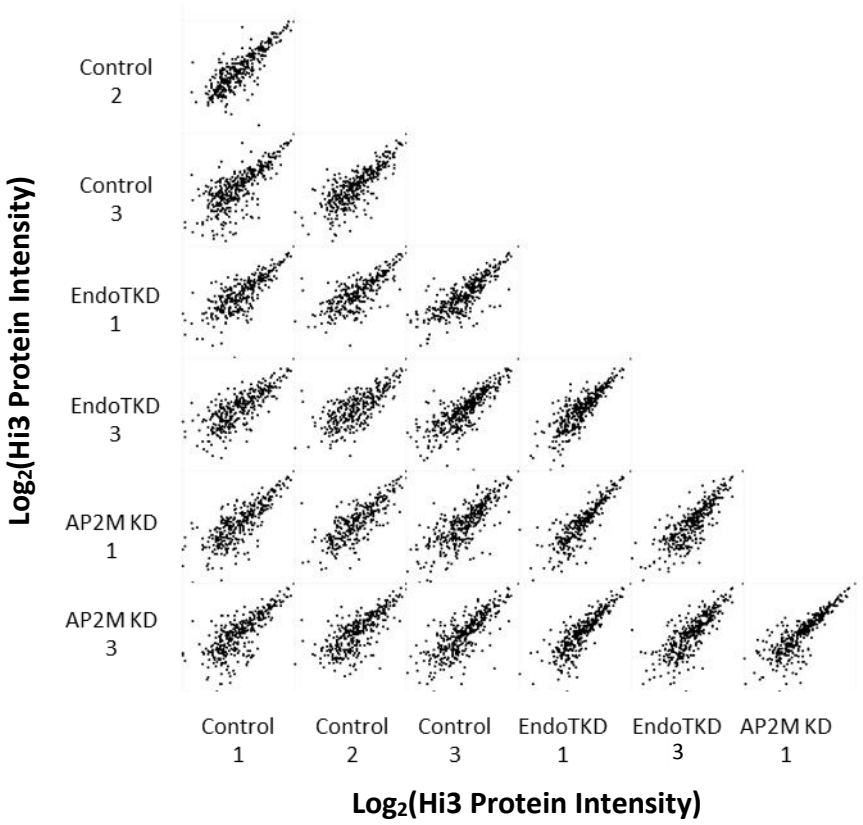

Figure 3.13 Scatter plot to show correlation in relative Hi3 protein intensity across biological replicates.

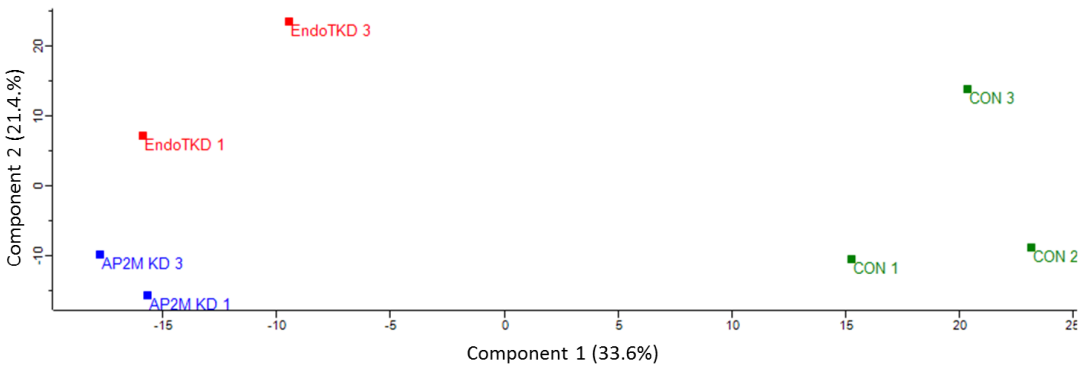

Figure 3.14 PCA plot to represent data similarity within minimal dataset in 2 dimensions. CON – Control, EndoTKD – Endophilin triple KD, AP2M KD – μ-Adaptin knockdown. Numbers indicate biological replicate identifier.

One-way ANOVA was performed to determine if protein abundance of a given protein in a given condition displayed a statistically significant difference from any other group. False discovery rate was controlled using a permutation-based FDR of 0.05. Two proteins, Junction Plakoglobin (JUP) and Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 (PLOD1) were denoted as significantly different in at least one pairwise condition. Tukey's Honest Significant difference (THSD) post-hoc test was performed and found differences in protein intensity were significant in all three pairwise comparisons for both proteins. To further interrogate the data, pairwise *t*-tests were performed between conditions, using an unpaired, 2-sided *t*-test with equal variances and correcting for multiple comparisons using a permutation-based FDR correction (FDR cut-off of 0.05). No hits reached significance using this method, likely due to loss of power incurred when reducing sample size. To glean more information from the dataset, the results of *t*-tests were plotted without correction for multiple testing, and the results were interpreted with the caveat of accepting a higher rate of type I errors.

Proteins were denoted as differentially expressed upon fulfilment of two criteria: *p*-value < 0.05 and a fold change (ratio of average protein intensity) greater than 1.3-fold in either direction. Quantitative measurements of peptides using MS^E have been shown to display an error in signal intensity of 10-15% (Silva *et al.*, 2006; Patel *et al.*, 2009). Thus, a fold change cut-off of 1.3

represents a value 2-3 times larger than that expected given normal technical variation. Using this definition 29 differentially expression proteins were identified in EndoTKD vs Control, 35 in Control vs AP2M and 18 in EndoTKD vs AP2M KD (Figure 3.15 and Tables 3.5, 3.6 and 3.7).

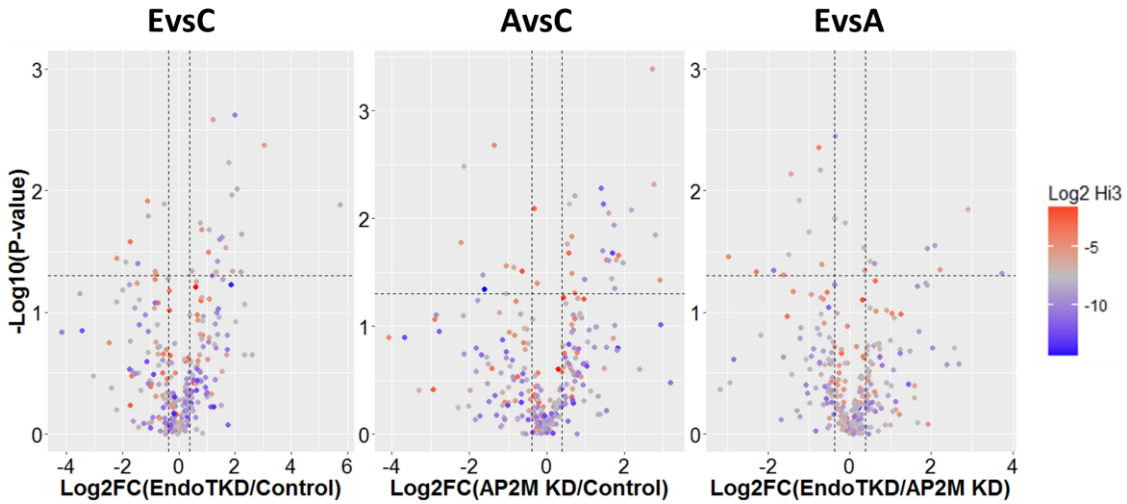

Figure 3.15 Volcano plots to visualise quantitative pairwise comparisons between each dataset. Plot colours are representative of Hi3 protein intensity for the highest expressed protein in each comparison. Fold changes (FC) are calculated by finding the ratio of average Hi3 protein intensity from each condition. Vertical dashed lines represent fold change cutoffs for significance. Horizontal dashed line represents -Log_{10}(*p*-value) cutoff for significance as calculated using a 2-sided t-test ($\alpha < 0.05$). Proteins with data plotted in the upper left and right quadrants are denoted as differentially expressed. EvsC – EndoTKD vs Control, AvsC – AP2M KD vs Control, EvsA – EndoTKD vs AP2M KD.

Data was analysed with reference to the hypothetical responses to treatment as described in section 3.1.2 (Figure 3.16). Receptors that internalised exclusively via FEME can be expected to accumulate at the plasma membrane when FEME is blocked relative to untreated cells but will not accumulate at the plasma membrane of CME-blocked cells (Figure 3.16), and vice-versa. Receptors which are internalised via multiple endocytic routes can be expected to display variable responses.

As of this writing, only 1 protein, β_1-adrenergic receptor (β_1-AR), has been shown to be internalised exclusively by FEME (Boucrot *et al.*, 2015). Unfortunately, this analysis did not detect β_1-AR successfully and it could not be used for validation. However, a number of other proteins for which relevant pathway information is known were detected and were used for dataset assessment.

Examining differentially regulated proteins between EndoTKD and AP2M KD conditions (EvsA in Figure 3.15), the highest enrichment for any protein in AP2M KD was the well validated CME cargo Transferrin receptor protein 1 (TFRC). Enrichment of the transmembrane proteins Matrix metalloproteinase-14 (MMP14) and V-type proton ATPase catalytic subunit A (ATP6V1A) was also observed in this comparison, both of which are known to enter cells via CME (Jiang *et al.*, 2001; Remacle, Murphy and Roghi, 2003). Furthermore, TFRC was strongly depleted (5-fold) in EndoTKD plasma membrane fractions relative to control (EvsC). Whilst these results appeared to validate the accumulation of CME cargoes upon AP2M KD, it was noted that TRFC did not appear enriched in plasma membranes from AP2M KD cells when compared to Control plasma membrane (AvsC in Figure 3.15). A further contradiction was the enrichment of CD44, a known CIE cargo (Eyster *et al.*, 2009), in the AP2M KD condition upon AvsC comparison. These effects are consistent with the incomplete knockdown of AP2 complex components noted by western blotting (Figure 3.11).

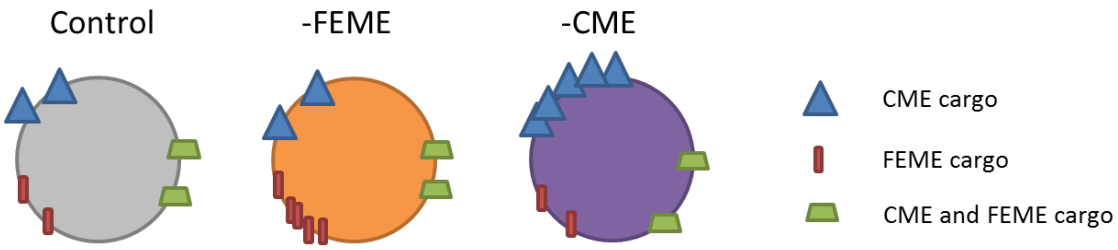

Figure 3.16 Expected effect of endocytic blockade on cell surface receptor populations at the plasma membrane. Circles represent mammalian cells. Other shapes represent cargo receptors which enter cells exclusively via FEME/CME or via both pathways.

The behaviour of several RTKs identified in this work with links to FEME was examined. The only protein with a link to known FEME cargoes enriched in EndoTKD in the EvsC comparison is the receptor tyrosine kinase Platelet-derived growth factor receptor beta (PDGFRB). FEME could be stimulated upon treatment of cells with platelet derived growth factor (PDGF), and the demonstration that ligand-receptor engagement is a requirement for uptake of FEME cargoes argues that PDGFRB is likely to be an FEME receptor (Boucrot *et al.*, 2015). However, this has not been specifically demonstrated. Furthermore, PDGFRB was found to be enriched in the

AP2M KD condition in an AvsC comparison, indicating that the PDGFRB may enter cells via CME as well. Two other RTKs known to be internalised by both FEME and CME were quantified in this work. Epidermal growth factor receptor (EGFR) was detected in all conditions but did not exhibit significant differences in any of the comparisons. Hepatocyte growth factor receptor (MET) was also detected in this analysis but was only enriched in AP2M KD plasma membranes in the AvsC comparison. Taken together, these results cannot unequivocally confirm the accumulation of FEME receptors at the plasma membrane as a cause of EndoTKD.

FEME receptors are expected to be observed in EndoTKD plasma membranes upon comparison to the other conditions (EvsC and EvsA). A group of 4 proteins enriched in EndoTKD (EvsC) were identified through network analysis using the STRING database search tool (Figure 3.17). These included Desmoglein-2 (DSG2), Cadherin-6 (CDH6), Cadherin-13 (CDH13) and JUP, which was previously identified in ANOVA analysis. These proteins all play roles at cell-cell junctions and are interesting candidates for future analysis due to the observation of high FEME activity at cell-cell junctions in confluent cells (*E. Boucrot*, unpublished results). Indeed, JUP is a

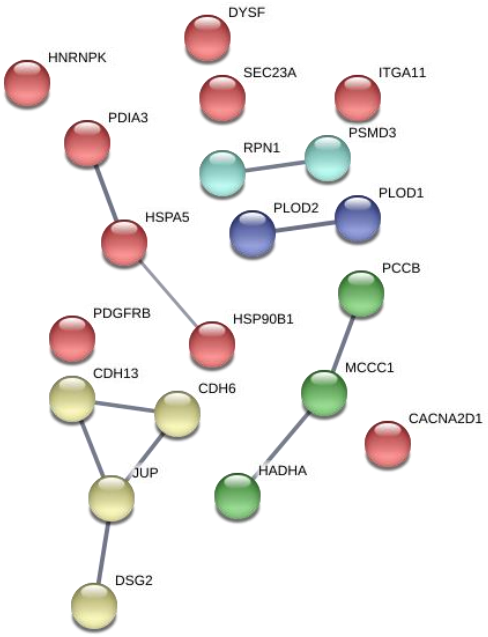

Figure 3.17 STRING network analysis to show interactome of proteins enriched in EndoTKD plasma membranes upon comparison of EndoTKD/Control samples (EvsC). Edges indicate interactions from experimental, databases and gene-fusion categories. The minimum interaction score required was 0.7 and clusters were demarcated using k-means clustering assuming 5 clusters. Note the cluster of cell-cell junction proteins highlighted in yellow.

cytosolic partner of the transmembrane glycoprotein DSG2 and contains a high confidence Endophilin SH3-target proline-rich motif (PRM). Whilst intriguing, these 4 proteins are unlikely to represent FEME-exclusive cargoes, as they were all found enriched in AP2M KD plasma membrane relative to control (AvsC).

In contrast, Dysferlin (DYSF) was identified as enriched in EndoTKD plasma membrane fractions relative to Control (EvsC) but was not differentially expressed upon CME blockade (AvsC), thus representing a candidate for exclusive uptake via FEME. DYSF is a single-pass type II transmembrane protein found to be 54-fold enriched in the plasma membrane of FEME- blocked cells relative to Control. Structurally, the protein consists of a small extracellular C-terminal domain and a large N-terminal cytoplasmic domain, which contains a number of proline rich motifs, one of which is a putative target for the endophilin SH3 domain.

Protein Name	Gene name	Fold Change (EndoTKD/Control)	*p*-value
Dysferlin	DYSF	53.94	0.0130
Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	8.15	0.0042
Cadherin-6	CDH6	4.69	0.0229
Propionyl-CoA carboxylase beta chain_ mitochondrial	PCCB	4.58	0.0463
Trifunctional enzyme subunit alpha_ mitochondrial	HADHA	4.19	0.0097
Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	PLOD1	4.01	0.0024
Procollagen-lysine_2-oxoglutarate 5-dioxygenase 2	PLOD2	3.76	0.0458
Desmoglein-2	DSG2	3.68	0.0108
Junction plakoglobin	JUP	3.42	0.0059
Platelet-derived growth factor receptor beta	PDGFRB	3.17	0.0294
Cadherin-13	CDH13	2.91	0.0242
Methylcrotonoyl-CoA carboxylase subunit alpha_ mitochondrial	MCCC1	2.76	0.0250
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	2.58	0.0381
Protein disulfide-isomerase A3	PDIA3	2.52	0.0462
Integrin alpha-11	ITGA11	2.33	0.0026
Protein transport protein Sec23A	SEC23A	2.13	0.0208
Endoplasmin	HSP90B1	2.06	0.0319
Endoplasmic reticulum chaperone BiP	HSPA5	1.75	0.0210
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	1.70	0.0182
26S proteasome non-ATPase regulatory subunit 3	PSMD3	1.38	0.0457
Ezrin	EZR	0.69	0.0129
Heat shock protein HSP 90-alpha	HSP90AA1	0.57	0.0482
Neuropilin-1	NRP1	0.56	0.0456
Matrix remodeling-associated protein 8	MXRA8	0.47	0.0162
Basigin	BSG	0.46	0.0122
Probable ATP-dependent RNA helicase DDX5	DDX5	0.36	0.0395
ATP-citrate synthase	ACLY	0.30	0.0262
Trifunctional purine biosynthetic protein adenosine-3	GART	0.27	0.0385
Transferrin receptor protein 1	TFRC	0.22	0.0358

Table 3.5 Differentially expressed proteins (EndoTKD/Control; EvsC). Shaded proteins are enriched in Control.

Protein name	Gene name	Fold Change (AP2M/Control)	*p*-value
Endothelin-converting enzyme 1	ECE1	7.49	0.0376
Thrombospondin-1	THBS1	6.98	0.0143
Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	6.79	0.0049
Junction plakoglobin	JUP	6.58	0.0004
Trifunctional enzyme subunit alpha_ mitochondrial	HADHA	4.52	0.0084
CD276 antigen	CD276	3.90	0.0260
5'-nucleotidase	NT5E	3.58	0.0220
Platelet-derived growth factor receptor beta	PDGFRB	3.43	0.0242
Cadherin-6	CDH6	3.35	0.0447
Ras GTPase-activating-like protein IQGAP1	IQGAP1	3.29	0.0116
Tubulin beta-1 chain	TUBB1	3.22	0.0209
Desmoglein-2	DSG2	3.20	0.0363
Prolyl endopeptidase FAP	FAP	3.03	0.0090
Dolichyl-diphosphooligosaccharide-- protein glycosyltransferase subunit 1	RPN1	3.01	0.0247
Integrin alpha-11	ITGA11	2.92	0.0241
Mitochondrial proton/calcium exchanger protein	LETM1	2.74	0.0073
Cadherin-13	CDH13	2.71	0.0199
Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	PLOD1	2.63	0.0052
Neuropilin-2	NRP2	1.95	0.0370
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	1.65	0.0061
Myoferlin	MYOF	1.65	0.0489
Endoplasmic reticulum chaperone BiP	HSPA5	1.56	0.0322
Matrix metalloproteinase-14	MMP14	1.56	0.0146
V-type proton ATPase catalytic subunit A	ATP6V1A	1.49	0.0073
Pyruvate kinase PKM	PKM	1.47	0.0208
Extended synaptotagmin-1	ESYT1	1.43	0.0174
CD44 antigen	CD44	0.64	0.0309
Hepatocyte growth factor receptor	MET	0.55	0.0285
Importin-7	IPO7	0.49	0.0499
D-3-phosphoglycerate dehydrogenase	PHGDH	0.48	0.0276
Basigin	BSG	0.39	0.0021
Catenin alpha-3	CTNNA3	0.33	0.0455
Eukaryotic translation initiation factor 3 subunit F	EIF3F	0.32	0.0335
Trifunctional purine biosynthetic protein adenosine-3	GART	0.23	0.0033
Inactive tyrosine-protein kinase 7	PTK7	0.21	0.0167

Table 3.6 Differentially expressed proteins (AP2M KD/ Control; AvsC). Shaded proteins are enriched in Control.

Protein name	Gene name	Fold Change (EndoTKD/AP2M)	*p*-value
Spectrin beta chain_ non-erythrocytic 4	SPTBN4	13.32	0.0475
Talin-2	TLN2	7.48	0.0141
Inactive tyrosine-protein kinase 7	PTK7	4.67	0.0447
Methylcrotonoyl-CoA carboxylase subunit alpha_ mitochondrial	MCCC1	4.22	0.0281
Heat shock protein 105 kDa	HSPH1	3.73	0.0299
Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	PLOD1	1.52	0.0396
ATP-dependent 6-phosphofructokinase_ platelet type	PFKP	1.41	0.0379
Matrix metalloproteinase-14	MMP14	0.62	0.0403
V-type proton ATPase catalytic subunit A	ATP6V1A	0.61	0.0068
5'-nucleotidase	NT5E	0.59	0.0044
Matrix remodeling-associated protein 8	MXRA8	0.50	0.0218
Tubulin beta-2A chain	TUBB2A	0.42	0.0120
Prolyl endopeptidase FAP	FAP	0.37	0.0073
Podocalyxin	PODXL	0.35	0.0333
Endothelin-converting enzyme 1	ECE1	0.32	0.0488
Putative heat shock protein HSP 90-beta 2	HSP90AB2P	0.27	0.0454
Tubulin alpha-1A chain	TUBA1A	0.20	0.0461
Transferrin receptor protein 1	TFRC	0.13	0.0347

Table 3.7 Differentially expressed proteins (EndoTKD/ AP2M KD; EvsA). Shaded proteins are enriched in AP2M KD.

3.3. Part B: Identifying candidate FEME cargoes through GST-Endophilin-SH3 pulldown from a plasma membrane extract

3.3.1. Selection of plasma membrane enrichment strategy

A workflow was developed employing recombinant GST-EndophilinA2-SH3 (GST-EndoA2-SH3) to screen plasma membrane fractions from human RPE1 cells for potential FEME cargoes by AP-MS (Figure 3.18). Pre-enrichment of plasma membranes, which aims to minimise non-specific interactions from other cellular compartments and assist peptide identification during MS analysis, was optimised first. One well-established approach to enrich a membrane fraction features mechanical lysis of cells in hypotonic fractionation buffer, followed by differential centrifugation (*e.g.* Lai, 2013). Affinity purification of plasma membranes following biotin labelling of surface proteins is an alternative technique (Deblaquiere and Burgess, 1999; Weekes *et al.*, 2010; Bitsikas, Corrêa and Nichols, 2014). Adapted protocols for both methods were created and assessed for suitability.

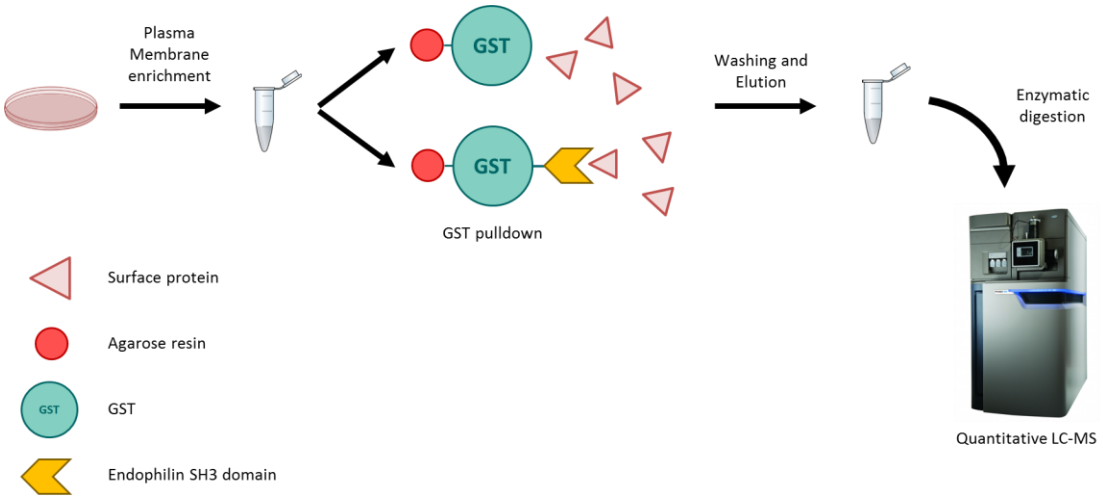

Figure 3.18 Proposed AP-MS workflow. Plasma membrane enriched fractions will be exposed to GST-only- and GST-EndoA2-SH3-coupled beads during the affinity chromatography step. Following washing and elution from beads, captured proteins will be coupled to LC-MS using appropriate sample preparation. Identification and quantitation of eluting peptides will enable sorting of specific and nonspecific interactions by comparison to control "GST-only" beads.

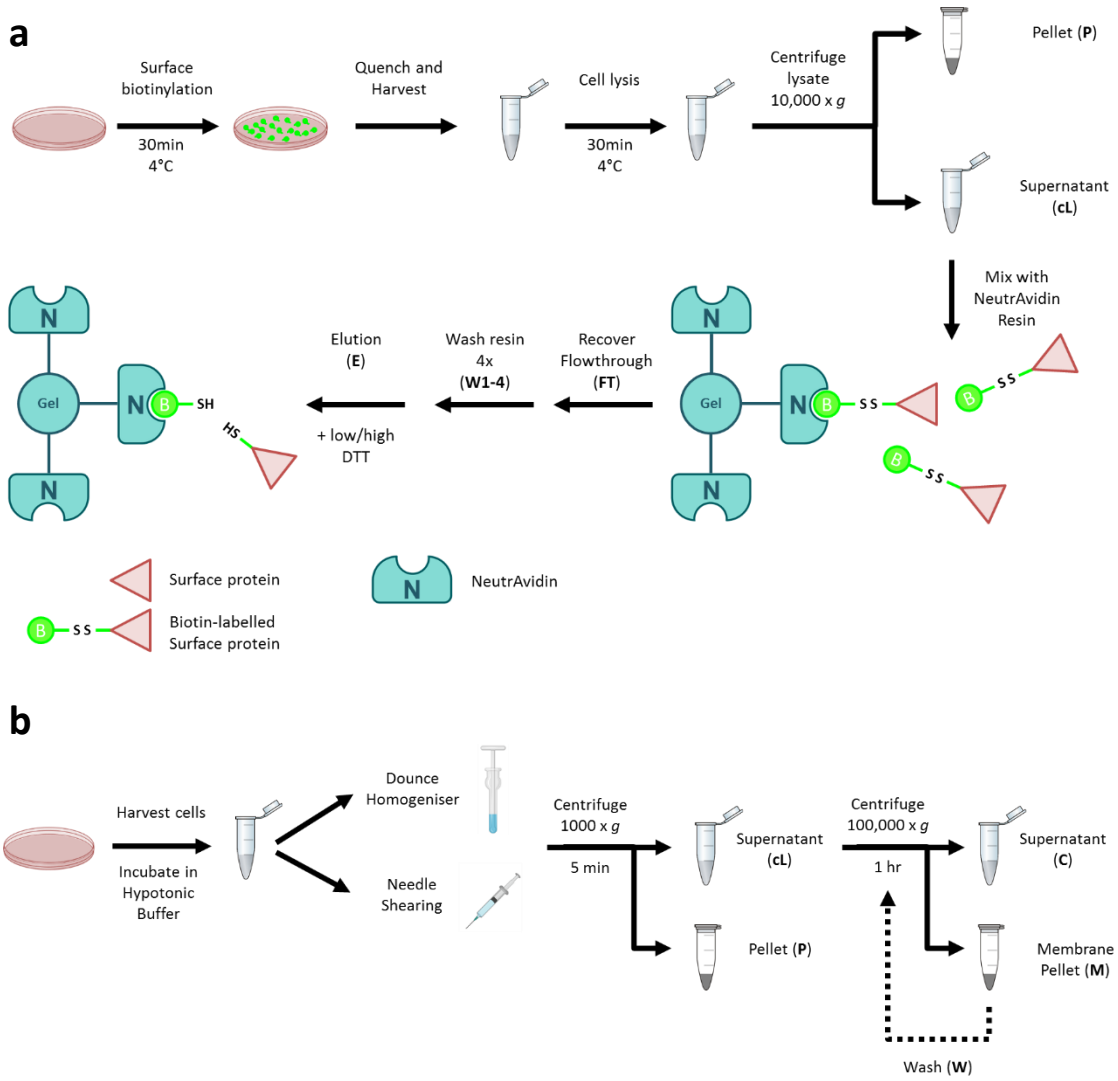

Figure 3.19 Techniques to enrich plasma membrane proteins. **a**, Schematic of surface biotinylation followed by affinity purification of plasma membrane protocol. Note the reduction of the disulphide linker in the elution step to release captured surface proteins. **b**, Schematic of crude membrane preparation workflow. Membrane pellets can be washed by resuspension in fresh lysis buffer and re-centrifuged multiple times.

A commercially available plasma membrane isolation kit utilising the cell surface biotinylation paradigm was tested (Pierce Cell Surface Protein Isolation Kit) (Figure 3.19a). Briefly, surface proteins on RPE1 cells grown to 80-90% confluence in 100 mm dishes were labelled with sulfo-NHS-SS-Biotin, a membrane impermeable reagent reactive against primary amines. The reaction was quenched and cells were lysed in a proprietary lysis buffer. Solubilised, biotinylated proteins were affinity purified from the lysate by exposure to agarose beads coated with NeutrAvidin monomers, which associate strongly with biotin. The bead resin was washed 4 times prior to elution. Cleavage of the disulphide linker by reduction in SDS-PAGE sample buffer

supplemented with 50 mM dithiothreitol (DTT) released captured surface proteins in the elution step. This protocol was found to solubilise and purify marker plasma membrane proteins (EGFR, HGFR, Integrins β1 and αV and the Transferrin receptor) whilst depleting contaminant markers for cytosol (GAPDH), nuclear (HH3) and ER (Calnexin) compartments (Figure 3.3a and see detailed evaluation in Section 3.2.1)

Optimisation of crude membrane preparation was then performed. RPE1 cells were subjected to mechanical lysis in hypotonic buffer (Figure 3.19b). RPE1 cells were found to visually round and swell within 15 min incubation in a hypotonic fractionation buffer on ice. Preparation of crude membrane fractions then was attempted using two mechanical lysis techniques: Dounce homogenisation and needle shearing (see schematic in Figure 3.19b). Briefly, RPE1 cells from 2 100 mm dishes were grown to 80% confluence, scraped into a sample tube and allowed to swell in hypotonic buffer. Cells were then lysed by 20 passes through a 23-gauge (23G) needle or 100 strokes of a Dounce homogeniser. Intact cells and nuclei were collected by centrifugation of the lysate at 1000 x *g* for 5 min and the supernatant was centrifuged for 1 hour at 100,000 x *g* to collect a membrane pellet and cytosolic fraction (supernatant). Western blotting indicated that neither mechanical lysis option dramatically outperformed the other, with similar a yield of EGFR and similar Membrane fraction purity achieved (Figure 3.20b, left). Both techniques displayed excellent exclusion of nuclear proteins, and clear enrichment of EGFR in the membrane instead of cytosolic fraction. As expected, the ER-resident protein Calnexin was not was excluded from the membrane fraction. However, membrane fractions from both techniques exhibited high levels of GAPDH, suggesting poor separation of membrane and cytosolic protein (Figure 3.20b, left).

The crude membrane preparation protocol was adjusted by performing lysis in a smaller volume with a reduced needle bore size and increasing the number of passes to 40. The membrane pellet was also washed 3 times to remove non-specifically bound cytosolic proteins. These optimisations appeared to improve membrane fraction purity, with reduced GAPDH signal in the membrane fraction relative to the original method (Figure 3.20b, right). However, these optimisations did not reduce the levels of ER protein in the membrane fraction. The successful elimination of ER proteins from the plasma membrane fractions using biotin-labelling of surface

protein was the key difference between these two workflows and resulted in selection of this technique for further optimisation.

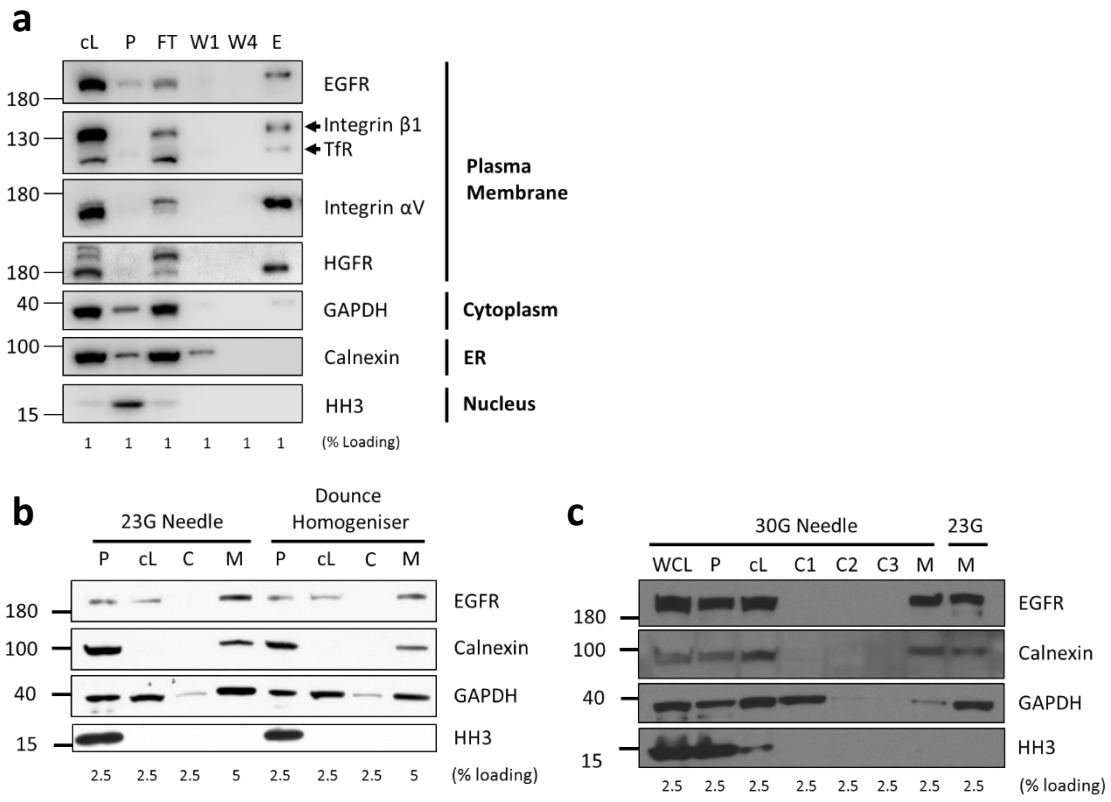

Figure 3.20 Affinity purification of biotin-labelled plasma membranes outperforms crude membrane extraction protocols. **a**, The presence of proteins from various cellular compartments in aliquots from different stages of the plasma membrane isolation workflow was assessed by western blotting. Elution was conducted in SDS-PAGE sample buffer with 50mM DTT. **b**, Microscopy images to show swelling of RPE1 cells following incubation in hypotonic fractionation buffer. **c**, Western blotting to assess Membrane fraction purity during optimisation of crude membrane extraction protocol. cL – Clarified Lysate, P – Pellet, FT – Flowthrough, W1-4 – Wash 1-4, E – Elution, WCL – Whole Cell Lysate, C1-3 – Cytosolic fraction 1-3, M – Membrane fraction.

3.3.2. Optimisation of mild elution conditions for release of plasma membrane proteins

Cell surface capture using sulfo-NHS-SS-Biotin is now well established (Deblaquiere and Burgess, 1999; Weekes *et al.*, 2010; Bitsikas, Corrêa and Nichols, 2014). In the literature, cleavage of the sulfo-NHS-SS-Biotin linker is often performed in high concentrations of DTT (>50 mM) to ensure timely reduction of all bound linkers. Elution is also usually performed in SDS to preserve solubility of the released plasma membrane proteins. Our experimental strategy requires the release of plasma membrane proteins in a state suitable for secondary GST pulldown, so optimisation to define non-denaturing, low DTT elution conditions was undertaken.

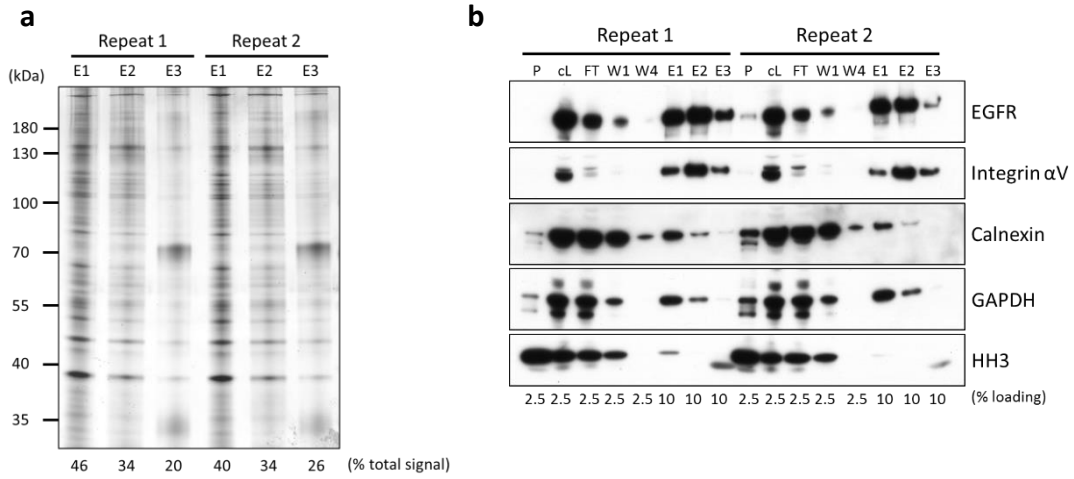

Figure 3.21 A significant proportion of captured plasma membrane protein is released under mild elution conditions. **a**, Silver staining to compare sequential elution of plasma membrane proteins following biotin labelling, capture and washing in three different elution buffers. E1 – low DTT CHAPS, E2 – High DTT Urea/CHAPS, E3 – SDS-PAGE sample buffer (see Table 3.2). Protein amount was quantified by densitometry and % total signal among all 3 lanes within a repeat is displayed. **b**, Western blot to assess the success of plasma membrane isolation workflow for the Elution samples presented in **a**. Note higher loading of eluates relative to other samples. cL – Clarified Lysate, P – Pellet, FT – Flowthrough, W1-4 – Wash 1-4, E1 – Low DTT CHAPS, E2 – High DTT Urea/CHAPS, E3 – SDS-PAGE sample buffer – see Table 3.8. Repeats shown are biological replicates.

The anionic detergent SDS strongly denatures protein secondary structure when present at concentrations above the critical micelle concentration (CMC). We therefore tested elution buffers containing the zwitterionic detergent CHAPS, which has been successfully employed to solubilise membrane proteins in AP-MS experiments (Hjelmeland, 1980; Kanellopoulos *et al.*, 2018). In addition, we reasoned that cleavage of the disulphide linker in sulfo-NHS-SS-Biotin could probably occur at much lower concentrations of DTT (1-2 mM). A Urea/CHAPS elution

buffer with high (50 mM) DTT was also tested in the ultimately futile hope that this buffer would be conducive to downstream UV absorbance protein quantitation (data not shown).

Plasma membrane proteins were captured using the workflow in Figure 3.19a, with a modified elution step. Following washing, proteins were sequentially eluted in equal volumes of 3 different buffers representing a range of DTT concentrations, detergent/ chaotrope levels, temperatures and elution step durations (Table 3.2).

Lysis/ Elution Buffer	Core Buffer components	Detergent/ Chaotrope	DTT concentration	Elution conditions
No DTT CHAPS	25mM Tris-HCl pH 7.4, 150mM NaCl, protease inhibitors	1% CHAPS	n/a	RT, 1 hour
Low DTT CHAPS	25mM Tris-HCl pH 7.4, 150mM NaCl, protease inhibitors	1% CHAPS	1mM	RT, 1 hour
High DTT urea/CHAPS	25mM Tris-HCl pH 7.4, 150mM NaCl, protease inhibitors	8M urea/ 1% CHAPS	50mM	RT, 1 hour
SDS-PAGE sample buffer	60mM Tris-HCl pH 6.8, 10% glycerol, 0.002% Bromophenol Blue	2% SDS	50mM	95°C, 5min
Kit Lysis Buffer	Proprietary	Proprietary	n/a	n/a

Table 3.8 Lysis and Elution buffer constitution. RT – room temperature.

Silver staining labels total protein with high sensitivity (Weiss, Weiland and Görg, 2009). Densitometric software can be used to obtain a semi-quantitative estimate of protein amount in a given area of pixels, with darker pixels indicating more intense signal. ImageStudio Lite (LI-COR) software was used to estimate protein amount in whole lanes in Figure 3.21a. On average, 43% of bound protein was released in low DTT, CHAPS buffer (E1), 34% of bound protein was released in high DTT, urea/CHAPS buffer (E2) and 23% of bound protein was released in the SDS-PAGE sample buffer elution (E3), with low deviation from these proportions between repeats.

Western blotting shown in Figure 3.21b gives confidence that the eluted material indeed represents a plasma membrane-enriched fraction, with cytosolic (GAPDH), ER (Calnexin) and nuclear (Histone H3) marker proteins depleted in the Elution fractions concomitant with retention of plasma membrane receptors EGFR and Integrin αV. Interestingly, the majority of plasma membrane protein appears to elute in the E2 fraction, despite indications from the silver stain that most protein overall elutes in the E1 fraction. This effect is particularly pronounced for Integrin αV, but less clear in the EGFR blot. These results indicate that although the majority of plasma membrane protein stays attached to the beads, a significant portion of these proteins are reproducibly eluted following incubation in Low DTT CHAPS elution buffer for 1 hour at room temperature.

3.3.3. Investigating if optimised elution conditions are conducive to a secondary GST pulldown using a recombinant β_1-AR construct

As a known cargo of FEME, β_1-AR represented a key positive control for optimising bait-prey binding conditions in the AP-MS workflow. Having established mild elution conditions for obtaining a plasma membrane fraction, overexpression of a recombinant murine β_1-AR tagged C-terminally with enhanced-green fluorescent protein (eGFP) in HEK293 cells (a cell line in which >80% transfection can be routinely achieved) was used to investigate the viability of a secondary GST pulldown step.

Recombinant GST-EndoA2-SH3 was expressed in *E. coli* and coupled to GSH-coated magnetic agarose beads. A β_1-AR-eGFP construct was transiently expressed in HEK293 cells. Upon visual confirmation of construct overexpression, plasma membrane fractions were obtained using Low DTT CHAPS elution conditions. Plasma membrane proteins were then exposed to GST-EndoA2-SH3 beads overnight, washed and eluted from beads by heating at 65°C for 15 min (to avoid membrane protein aggregation) in SDS-PAGE sample buffer.

Immunoblotting revealed successful enrichment of the plasma membrane protein EGFR in the Elution fraction (E), with low levels of cytosolic (GAPDH), ER (Calnexin) and nuclear (HH3) marker protein signal (Figure 3.22a). Two bands (~80 kDa and ~100 kDa) were visibly enriched in the anti-eGFP Elution fraction (E) and represent good candidates for surface-resident β_1-AR-eGFP. Firstly, the lower band is close in size to the theoretical mass of the construct (81 kDa). In addition, a highly intense species observed at ~60kDa in the cL fraction was not retained well in the Elution fraction (Figure 3.22a, right, red arrowhead), arguing for specific cell surface labelling of the ~80 kDa and ~100 kDa species. The ~100 kDa band may represent a post-translationally modified form of the construct, as β_1-AR is known to be glycosylated at several extracellular sites (Park *et al.*, 2017) and phosphorylated at a number of intracellular sites (Hinz *et al.*, 2017). Neither protein was retained in the secondary pulldown Bound fraction (B), remaining in the Flowthrough (FT) following exposure to GST-only and GST-EndoA2-SH3 beads (Figure 3.22a). This result indicated that no interaction occurs between B1-AR-eGFP and GST-EndoA2-SH3 under the conditions tested.

A number of factors could disrupt the interaction between β_1-AR-eGFP and GST-EndoA2-SH3. The constructs may not fold correctly, or the eGFP domain may interfere with binding. Alternatively, the constructs may not retain structure upon solubilisation in the detergents present in the Kit lysis buffer and/or in low DTT CHAPS elution buffer.

GST construct folding was assessed using recombinant Dynamin-1-eGFP (Dyn1-eGFP), a known endophilin binding partner. When cells were lysed in a Triton X-100 lysis buffer, Dyn1-eGFP displayed low binding to a GST-only construct but was highly enriched following pulldown with GST-EndoA2-SH3, indicating that the GST constructs were correctly folded (Figure 3.22b).

To assess if the Low DTT elution buffer was detrimental to secondary pulldowns, HEK293 cells expressing β_1-AR-eGFP were lysed directly in Low DTT CHAPS buffer and a pulldown was conducted without pre-enrichment of the plasma membrane (Figure 3.22c). No specific binding was observed, raising two possibilities: i) overexpressed β_1-AR-eGFP cannot be solubilised in a

functional state in Low DTT CHAPS buffer or, ii) the β_1-AR-eGFP construct suffers from overexpression-related artifacts.

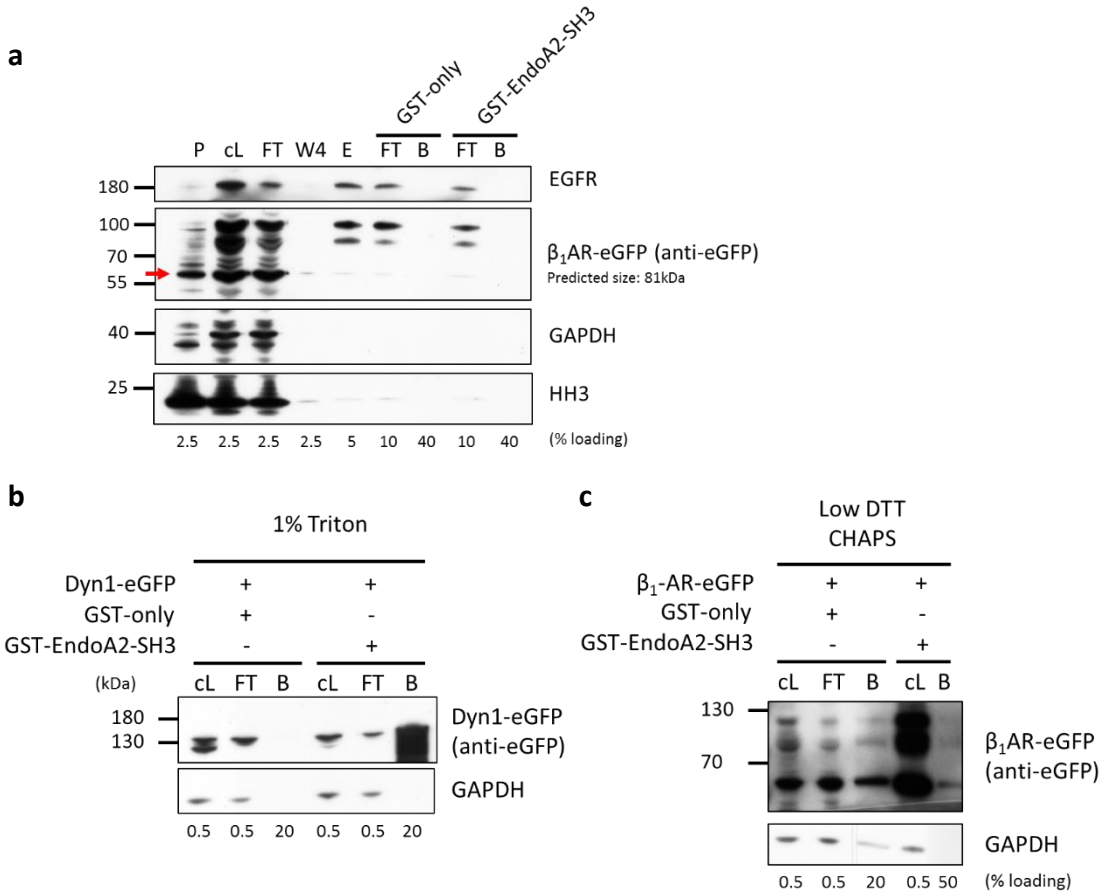

Figure 3.22 A β_1-AR-eGFP construct expressed in HEK293 cells does not copurify with GST-EndoA2-SH3 following plasma membrane elution in Low DTT CHAPS buffer. **a**, Left, β_1-AR-eGFP construct expression in HEK293 cells following transient transfection. Right, Western blot to assess plasma membrane capture and secondary GST pulldown in low DTT CHAPS buffer. Red arrowhead indicates ~60 kDa band of interest. **b**, Western blots to assess GST pulldown of Dyn1-eGFP constructs overexpressed in HEK293 cells in Triton X-100 Lysis buffer. **c**, Western blotting to assess GST pulldown of β_1-AR-eGFP constructs overexpressed in HEK293 cells in Low DTT CHAPS buffer. P – Pellet, cL – Clarified lysate, FT – Flowthrough, W4 – Wash 4, E – Elution, B – Bound Fraction.

3.3.4. A small proportion of endogenous β_1-AR is captured by GST-EndoA2-SH3 beads

Transient overexpression of recombinant proteins is subject to well documented artefacts such as the promotion of non-native protein-protein interactions and mislocalisation, even when performed in a mammalian expression system (Reaves and Banting, 1994; Sakaue-Sawano *et al.*, 2008; Chudakov *et al.*, 2010). To avoid these issues, the secondary pulldown workflow was evaluated using endogenous β_1-AR from RPE1 cells.

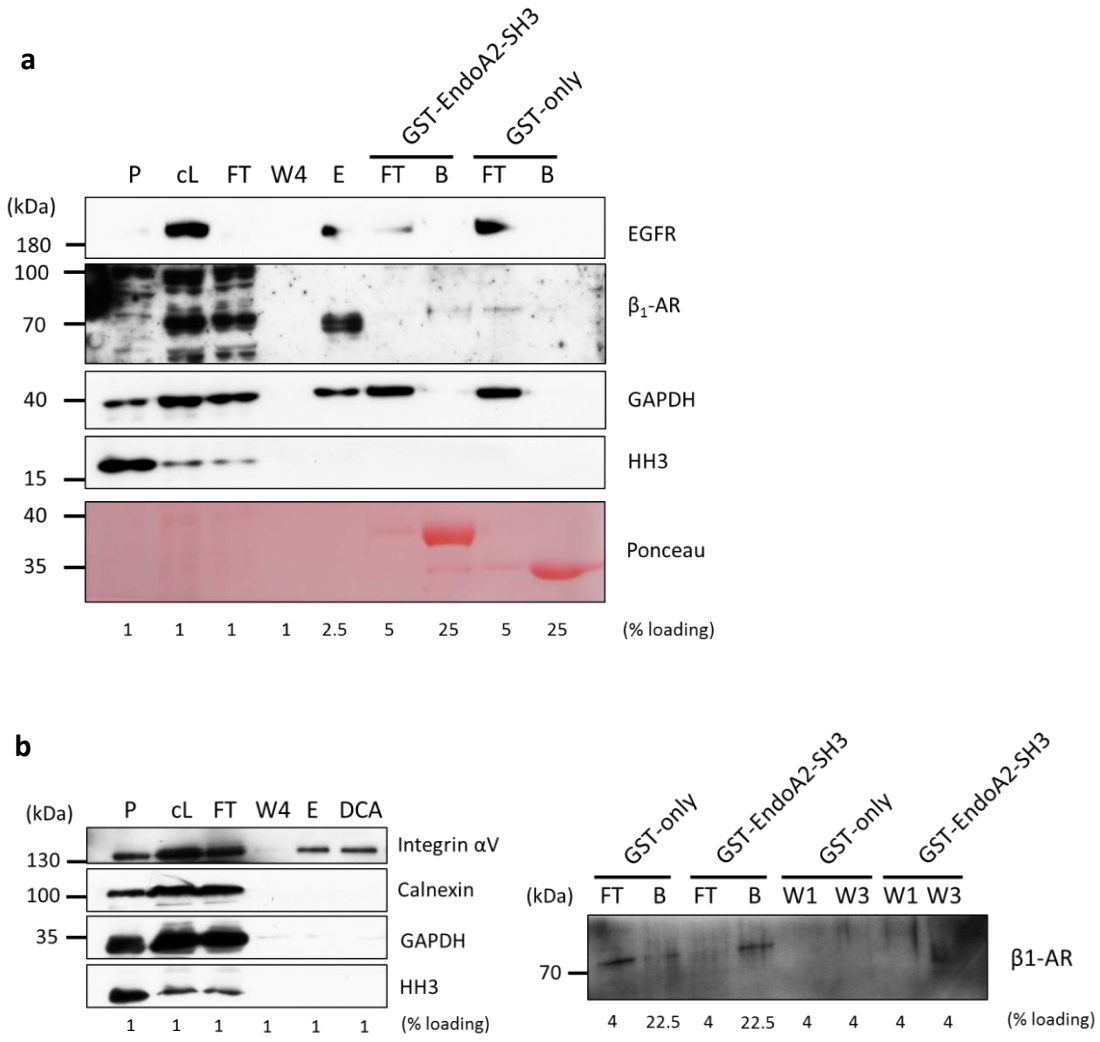

Figure 3.23 A small proportion of endogenous β_1-AR is captured on GST-EndoA2-SH3 beads by secondary elution in Low DTT CHAPS buffer. **a**, Western Blot analysis of a complete secondary GST-pulldown from plasma membrane fraction workflow conducted on RPE1 cells. Ponceau stain panel indicates the relative sizes and intensity of both GST constructs in Bound fraction (B). **b**, Western blot analysis of an improved secondary GST-pulldown workflow. P – Pellet, cL – Clarified lysate, FT – Flowthrough, W1-4 – Wash 1-4, E – Elution, DCA – degradation control aliquot, B – Bound Fraction.

In accordance with the fact that endogenous β_1-AR is expressed at relatively low levels in RPE1 cells (Boucrot *et al.*, 2015), starting material for the secondary GST-pulldown workflow was increased 12-fold (12 x 100 mm dishes). Western blot analysis indicated successful enrichment of a 60-70 kDa doublet believed to represent surface-resident β_1-AR in the plasma membrane fraction (Figure 3.23a, Lane E). As expected, the blots reveal no interaction between either GST construct and GAPDH or EGFR. By contrast, a small amount of β_1-AR appears to copurify with GST-EndoA2-SH3 but not with GST-only, suggesting that a limited pool of endogenous β_1-AR can interact specifically with GST-EndoA2-SH3 in Low DTT CHAPS buffer. In support of this notion, β_1-AR appears to be depleted in the GST-EndoA2-SH3 flow-through but not in the GST-only flow-through (Figure 3.23a). However, the low yield of β_1-AR was puzzling given its abundance in the Elution fraction. It was established that no β_1-AR protein remains on the beads following heating in SDS-PAGE sample buffer (data not shown), so an incomplete elution step is not the cause. The interaction between β_1-AR and EndoA2-SH3 may be mediated by phosphorylation sites on β_1-AR which are not preserved as phosphatase inhibitors were not used in these trials. Another possibility is that the majority of β_1-AR is captured but does not withstand the subsequent GST-pulldown wash steps, which were not collected for evaluation in this instance. Finally, the proteins may degrade during the overnight secondary pulldown step. An improved workflow was performed to address these possibilities.

Immunoblots of marker proteins show that plasma membrane enrichment was achieved successfully (Figure 3.23b, left). An aliquot of the Elution fraction was subjected to the same handling steps for the secondary GST-pulldown protocol (end-over-end agitation overnight at 4°C) but without exposure to the GST beads, with no noticeable degradation of the membrane protein proxy Integrin αV observed (degradation control aliquot - DCA lane, Figure 3.23b, left). The use of phosphatase inhibitors did not cause a dramatic improvement in β_1-AR signal in the Bound/ FT fractions following secondary GST pulldown, with very faint bands observed (Figure 3.23b, right). However, β_1-AR's binding pattern was consistent with previous results. A more intense β_1-AR signal was observed in the GST-EndoA2-SH3 Bound fraction than in the GST-only Bound fraction. Furthermore, β_1-AR is dramatically depleted in the GST-EndoA2-SH3 flowthrough in comparison to the clear β_1-AR signal observable in the GST-only flow-through.

Although pulldown wash fractions were included, they did not show definitive evidence of β_1-AR loss (Figure 3.23b, right). Together, these results suggest that proteins in the plasma membrane fraction are not degraded by temperature or mechanical handling during the secondary GST-pulldown and that the use of phosphatase inhibitors does not have a large impact on pulldown efficacy. Directions for future work are discussed in Section 3.4.

3.4. Discussion

Optimisation of two workflows aimed at discovering novel receptor cargoes for fast endophilin-mediated endocytosis was performed. Section 3.2 (Part A) detailed optimisation of a workflow for comparing quantitative differences between proteins in plasma membrane extracts from cells following RNAi-induced endocytic inhibition. Section 3.3 (Part B) described optimisation of a workflow to identify candidate FEME cargoes through GST-Endophilin-SH3 pulldown from a plasma membrane extract.

Plasma membrane enrichment was a key step in both workflows and was achieved using cell surface biotinylation with sulfo-NHS-SS-Biotin (which targets primary amines) and affinity purification of membrane proteins. A key limitation for both workflows, therefore, is the restriction of possible targets for affinity purification to the subset of surface proteins which have exposed lysine residues in their extracellular regions or an extracellular N-terminus. At present, the extent to which this limitation affected the scope of this work, in terms of yield of plasma membrane proteins and subsequent depth of LC-MS protein coverage, is unknown. Indeed, studies have shown that the use of different surface labelling methods can have dramatic effects on purity and quality of membrane protein data obtained by LC-MS (Weekes *et al.*, 2010). Future trials of alternative methods for plasma membrane labelling would be a desirable step in answering this question. One method which has been used to map the cell surface with great effect Cell Surface Capture (CSC). Introduced by Wollscheid *et al*. (2009), this technique involves chemical labelling of *N*-linked glycoproteins at the cell surface with biotin. Periodate is used to oxidise sugar moieties on *N*-glycosylated proteins at the cell surface to generate aldehyde groups for targeting by biotin hydrazide, forming a covalent linkage between biotin and the target glycoprotein. Cells are then lysed, and the contents subjected to enzymatic digestion. Biotin-conjugated glycopeptides are then purified from the sample using an avidin resin. The resin is washed extensively and subjected to enzymatic digestion by peptide *N*-glycosidase F (PNGase F), which cleaves the *N*-glycosidic bond and releases peptides for LC-MS analysis. This 2-stage affinity purification was shown to be highly selective for cell surface proteins, and was further developed

to perform cell surface profiling of 41 human and 31 mouse cell types, identifying a pool of 1492 human and 1296 mouse cell surface glycoproteins (Bausch-Fluck *et al.*, 2015). This technique suffers from the limitation that only *N*-glycosylated proteins can be detected, but this pool may outnumber the pool of receptors with accessible primary amine residues and should be tested. A similar glycoprotein-targeting technique which could trialled is lectin-based affinity purification of plasma membrane (Deeb *et al.*, 2014).

The quality of data obtained from the workflow in Part A was compromised by the inconsistent siRNA knockdown of μ-Adaptin and technical issues with mass spectrometry sample preparation. With respect to improving the quality/robustness of endocytic inhibition in RPE1 cells, gene editing using CRISPR-Cas9 technology could be attempted. The advantage of using this technique is that generation of a complete knockout of μ-Adaptin could be achieved. Drawbacks would include the length of time required to select and expand suitable clones, and the dangers of selecting clones which have reverted to a wild-type phenotype stochastically (so-called "escapers"). With respect to improving the performance of the sample preparation for LC-MS, focusing on fractionation of the sample would likely yield the best immediate improvements, given that the yield of protein from plasma membrane isolation is sufficient for microproteomics. In the In-gel-SCX workflow, only 3 fractions were used, and no optimisation of the SCX intervals in ammonium acetate concentration was performed. Increasing the number of fractions and optimisation of ammonium acetate intervals could improve the workflow performance. Indeed, it would be interesting to test an alternative method of peptide-level fractionation such as High pH reversed-phase fractionation, given the relatively poor performance of SCX as highlighted in Section 3.2.3.

A number of interesting candidates for involvement in FEME were highlighted for future analysis from the workflow in Part A. Of these, JUP is a particularly exciting candidate due to the presence of a proline-rich motif suitable for targeting by the Endophilin-SH3 domain in its primary sequence. The co-enrichment of several predicted and known binding partners for JUP upon inhibition of FEME, and the identification of JUP as a differentially regulated protein in ANOVA,

lend weight to the classification of JUP as a potential adaptor for FEME-mediated uptake of proteins at cell-cell junctions. Indeed, higher FEME activity has been observed at cell-cell junctions in confluent cells (*E. Boucrot*, unpublished results). DYSF also contains a putative Endophilin SH3 target proline-rich motif and was found to display 54-fold enrichment in FEME-blocked cells. Previous characterisation of DYSF has focused on its function in membrane repair as a mediator of intracellular membrane fusion at sites of plasma membrane lesions in a Ca^{2+} dependent manner (Davis *et al.*, 2002; Bansal *et al.*, 2003). The dramatic upregulation of DYSF at the plasma membrane in EndoTKD cells could be due to a higher incidence of membrane lesions in EndoTKD cells, or due to dysfunctional Ca^{2+} regulation in these cells, both avenues worth investigating. To establish if JUP or DYSF are bona fide FEME adaptors/cargoes, a prudent first step would be to investigate their interaction with EndoSH3 through pull-down assays. Promising results can be followed with antibody-based confirmation of in vivo associations with endophilin through immunoprecipitation with western blotting and/or fluorescent microscopy.

A key limitation introduced by the selection of the biotinylation strategy as a means of purifying plasma membrane proteins for the workflow in Part B was the low yield of plasma membrane proteins from this technique: calculated to lie between 30 µg – 80 µg (Section 3.2.1). Gentle elution conditions in 1mM DTT were predicted to release ~43% of bound protein, therefore an expected yield from elution in low DTT CHAPS is between ~13 µg - 34 µg of protein. At these yields, generation of a plasma membrane fraction at a concentration of 5 mg/ml (recommended for EndoSH3 pulldowns) at sufficient volume to conduct an AP-MS workflow would have been prohibitively expensive. Future attempts to obtain a plasma membrane extract for secondary GST-SH3 pulldowns should begin with trials to identify a workflow which can produce sufficient amounts of plasma membrane starting material at an acceptable cost. Future work would include testing alternative plasma membrane isolation strategies such as CSC and gradient density centrifugation to assess purity and yield of plasma membrane proteins prior to GST-pulldown optimisation.

An important limitation for the workflow presented in Part B is the difficulty of solubilising different membrane proteins in a single detergent. Functional solubilisation of the full range of membrane proteins is impossible with a single detergent owing to the heterogeneity of protein structure and the differences between a bilayer and micellar environment. Previous AP-MS studies targeting membrane proteins have addressed this issue by extensive detergent screening and the use of multiple overlapping datasets from cells solubilised in different detergents (Babu et al., 2012). Indeed, Babu *et al*. found that CHAPS detergent was not ideal for solubilisation of a large panel of tandem affinity purification (TAP)-tagged membrane proteins, preferring the detergents Triton X-100, n-dodecyl- b-D-maltopyranoside (DDM), and octaethylene glycol monododecyl ether (C12E8). In order to maximise the chances of solubilising FEME cargoes, membranes should be solubilised in a range of detergents in future studies.

4. Investigating FEME regulation through affinity purification-MS of endophilin

4.1. Introduction

4.1.1. FEME is differentially regulated in different cell types

FEME is a rapid pathway for internalisation of specific cargo and is regulated at several levels. Firstly, cargoes in the plasma membrane are only internalised upon receptor-ligand interaction (Boucrot *et al.*, 2015). Secondly, the pre-enrichment of endophilin at the plasma membrane promotes prompt FEME responses (Chan Wah Hak *et al.*, 2018). Studies of FEME in a range of cell types have shed further light on how FEME is regulated.

FEME activity is measured by quantifying the average number of endophilin-positive assemblies (EPAs) proximal to the cell's leading edge in confocal microscope images. Different cell types were found to display different basal rates of FEME activity when growing in their respective complete media. HEK293, HeLa and BSC1 cells displayed a low number of EPAs (1 to 3 per 100 μm^2), whilst RPE1, human dermal fibroblasts (hDFA) and human umbilical vein endothelial cells (HUVEC) displayed higher numbers of EPAs (5 to 15 per 100 μm^2) (Ferreira *et al., submitted*). There was also a difference in the fraction of cells displaying FEME between cell types: HUVEC cells displayed the highest rate of FEME-competent cells, with an average of 60% of cells displaying EPAs in their cytoplasm.

In addition to cell type, another factor which affected FEME rate was the concentration of serum in cell media. Addition of 10% serum to cell media for a few minutes (5 to 20 min) robustly increased the frequency of EPAs observed all cell types tested. However, FEME was never observed in every single cell in a culture, even upon stimulation. Consistently, exchanging cells

into serum-free media suppressed FEME. These observations implied the existence of a mechanism by which individual cells in a culture are either FEME-blocked or FEME-competent.

4.1.2. GSK3β and CDK5 are negative regulators of FEME

The modulation of serum concentration and therefore growth factor signalling appeared to influence the FEME-competency mechanism. Kinases play an important role in growth factor signalling cascades and are known to modulate other endocytic pathways (Liang *et al.*, 2007; Clayton *et al.*, 2010; Smillie and Cousin, 2012; Reis *et al.*, 2015). A kinase inhibitor screen performed in our lab identified Glycogen synthase kinase 3 (GSK3) and Cyclin dependent kinase 5 (CDK5) as negative regulators of FEME in RPE1 cells (*Ferreira et al., submitted*). A range of small molecule inhibitors were chosen for testing on the basis of their reported roles in cytoskeletal regulation. Growing cells were treated with each inhibitor for 10 minutes at 37°C, then fixed in paraformaldehyde to preserve FEME carriers. Confocal microscopy was used to assess the effects of acute inhibition on the numbers of EPAs near the cell leading edge. The effects were scored with reference to EPA numbers observed under conditions of "Decreased", "Normal" or "Increased" FEME (Figure 4.1a). "Decreased" FEME was achieved by inhibition of PI3K and "Increased" FEME was induced by stimulation with the β_1-AR agonist dobutamine. "Decreased" FEME was assigned for samples with >80% reduction in the number of EPAs, in at least 50% of the cells. "Increased" FEME was attributed to samples with >200% elevation in the number of EPAs, in at least 50% of the cells.

As shown, inhibition of several CDKs, GSK3 and p38 was found to significantly increase FEME in RPE1 cells, whilst inhibition of CAMKK1/2, SYK, FAK and mTORC1/2 kinases significantly decreased FEME (Figure 4.1b). Individually, inhibition of CDKs 1 and 2 did not cause significant changes to FEME rate, and the nuclear localisation of CDK7 and 9 argued against their role in acute modification of endocytic events at the leading edge (Malumbres, 2014) (Figure 4.1b). Therefore, CDK5 appeared to be a negative regulator of FEME. Further work in our lab was able

to show that GSK3 and CDK5 inhibition enhanced FEME in a dose-dependent manner, confirming the results of the screen (Figure 4.1c).

Kinases identified in this screen could regulate FEME at various stages of the pathways and in a variety of modes. Endophilin presented an ideal candidate to define this regulatory network owing to its central role in FEME. We designed experiments to assess if endophilin binding partners changed upon growth factor/ kinase activity modulation through AP-MS.

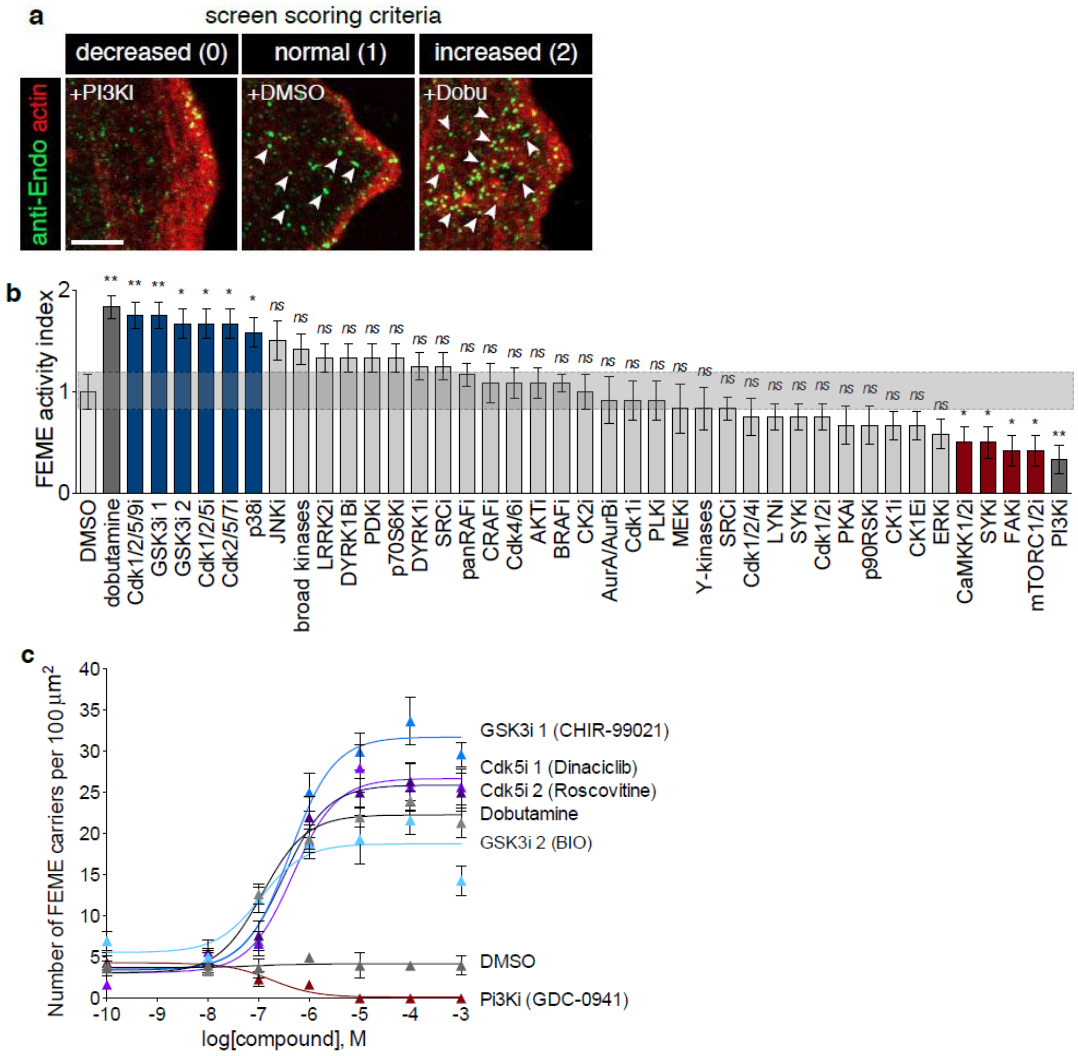

Figure 4.1 CDK5 and GSK3 negatively regulate FEME. **a**, Scoring criteria used in the kinase screen. Representative images of 'decreased', 'normal' and 'increased' FEME in resting human RPE1 cells treated with 10µM dobutamine, 10µM DMSO and 10 nM GDC-0941 (PI3Ki), respectively. Arrowheads point at FEME carriers. 'Decreased' FEME was assigned for samples with >80% reduction in the number of EPAs, in at least 50% of the cells. 'Increased' FEME was attributed to samples with >200% elevation in the number of EPAs, in at least 50% of the cells. The corresponding scoring marks were 0, 1 and 2, respectively. **b**, Kinase screen using small compound inhibitors. **c**, Number of FEME carriers (cytoplasmic Endophilin-positive assemblies, EPAs) upon titration of CHIR-99021, BIO, Roscovitine and Dinaciclib. Dobutamine and GDC-0941 were used as positive and negative controls, respectively. All experiments were repeated at least three times with similar results. Statistical analysis was performed by one-way ANOVA: NS, non-significant; *, P<0.05, **, P <0.01, ***, P <0.001. Scale bars, 5µm.

(adapted from Ferreira et al., submitted)

4.2. Comparison of the endogenous endophilin interactome during high and low FEME activity by affinity purification-MS

4.2.1. Experiment aim

Endophilin is involved at many stages of the novel endocytic pathway, FEME. During FEME carrier formation, endophilin plays a role in cargo selection and membrane curvature stabilisation. Endophilin recruits the GTPase dynamin to effect carrier scission and coats the carrier as it is trafficked into the cell interior, possibly serving as a link to the cell's motor protein machinery. Work in our lab determined that FEME can be enhanced by addition of serum to cell culture media and suppressed by serum withdrawal, implicating growth factor signalling in FEME regulation. Regulation of FEME might be achieved by modification of endophilin or of its binding partners at any of the aforementioned stages. We designed an experiment to investigate how growth factor signalling impacts FEME.

Affinity purification mass spectrometry (AP-MS) was used to identify proteins in endogenous endophilin A2 complexes purified from RPE1 cells with and without prior serum stimulation. Any changes in the endophilin interactome between these conditions could represent regulatory targets downstream of growth factor signalling (Figure 4.2).

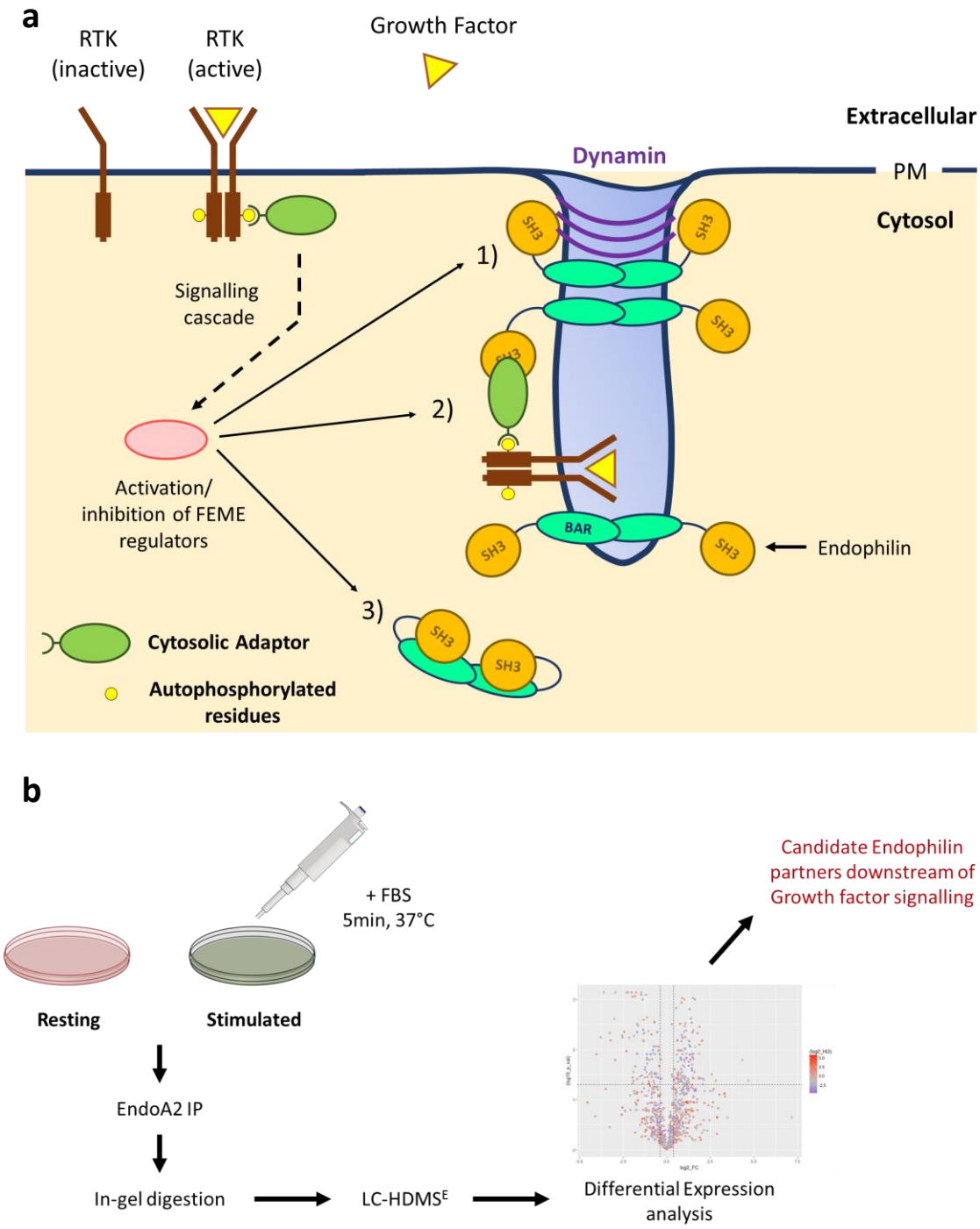

Figure 4.2 Investigating changes in the endophilin interactome upon serum modulation: experimental concept and workflow. **a**, Diagram to illustrate possible modes of FEME regulation via modification of endophilin. Activation/inhibition of FEME regulators could promote/inhibit endophilin functions at various stages: 1) Recruitment of dynamin to vesicle necks, 2) Cargo sorting via direct/indirect interaction with activated receptors (indirect interaction via cytosolic adaptors is shown), 3) Release of endophilin autoinhibition. **b**, Schematic to show experimental workflow. RPE1 cells were harvested under resting conditions or following stimulation with Foetal Bovine Serum (FBS). Endophilin A2 (EndoA2) complexes were immunoprecipitated (IP) and subjected to in-gel digestion. Peptides were analysed by LC-HDMS^{E}. Differential expression analysis was used to identify candidate endophilin binding partners and compare their affinity upon growth factor modulation.

4.2.2. Cell culture, sample preparation by in-gel digestion and LC-MS analysis of peptides

RPE1 cells were grown on 100 mm dishes to a confluence of 80% before harvest. For the stimulated condition, cell media was supplemented with extra 10% FBS (20% final) for 5 min at 37°C. For the resting condition, no additional treatment was performed prior to cell lysis. Cells were washed 2 times in ice-cold PBS before being gently scraped into lysis buffer, incubated on ice for 30 minutes and centrifuged. Anti-endophilin A2 beads were incubated with the clarified cell lysate overnight at 4°C with end-over-end rotation. The beads were washed to remove/reduce non-specific interactors and boiled in SDS sample buffer to release the captured protein complexes.

Aliquots of the resulting eluates were subjected to western blotting and displayed successful IP of the known endophilin binding partner Dynamin (Figure 4.3). Samples prepared for LC-MS analysis 1D SDS-PAGE separation and in-gel digestion with trypsin, cutting lanes into 10 fractions. Cell culture, immunoprecipitation and in-gel digestion were performed by J. Panambalana. The resulting peptides were resuspended in LC-MS running buffer and samples were spiked uniformly with peptides from *E. coli* ClpB protein to act as a standard for Hi3 quantitation. Peptides were separated using a 60-minute gradient (0-40% acetonitrile) by reversed phase LC coupled online to a quadrupole time-of-flight mass spectrometer (Synapt-G2-Si, Waters) via a nanoESI sprayer. The instrument was operated in $HDMS^E$ mode. Samples were analysed in biological and technical duplicate.

Raw data was analysed using Progenesis v4.0 (Waters, UK) using a Human UniProt FASTA database to match ions to peptides and proteins. Samples were normalised to Endophilin peptide abundance and fractions were combined *in-silico* in Progenesis to obtain absolute protein abundances for differential expression analysis.

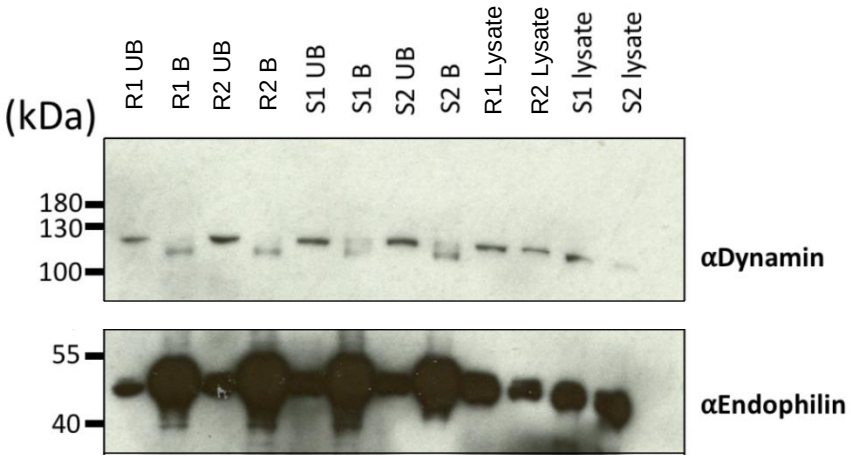

Figure 4.3 Western blot to assess efficacy of Endophilin IP. Bound fractions shown represent 20% sample loading. All other lanes represent 2% sample loading. As shown, a subset of the available dynamin co-immunoprecipitates with Endophilin A2, indicating that IP was successful. R – Resting, S – Stimulated, UB – unbound, B – Bound fraction. Numbers indicate biological sample 1 or 2.

4.2.3. Data quality control

Protein groups identified with a minimum of 3 peptides were quantified using the Hi3 method to give an absolute protein abundance in femtomoles (Silva *et al.*, 2006). After removing contaminants, a total of 3197 protein groups were quantified. Quantitative reproducibility was assessed at the level of combined technical replicates (protein amounts from all 10 fractions summed in Progenesis). Samples displayed a high degree of quantitative reproducibility, with a mean within-condition Pearson correlation of 0.96 for resting samples and 0.97 for stimulated samples (Figures 4.4a and 4.4b) indicating reproducible sample preparation and instrument performance. Interestingly, the correlation in protein abundance across conditions was also very high, with a mean between-condition Pearson correlation coefficient of 0.93. This suggests that the endophilin interactome did not change dramatically between the 2 conditions. As correlation coefficients can only describe sample similarity insofar as protein groups are shared between samples, it could be argued that the high scores across conditions do not account for proteins not shared between conditions (unique hits). However, as Progenesis assigns a protein abundance to all features in all runs (sourcing from background noise if no peptides are present), there are no missing values present in the dataset and thus no proteins unaccounted for when calculating correlation coefficients in this case.

A principle component analysis (PCA) was performed using the Perseus software platform to represent similarity in protein abundance among the samples in 2 dimensions (Figure 4.4c). The tight clustering of technical replicates indicates good technical precision in the data. In agreement with the previously described correlations, the stimulated samples display higher similarity in protein abundance, clustering closely in the bottom left of the plot. Samples from the resting condition are spread further apart in the plot, but are still well separated from the stimulated samples, arguing for the biological validity of conclusions drawn from this data.

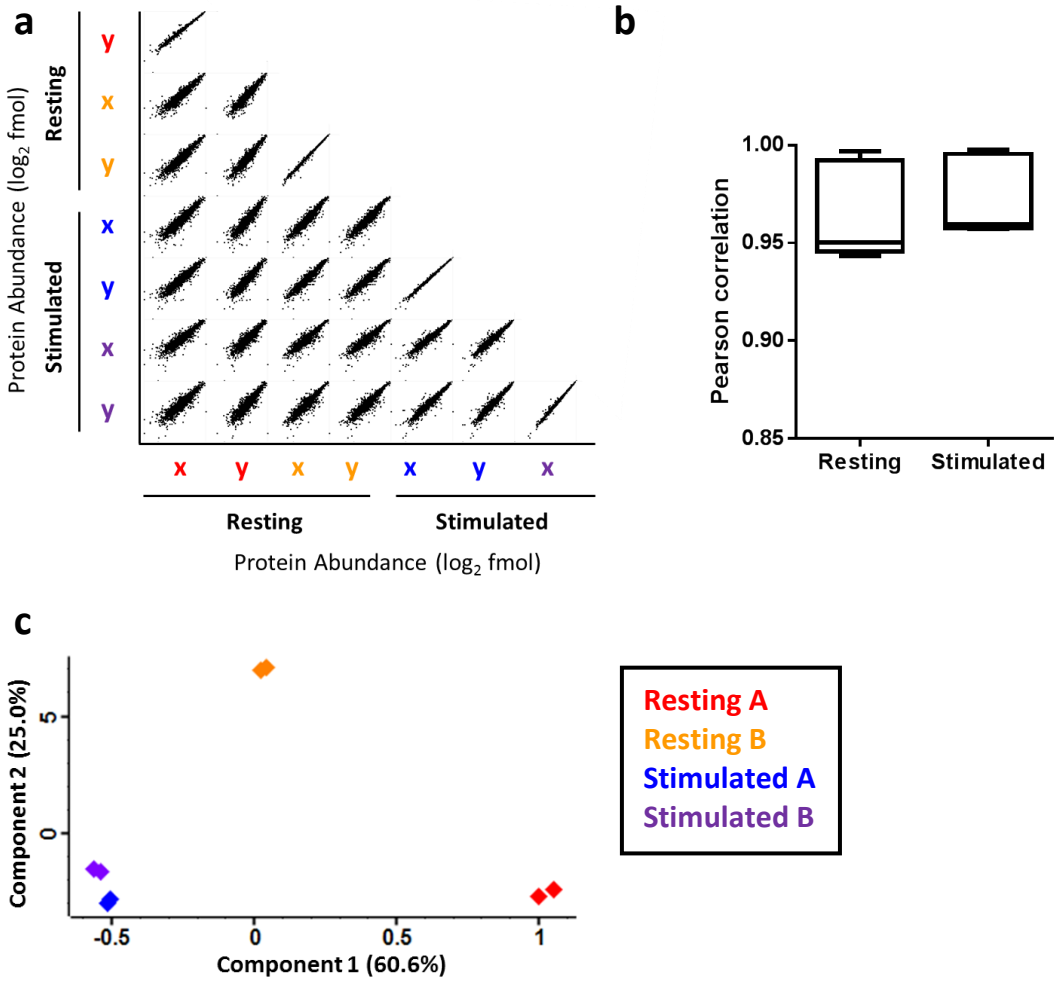

Figure 4.4 Evaluation of instrument performance and data structure. **a**, Multiscatter plot to illustrate correlation in protein abundance across technical replicates and experimental conditions. Technical replicates are denoted x or y, colours correspond to biological sample. Red – Resting A, Orange – Resting B, Blue – Stimulated A, Purple – Stimulated B. **b**, Boxplot to show distribution of Pearson correlation coefficients between all samples within each condition. Whiskers represent minimum and maximum values. **c**, PCA analysis to illustrate similarity among technical replicates in 2 dimensions. Filled diamonds represent individual technical replicates, colours represent biological samples as in **a**.

4.2.4. Differential expression analysis

Absolute protein abundances as estimated by the Hi3 method in recombined samples were averaged across technical replicates. Proteins were classified as differentially expressed (implying a significant change in protein occupancy in endophilin A2 immunoprecipitates between conditions) upon fulfilment of two criteria: 1) protein abundance fold-change greater than 1.3-fold in either direction and 2) q-value < 0.05. So-called q-values were calculated from *p*-values from Student's *t*-tests comparing mean protein abundance between the 2 conditions (2-sided, unpaired, assuming equal variance). The obtained *p*-values were corrected for dataset-specific false discovery rate through a permutation-based methodology implemented in the Perseus platform to convert them into q-values (Tyanova *et al.*, 2016). Encouragingly, Endophilin A2 itself was not differentially expressed across conditions, arguing for the validity of our normalisation strategy. Five proteins of the 3197 quantified were found to be differentially expressed between the resting and stimulated conditions (Table 4.1). Of these, Cutaneous T-Cell Lymphoma-Associated Antigen 6 (CTAGE6) is the most interesting. Observed at higher levels in the stimulated condition, this single-pass transmembrane protein has predicted functions in intracellular trafficking and contains a long proline-rich C-terminal region suitable for engagement by the Endophilin A SH3 domain. The other identifications are either nuclear-resident proteins (Ribosomal RNA processing protein 1 homolog B and Sister chromatid cohesion protein PDS5 homolog A; both enriched in the stimulated condition) or do not have an obvious connection to FEME (a cuticular Keratin variant and the protease inhibitor Serpin B4; both enriched in resting condition). A possible reason for the observation of nuclear proteins with endophilin in the IPs presented here is the loss of organelle integrity upon cell lysis, resulting in the interaction of proteins which would normally be localised in different parts of the cell *in vivo*. This analysis did not identify any high-confidence candidate regulators of FEME.

Owing to the use of only 2 biological replicates, the analysis lacks power. In order to gain more information from the dataset, the criteria for differential expression were relaxed and the raw *p*-values were used in place of the corrected q-values. This expanded pool of differentially expressed proteins was examined with the caveat of accepting a higher rate of type I errors.

Under these criteria, 141 proteins were enriched in the resting condition and 139 proteins were enriched in the stimulated condition (Figure 4.5 and listed in Appendix Table 8.3). The behaviour of known endophilin binding partners was assessed. The E3 ubiquitin-protein ligase CBL (CBL) was identified as an adaptor for RTK sorting into FEME carriers and interacts directly with endophilin (Soubeyran *et al.*, 2002; Boucrot *et al.*, 2015). In agreement with this, it is enriched in the stimulated condition, possibly reflecting increased engagement with endophilin at the plasma membrane and on intracellular FEME carriers. By contrast, the known FEME cargo vascular endothelial growth factor receptor 2 (VEGFR2) did not appear to preferentially associate with EndophilinA2 complexes in either condition. Unexpectedly, two proteins which directly bind to endophilin and play obligate roles in FEME, Dynamin-2 and lamellipodin, (Boucrot *et al.*, 2015) were not enriched in either condition.

Of the proteins enriched in the stimulated condition, NEDD4-like E3 ubiquitin-protein ligase (WWP1) displayed the second highest fold-change enrichment and represents an interesting potential target for future study. WWP1 can associate with the plasma membrane via its N-terminal C2 domain and monoubiquitinates the receptor tyrosine-protein kinase erbB-4 (ERBB4), promoting its degradation in lysosomes and proteasomes (Feng *et al.*, 2009; Y. Li *et al.*, 2009). It's strong enrichment in Endophilin A2 immunoprecipitates following stimulation of cells raises the possibility that it forms a degradative complex in a similar manner to the Endophilin-CIN85-CBL complex which participates in EGFR and HGFR downregulation (Petrelli *et al.*, 2002; Soubeyran *et al.*, 2002).

The lack of FEME activity in cells grown in serum-free conditions implies the involvement of protein kinases as master regulators of FEME. The dataset was specifically probed to identify candidate kinases for this role. To do this, *t*-tests were performed on the technical replicates without averaging and corrected using the Benjamini-Hochberg procedure to obtain corrected q-values. A list of kinases with q-values below 0.05 and with fold changes greater than 1.3-fold in either direction was compiled, noting that this dataset represents a detailed comparison between 2 biological replicates and extreme caution should be taken when generalising to RPE1 cell

populations (Table 4.2). Of the 27 kinases identified this way, Glycogen synthase kinase 3β (GSK3β) represented an interesting candidate for future investigation. In a kinase inhibitor screen performed in our lab, GSK3β was found to negatively regulate FEME in RPE1 cells (*Ferreira et al., submitted*). As GSK3β can be inactivated by growth factor signalling (Cross *et al.*, 1995; Patel and Woodgett, 2017), the enrichment of GSK3β in resting rather than stimulated endophilin IPs raised the possibility that endocytic flux through the FEME pathway could be controlled by binding and likely phosphorylation of endophilin by GSK3β in conditions of low growth factor stimulation.

Protein name	Gene name	Fold Change (Stimulated/Resting)	q-value
cTAGE family member 6	CTAGE6	2.317	0.0414
Ribosomal RNA processing protein 1 homolog B	RRP1B	2.233	0.0414
Sister chromatid cohesion protein PDS5 homolog A	PDS5A	1.544	0.0456
Keratin_ type I cuticular Ha1	KRT31	0.679	0.0414
Serpin B4	SERPINB4	0.382	0.0414

Table 4.1 List of proteins differentially expressed in EndophilinA2 immunoprecipitates from Stimulated and Resting conditions, stringent hypothesis testing. Hypothesis testing was performed using an unpaired, 2-sided student's t-test on mean protein abundance from both technical replicates per sample. To correct for multiple testing, a permutation-based estimate of dataset FDR was calculated and used to correct *p*-values to q-values. Proteins enriched in Resting condition are shaded in grey.

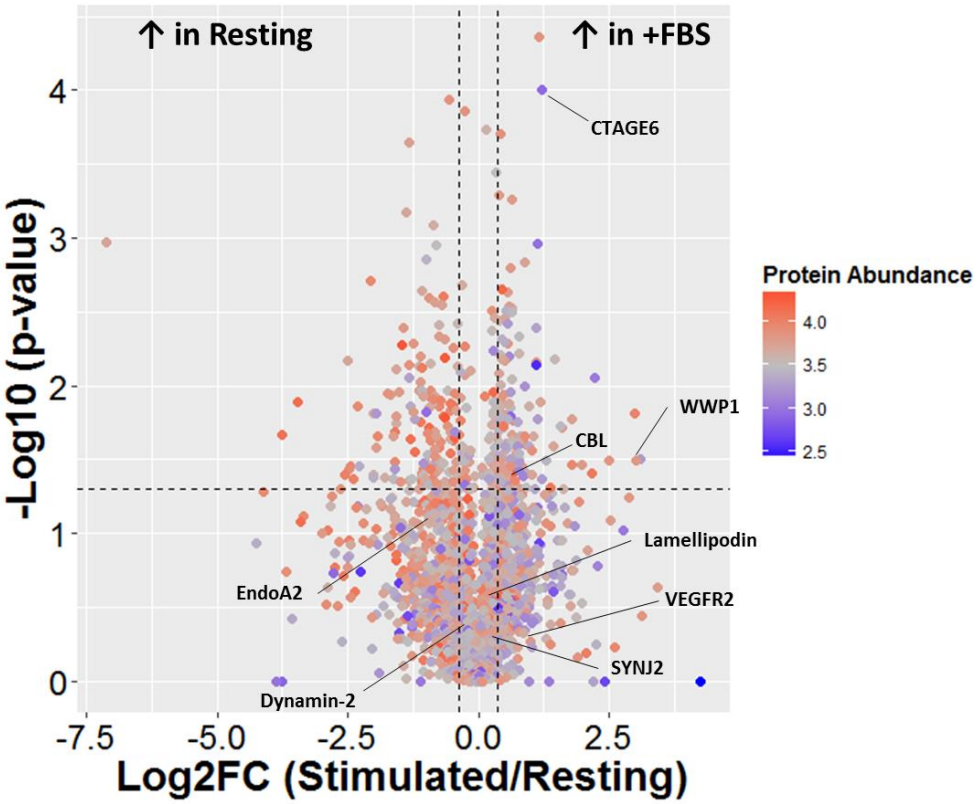

Figure 4.5 Volcano plot to illustrate quantitative differences between proteins in endophilin A2 immunoprecipitates under Resting and Stimulated (+FBS) conditions. For each protein, mean abundance and standard deviation (n=2) was found and used to perform unpaired, 2-sided t-tests and to calculate fold change in protein abundance between conditions. These values were log transformed such that proteins with data in the top left and top right quadrants are enriched in Resting and Stimulated conditions respectively. Dashed vertical lines represent fold change cut-offs for significance of 1.3-fold in either direction. The dashed horizontal line represents *p*-value cut-off for significance of 0.05. Plot colours reflect the highest protein abundance value from each pairwise comparison.

Protein name	Gene name	Fold Change (Stimulated/Resting)	Benjamini Hochberg q-value
SRSF protein kinase 2	SRPK2	4.528	0.0072
Diacylglycerol kinase zeta	DGKZ	3.458	0.0024
Kinase suppressor of Ras 1	KSR1	2.058	0.0232
Mitogen-activated protein kinase kinase kinase kinase 5	MAP4K5	1.662	0.0053
LIM domain kinase 2	LIMK2	1.633	0.0379
Serine/threonine-protein kinase RIO2	RIOK2	1.552	0.0040
Serine/threonine-protein kinase OSR1	OXSR1	1.497	0.0006
Cyclin-dependent kinase 17	CDK17	1.486	0.0009
Cyclin-dependent kinase 15	CDK15	1.454	0.0433
5'-AMP-activated protein kinase catalytic subunit alpha-1	PRKAA1	1.410	0.0182
Dual specificity protein kinase TTK	TTK	1.376	0.0143
Serine/threonine-protein kinase ICK	ICK	1.341	0.0356
Serine/threonine-protein kinase 4	STK4	1.313	0.0153
Mitogen-activated protein kinase kinase kinase kinase 3	MAP4K3	1.308	0.0299
Protein kinase C iota type	PRKCI	1.298	0.0136
Serine/threonine-protein kinase LATS1	LATS1	0.764	0.0455
Cyclin-dependent kinase 3	CDK3	0.745	0.0067
Glycogen synthase kinase-3 beta	GSK3B	0.628	0.0314
Tyrosine-protein kinase JAK1	JAK1	0.608	0.0174
Tyrosine-protein kinase Blk	BLK	0.562	0.0126
Serine/threonine-protein kinase PRP4 homolog	PRPF4B	0.479	0.0051
Casein kinase II subunit alpha	CSNK2A1	0.426	0.0018
Serine/threonine-protein kinase TAO2	TAOK2	0.407	0.0221
Cyclin-dependent kinase 18	CDK18	0.386	0.0040
Serine/threonine-protein kinase Nek1	NEK1	0.347	0.0012
Serine/threonine-protein kinase B-raf	BRAF	0.317	0.0070
FAST kinase domain-containing protein 2_ mitochondrial	FASTKD2	0.201	0.0045

Table 4.2 List of protein kinases differentially expressed between Stimulated and Resting conditions, permissive hypothesis testing. Tests was conducted on technical replicates without averaging. The Benjamini Hochberg correction for multiple testing was used to calculate q-values. Proteins enriched in Resting condition are shaded in grey.

4.3. Endogenous GSK3β interacts with the Endophilin A2 SH3 domain under conditions of serum starvation

The serine/threonine kinases GSK3α and β are ubiquitously expressed paralogs distinguished by their wide variety of intracellular targets (Lau *et al.*, 1999; Patel and Woodgett, 2017). In a kinase inhibitor screen performed in our lab, GSK3α/β were shown to negatively regulate FEME activity in RPE1 cells (Figure 4.1). Negative regulation of FEME activity through GSK3 can proceed via several routes: 1) phosphorylation of endophilin resulting in blockage of partner binding sites/ adoption of an autoinhibited fold 2) phosphorylation of an endophilin inhibitor which potentiates inhibitor binding to endophilin 3) phosphorylation of endophilin partners which abrogate binding and 4) phosphorylation of proteins which promote sequestration of endophilin binding partners. Casting light on such mechanistic details could help understand the control of FEME competency in cells and may provide a tool for selective inhibition of the pathway for future study.

AP-MS data presented in Section 4.2 hinted that GSK3β was more strongly associated with endogenous Endophilin in cells undergoing lower growth factor stimulation. GSK3 phosphorylates serine (S) or threonine (T) residues 4 amino acids N-terminally to a pre-phosphorylated S or T residue (Cohen and Frame, 2001). It therefore requires targets proteins to be initially modified by a different kinase before it propagates this signal to nearby S/T sites. GSK3 activity can be regulated by phosphorylation of S9 at the protein's N-terminus. Phosphorylated S9 (pS9) can occupy the docking site which recognises pre-phosphorylated residues on target proteins, inhibiting kinase activity (Cross *et al.*, 1995; Frame, Cohen and Biondi, 2001). Phosphorylation of S9 is downstream of several growth factor signalling cascades (Cross *et al.*, 1995; Patel and Woodgett, 2017), and GSK3 activity is therefore enhanced under conditions of low growth factor stimulation. The results of the kinase screening together with the detection of GSK3β in endophilin IPs from resting but not stimulated cell extracts led us to wonder if GSK3β negatively regulates FEME by binding to and phosphorylating Endophilin.

To determine if Endophilin and GSK3β could interact in the cell, and if this interaction was subject to growth factor signalling, we performed GST pulldowns, using the Endophilin A2 SH3 domain as a bait, on RPE1 cell extracts from 3 conditions: Serum-free, resting and stimulated. For preparation of extracts for the serum-free condition, RPE1 cells growing in complete media were exchanged into serum-free media for a minimum of 2 hours at 37°C prior to harvest. For the stimulated condition, RPE1 cell media was supplemented with Foetal Bovine Serum (FBS) to a final concentration of 20% v/v for 5 minutes at 37°C prior to harvesting cells. For the Resting condition, cell extracts were harvested without pre-treatment (growing in media supplemented with 10% v/v FBS).

GST and GST-EndophilinA2 SH3 (GST-EndoSH3) magnetic agarose beads were exposed to 500 µl clarified RPE1 cell lysate overnight at 4°C with agitation. Cell culture, treatment and pulldown assays were performed by A. Casamento. Following overnight incubation beads were washed and the captured proteins were eluted by boiling in SDS sample buffer. Western blotting was performed to assess the amount of GSK3α/β in the bound fraction of each sample. The anti-GSK3 antibody chosen recognises an epitope present in both GSK3α and GSK3β, enabling quantification of both proteins on the same blot (Figure 4.6a).

Western blot signal was analysed by densitometry. Neither GSK3 protein copurified strongly with the GST-only beads, indicating low non-specific interaction between GSK3 proteins and the GST tag. To assess if GSK3α/β binding to Endophilin A2 SH3 was altered by the cell treatments, GSK3β signal in each bound fraction was expressed as a percentage of input signal, and this % recovery was normalised to the % recovery of GSK3α/β in GST-only Resting pulldowns. Statistical analysis was performed using two-way ANOVA. Differences in mean normalised recovery arising from GST constructs were found to be significant whilst differences in means arising from treatment conditions were non-significant. Pairwise comparisons using Tukey's multiple comparison test found a significant difference between GSK3β levels in GST versus GST-EndoSH3 eluates from cells grown in serum-free media prior to harvest, but not between GSK3β levels in GST vs GST-EndoSH3 eluates in the other 2 conditions (Figure 4.6b, top).

Interestingly, pairwise comparisons using Tukey's multiple comparison test did not show a significant difference in GSK3α levels between GST and GST-EndoSH3 eluates in any condition (Figure 4.6b, bottom). This data suggests that GSK3β rather than GSK3α engages endophilin via the endophilin SH3 domain under conditions of low growth factor signalling.

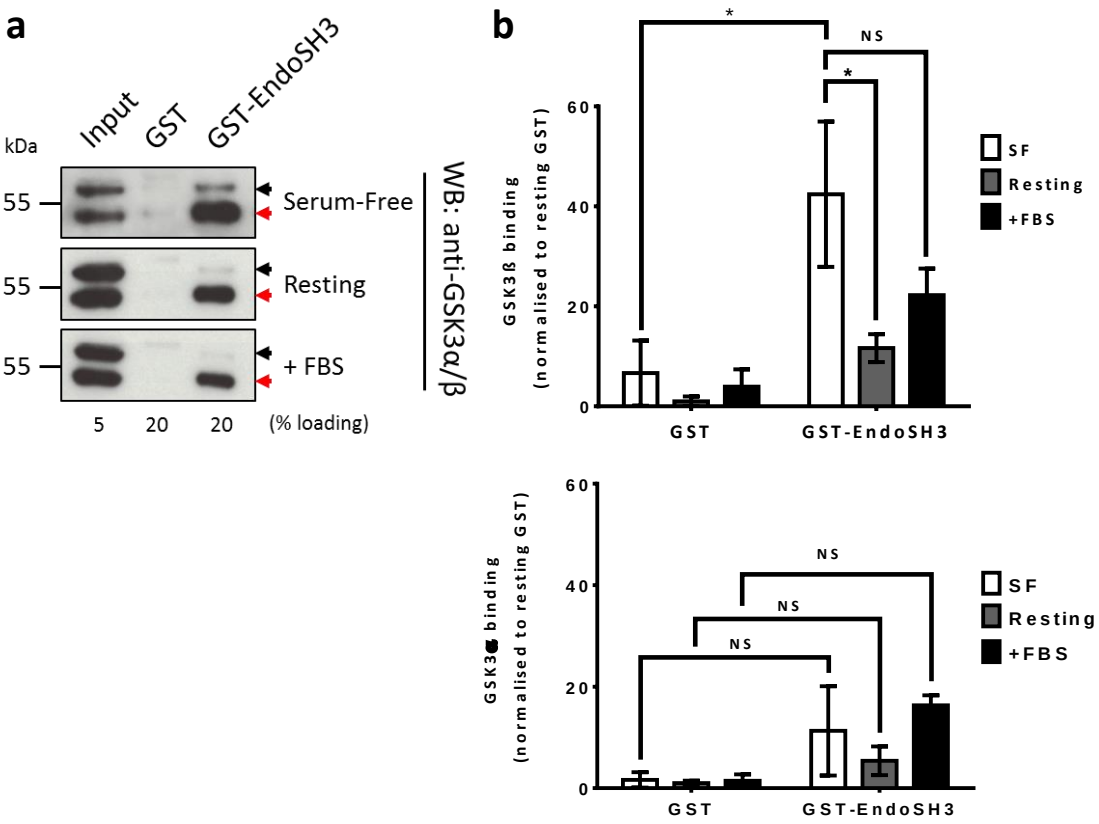

Figure 4.6 The association of endogenous GSK3β with Endophilin A2 SH3 domain in GST pull-down assays is altered by serum levels in cell media. **a**, Representative western blots to show GSK3α/β levels in different fractions of GST pulldown experiments conducted on RPE1 cell extracts. Cells were grown in serum-free (SF), FBS stimulated (+FBS) or normal (Resting) media prior to harvesting lysates. A GST-Endophilin A2 SH3 contruct (GST-EndoSH3) was used as bait and GST beads were included as a negative control. Black arrowheads indicate GSK3α signal, Red arrowheads indicate GSK3β signal. **b**, Bar charts to show densitometric quantitation of GSK3β (top) and GSK3α (bottom) signal in western blots (means ± s.e.m. from three independent experiments). Statistical analysis was performed by 2-way ANOVA: NS, non-significant, *, $P < 0.05$.

4.4. Profiling of the endophilin interactome following kinase inhibition by affinity purification-MS

4.4.1. Experiment aim and strategy

Endophilin is involved at many stages of the novel endocytic pathway, FEME. During FEME carrier formation, endophilin plays a role in cargo selection and membrane curvature stabilisation. Endophilin recruits the GTPase dynamin to effect carrier scission and coats the carrier as it is trafficked into the cell interior, possibly serving as a link to the cell's motor protein machinery.

Work in our lab established a link between inhibition of the kinases CDK5 and GSK3β and upregulation of FEME (*Ferreira et al., submitted*). CDK5 and GSK3β negatively may regulate FEME by modifying the binding partners of proteins involved in FEME, including endophilin. Negative regulation of FEME activity through a kinase targeting endophilin can proceed via 4 routes: 1) phosphorylation of endophilin resulting in blockage of partner binding sites/ adoption of an autoinhibited fold 2) phosphorylation of an endophilin inhibitor which potentiates binding to endophilin 3) phosphorylation of endophilin partners which abrogate binding and 4) phosphorylation of proteins which promote sequestration of endophilin binding partners. In all scenarios, we hypothesised that acute inhibition of CDK5 and GSK3β activity would result in increased numbers of endophilin molecules in complex with proteins that act in the FEME pathway. Affinity purification mass spectrometry (AP-MS) can be used to identify and quantify proteins in these complexes, identifying potential FEME proteins. We designed a quantitative AP-MS experiment to identify proteins which are found in complex with endophilin upon inhibition of CDK5 and GSK3β.

Ideally, quantitative measurements of proteins in endophilin complexes would be compared following immunoprecipitation (IP) of complexes obtained from with and without kinase inhibitor

treatment. Due to time constraints, we were only able to perform experiments on cells treated with kinase inhibitor (Figure 4.7). This represents the widest possible net of FEME-relevant endophilin interactors but forgoes the opportunity to screen out cytosolic endophilin partners which participate in non-FEME related functions. Relatedly, IP of endogenous endophilin is predicated on availability of beads coated in α-endophilin antibodies, which were not available for this experiment. As a compromise, we opted to perform IPs of GFP-tagged endophilin A2 (EndoGFP) expressed transiently in human RPE1 cells. To identify cytosolic proteins which non-specifically bind to the eGFP tag, we also performed IPs from cells overexpressing the eGFP tag alone (hereafter GFP). An on-bead digest coupled to strong cation exchange ('On-bead-SCX') protocol was performed to facilitate faster sample preparation and MS analysis and to reduce NP-40 detergent contamination of peptide eluates following trypsin digestion.

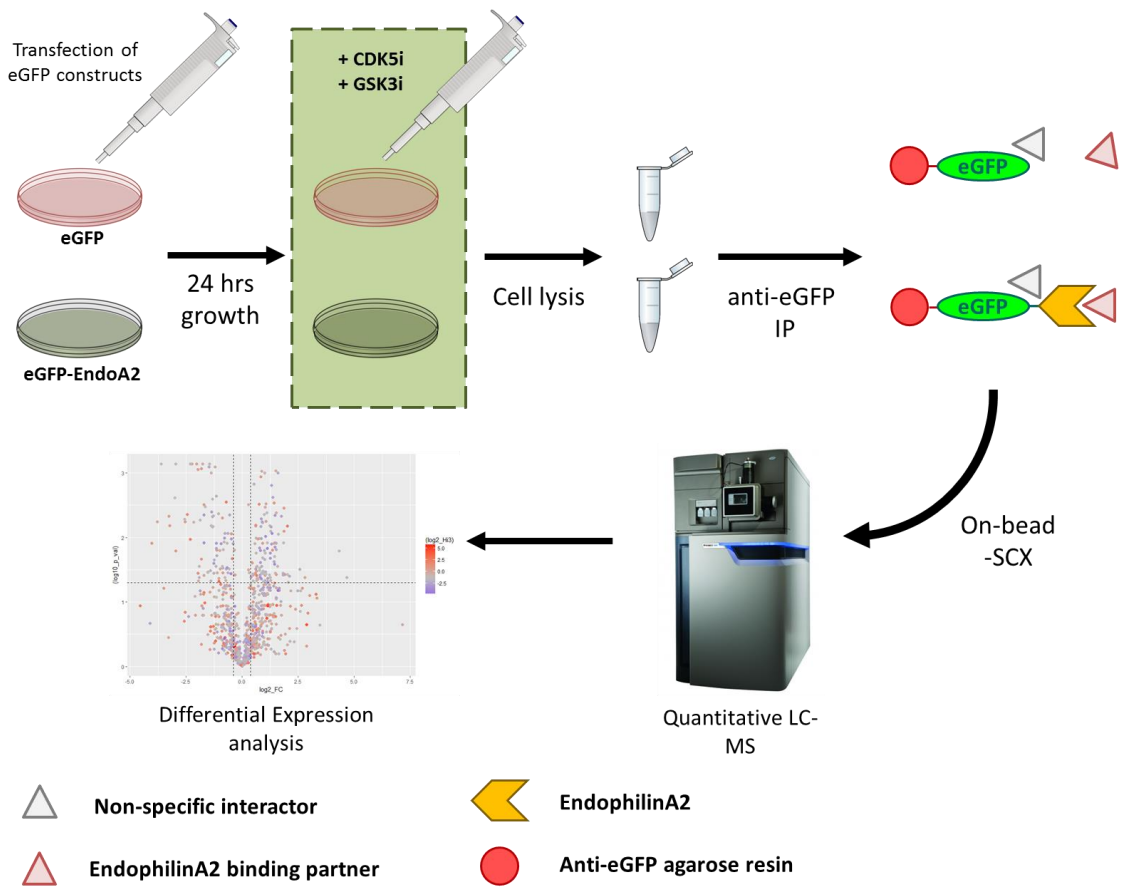

Figure 4.7 Experimental workflow to profile the endophilin interactome following kinase inhibition.

4.4.2. IP-MS of EndoGFP complexes and sample preparation for MS

Four dishes of RPE1 cells were seeded for each sample to generate enough material for MS sample preparation. Each 100 mm dish of RPE1 cells was transfected with 3 µg of plasmid encoding eGFP-only or eGFP-EndophilinA2 and grown for 24 hours. Cells were treated with the CDK1/2/5/9 inhibitor Dinaciclib and the GSK3α/β inhibitor CHIR-9902 at a working concentration of 1 µM for 10 min at 37°C in an incubator prior to cell lysis. Following treatment, cells were washed in cold PBS and harvested by gently scraping into lysis buffer. Samples were incubated on ice to facilitate complete cell lysis then centrifuged to remove the nuclear fraction and insoluble material. The resulting supernatant was mixed with 20 µl of GFP-Trap® magnetic agarose slurry and incubated for 3 hours at 4°C with end-over-end mixing to promote capture of eGFP-tagged proteins in complex with cytosolic binding partners. Using a magnetic rack, beads were washed 3 times in 500 µl Lysis buffer prior to On-bead-SCX digestion.

Following immunoprecipitation, GFP-Trap® beads were washed ice-cold Tris-buffered saline (TBS) solution and resuspended in Tris-Urea buffer containing 400 ng of trypsin and incubated at 27°C for 30 min. Reduction in DTT and trypsinisation were performed simultaneously in an overnight step, following by alkylation of peptides.

Previous MS-analysis of on-bead digests from cell lysates using an NP-40 lysis buffer displayed evidence of contamination by a PEGylated substance despite C18 SPE cleanup, suggesting that NP-40 is not removed effectively using this method. Samples were therefore desalted using C18-SCX StageTips in an effort to reduce the amount of non-ionic NP-40 detergent in the final sample for injection onto the LC column. Peptides were eluted from the SCX bed using 1000mM ammonium acetate solution, dried in a SpeedVac and resuspended in LC-MS running buffer for analysis. Peptides were separated using a 60-minute gradient and ionised by nanoESI. Data acquisition was performed using ion-mobility assisted tandem MS/MS ($HDMS^E$) on a Waters Synapt G2-Si q-TOF instrument.

4.4.3. LC-MS analysis, database searching, quality control and normalisation

Immunoprecipitated complexes from 2 RPE1 overexpression constructs (GFP and EndoGFP) were analysed in 3 biological and technical triplicate. HDMS^E data was processed using Progenesis software (Waters Corp.) which performed retention time alignment and peak picking, followed by noise subtraction and deisotoping. Peptide precursor and fragment ions were matched on the basis of retention and ion mobility drift times, and used for database searching against a canonical Homo sapiens Uniprot FASTA database concatenated the sequences of eGFP, *E. coli* ClpB (Hi3 standard) and a list of common MS protein contaminants obtained from the Global Proteome Machine (The Global Proteome Machine, 2016).

The average intensity of the 3 best ionising peptides for each identified protein (Top3 intensity) in a given sample was summed to obtain an estimate of peptide loading for each technical replicate. Within-sample technical reproducibility was shown to be high (with the exception of the biological replicate A runs from both conditions), but inter-sample summed Top3 intensity was found to be highly heterogeneous (Figure 4.8a). An assessment of instrument performance was carried out to explain these discrepancies. Instrument mass calibration was assessed by plotting the distribution of peptide mass errors (all samples) (Figure 4.8b). The density plot shows a slight right-shift with a median ppm error of +2.2, indicating suboptimal calibration. However, the majority of peptide ions were identified with good mass accuracy (83.1% of ion ±10 ppm). Variation in LC performance was also assessed. Visual inspection of TIC and BPI chromatograms did not reveal major failures in LC performance. A more systematic assessment was performed by plotting the distribution of LC peak widths (FWHM) for all peptides in the experiment (Figure 4.8c). Whilst a large proportion of peptides eluted in the 0 - 20 second range (40.1%), most peptides displayed longer peak widths (59.9% elute in 20+ seconds), indicating a possible issue with LC system performance.

Technical errors incurred during sample preparation could also account for the large differences in Top3 intensity observed. An assessment of on-bead digest efficacy was conducted by plotting the proportion of peptide IDs which feature a missed cleavage site grouped by condition (Figure

4.8d). GFP samples displayed a mean of 39% ± 4% missed cleavages (mean ± SD) and EndoGFP samples displayed a mean of 42% ± 7% missed cleavages (mean ± SD), indicating similar digestion efficacy across all samples. These values compare unfavourably with the digest efficacies achieved for similar on-bead digestion protocols presented in Section 3.2.3 of this report (average 14% missed cleavages), indicating an incomplete digest.

Further issues with sample preparation were identified when comparing construct expression across biological replicates. As shown in Figure 4.9, expression of GFP-tagged proteins was heterogeneous for all samples. It was therefore concluded that major issues with sample preparation and suboptimal instrument performance could not be ruled out as a cause for the high variability observed in summed Top3 intensity across samples.

Normalisation can ameliorate technical errors encountered during sample preparation and instrument operation. Two normalisation methods were compared for their ability to reduce intra-sample variability in Top3 intensities. Normalisation to Global intensity (GI) was performed by dividing the Top3 intensity for a given protein in a given technical replicate by the sum of Top3 intensity of all proteins in that technical replicate, then multiplying this by the median of summed Top3 intensity among all replicates. Data was then $\log_2$ transformed. Normalisation to Median intensity was performed by dividing the Top3 intensity for a given protein in given technical replicate by the median of Top3 intensity for all proteins in that technical replicate, then multiplying this by the median of summed Top3 intensity among all replicates. Data was then $\log_2$ transformed. Normalisation of data was implemented using the NormalyzerDE platform (Willforss, Chawade and Levander, 2019).

Normalisation efficacy at the intra-sample level was assessed by plotting the distribution of pooled CV (average CV for all proteins across intra-sample replicates) and mean Pearson correlation among all intra-sample replicates (Figure 4.10). In comparison to simple Log_2 transformation of the data, both GI and Median normalisation reduced the median and tightened the spread of pooled CV values. Median normalisation was the most effective, reducing median

pooled CV from 3.9% to 3.2% and reducing the interquartile range in CV values from 4.4 to 2.7 in comparison to Log_2 transformed samples (Figure 4.10a). By contrast, none of the normalisation methods had a marked impact on the distribution of mean Pearson correlations (Figure 4.10b).

Normalisation was also assessed at the level of individual technical replicates by plotting the distribution of relative log expression (RLE) values for all proteins in each technical replicate (Figure 4.11). RLE is defined as the ratio between Top3 intensity of a given protein in a given run and the median Top3 intensity of that protein across all runs. Well-normalised samples exhibit a distribution of RLE values centred around 0. Without normalisation, RLE distributions are not well centred at 0, especially for biological replicate A samples (Figure 4.11, top). Samples normalised to Global Intensity display distributions of RLE which are well centred about 0 with the exception of EndoGFP biological replicate B samples, which show a skew toward negative RLE values (Figure 4.11, middle). Median normalisation performed best, with all samples exhibiting RLE distributions centred around 0 (Figure 4.11, bottom). Based on these tests, Median normalisation was applied to the data and protein intensities were averaged across technical replicates prior to further statistical testing.

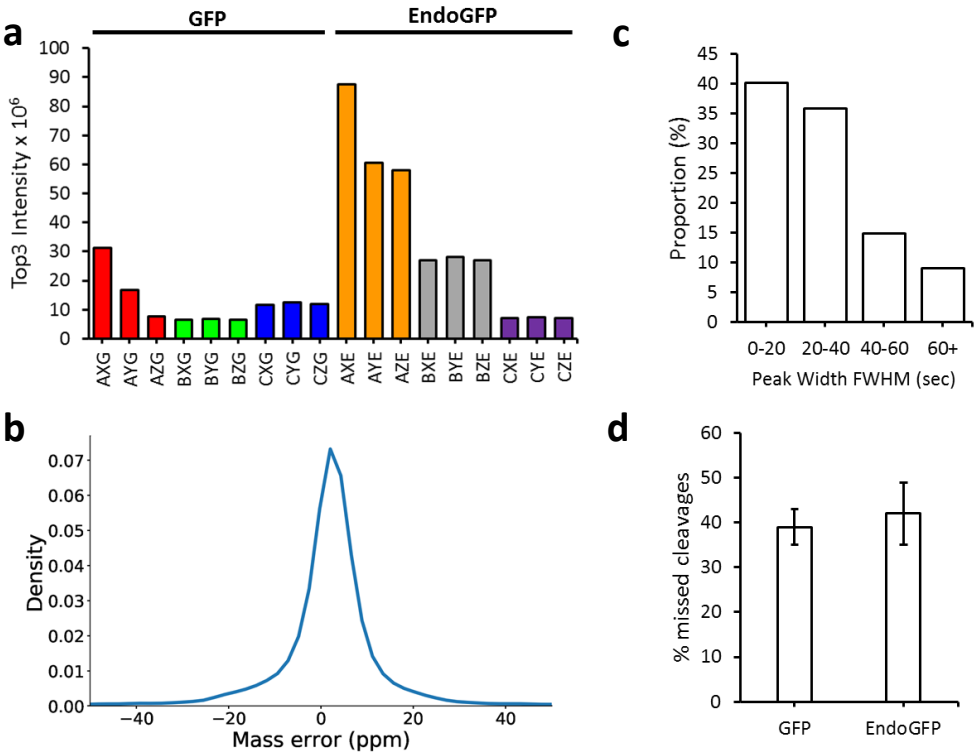

Figure 4.8 Outlier detection and evaluation of LC-MS instrument performance. **a**, Bar chart to show summed intensity of the 3 most intense peptides (Top3 intensity) for all protein groups in each technical replicate, without normalisation. Replicates identifiers first indicate biological replicate A-C, then technical replicate X, Y or Z. The final letter indicates experimental condition (G – GFP, E – EndoGFP). **b**, Density plot to show Peptide mass accuracy distribution. The distribution for all samples is shown. **c**, Histogram to show LC peak width FWHM distribution for all peptides in all samples. **d**, Bar chart to show % of all peptides featuring missed cleavages grouped by condition. Error bars represent ± 1 SD.

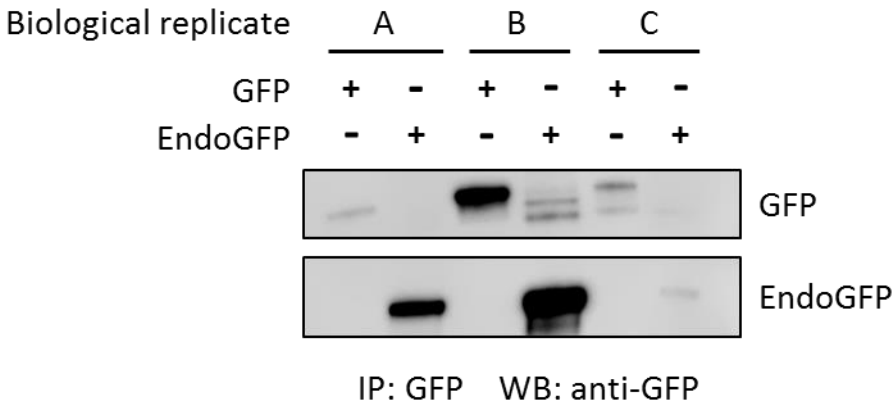

Figure 4.9 GFP construct capture during immunoprecipitation and quantitation by LC-MS was heterogeneous. Western blotting against eGFP was performed on equal volumes of the bound fraction of each IP sample.

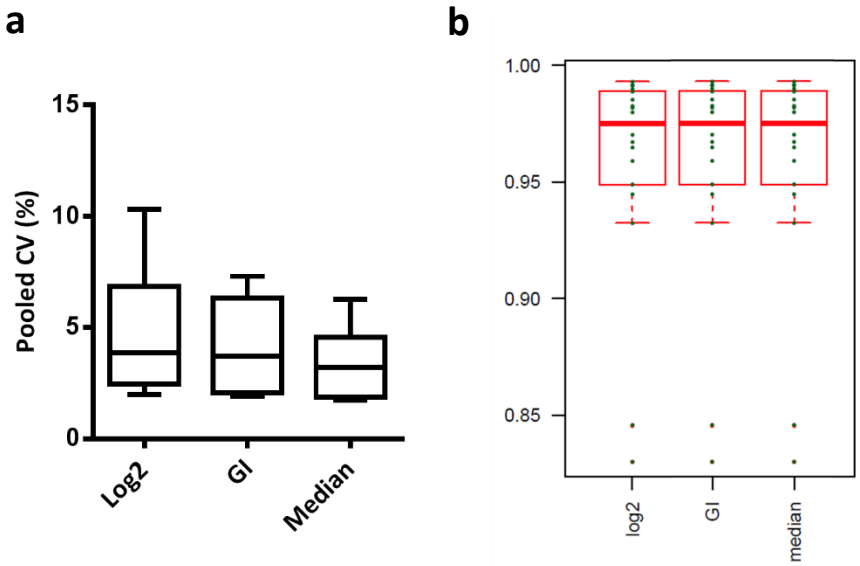

Figure 4.10 Global evaluation of normalisation techniques. **a**, Box and whisker plot to show the distribution of Pooled CV across all samples upon Log_2, transformation and GI or Median normalisation. Whiskers represent 5^{th} and 95^{th} percentiles. **b**, Box and whisker plot to show distribution of mean Pearson correlation within samples for Log_2, GI and Median normalisation. Whiskers represent 5^{th} and 95^{th} percentiles.

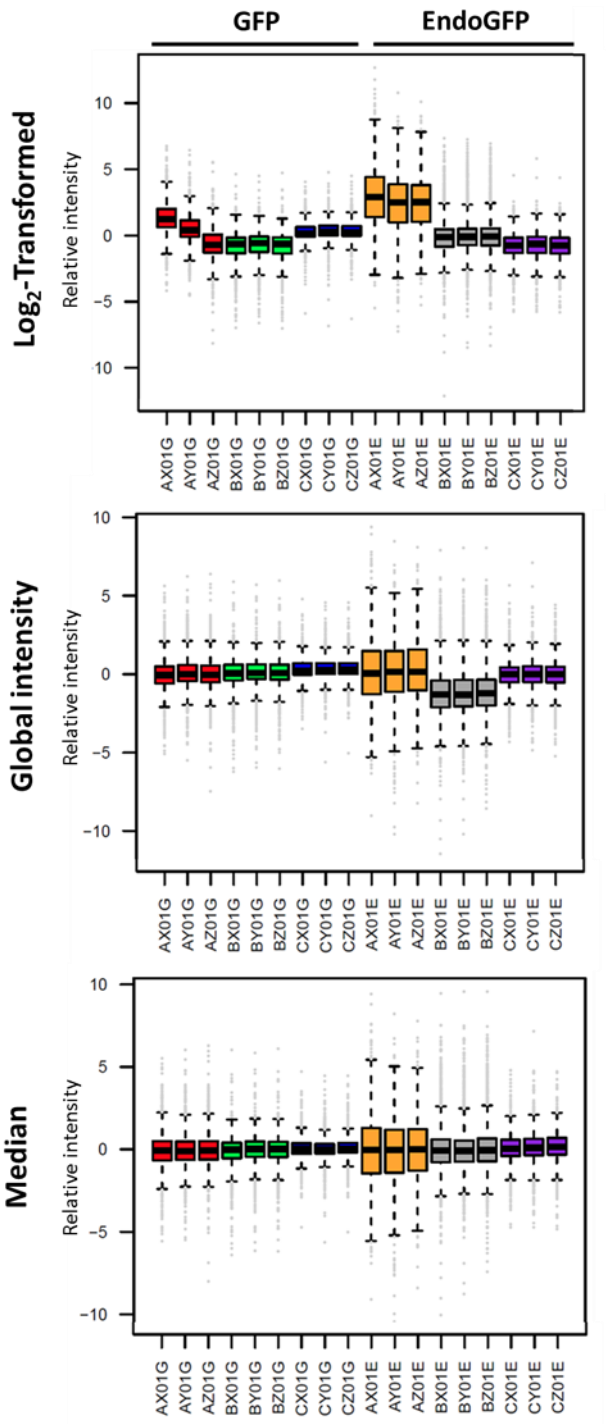

Figure 4.11 Relative log expression (RLE) plots to evaluate normalisation. **a**, Box and whisker plot to show the distribution of Pooled CV across all samples upon Log_2 transformation (top), GI (middle) or Median (bottom) normalisation. Whiskers represent 5^{th} and 95^{th} percentiles.

4.4.4. Differential expression analysis

Normalised protein intensities were averaged across technical replicates. To assess if proteins were differentially expressed in either condition two-sided, unpaired Student's *t*-tests were performed to compare Top3 protein intensity of biological replicates. The calculated *p*-values were corrected for multiple testing using a permutation-based estimation of dataset FDR to obtain q-values. No proteins displayed a q-value below the significance threshold of 0.05 and thus no proteins were identified as differentially expressed with high confidence.

This appears to be primarily a result of high variability between EndoGFP samples despite normalisation. Illustrating this, a principle component analysis performed on the normalised data indicates that, whilst technical replicates from the same sample cluster well together, EndoGFP samples are poorly clustered (Figure 4.12a). Indeed, EndoGFP biological replicate 3 appears to cluster with the GFP samples at the bottom left of the plot. Similarly, boxplots illustrating the distribution of inter-sample within-condition protein CV show two-fold higher variability in EndoGFP samples (median CV of 10.4% in EndoGFP samples versus 5.2% in GFP samples) (Figure 4.12b).

In an attempt to extract tentative future directions of enquiry from the data, the results of *t*-tests were plotted without correction for multiple testing, and the results were interpreted with the caveat of accepting a higher rate of type I errors (Figure 4.13). Proteins were denoted as differentially expressed upon fulfilment of two criteria: *p*-value < 0.05 and a fold change (ratio of average protein intensity) greater than 1.3-fold in either direction. Quantitative measurements of proteins using MS^E have been shown to display an error in signal intensity of 10-15% (Silva et al., 2006; Patel et al., 2009). Thus, a fold change cut-off of 1.3 represents a value 2-3 times larger than that expected given normal technical variation. Differentially expressed proteins are listed in Table 4.3.

Endophilin-A2-eGFP was overexpressed in RPE1 and then enriched through IP. As expected, the protein is highly enriched in the EndoGFP condition. Cadherin-13 was also enriched in EndoGFP immunoprecipitates and is an interesting candidate for future analysis due to the observation of high FEME activity at cell-cell junctions in confluent cells (*E. Boucrot*, unpublished results).

We assessed the behaviour of candidates endophilin partners identified in previous analyses. Melanoma inhibitory activity protein 2, also known as CTAGE5 is a cargo sorting receptor involved in ER-golgi transport (Saito *et al.*, 2011; Santos *et al.*, 2016; Tanabe *et al.*, 2016). It is also homologous to the CTAGE6 protein identified as a potential endophilin partner as reported in Section 4.2, sharing a long proline-rich domain at their c-terminus which could serve as a binding site for the endophilin SH3 domain. The candidate endophilin partner WWP1 was also identified in this analysis but was not differentially expressed.

Of the known endophilin partners identified, Dynamin-1, Synaptojanin 1, Synaptojanin 2 and vascular endothelial growth factor receptor 2 (VEGFR2) were not differentially expressed, but vascular endothelial growth factor receptor 1 (VEGFR1) was unexpectedly enriched in the GFP bound fractions. Interestingly, clathrin light chain (CLC) but not clathrin heavy chain (CHC) proteins were also found enriched in the EndoGFP bound fractions.

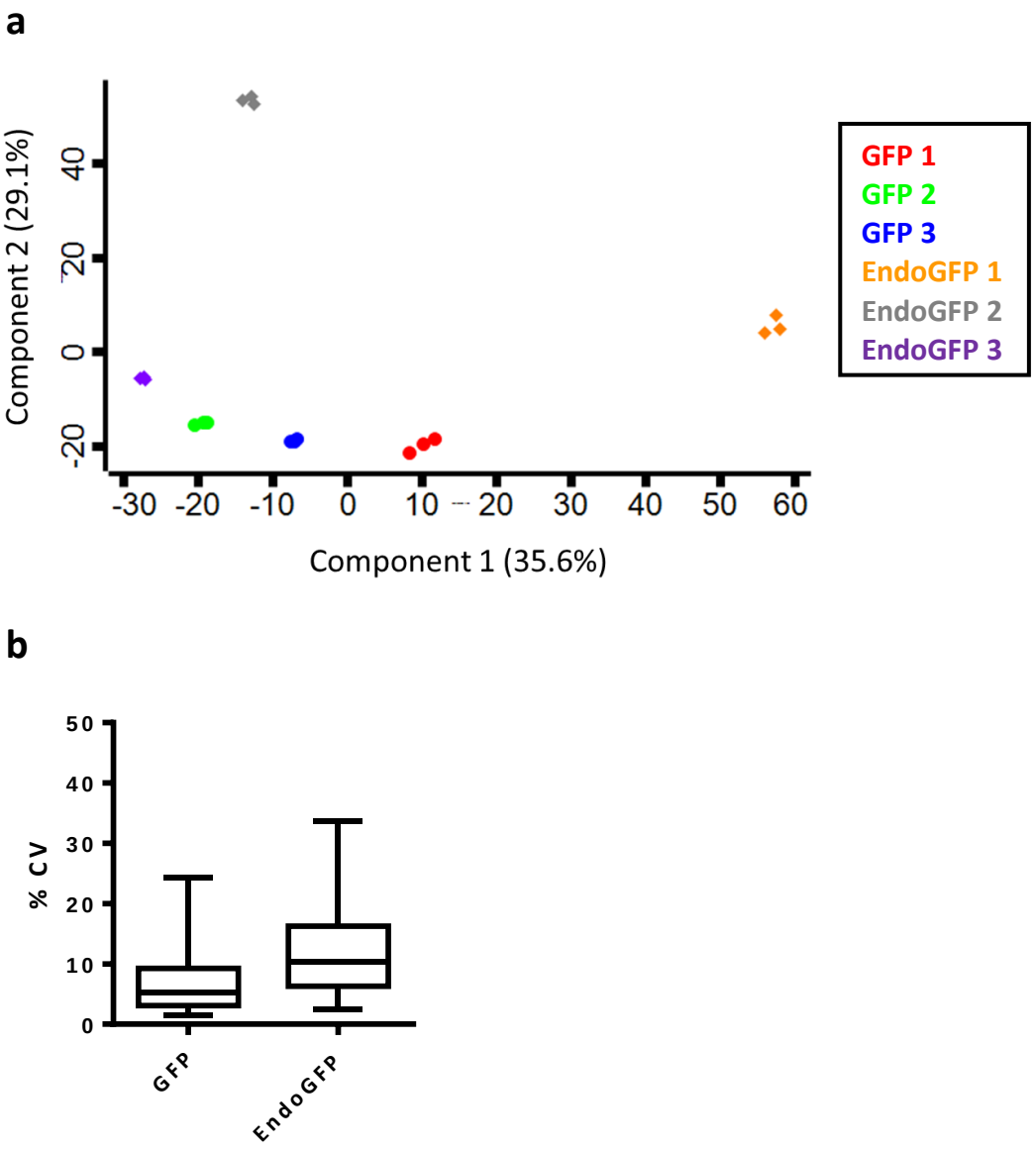

Figure 4.12 EndoGFP samples display high variability post-normalisation. **a**, PCA plot to represent sample relatedness in 2 dimensions. Plots colours indicate biological sample of origin as indicated. **b**, Boxplots to show distribution of inter-sample within-condition protein CV. Protein Top3 intensity was averaged across technical replicates for each sample prior to calculation of CV.

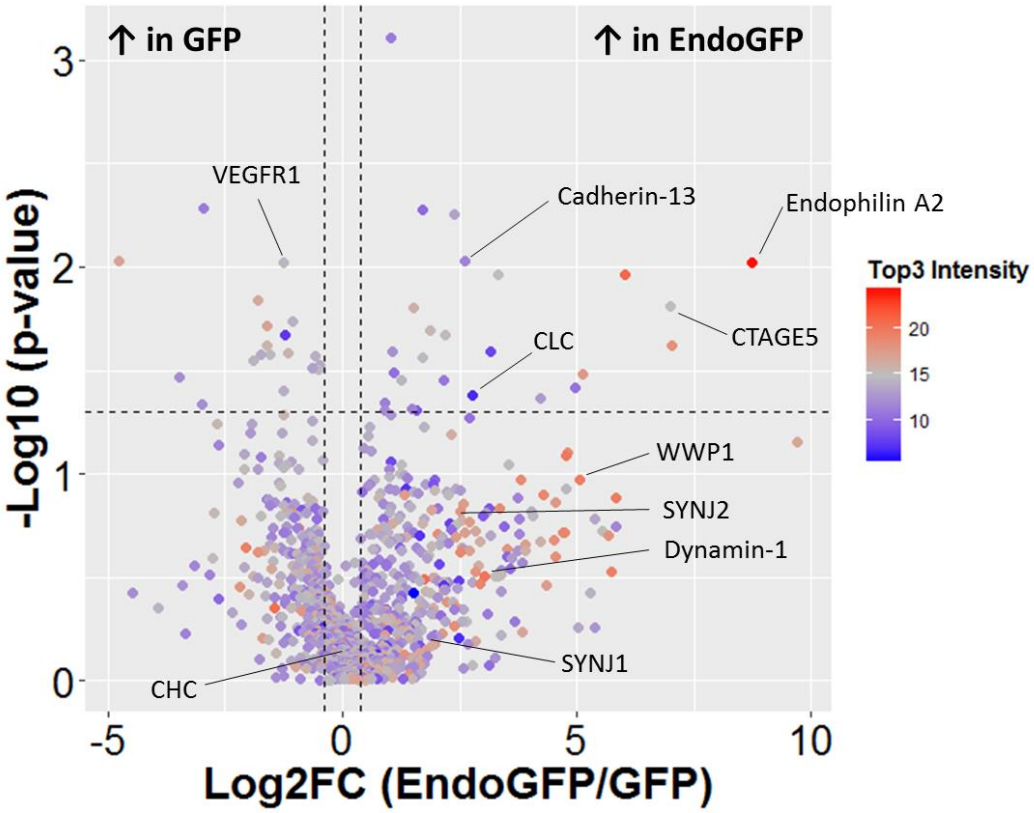

Figure 4.13 Volcano plot to illustrate differentially expressed proteins between GFP and EndoGFP immunoprecipitates. Fold change (FC) was calculated by finding the ratio of average Top3 protein intensity in each condition. Plot colours are representative of Top3 protein intensity for the highest expressed protein in each comparison. Vertical dashed lines represent fold change cut-offs for significance (± 0.369). Horizontal dashed line represents $-\text{Log}_{10}(p\text{-value})$ cutoff for significance as calculated using a 2-sided, unpaired t-test ($\alpha < 0.05$). Proteins with data plotted in the upper left and right quadrants are denoted as differentially expressed.

Protein name	Gene name	Fold Change (Stimulated/Resting)	Benjamini Hochberg q-value
Pre-mRNA-splicing factor SPF27	BCAS2	2.019	0.038
Endoplasmic reticulum export factor CTAGE5	CTAGE5	1.903	0.015
Clathrin light chain A	CLTA	1.812	0.042
Superkiller viralicidic activity 2-like 2	SKIV2L2	1.713	0.026
Nucleolar complex protein 3 homolog	NOC3L	1.567	0.024
Endophilin-A2	SH3GL1	1.520	0.010
RNA-binding protein 33	RBM33	1.485	0.043
60 kDa heat shock protein_ mitochondrial	HSPD1	1.403	0.011
Pre-rRNA-processing protein TSR1 homolog	TSR1	1.379	0.033
Cadherin-13	CDH13	1.330	0.009
Eukaryotic translation initiation factor 4B	EIF4B	0.763	0.046
Threonylcarbamoyladenine tRNA methylthiotransferase	CDKAL1	0.722	0.009
60S acidic ribomal protein P2	RPLP2	0.715	0.005
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	0.668	0.034

Table 4.3 List of proteins differentially expressed between GFP and EndoGFP conditions. Proteins enriched in GFP condition are shaded in grey.

4.5. Discussion

Work presented in this chapter addressed the issue of FEME regulation. Two pieces of evidence suggested that FEME competence and rate are regulated. Firstly, FEME rate as measured by the number of EPAs budding from the leading edge of growing cells varies between cell types. Secondly, the proportion of growing cells in culture which display any FEME at all is never 100%. Work in our lab established that increasing the concentration of serum in cell media increased FEME rate, whilst removal of serum reduced FEME. We speculated that growth factor signalling plays a role in regulating FEME beyond supplying ligands for activation of cargo receptors.

Growth factor signalling could regulate FEME at any of the stages of an endocytic event. The current model of an FEME endocytic event proceeds via the following sequence: 1) GTP-loaded Cdc42 at the leading edge of a cell recruits lamellipodin and endophilin via the F-BAR proteins CIP4 and FBP17 to form endophilin clusters at the leading edge ready to internalise activated cargo receptors, 2) activated receptors are sorted into nascent FEME vesicles by endophilin or endophilin-binding adaptor proteins, 3) oligomerisation of endophilin bends the membrane to form a tubulovesicular invagination, 4) Endophilin recruits the GTPase dynamin to constrict the vesicle neck and separate the vesicle from the plasma membrane (scission), 5) Endophilin-coated FEME carriers are transported into the cell interior through the action of Dynein. As illustrated, endophilin is involved in all stages of FEME after its initial recruitment by lamellipodin and is therefore a strong candidate for regulation downstream of growth factor signalling.

AP-MS experiments were performed to assess the interactome of endophilin following changes in growth factor levels and thus FEME rate in RPE1 cells. This analysis identified 27 kinases with potential involvement in FEME. Experiments conducted in our lab identified GSK3α/β as negative regulators of FEME. Interestingly, GSK3β was enriched in endophilin immunoprecipitates from Resting rather than Stimulated cells, raising the possibility that GSK3β binds to endophilin when growth factor signalling is low. Pulldowns performed to assess this showed that GSK3β from serum starved extracts but not from Resting or Stimulated cell extracts binds to the endophilin

A2 SH3 domain. Interestingly, GSK3α did not bind to the endophilin A2 SH3 domain under any of the conditions tested.

The recruitment of GSK3β but not GSK3α by endophilin SH3 may be explained by the absence of a suitable target PRM in GSK3α. A review of the primary sequence of both proteins reveals a single probable EndoSH3 target PRM (amino acids 307-311) in GSK3β which is not present in GSK3α.

The detection of GSK3β in complex with endogenous endophilin in Resting cells by AP-MS conflicts with the failure to detect EndoA2SH3-GSK3β binding by pulldown from Resting cells extracts. It is possible that GSK3β makes additional contacts with endophilin outside of the SH3 domain, the absence of which lowered binding affinity in the pulldown experiment. Furthermore, the higher sensitivity and dynamic range of LC-MS analysis may have contributed to identification of GSK3β binding in resting cells. Whatever the true nature of this interaction, both experiments agree that EndophilinA2-GSK3β binding is not favoured upon growth factor stimulation.

Growth factor signalling has known effects on GSK3β activity. When phosphorylated, serine 9 of GSK3β can occupy the protein's own substrate docking site and compete with true substrates, inhibiting kinase activity (Cross *et al.*, 1995; Frame, Cohen and Biondi, 2001). Several protein kinases downstream of growth factor signalling pathways can phosphorylate GSK3β at S9 including Protein Kinase B (as known as Akt) downstream of Insulin growth factor 1 (IGF-1) signalling and p90 S6 kinase (S6K) downstream of Epidermal growth factor (EGF) signalling (Cross *et al.*, 1995; Eldar-Finkelman *et al.*, 1995). It is possible that the loss of Endophilin-GSK3β binding upon growth factor stimulation is related to this modification. This would be true if GSK3β targets phosphoserine/ phosphothreonine sites on the endophilin SH3 domain for docking. Whilst there are several serine/threonine residues in the SH3 domain, they do not match the canonical GSK3 consensus (S/T-X-X-X-S/T). To shed light on this, an assessment of S9 phosphorylation upon serum modulation could be easily performed by western blotting, as an anti-pS9-GSK3β antibody is commercially available.

Endophilin A2 may be a direct target for GSK3β outside the SH3 domain, possessing at least 4 GSK3 consensus sequences at other sites in the protein. An attempt was made to detect and quantify changes in phosphopeptides in both AP-MS experiments reported here, but the phosphopeptides detected were of poor confidence. This was expected as a relatively small amount of protein was analysed and no effort to enrich the sample for phosphopeptides was made. There is no guarantee that a majority of the endophilin captured by IP would be modified by GSK3β. Indeed, many cells do not exhibit FEME in cell extracts from which IPs were conducted (no FEME is observed in ~50% of cells in Resting and ~30% of cells in +10% FBS conditions). The proportion of captured endophilin monomers which bear phosphorylations relevant to FEME is therefore expected to be small despite enrichment by IP. Further technical concerns related to MS analysis of phosphopeptides include ion suppression of phosphorylated peptides when analysis is conducted in positive mode and the tendency of collision-induced dissociation to remove phosphoric acid during fragmentation. To assess changes in the phosphorylation of endophilin upon serum withdrawal or stimulation, a TiO_2 phosphopeptide enrichment step could be added to the workflow following tryptic digest. This step is widely used to exclude non-phoshorylated peptides from a sample and thus improve workflow sensitivity for phosphopeptides (Thingholm *et al.*, 2006). The use of larger amounts of starting material (>1mg) is essential to ensure that adequate amounts of phosphopeptide will be present following enrichment for detection by MS, given their low expected prevalence in the sample.

Even if endophilin itself is not modified by GSK3β, its recruitment to the plasma membrane by endophilin would assist FEME regulation. Given GSK3's large variety of potential targets, any pathway in which it is involved must address the issue of specificity. An Endophilin-GSK3β interaction would solve this by helping to localise GSK3β to the leading edge and would allow for modification of other FEME proteins and cargoes to inhibit the pathway. For example, Dynamin-1 is a known target of GSK3β at position S774 following CDK5 phosphorylation of S778 (Tan *et al.*, 2003). The impact of GS3Kβ inhibition on Dynamin-1 recruitment by endophilin could be assessed by pulldowns and by microscopy (± GS3Kβ inhibitors).

The discovery that GSK3 and CDK5 act as negative inhibitors of FEME prompted investigation of their specific impact on the endophilin interactome. This was assessed by IP of a transiently expressed eGFP-EndophilinA2 construct from RPE1 lysates following treatment with GSK3 and CDK5 inhibitors. An interesting candidate from this study is CTAGE5, which interacts with and shares a high degree of sequence similarity with CTAGE6 (identified in endogenous IPs from Stimulated cell extracts) (Huttlin *et al.*, 2017). These proteins are known to play roles in aiding the assembly of large cargoes such as collagen VII and very low density lipoproteins (VLDLs) into carriers for onward trafficking from the ER (Saito *et al.*, 2011; Santos *et al.*, 2016). It is therefore unclear what role they would play in FEME. Both proteins feature long C-terminal PRMs containing several suitable EndoSH3 target sites. A first step in investigating their role in FEME would be to assess if an EndoSH3-CTAGE5/6 interaction is detectable by GST-pulldown. If successful, more detailed profiling of protein localisation relative EPAs could be performed by microscopy (commercial antibodies against both proteins are available).

Neither AP-MS experiment identified any high confidence regulators of FEME. Whilst technical issues were minimal for the AP-MS study of endogenous endophilin, the experiment was critically underpowered as only 2 biological repeats were performed, and serum starvation was not included as a third condition. This was largely due to time constraints in MS instrument availability. In hindsight it was overenthusiastic to divide samples from IP (i.e. reduced complexity) into 10 bands during in-gel digestion and to analyse these samples by 60 min gradients. It is likely that similar digestion efficacy and instrument performance could have been achieved from excision in single gel bands following a short PAGE separation, with a compensatory longer gradient. This would have enabled analysis of a larger number of biological samples to achieve increased power and more confidence in candidates.

The technical issues facing AP-MS of EndoGFP complexes were extensive, including poor digestion and chromatography. The On-bead-SCX workflow employed could potentially be improved by incubating digests at 37°C overnight instead of at room temperature. The largest

contributor to poor sample-to-sample reproducibility is likely to have been the lack of a normalisation step following transient overexpression of GFP constructs before IP. Normalisation could have been performed by western blotting, making roughly equivalent amounts of GFP constructs available for capture in all samples. Ultimately, the reliance on transient overexpression was a suboptimal strategy that was adopted due to the unavailability of anti-endophilin A2-coated magnetic beads. A future repeat of this experiment would be best attempted by targeting endogenous endophilin and should also feature IP of endophilin from resting and serum-starved extracts.

The description of CDK5 and GSK3β as negative regulators of FEME has a number of general implications for the control of FEME competence and rate in individual cells. Furthermore, it is not clear if CDK5 is the only priming kinase for GSK3β with regards to control of FEME. Control of FEME by kinases could also target the N-BAR protein Bridging Integrator 1 (BIN1), which was found to colocalise strongly with endophilin on FEME carriers (Chan Wah Hak *et al.*, 2018). Future AP-MS experiments focussing on the BIN1 interactome and phosphorylation state hold great promise for further elucidation of FEME pathway regulation.

5. Quantitative LC-MS profiling of spinal cord samples in a mouse model of motor neuron degeneration

5.1. Introduction

Neurodegenerative diseases feature the progressive dysfunction and eventual death of neuronal cells resulting in significant suffering. Worldwide, these diseases represent a major cause of premature death. Many neurodegenerative disorders such as Huntington's disease, Alzheimer's disease and Parkinson's disease feature deposits of aggregated protein called inclusions in affected neurons (Soto, 2003). While the toxicity of these inclusions is debated, the dysfunction of protein folding pathways is known to play a role in disease progression.

Mutations in the locus encoding DnaJ Heat Shock Protein Family Member B2 (DNAJB2), also known as Heat Shock Protein, Neuronal DNAJ-like 1 (hereafter HSJ1) have been causatively linked to distal hereditary motor neuropathy (dHMN) and Charcot-Marie-Tooth syndrome type 2 (CMT2) in humans (Gess *et al.*, 2014; Tazir *et al.*, 2014). Both conditions feature the progressive dysfunction of motor neurons, with concomitant loss of sensory neurons in CMT2 only. HSJ1 is a member of the Hsp40 family of molecular co-chaperones. These proteins regulate the activity and specificity of the chaperone Hsp70 by binding to hydrophobic regions of misfolded proteins and recruiting them to Hsp70. All members of the family feature a highly conserved J-domain, a ~70-amino acid region which is required for stimulation of Hsp70 ATPase activity. Two HSJ1 isoforms are preferentially expressed in neurons (Cheetham, Brion and Anderton, 1992). The smaller HSJ1a (36kDa) lacks a C-terminal extension found in HSJ1b (42 kDa) which features a prenylation motif. Isoprenylation of this motif facilitates targeting of HSJ1b to the cytosolic face of the ER whilst HSJ1a displays cytoplasmic and nuclear localisation (Chapple and Cheetham, 2003).

HSJ1 is unique among co-chaperones in that it possesses multiple ubiquitin-interacting motifs in addition to its client protein binding domain (Chapple *et al.*, 2004; Westhoff *et al.*, 2005), and thus participates in the ubiquitination and proteasomal targeting of a number of its client proteins (Westhoff *et al.*, 2005). Previous work has shown that HSJ1 activity promotes disaggregation of Huntingtin (Htt) aggregates (Westhoff *et al.*, 2005; Borrell-Pagès *et al.*, 2006; Howarth *et al.*, 2007; Labbadia *et al.*, 2012) and can reduce aggregation of mutant Superoxide dismutase 1 (SOD1) in cell and mouse models of Huntington's Disease and Amyotrophic lateral Sclerosis respectively (Novoselov *et al.*, 2013).

This work aims to gain an understanding of the mechanisms through which functional HSJ1 protects motor neurons. Specifically, we aim to define potential client proteins of HSJ1 which contribute to cell maintenance and describe which cellular pathways are dysregulated in the absence of HSJ1.

5.1.1. Experimental model and design

The Cheetham lab generated a HSJ1$^{-/-}$ knockout (KO) mouse for use in the elucidation of HSJ1's functional roles in motor neuropathies. The mice did not exhibit any congenital disabilities, lived average lifespans and did not manifest any behavioural signs of motor neuropathy sooner than control mice. Comparison of motor neuron number in lumbar spinal cords of HSJ1 wild-type (WT) and KO mice revealed a consistent 11.6% decrease in motor neuron counts by postnatal day 20 (P20), starting after P15 (Heather Smith, unpublished work). At P120, HSJ1 KO mice showed a ~28% reduction in lumbar spinal cord motor neuron counts compared to WT mice. Further analysis revealed that heightened levels of phosphorylated PRKR-like endoplasmic reticulum kinase (Perk) and phosphorylated Eukaryotic translation initiation factor subunit α (eIF2α) were detectable by P10 in the spinal cord of HSJ1 KO animals, indicating activation of the unfolded protein response (UPR) (Figure 5.1a). Other indicators of a defect in protein folding included heightened levels of ubiquitinated proteins and autophagy marker proteins such as LC3-II in lumbar spinal cord extracts from HSJ1 KO mice relative to HSJ1 WT mice. (Cheetham Lab, unpublished work) (Figure 5.1a).

Two mechanisms by which loss of HSJ1 function can lead to motor neuron death are proposed: Loss of HSJ1 may lead to death of motor neurons through a failure in the biogenesis or quality control of a specific client protein or, loss of HSJ1 might lead to a failed motor neuron stress response and thus result in aberrant expression of essential proteins. These mechanisms are not mutually exclusive. To address these hypotheses, label-free quantitative mass spectrometry was performed on lumbar spinal cord extracts from HSJ1 WT and HSJ1 KO mice (Figure 5.1b). We hypothesised that HSJ1 client proteins would misfold in the KO condition and either aggregate or be degraded and therefore lost from the soluble fraction upon protein extraction. These phenomena would be detectable as a protein enrichment in the WT sample. HSJ1 has been shown to play a role in targeting misfolded plasma membrane components to the proteasome (Chapple and Cheetham, 2003; Westhoff *et al.*, 2005; Borrell-Pagès *et al.*, 2006) Thus, loss of HSJ1 could also result in an accumulation of misfolded transmembrane proteins in the plasma membrane of KO samples (Figure 5.1b).

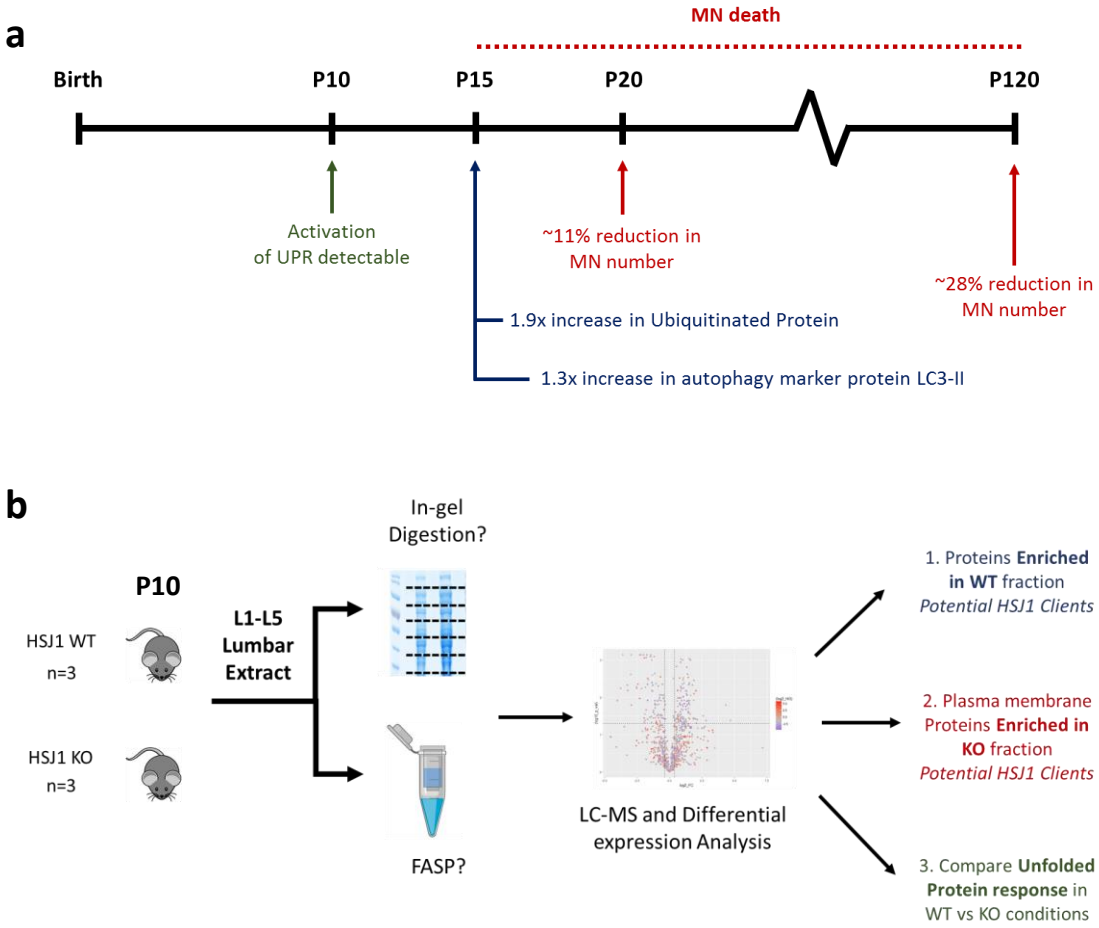

Figure 5.1 Lumbar spinal cord phenotype of the HSJ1 knockout mouse and experimental design. **a**, Timeline to illustrate postnatal events occurring in HSJ1 KO lumbar spinal cord relative to wild-type animals. At postnatal day 10 (P10) markers of the unfolded protein response (UPR) are detectable. At P15, increased levels of ubiquitinated proteins and autophagy markers are present, but motor neuron counts are not significantly different from WT mice. From P20 onwards, a significant reduction in motor neuron number is observed in lumbar spinal cords from HSJ1 KO mice relative to WT mice. **b**, Schematic to show proposed experimental design to compare P10 Lumbar spinal cord extracts from HSJ1 WT and HSJ1 KO mice. Following Lumbar spinal cord excision and protein extraction, sample preparation will be performed by in-gel digestion or filter-aided sample preparation (FASP). Peptides will then be analysed by quantitative LC-MS and subjected to differential expression analysis. Three classes of interesting proteins could be observed in this analysis as described. MN – motor neuron.

5.2. Optimisation of mass spectrometry sample preparation and LC-MS acquisition for whole proteome samples

Our experimental strategy represents an effort to characterise quantitative changes in the whole proteome of tissue extracts from mouse lumbar spinal cord. The greatest barrier to obtaining good quantitative measurements in such a format is the limited dynamic range of the mass spectrometer relative to the proteome. Samples must, therefore, be fractionated to reduce complexity and the instance of highly abundant peptides masking the signal from peptides of lower abundance. We tested two methods of sample fractionation (Figure 5.2).

Conventional “in-gel” digestion following SDS-PAGE-mediated protein-level fractionation has been effective in generating maps of whole proteomes (M. S. Kim *et al.*, 2014; Wilhelm *et al.*, 2014). However, the technique is time-consuming (both in labour and instrument time) and suffers from issues with variable efficiency of peptide recovery from the gel matrix (Speicher *et al.*, 2000). Additionally, the manual excision of gel bands is inevitably prone to experimental error, a major concern in quantitative workflows.

Filter-aided sample preparation (FASP) is an on-filter digestion technique which has been successfully utilised to achieve deep proteome coverage when combined with secondary peptide level fractionation (Wiśniewski *et al.*, 2009, 2012). Samples can be prepared in SDS, ensuring near-complete solubilisation and denaturing with subsequent desalting and detergent removal through centrifugation and buffer exchange in a spin-filter prior to on-column tryptic digest (Wiśniewski *et al.*, 2009).

In-gel digestion and FASP coupled to strong cation exchange (FASP-SCX) were tested on extracts of RPE1 cells grown in single 100 mm dishes (95% confluence). Cells were lysed in 8M Urea Lysis buffer (8M Urea in 0.1 M Tris pH 7.8), sonicated (3 x 10-second pulses, 10 μm) and centrifuged to remove insoluble particles.

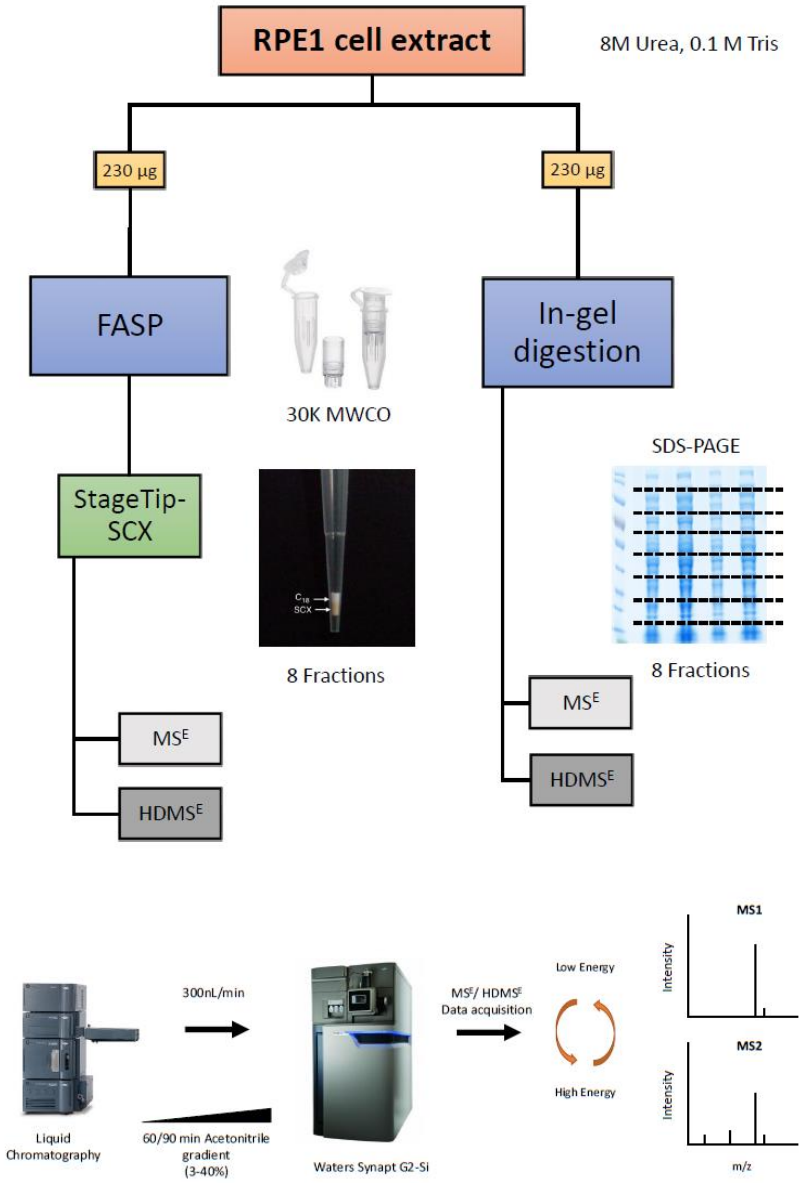

Figure 5.2 Schematic representation of whole cell sample preparation trial. 230 µg of RPE1 cell extract was digested via FASP (left) or in-gel digestion (right.) SDS-PAGE gel shown is illustrative. Samples were split into an equal number of fractions at the protein (In-gel) or peptide (FASP) level. Samples were analysed using 2 different Instrument acquisition modes to compare workflow sensitivity.

RPE1 cells were lysed in Tris-Urea buffer and samples were quantified by UV absorbance. 230 µg of protein was used for each sample preparation workflow. For in-gel digestion, 230 ug aliquots of RPE1 extract were separated on a 4-12% Bis-Tris polyacrylamide gel and lanes were cut into 8 fractions. Proteins in gel pieces were reduced and alkylated before digestion overnight at 37°C with 10 ng/µl trypsin (Promega) in 50 mM ammonium bicarbonate. Peptides were extracted from gel pieces and dried in a vacuum concentrator prior to resuspension in LC-MS running buffer for analysis.

For FASP-SCX, 230ug aliquots of RPE1 extract were concentrated in Vivaspin 500 (30kDa MWCO) filter units, reduced using 50mM DTT and alkylated with 50mM Iodoacetamide then digested overnight at 37°C with trypsin (Promega) in a 1:100 enzyme/protein ratio. Peptides were collected by centrifugation and dried in a vacuum concentrator. SCX was performed using C18-SCX StageTips, eluting the sample into eight fractions. Fractions were collected by serial elution in buffers containing an increasing concentration of ammonium acetate (0, 20, 50, 80, 150, 250, 500 and 1000 mM) and dried in a vacuum concentrator.

Peptides were analysed on a Waters Synapt G2 Si QTOF instrument operating in Resolution mode via a NanoESI source. Data were acquired in MS^E or high-definition MS^E ($HDMS^E$) mode with the following settings: Low energy scan (4 eV), High energy scan (15 – 35 eV in MS^E mode or 25 – 55 eV in $HDMS^E$ mode). Measurements were made over an m/z range of 50 – 2000 Da with a scan time of 0.6s. Raw data were processed using PLGS v3.0.2. Data were queried against a Homo sapiens protein FASTA database concatenated with a list of contaminant FASTA sequences (The Global Proteome Machine, 2016). Protein identifications had to fulfil the following search criteria: a maximum of 2 missed cleavages, a minimum of 3 fragments required per peptide identification and a minimum of 7 fragment ions required per protein. A protein level false discovery rate of 1% was implemented using a decoy database.

In terms of protein identifications, in-gel digestion with $HDMS^E$ acquisition performed returned the highest number of hits, with 1685 protein groups identified. In-gel digestion with MS^E

acquisition returned the second-most protein identifications with 1269 identifications. FASP-SCX with HDMS^E and MS^E acquisition identified the fewest protein groups (674 and 572 respectively). On average, 19% more protein groups were identified per fraction when using HDMS^E acquisition (Figures 5.3 a and b).

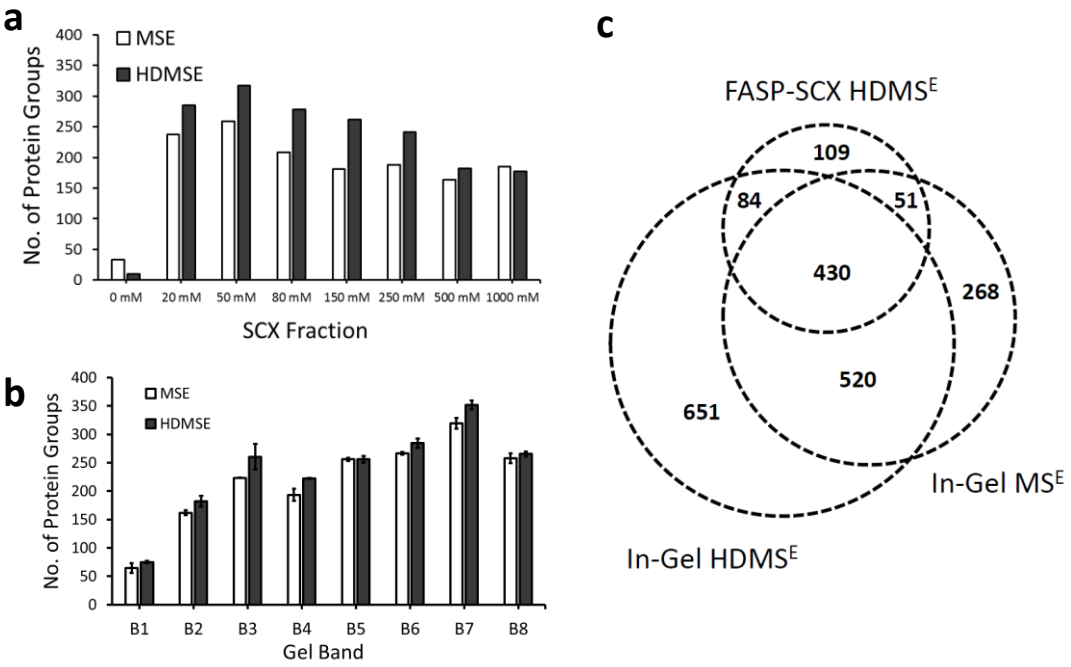

Figure 5.3 Comparison of sample preparation workflow and LC-MS acquisition mode performance. **a**, Barchart to show no. of protein groups identified per fraction using MS^E or HDMS^E using identical samples and injection volumes. **b**, Barchart to show no. of protein groups identified per gel band using MS^E or HDMS^E acquisition on identical samples (means ± s.e.m of 2 technical replicates). **c**, Venn diagram to illustrate overlap in unique protein IDs between datasets with the highest performance.

The high performance of in-gel digestion workflows could be attributed to better fractionation of the sample, with an average of >10-fold more unique protein groups identified per fraction (Figure 5.4, in-gel HDMS^E vs FASP-SCX HDMS^E). Similarly, in-gel workflows exhibited an average three-fold increase in percentage of unique peptides per fraction. Based on these results, tissue digestion of murine lumbar spinal cords was performed using in-gel digestion.

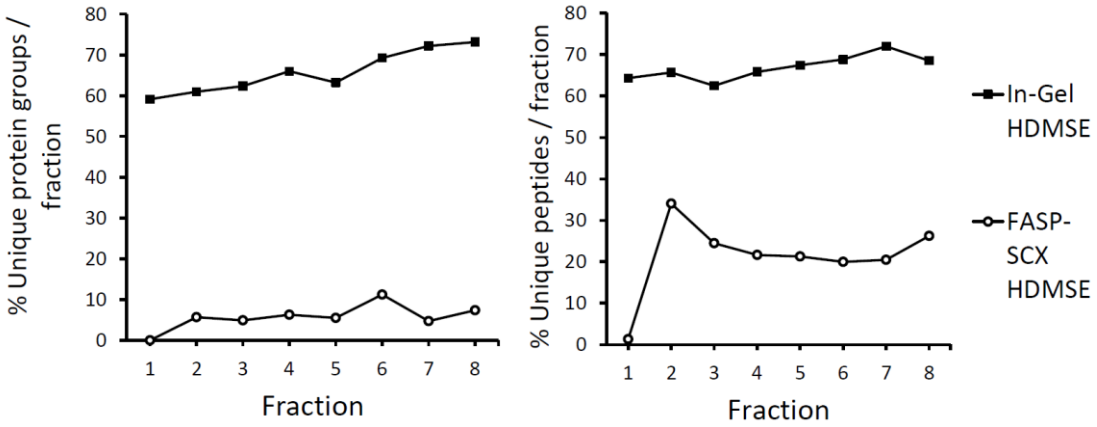

Figure 5.4 In-gel workflow exhibits superior fractionation of RPE1 cell lysate. Left, Line chart to show frequency of unique protein groups/fraction. Right, Line chart to show frequency of unique peptides/fraction.

5.3. Comparative analysis of mouse lumbar spinal cord extracts from HSJ1 wild-type and knockout mice

5.3.1.Preparation of $Hsj1^{+/+}$ and $Hsj1^{-/-}$ lumbar spinal cord samples

HSJ1 knockout mice (*$Hsj1^{-/-}$*) were generated by genOway corporation (Lyon, France) and subsequently bred and maintained at the UCL Institute of Neurology Biological Services. Dr D. Parfitt (Cheetham Lab, UCL) prepared lumbar spinal cord extracts and performed in-gel digestion. Sections of lumbar spinal cord, spanning L1-L5 vertebrae, weighing between 20–25 mg were prepared from three *$Hsj1^{+/+}$* (WT) and three *$Hsj1^{-/-}$* (KO) mice ten days after birth (P10), coincident with a detectable unfolded protein response (UPR), but before neuronal death, and lysed in a Tris-SDS lysis buffer. Samples were boiled and centrifuged, and the cleared lysate was mixed with SDS-PAGE sample buffer. Proteins in each sample were separated on a 1D SDS-PAGE gel and prepared for LC-MS analysis by in-gel digestion (6 bands) followed by C18-StageTip clean-up. Peptides were resuspended in LC-MS running buffer and were analysed in technical triplicate by nanoLC-ESI-HDMS^E. Peptide and protein identification were performed using PLGS (Waters Corp.) and identified protein groups were quantified using the Hi3 method to obtain absolute protein abundance measurements in femtomoles (fmols). The average protein abundance from all technical repeats for a given protein from a given fraction was summed across all fractions to generate a single protein abundance value for each protein in each "reconstituted" sample.

5.3.2. Data processing and quality control

A large multi-fraction LC-MS experiment using relatively long gradients such as the one reported here is subject to many sources of technical error. These include issues with sample preparation including quality of protease digestion or losses/inconsistencies owing to sample handling. Inconsistencies in LC-MS instrument performance such as variable chromatography and mass calibration can also cause data analysis problems. Label-free LC-MS experiments rely on

accurate measurements of signal intensity and are especially prone to problems with instrument performance, as samples must be analysed separately.

Data was examined to determine if it was of sufficient quality to proceed with statistical testing and draw meaningful biological inferences. An examination of total protein and peptide identifications per sample was undertaken. Samples from WT mice displayed 2160 ± 18% (mean ± CV) total protein groups and samples from KO mice displayed 2012 ± 52% (mean ± CV) total protein groups. A similar pattern was observed when comparing peptide identifications, with 115928 ± 11% (mean ± CV) peptides in WT samples and a 99995 ± 56% (mean ± CV) peptides observed in KO samples (before combining intensities across fractions). The large spread in identification numbers for KO samples is primarily due to low protein and peptide identification rates in the KO sample from biological repeat B (BKO) and relatively high numbers of protein and peptide identifications for the KO sample from biological repeat C (CKO) (Figures 5.5a and 5.5b). Several parameters were assessed to determine if this spread in peptide and protein identifications reflected genuine biological variation or technical issues.

A fundamental parameter in the performance of an LC-MS workflow is instrument calibration. The SYNAPT-G2Si used in this study was calibrated using a lockspray infusing a peptide of known mass ([Glu-1] Fibrinopeptide B; doubly charged m/z of 785.8426) which was sampled every 30 seconds. During database searching, mass spectra are calibrated via reference to the sampled mass of [Glu-1] Fibrinopeptide B fragments and a calibration file created prior to starting all runs to correct for drift in instrument performance. Over the course of an experiment, changes in ambient temperature and build-up of residue from analytes on the internal components of the mass spectrometer can cause the calibration file to be unrepresentative, compromising mass accuracy. As illustrated in Figure 5.5d, the distribution of MS1 peptide mass errors for all samples, including BKO samples, lies almost entirely between ± 10 ppm. The distributions are overlapping with BKO samples displaying a tighter distribution with a slight offset: the average mass error for all samples and the subset of BKO samples is similar at 0.13 ppm and 0.21 ppm, respectively. Overall, this indicates good mass calibration during all runs, including those for BKO samples.

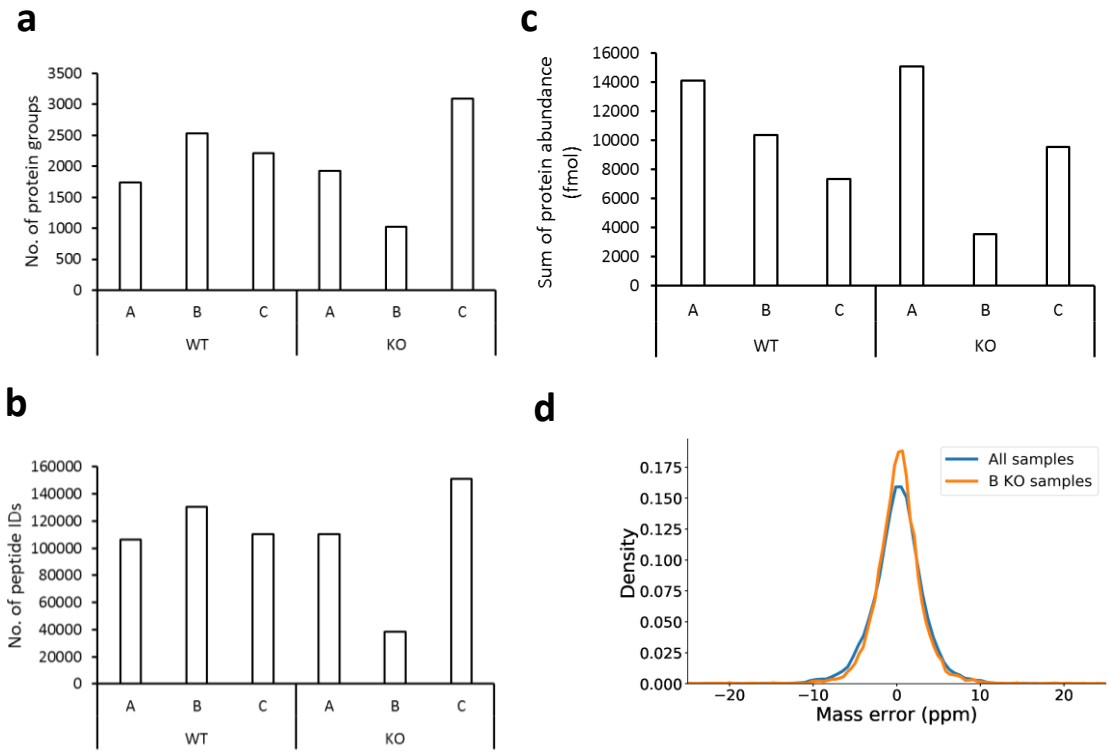

Figure 5.5 Evaluation of instrument performance. **a**, Bar chart to show the total number of protein groups identified in each sample (sum of all fractions). **b**, Bar chart to show the total number of peptides identified in each sample (sum of all fractions). **c**, Bar chart to show summed protein abundance (as quantified with reference to the Hi3 standard) for all protein groups across all fractions for each sample, without normalisation. **d**, Density plot to show peptide mass accuracy distribution. The distribution for all samples (blue) and for B KO samples (orange) only is shown.

The disparity in protein and peptide identifications described above could also be due to issues with liquid chromatography performance. Visual inspection of the base peak intensity (BPI) and total ion current (TIC) chromatograms for all runs did not indicate any significant failures in LC performance, blockages or evidence of sputtering.

A coarse measure of sample preparation and instrument reproducibility can be gained through comparison of total protein amounts per sample. Hi3 quantitation relies on the observation that the summed intensity of the three most abundant peptide ions for a protein is proportional to the concentration of that protein in the sample. If equal amounts of peptide were present in each sample, one would expect that the total protein amount for all protein groups would be approximately equal across samples. As shown in Figure 5.5c, total protein amount was highly

variable across biological replicates in both conditions. WT samples displayed a higher average protein abundance and a smaller spread in abundance values (14136 fmol ± 32%; mean ± CV) than KO samples (9405 fmol ± 62%; mean ± CV), with the highest intensity observed in the AKO sample and the lowest intensity observed in the BKO sample. Notably, the BKO sample also displayed the lowest rate of protein and peptide identifications (Figure 5.5a and 5.5b), raising the possibility of an instrument or sample preparation issue for the sample. BWT samples were run in alternate with BKO samples and thus shared the most similar instrument running conditions. BWT samples do not display a systematic reduction in protein abundance or identifications, thus ruling out instrument error as a cause for the low numbers of protein IDs in BKO samples. Together, these data indicate that error introduced by LC-MS instrumentation was very small and is unlikely to be the main reason for the differences observed among WT and KO samples.

The quality of sample preparation was assessed to determine if an error was introduced at this stage. Tryptic digest efficiency could vary among samples due to pipetting error or the use of an old batch of trypsin during sample preparation. Incomplete digestion could result in a high proportion of generated peptides being too large for detection using the 50-2000 m/z mass range of the quadrupole and thus reduce workflow sensitivity. Digest efficiency was assessed indirectly by measuring the proportion of peptide intensity in a given sample originating from peptides with missed cleavages (peptides which contain lysine or arginine residues within the peptide sequence in addition to the C-terminal basic residue). As shown in Figure 5.6a, this proportion is similar for all samples (7.5%), and the subset of BKO (8.1%) and CKO samples (7%), indicating that inefficient trypsin digest is also unlikely to explain the disparity in protein amount and identifications between KO samples.

If protein samples were not well normalised at the level of loading for gel electrophoresis, issues with the total amount of peptide yield would necessarily be incurred. Prior to tryptic digest, proteins were separated by 1D-PAGE and stained with Coomassie dye. This gel was imaged, and densitometric quantitation reveals that total sample load was variable for samples from biological replicates B and C at the point of gel loading (Figure 5.6 b and c). A comparison of on-

gel loading amount (using Coomassie signal as a proxy measure) with summed total absolute protein abundance for biological repeats B and C displayed a moderate correlation (R^2 =0.65) (Figure 5.6c). For both measures, the lowest signal values were observed for the BKO sample. Taken together, these results indicate that the variation in KO sample protein identification and total protein abundance are not a result of instrument error or a poor-quality tryptic digest but rather a reflection of poor sample normalisation at the point of gel loading for BKO samples in particular.

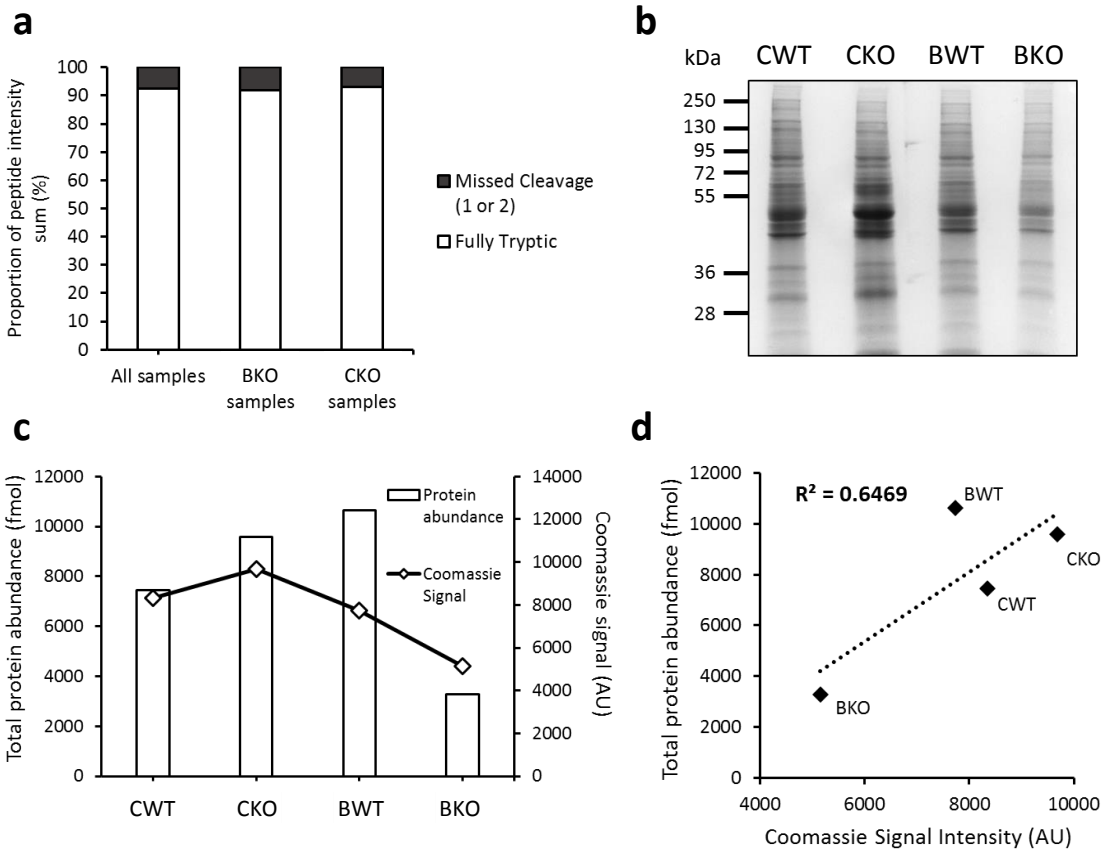

Figure 5.6 Evaluation of sample preparation steps. **a**, Bar chart to compare the proportion of total peptide intensity from peptides with 1 or 2 missed cleavages among all samples, BKO samples and CKO samples. **b**, Coomassie staining of indicated samples following 1D SDS-PAGE prior to band excision and in-gel digestion. **c**, Comparison of densitometric quantification of signal from corresponding lanes in **b** with summed protein abundance for the indicated samples. **d**, Scatterplot to indicate positive correlation between protein abundance and Coomassie signal. AU – arbitrary units. The 3-letter sample identifiers first indicate biological replicate (A, B or C) then condition (WT or KO).

5.3.3. Data filtering and normalisation

A total of 3094 proteins were quantified in this analysis. Only proteins which were observed in at least two biological replicates were taken forward for comparison. In addition, proteins were designated as unique to a condition if they satisfied the above criteria and were not observed in any of the replicates in the other condition. This filtering process resulted in a dataset of 1538 well-replicated proteins, 88 of which were unique to WT samples and 76 unique to KO samples.

Data were normalised at the level of reconstituted biological samples to account for variation introduced during sample preparation. Three normalisation strategies were compared: simple log transformation of the summed protein abundances (Log2), normalisation to global intensity (GI) and normalisation to median intensity (Median). Normalisation was implemented using the NormalyzerDE platform (Willforss, Chawade and Levander, 2019). All normalisation steps were performed on absolute protein abundances as calculated using the Hi3 method. GI normalisation is accomplished by dividing measured protein abundance by the total abundance of all proteins in the sample, then multiplying these values by the mean of summed protein abundance of all samples. Values are then log2 transformed. Normalisation to median intensity was performed by dividing the measured protein abundance for a given protein in given technical replicate by the median of protein abundance for proteins in that sample, then multiplying this by the median protein abundance of all proteins in all replicates. Data was then log2 transformed.

To assess the suitability of each normalisation method, the distribution of within-condition % CV for all proteins was plotted. The median CV observed when using simple log2 transformation (30%) was reduced by both GI normalisation (25%) and normalisation to median intensity (24%), but normalisation to median intensity was more effective at tightening the spread of the data (Figure 5.7a). Inspection of relative log expression (RLE) plots indicates that the BKO sample is best normalised (displays a median close to 0) under median normalisation as opposed to GI normalisation (Figure 5.7b).

Using all normalisation methods, samples displayed an average within-condition Pearson correlation coefficient of 0.64, indicating a moderate correlation in protein abundance across biological replicates. Principle component analysis was performed to assess the similarity between biological samples in two dimensions. Samples did not appear to cluster by condition (WT or KO) when using any normalisation method (Figure 5.8), indicating low within-condition similarity. Based on these tests, normalisation to median intensity was implemented prior to statistical testing.

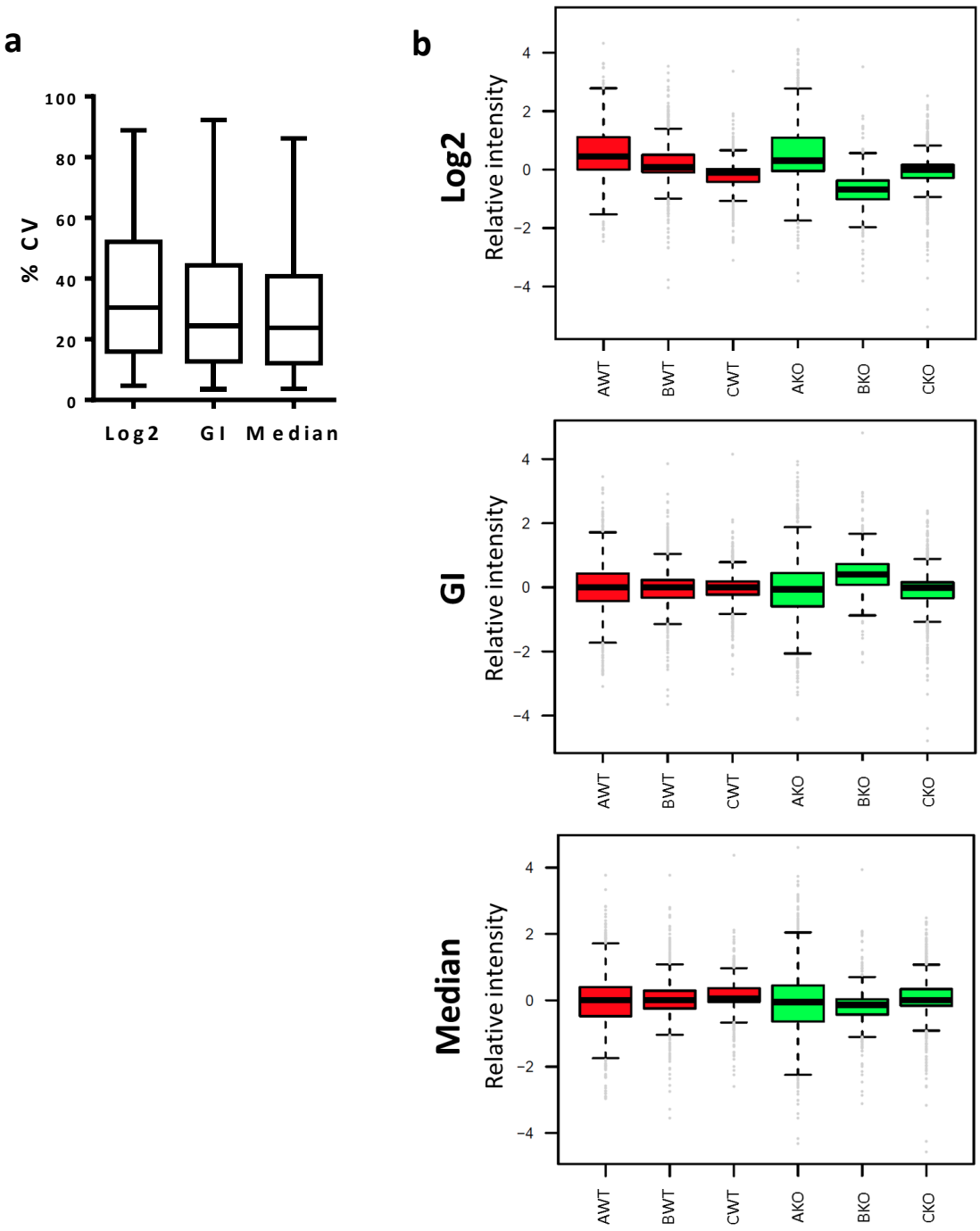

Figure 5.7 Evaluation of normalisation methods. **a**, Box and whisker plot to show the distribution of % CV for all proteins in Log_2-transformed and GI normalised samples. Whiskers indicate 5th and 95th percentiles. **b**, Relative log expression (RLE) plots to show the distribution of deviations from the population median under different normalisation strategies. Plots with medians deviating from 0 are poorly normalised. Log2 – log_2-transformed, GI – normalised to global intensity, Median – normalised to median intensity. The 3-letter sample identifiers first indicate biological replicate (A, B or C) and then condition (WT or KO).

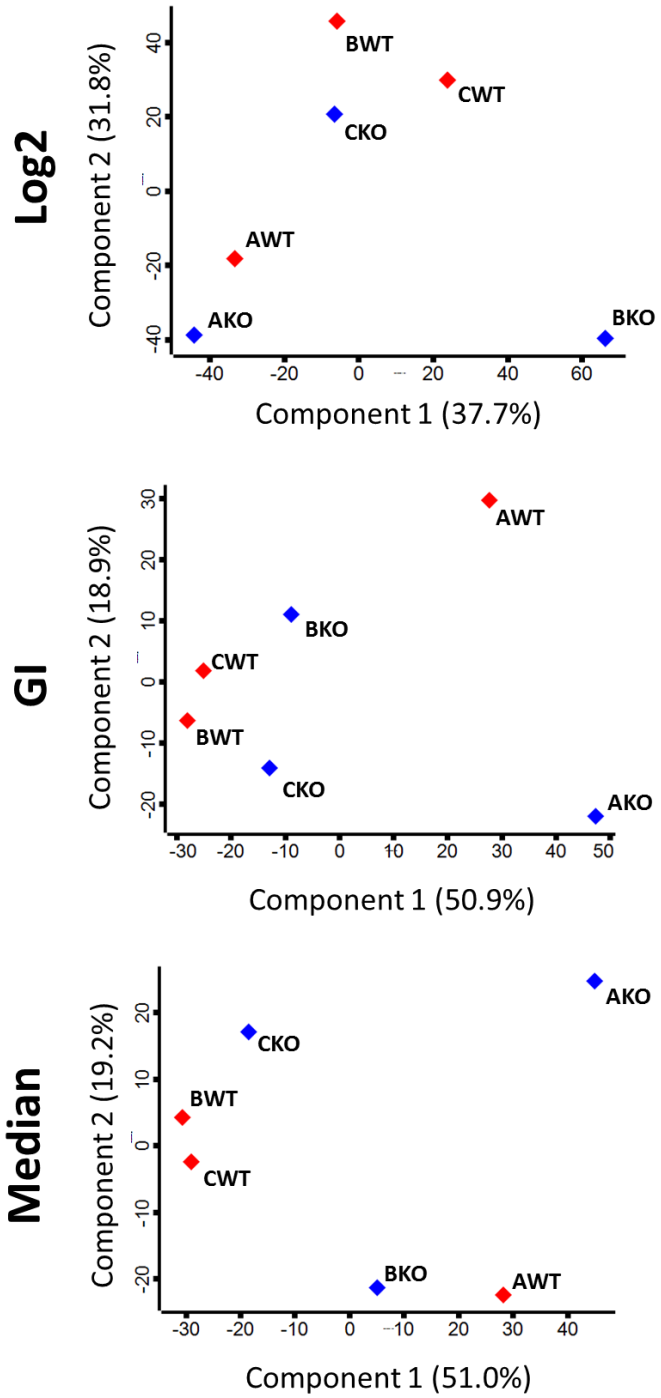

Figure 5.8 Samples do not cluster by condition under any normalisation method tested. PCA illustrates changes in data structure in 2 dimensions upon using different normalisation methods. Red diamonds indicate WT samples and blue diamonds indicate KO samples. The 3-letter sample identifiers first indicate biological replicate (A, B or C) and then condition (WT or KO).

5.3.4. Comparison of Lumbar extract dataset with published motor neuron and spinal cord datasets

To assess dataset validity, comparisons were performed to 4 published murine datasets from different backgrounds (Table 5.1). P10 lumbar spinal cords (Lumbar SC) displayed poor correlation with Hepa 1-6 cells (Pearson correlation 0.17), a slightly better correlation to the Embryonic spinal cord (0.36), and a moderate correlation with primary motor neurons (0.50) and adult spinal cord (0.55) (Figure 5.9a). Given the large disparities in animal age and culture conditions between these sample types, the presence of even a moderate correlation supports the notion that our dataset is of spinal cord origin. Specific markers of neuronal lineage in our data were examined next.

Dataset	Sample type	Publication	Proteome size
Hepa 1-6	Mouse Hepatoma cell line	(Hornburg *et al.*, 2014)	6201
Primary MN	Mouse primary motor neurons (3-7 days culture, following enrichment of neural progenitors from mouse embryos)	(Hornburg *et al.*, 2014)	6670
Embryonic SC	Embryonic Spinal cord extract (E12.5)	(Hornburg *et al.*, 2014)	6366
Adult SC	Whole spinal cord extract from adult mice (15 weeks)	(Hasan *et al.*, 2019)	7211

Table 5.1 Source and description of datasets used for benchmarking of lumbar spinal cord samples. In each case the control dataset was used for comparison to the WT lumbar spinal cord sample.

Overlap in protein identifications between our lumbar spinal cord dataset and the published primary motor neuron and adult spinal cord datasets was assessed. The majority of proteins detected in our lumbar spinal cord dataset (82%) are shared with both of these datasets. Consistent with the notion that our lumbar spinal cord dataset represents a mixed population of spinal cord tissue, the adult spinal cord dataset shares the largest number of pairwise IDs with our lumbar spinal cord samples, including glial-specific proteins such as Glial fibrillary acidic protein (Gfap) and Myelin basic protein (Mbp) (Fig 5.9b).

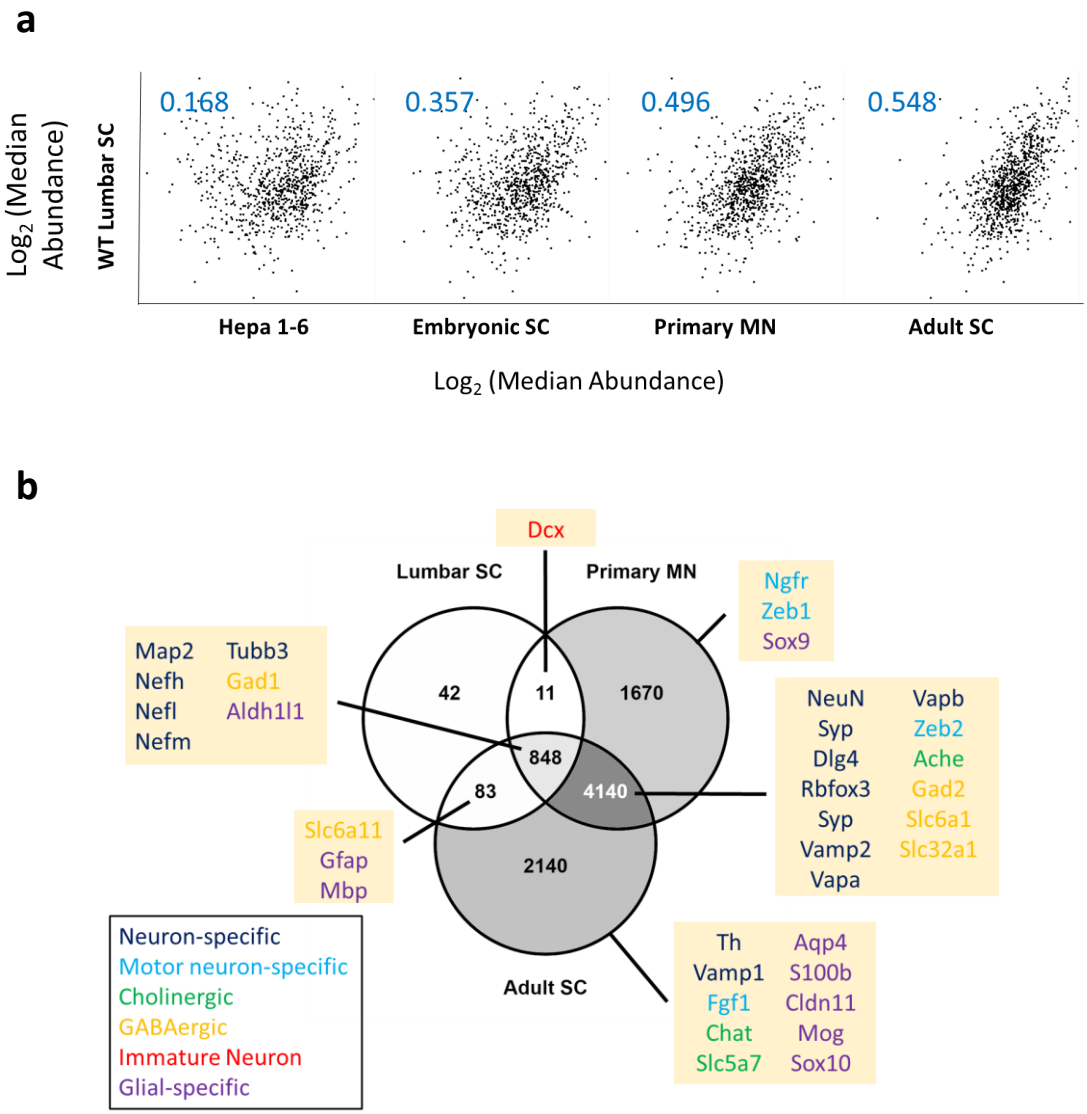

Figure 5.9 Global comparison of wild-type neuronal datasets. **a**, Scatter plots to show correlation in protein abundance across datasets. Numbers inset indicate Pearson correlation. **b**, Venn diagram to show overlap in protein IDs across datasets. Selected lineage specific marker proteins of interest are indicated. Lumbar SC – Lumbar Spinal Cord, Hepa 1-6 – Mouse hepatoma cell line, Embryonic SC – embryonic spinal cord, Primary MN – primary motor neuron, Adult SC – adult spinal cord. For all comparisons shown, the wild-type datasets were compared.

Six of the fourteen neuron-specific markers tested were found in the lumbar spinal cord dataset (Figure 5.9b). All three isoforms of the neuron-specific protein neurofilament heavy (Nefh), neurofilament medium (Nefm) and neurofilament light (Nefl) were detected, along with microtubule-associated protein 2 (Map2), tubulin β3 (Tubb3) and synaptophysin (Syp). The commonly used marker of mature neurons NeuN (Rbfox3) was not detected in our dataset but was found in both primary motor neuron and adult spinal cords.

None of the 16 motor neuron markers tested were observed in our dataset. Notably, only one of these markers was observed in the primary motor neuron and adult spinal cord datasets: the homeobox domain protein Zinc Finger E-Box Binding Homeobox 2 (Zeb2). Additionally, neuronal growth factor receptor (Ngfr) and Zinc Finger E-Box Binding Homeobox 1 (Zeb1) were observed in primary motor neuron, with choline acetyltransferase (Chat) and fibroblast growth factor (Fgf1) observed in adult spinal cords (Figure 5.9b). The commonly used motor neuron markers Islet-1 (Isl-1) and Motor Neuron and Pancreas Homeobox 1 (HB9) were not observed in any datasets. This result is perplexing given the depth of proteome coverage achieved in the published datasets and the fact that both proteins can generate >10 tryptic peptides suitable for MS detection, suggesting that these proteins are expressed at very low levels in neurons.

The microtubule-associated protein doublecortin (Dcx) is commonly used as a marker of immature neurons as it appears to be involved in neuron differentiation and migration during embryonic development and postnatally (Des Portes *et al.*, 1998; Francis *et al.*, 1999; Brown *et al.*, 2003). Interestingly, Dcx is detected in our dataset and in the primary motor neuron dataset, but not in the adult spinal cord dataset (Figure 5.9b).

As almost all motor neurons (including those connecting to skeletal muscle on the lumbar spinal cord) are cholinergic, datasets were assessed for the presence of proteins involved in the synthesis and/or neurotransmitter cycling of acetylcholine. Similar tests were performed for the other major neurotransmitter types. None of the cholinergic neuron marker proteins were detected. Five markers of GABAergic neurotransmission were detected; two of which were

present in the lumbar spinal cord dataset, indicating the presence of GABAergic neurons in our samples (Figure 5.9b).

Lastly, several glial cell markers were detected in the lumbar spinal cord samples, including Glial-fibrillary acidic protein (Gfap), the oligodendrocyte-specific Myelin basic protein (Mbp) and the pan-astrocyte marker 10-formyltetrahydrofolate dehydrogenase (Aldh1l1) (Neymeyer, Tephly and Miller, 1997; Cahoy *et al.*, 2008; Tong *et al.*, 2014), indicating the presence of a mixed population of neuronal and glial cells in our lumbar spinal cord dataset. These results, including all marker proteins used, are provided in Appendix Table 8.5.

To obtain a better understanding of the depth of proteome coverage, the dynamic range of quantified proteins in the lumbar dataset was compared to those in the adult spinal cord dataset (Hasan *et al.*, 2019). Proteins quantified in the lumbar dataset span a range of ~3.5 orders of magnitude, whereas the adult spinal cord dataset spans a dynamic range of ~5.5 orders of magnitude. This difference is likely due to the authors' use of an extensive 40-fraction HPLC fractionation step.

Taken together, these data suggest that the lumbar dataset obtained originates from a mixed pool of neuronal and glial cells, sharing highest similarity in protein abundance and identifications with an adult spinal cord dataset, validating our sample preparation pipeline. Unfortunately, the data do not confirm the presence of motor neurons. Notably, many motor neuron markers were not detected in datasets with much deeper coverage, suggesting that these proteins are expressed at low abundance. Their absence in our dataset does not, therefore, exclude the presence of motor neurons in our lumbar extract samples.

5.3.5. Differential expression analysis

Unpaired, two-sided *t*-tests were performed to compare the average normalised absolute protein abundance between conditions, assuming equal variances and correcting for multiple testing

using an FDR cut-off of 0.05. These tests did not identify any significant differences between the conditions. We turned to our qualitative data and performed functional analysis on the subset of proteins which were defined as unique in WT or KO conditions. Notably, HSJ1 itself was detected uniquely in WT samples, lending credibility to this approach. A full list of these proteins is given in Appendix Table 8.4.

The behaviour of proteins known to aggregate in various neurodegenerative diseases were evaluated. Htt, Sod1 and α-synuclein were not detected in our dataset, whilst Tau and TAR DNA-binding protein 43 (TDP-43), which is linked to ALS (Arai *et al.*, 2006; Ling, Polymenidou and Cleveland, 2013; Tan *et al.*, 2017), were detected in our data but were not differentially expressed.

Three proteins observed uniquely in WT samples stood out as candidate HSJ1 client proteins. Ubiquilin-1 (Ubqln1) is a component of the ubiquitin-proteasome system, and is known to interact with HSJ1a (Heir *et al.*, 2006). In addition, Ganglioside Induced Differentiation Associated Protein 1 (Gdap1) and Ataxin-10 (Atxn10) were observed at relatively high abundance in all 3 WT samples but in none of the KO samples, and therefore represent high confidence candidates. Intriguingly, mutations in the human homologue of Gdap1 have been shown to cause CMT2 (Cuesta *et al.*, 2002), a neurodegenerative disease bearing very similar clinical features to that caused by some HSJ1 mutations.

Gene Ontology (GO) enrichment analysis was performed using the DAVID functional annotation suite assessing biological process and cellular compartment terms for enrichment. The software calculates the chance that a given GO term is over-represented in a query gene list relative to a background gene list. Enrichment analysis of the proteins defined as unique in WT or KO conditions was performed using all 1538 well-replicated proteins as a background. Enrichment was defined as a p-value ≤ 0.05, corrected for multiple testing using an FDR cut-off of 0.05 (Benjamini-Hochberg procedure). No GO functional annotations were significantly enriched relative to the background dataset among the unique proteins in either condition.

Analysis of functional annotation enrichment does not account for protein-protein interactions (PPI). Network analysis was performed on the unique IDs to identify patterns in the data. PPI information from published sources and curated databases were ranked and scored by STRING (Szklarczyk *et al.*, 2019) and high confidence interactions (score > 0.7) were used to construct a PPI network (Figure 5.10) This analysis identified several interesting clusters of proteins which differ in expression between P10 lumbar spinal cords. Several proteins which form the large ribosomal subunit were identified in KO samples only, indicating a possible upregulation of protein synthesis. Three proteins linked to lipolysis in adipocytes, Perilipin-1 (Plin1), Hormone-sensitive factor (Lipe) and 1-acylglycerol-3-phosphate O-acyltransferase (Abhd5) were found exclusively in KO samples, suggesting impaired lipid homeostasis. A group of proteins involved in phosphatidylinositol 3-kinase (PI-3K) and mammalian target of rapamycin complex 1 (mTORC1) signalling were not detected in KO samples. Another group of related proteins unique in WT samples included Plexins A2 and A3 and their co-receptor Neuropilin 2 (Nrp2), possibly indicating a loss of sensitivity to semaphorin signalling, which has roles in axon guidance (Pasterkamp and Giger, 2009).

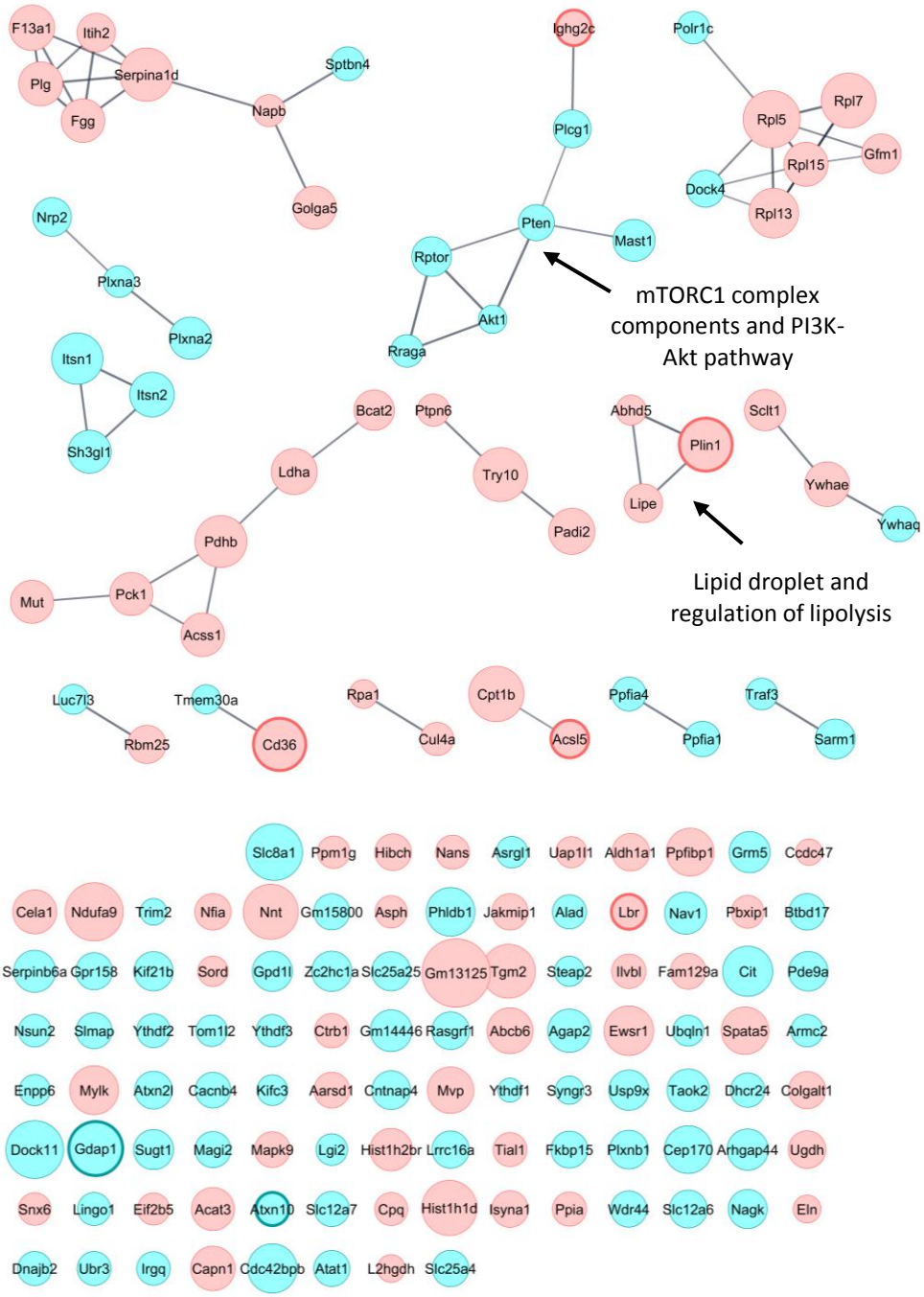

Figure 5.10 Protein-protein interaction network to show connectivity among proteins uniquely identified in HSJ1 WT or HSJ1 KO lumbar spinal cord extracts. Proteins detected in ≥2 biological samples from one condition and in none of the samples from the other condition are denoted unique in WT/KO. Proteins detected in all 3 biological replicates are highlighted with a thick border. Nodes represent proteins and edges represent protein-protein interactions as annotated and scored in the STRING network analysis tool. Node colours indicate the condition in which the protein was uniquely identified (Blue – WT, Pink – KO). Node size is scaled to mean absolute protein abundance and edge thickness to STRING interaction score. Nodes without high confidence interactions to other nodes in the network are shown below for completeness.

5.4. Discussion

A large network of chaperones and co-chaperones actively maintain protein copy number and folding state in the cell. Failures in the chaperone network often result in neurodegenerative disease (Hartl, Bracher and Hayer-Hartl, 2011). In this work, we attempted to identify the molecular clients of the HSJ1 cochaperone and characterise cellular dysregulation in its absence. A HSJ1$^{-/-}$ mouse was used as a model system for the study of motor neurons. Motor neurons innervating the lower limbs are located in the lumbar spinal cord. Previous work in the Cheetham group established that motor neuron numbers in the lumbar spinal cord of HSJ1 KO mice are reduced by 11.6% relative to their wild-type counterparts, and that this loss occurs between postnatal day 15 (P15) and postnatal day 20 (P20). Also, the elevated levels of total ubiquitinated proteins (1.9-fold increase versus WT) and the autophagy marker protein LC3-II (1.3-fold higher versus WT) were observed in lumbar spinal cord extracts at least as early as P15. Finally, elevated levels of phosphorylated eIF2α and phosphorylated IRE-1, were detectable in P10 lumbar spinal cord extracts from HSJ1 KO mice indicating an active unfolded protein response (UPR). Together, these markers indicate a defect in proteostasis in the cells of the lumbar spinal cord as early as P10 involving high levels of misfolded proteins, an increase in autophagy and induction of the UPR. This defect appears to cause the death of ~11% of neurons in the lumbar spinal cord by P20. Whilst HSJ1 KO mice are behaviourally normal, investigating the mechanism by which HSJ1 loss induces a stress response in the lumbar spinal cord could shed light on motor neuron disease in humans.

Towards this end, a sample preparation and data acquisition workflow was optimised, identifying in-gel digestion with HDMS^{E} acquisition as the most sensitive protocol for future experiments to characterise complex samples. When applied to murine lumbar extracts, this protocol achieved good quantitative data for 1538 proteins. However, total protein amounts in each sample, as measure by Hi3 quantitation, were found to be heterogeneous. This variability could not be attributed to poor mass calibration, chromatography or digest quality. Densitometric quantitation gel loading amounts indicated that the lack of a pre-digestion normalisation step was the most

likely source of KO sample variability. In future, this issue could be avoided by measuring protein amount in aliquots of each extract using a detergent compatible protein quantification method and normalising protein amounts prior to gel loading.

The depth of proteome coverage achieved in this study compares poorly with recent studies in the literature (e.g. Hasan *et al.*, 2019), and suggests that the identifications reported in this analysis are likely to be biassed towards abundant proteins. Modifications to the sample preparation protocol could improve this. The simplest change would be to increase sample loading. During sample preparation for the experiments reported in this work, ~6% of the total lumbar sample extract was loaded onto gels for in-gel digestion. Coomassie staining following 1D-PAGE showed that gel lanes were far from overloaded, emphasising the scope for improving workflow sensitivity at this stage. Furthermore, the use of six fractions is relatively low, so increasing the degree of sample fractionation can be expected to improve coverage. This is exemplified by Hasan *et al.* (2019), who used a 40-fraction HPLC step to measure a larger proportion of the adult mouse spinal cord proteome. An inherent technical challenge when performing in-gel sample preparation is the reproducibility of SDS-PAGE separation and manual gel band excision. Gel-free sample preparation techniques such as in-StageTip (iST) and single pot solid phase-enhanced sample preparation (SP3), coupled with peptide fractionation, have achieved deep proteome coverage from low starting sample material and represent a promising alternative to the in-gel digestion workflow (Hughes *et al.*, 2014; Kulak *et al.*, 2014; Sielaff *et al.*, 2017).

Protein expression in WT lumbar extracts correlated well with published datasets from primary motor neuron and adult spinal cord samples but correlated poorly with proteomes from embryonic spinal cords and a mouse hepatoma cell line. This correlation, together with the observation of neuronal and glial lineage markers in WT lumbar extracts, indicates that samples analysed here originate from a mixed population of neuronal and glial cells.

Despite being unable to identify differentially expressed proteins via statistical analysis, qualitative information in the form of proteins identified as unique in a single condition could be used as a proxy for enrichment. It should be noted that, due to the stochastic nature of peptide ionisation and detection in ESI-MS studies, failure to detect protein in this study cannot be used as proof of absence in the sample. Functional enrichment analysis did not indicate enrichment of any GO terms in either set of unique proteins relative to the background. The failure to identify enriched GO terms is consistent with the results of principal component analysis, which showed poor separation between samples from different conditions. Together, these factors indicate that HSJ1 deficiency does not induce major alterations in proteome composition at the depth of proteome coverage attained and at this time point (P10).

As expected, HSJ1 protein was uniquely detected in WT spinal cord extracts. HSJ1 client proteins are predicted to misfold and be degraded or lost in the detergent-insoluble fraction upon protein extraction. In either case, this would be observable as an enrichment in WT samples. Among the 88 proteins defined as unique in WT samples, two were detected in all 3 WT samples: Atxn10 and Gdap1.

Ataxin-10 is linked to spinocerebellar ataxia type 10 (SCA10), a hereditary neurological disorder. Disease onset has been catalogued as early as age 12 and symptoms include uncoordinated muscle movement and seizures (Grewal *et al.*, 1998; Rasmussen *et al.*, 2001; Teive and Ashizawa, 2013). Interestingly, a subset of SCA10 patients also display motor neuropathy. Whilst the SCA10 disease mechanism is currently unclear (Wakamiya *et al.*, 2006; Keren *et al.*, 2010; White *et al.*, 2010, 2012), knockdown of Atnx10 in cultured rat cerebellar neurons resulted in neuronal death (März *et al.*, 2004) and Atnx10-null mice displayed embryonic lethality (Wakamiya *et al.*, 2006), suggesting that Atxn10 loss can cause severe disruption of cell function. To investigate if Atnx10 folding is assisted by HSJ1, pulldown assays or immunohistochemistry could be used to determine if the proteins binding partners and/or if they display overlapping expression patterns.

Gdap1 represents a particularly interesting candidate for future study, as mutations in this protein have been linked to CMT2 (Cuesta *et al.*, 2002; Tazir *et al.*, 2014). This disease closely resembles the pattern of motor neuron loss without demyelination observed in HSJ1-linked CMT2, raising the possibility that a failure to correctly fold Gdap1 underlies HSJ1-linked CMT2 disease mechanistically. Gdap1 is a relatively short 367-amino acid protein which is highly expressed in neurons (Cuesta *et al.*, 2002). It is localised to the outer mitochondrial membrane, where its exact function is unclear. Study of Gdap1 in cell and mouse models have proposed a role for Gdap1 in mitochondrial fission (Niemann *et al.*, 2009) and positioning (Pla-Martín *et al.*, 2013), resistance to reactive oxygen species (ROS) (Noack *et al.*, 2012; Niemann *et al.*, 2014; Huber *et al.*, 2016) and Calcium homeostasis (Pla-Martín *et al.*, 2013; Barneo-Muñoz *et al.*, 2015; González-Sánchez *et al.*, 2017).

Indeed, studies of Gdap1 deficiency in a neuroblastoma cell line and in a mouse model demonstrated that loss of Gdap1 resulted in lowered ER calcium levels (Barneo-Muñoz *et al.*, 2015; González-Sánchez *et al.*, 2017). As depletion of ER calcium is a well-documented UPR initiator (Shore, Papa and Oakes, 2011), Gdap1 misfolding in the absence of HSJ1 may underlie activation of the UPR in P10 lumbar spinal cord extracts.

The known HSJ1 interaction partner Ubqln1 was identified uniquely in WT samples, indicating that Ubqln1 levels in HSJ1-/- samples are strongly reduced by P10 in the lumbar spinal cord. Ubqln1 is a ubiquitously expressed member of the ubiquilin protein family, all members of which possess a C-terminal ubiquitin-like domain (UBL) and an N-terminal ubiquitin-associated domain (UBA). The UBA domain facilitates binding to polyubiquitin (Zhang, Raasi and Fushman, 2008) and the UBL domain mediates interactions with HSJ1a and with the proteasome (Kleijnen *et al.*, 2000). This domain composition allows ubiquilins to serve as shuttling factors, delivering ubiquitinated substrates to the proteasome for degradation. It is conceivable that Ubqln1 is a client of HSJ1, as an interaction between the HSJ1 ubiquitin-interacting motifs (UIMs) and the Ubqln1 UBL domain has been demonstrated (Heir *et al.*, 2006). However, no explicit link between Ubqln1 folding and HSJ1 has been shown in the literature. If Ubqln1 is indeed misfolded at a

higher rate in the absence of HSJ1, loss of Ubqln1's proteasome shuttling function may result in less efficient clearance of ubiquitinated proteins. This could explain the 1.9-fold increase in ubiquitinated proteins observed in HSJ1-/- lumbar spinal cord extracts at P15. In the context of neurodegeneration, overexpression of Ubqln1 has been shown to be beneficial in mouse and cell models of Alzheimer's and Huntington's disease (Wang *et al.*, 2006; Wang and Monteiro, 2007; Safren *et al.*, 2014). Furthermore, the ubiquilins were recently proposed to act as the primary cytosolic chaperones for transmembrane proteins destined for insertion in the mitochondrial membrane (Itakura *et al.*, 2016). Loss of these functions may compromise proteostasis in HSJ1-deficient motor neurons, possibly causing the observed UPR induction by P10 in HSJ1-/- spinal cords. It would be prudent to confirm the reduction of Ubqln1 levels in HSJ1-/- spinal cord extracts using an using an orthologous method such as western blotting or immunohistochemistry. If confirmed, experiments to explicitly test if Ubqln1 stability is affected by loss of HSJ1 should be pursued.

PPI network analysis of the unique IDs helped glean deeper insights into the dataset. One group of interconnected proteins found exclusively in KO samples play roles in lipid storage: Perilipin-1 (Plin1), Hormone-sensitive factor (Lipe) and 1-acylglycerol-3-phosphate O-acyltransferase (Abhd5). The upregulation of this cluster may indicate a general dysregulation of lipid homeostasis in HSJ1-deficient motor neurons. Defects in lipid metabolism have been proposed as major pathogenic drivers in motor neuron diseases such as ALS and spinal muscular atrophy (SMA) (Schmitt *et al.*, 2014; Tracey *et al.*, 2018). Plin1, identified in all 3 KO samples, is a transmembrane protein found in the unilamellar coat of lipid droplets. Lipid droplets are intracellular stores of neutral lipids which have been shown to exist in most eukaryotic cells including neurons and glial cells (Savage, Goldberg and Schacher, 1987; Papadopoulos *et al.*, 2015; Renvoisé *et al.*, 2016; Olzmann and Carvalho, 2019). Increased levels of Plin1 may indicate an increase in the biogenesis of lipid droplets. Lipid droplet biogenesis in glial cells was recently reported to protect neural stem cells from ROS-induced peroxidation of polyunsaturated fatty acids (PUFAs) (Bailey *et al.*, 2015). This raises a possible link to Gdap1, which has been shown to play a role in the mitigation of ROS-induced stress (Noack *et al.*, 2012; Niemann *et al.*, 2014). If Gfap1 misfolding is indeed a consequence of HSJ1 deficiency, the upregulation of Plin1

in HSJ1-deficient lumbar spinal cords may signal increased lipid droplet biogenesis to mitigate oxidative stress.

PPI network analysis also highlighted a seven-member cluster of proteins involved in Protein kinase B (Akt) and mammalian target of rapamycin complex 1 (mTORC1) signalling, which were mostly downregulated in KO samples. Akt and Phosphatase and tensin homologue deleted in chromosome 10 (Pten) were not detected in any of the KO samples. These proteins are key components of the Phosphatidylinositol 3 kinase (PI3K)-Akt signalling cascade, which promotes growth and neuronal survival when active (Brunet, Datta and Greenberg, 2001; Hanada, Feng and Hemmings, 2004; Endo *et al.*, 2006; Manning and Cantley, 2007). Akt mediates its anti-apoptotic effects in neurons through phosphorylation and inactivation of the pro-apoptotic protein B-cell CLL/lymphoma (Bcl-2) associated agonist of cell death (Bad) (Datta *et al.*, 1997), indirect inhibition of Caspase-9 activation (Zhou *et al.*, 2000) and inhibition of Forkhead Box O3 (FKHRL1) nuclear translocation (Brunet *et al.*, 1999). If Akt were a client protein of HSJ1, its misfolding and loss would promote apoptotic pathways and could explain the loss of motor neurons identified in the lumbar spinal cord of HSJ1 KO mice. However, one would have to explain the relatively mild phenotype (11.6% motor neuron loss), given the central role of Akt in anti-apoptotic signalling in neurons.

Another interesting group within the seven-member cluster are the Regulatory-associated protein of mTOR (Rptor) and Ras-related GTP-binding protein A (Rraga) proteins, both of which play roles in mTORC1 signalling. The mTORC1 complex acts as a nutrient sensor in the cell, promoting or inhibiting protein synthesis in response to nutrient/ energy levels in the cell as well as extracellular cues such as growth factors (Laplante and Sabatini, 2012). Rptor is an important member of mTORC1, acting as a scaffold and protein-protein interaction hub for assembly, regulation and localisation of the complex (Sancak *et al.*, 2008; Dunlop *et al.*, 2009). Rraga is a GTP-binding protein which directly functions in the activation of mTORC1 in response to increased amino acid availability (Sancak *et al.*, 2010). Therefore, the loss of Rptor and Rraga proteins could indicate a defect in mTORC1 assembly and/or activation. When active, mTORC1

antagonises macroautophagy (hereafter autophagy) through phosphorylation of Autophagy and Beclin 1 Regulator 1 (AMBRA1) (Noda and Ohsumi, 1998; Nazio *et al.*, 2013). Inhibition of mTORC1 signalling, caused by loss of Akt signalling or loss of HSJ1 co-chaperone activity, could stimulate chronic autophagy in motor neurons. This could explain the increased levels of autophagosome marker LC3-II in HSJ1 lumbar spinal cords (Heather Smith, unpublished results).

The connections proposed above are purely speculative and require substantial validation. A logical next step would be to confirm the phenotypes described above using an orthogonal measure of protein abundance such as western blotting or immunohistochemistry on spinal cord extracts or sections, respectively. If the loss of Gdap1 is confirmed, characterisation of mitochondrial distribution and ER calcium levels in motor neurons would be interesting.

Mass spectrometry experiments utilising a different experimental paradigm could also be brought to bear to find novel candidates HSJ1 clients. Affinity-purification MS would be an ideal tool to retrieve HSJ1 binding partners from a lumbar spinal cord extract. An alternative experimental design would be to profile the detergent-insoluble fraction by LC-MS to positively identify any aggregated proteins, instead of relying on non-detection to assign candidates. This approach has the advantage of avoiding the issue of stochastic peptide ionisation intrinsic to MS analyses. Spatial proteomics also presents an exciting methodology for identifying client HSJ1 proteins. This experimental strategy involves tagging a protein of interest with an enzyme that chemically labels nearby proteins in the cell. Two popular methods are proximity-dependent Biotin identification (BioID) and APEX2. In BioID, a protein of interest is tagged with the BirA enzyme and heterologously expressed in cultured cells. When Biotin monomers are provided in cell media, BirA ligase attaches Biotin to lysine residues ~10 nm distant from the BirA ligase (Kim *et al.*, 2014). These proteins can then be purified affinity chromatography and identified by MS. Whilst such experiments have mostly been performed in cell culture, BioID was recently applied to the study of mouse brains *in vivo*. These studies have enabled the characterisation of at least two synaptic proteins with previously unknown functions (Uezu *et al.*, 2016; Spence *et al.*, 2019).

A major limitation faced in this work was cross-contamination from non-neuronal cell types. An exciting opportunity to obtain more homogenous samples for MS profiling is the use of patient-derived induced pluripotent stem cells (iPSCs), which have been successfully generated for studies of motor neuron disease. The use of iPSCs incurs many benefits in addition to sample homogeneity such as the ability to make targeted interventions in motoneuron biology such as protein overexpression and knockdown. For example, iPSCs sourced from CMT2 patients were stimulated to differentiate into motor neurons and used to test the potential of therapeutic intervention of Histone deacetylase 6 (HDAC6) inhibition. The patients had mutations in Heat shock protein B1 (HSPB1), which features disruption of mitochondrial transport in axons (Evgrafov *et al.*, 2004; Irobi *et al.*, 2004; Ackerley *et al.*, 2006; Kalmar *et al.*, 2017). HDAC6 is a major regulator of microtubule accessibility, as deacetylation of tubulin prevents attachment of motor proteins to the microtubule and hence axonal transport of mitochondria (Reed *et al.*, 2006). Treatment with HDAC6 inhibitors ameliorated the mitochondrial transport defects in iPSCs and was later shown to improve motor and sensory issues in a mouse model of HSPB1-related CMT2 (Kim *et al.*, 2016). Quantitative proteomic profiling of induced pluripotent stem cells derived from patients with mutations in HSJ1 could provide cross-validation of the patterns identified here and with greater relevance to human disease. Candidate HSJ1 clients identified in this way would then be subjected to extensive validation through mutagenesis or overexpression studies. The tissue extracts profiled in this work provide a reference point for future studies.

6. General Discussion

The analysis of proteins by mass spectrometry is now a mature field and, as exemplified in this work, has great utility when applied to a variety of biological problems. In the field of bottom-up proteomics, a number of exciting milestones have been reached in recent years, including the first draft of the human proteome (Kim *et al.*, 2014), in-depth mapping of subcellular proteomes (Bausch-Fluck *et al.*, 2015; Thul *et al.*, 2017), important advances in sample preparation for microproteomics (Hughes *et al.*, 2014; Kulak *et al.*, 2014; Sielaff *et al.*, 2017), and large scale efforts to compare proteomes across the kingdoms of life (Müller *et al.*, 2020). Relative to genomics technology, there is still great scope for improvement, as bottom-up proteomics experiments are still more labour intensive and require longer instrument times per sample. In this work, quantitative proteomics was used to study the recently described endocytic pathway, FEME, and to help elucidate the causes of neuron death in a murine model of motor neuron degeneration.

6.1. Identifying and validating new cargoes for FEME

Boucrot *et al.* (2015) recently reported the discovery of a novel clathrin independent endocytic route in mammalian cells. The pathway featured rapid uptake of ligand stimulated cargo in tubular carriers. This was dependent on dynamin, Rac, phosphatidylinositol-3-OH kinase, PAK1 and actin polymerization. The authors also demonstrated a regulatory mechanism involving phosphatidylinositol phosphate patterning of the plasma membrane. The pathway showed heterogeneity in its cargo receptors, transporting GPCRs, RTKs and the cytokine receptor IL-2R. Like many CIE pathways, a major impediment for further study of the pathway was the lack of marker cargoes (cargoes exclusively trafficked by FEME). Due to its novelty, FEME's physiological roles were poorly understood. The main thrust of my research was to discover and validate new FEME cargoes in the hopes of understanding the role of FEME and to find useful marker cargoes for pathway elucidation.

New cargoes for FEME had been discovered by rational selection of candidates: the amine GPCRs were tested because β_1-AR had been the first cargo identified and, similarly, the discovery of EGFR as a FEME cargo prompted the investigation of a panel of other RTKs. I set out to discover new FEME cargoes in an unbiased manner. The pool of receptors at the plasma membrane is maintained by a homeostatic balance between secretory/ recycling pathways which localise receptors to the plasma membrane, and endocytic pathways which remove them. Blockade of an endocytic pathway can result in accumulation of receptors which are normally internalised via that pathway at the plasma membrane. Recently published studied had taken advantage of this, using quantitative proteomics to assess changes in the abundance of plasma membrane proteins in response to genetic interventions (Steinberg *et al.*, 2013; Bitsikas, Corrêa and Nichols, 2014). I hypothesised that RNAi-mediated disruption of endophilin would cause accumulation of many FEME cargoes at the cell surface. These cargoes could be then identified in an unbiased manner using quantitative mass spectrometry and later validated using microscopy and biochemical assays.

I optimised conditions for enrichment of the plasma membrane, RNAi-mediated knockdown of endophilins A1, A2 and A3 (EndoTKD) and that of the µ2-subunit of adaptor protein complex 2 (AP2M KD) and, finally, sample preparation for proteomics. The final workflow identified 129 plasma membrane proteins. While differential analysis identified junction plakoglobin (JUP) and dysferlin (DYSF) as interesting candidate FEME proteins, I did not have time to validate them, so this study failed to answer my research questions.

Nevertheless, these candidates do represent exciting opportunities for follow up. JUP is an interesting candidate due to the presence of a proline-rich motif (PRM) suitable for targeting by the endophilin-SH3 domain in its primary sequence. The co-enrichment of several predicted and known binding partners for JUP upon inhibition of FEME, and the identification of JUP as a differentially regulated protein in ANOVA, lend weight to the classification of JUP as a potential adaptor for FEME-mediated uptake of proteins at cell-cell junctions. Indeed, higher FEME activity has been observed at cell-cell junctions in confluent cells (*E. Boucrot*, unpublished results).

Dysferlin (DYSF) was identified as enriched in EndoTKD plasma membrane fractions relative to resting extracts but was not enriched upon CME blockade, thus representing a candidate for exclusive uptake via FEME ("marker cargo"). DYSF is a single-pass type II transmembrane protein found to be 54-fold enriched in the plasma membrane of FEME- blocked cells relative to Control. Structurally, the protein consists of a small extracellular C-terminal domain and a large N-terminal cytoplasmic domain containing a number of PRMs, one of which is a putative target for the endophilin SH3 domain.

A limitation of this study is that none of the cargoes known to accumulate in the plasma membrane upon EndoTKD could be used for verification. β_1-AR was not detected, and EGFR was not observed to accumulate in any of the conditions, casting doubt on the validity of the identified candidate proteins. The failure to detect β_1-AR may be due to its relatively low abundance in RPE1 cells (Boucrot *et al*., 2015). A possible remedy would be to try different methods of surface protein labelling. The strategy I used relies on the availability of primary amines in receptor ectodomains, possibly limiting the number of proteins available for affinity purification. Alternative strategies for surface protein labelling and enrichment have had success (Deeb *et al.*, 2014; Bausch-Fluck *et al.*, 2015), and could be employed to widen the scope of the analysis.

A further weakness was the incomplete knockdown observed for the AP2M KD condition (~36%) in the full-scale experiment. Given that clathrin-coated pits could still form in cells with no detectable μ2-adaptin (by western blotting) it is possible that the knockdown achieved was insufficient to cause strong accumulation CME cargoes at the cell surface, which could explain the lack of EGFR accumulation (Motley *et al.*, 2003; Boucrot *et al.*, 2010). In future studies, it may be useful to generate knockout mutants of μ2-adaptin and endophilin using CRISPR-Cas9 technology. Drawbacks of this approach include the length of time required to select and expand suitable clones, and the danger of selecting clones which have reverted to a wild-type phenotype stochastically (so-called "escapers"). There is also a risk that cells with complete knockout of endophilin/ μ2-adaptin would display aberrant phenotypes or be non-viable.

An unexpected observation was the strong improvement in µ2-adaptin knockdown upon switching transfection reagent (Oligofectamine to RNAiMAX). The composition of these transfection reagents is proprietary, so it is hard to explain why the improvement in knockdown quality was observed, but preliminary immunostaining confocal microscopy analysis revealed tentative clues. RNAiMAX appeared to induce some knockdown in most of the cells in a culture, whilst Oligofectamine only induced strong knockdown in some cells, with many cells unaffected (*E. Boucrot*, unpublished results). These observations underscore the importance of testing multiple transfection reagents, and verifying their effects using microscopy in addition to western blotting, when optimising RNAi experiments.

In my effort to identify novel FEME cargoes, I attempted an affinity purification mass-spectrometry experiment from a membrane protein fraction using an endophilin A2 SH3 domain bait. After establishing conditions for release of biotin-labelled proteins for secondary pulldown, I found that the cost of generating a plasma membrane fraction using biotin-labelling at sufficient volume and concentration to conduct an AP-MS workflow would have been prohibitive, with the biotinylation reagent representing the highest expense. An alternative strategy worth pursuing would be to use centrifugation to generate a crude membrane fraction for affinity purification (e.g. as in Lai, 2013). A sufficient quantity of plasma membrane fraction could also be generated using gradient density centrifugation (Duve, 1975; Dormeyer *et al.*, 2008; Lewandrowski *et al.*, 2009). These options rely on high starting numbers of cells and may result in a less pure plasma membrane fraction (high contamination from ER proteins would be expected) but avoids the high cost of reagents.

Another exciting alternative would be the use of spatial proteomics for identification of potential cargoes. Two popular methods are proximity-dependent Biotin identification (BioID) and APEX2. In BioID, a protein of interest is tagged with the BirA ligase and heterologously expressed in cultured cells. When biotin monomers are provided in cell media, BirA ligase attaches biotin to lysine residues ~10 nm distant from the BirA ligase (Kim *et al.*, 2014). These proteins can then be purified using affinity chromatography and identified by MS. As the central FEME component,

endophilin is a perfect candidate for tagging with BirA ligase. Whilst this experiment is highly dependent on the expressed construct being correctly localised, subsequent purification would not require a plasma membrane enrichment step.

6.2. Investigating how growth factor signalling regulates FEME

In addition to expanding knowledge of FEME's functional roles through identification of candidate cargoes, I also had the opportunity to elucidate details of FEME regulation. Boucrot *et al.* (2015) identified phosphatidyl inositol phosphate patterning of the inner leaflet of the plasma membrane at the leading edge of cells as a key mechanism through which FEME was controlled. Later work undertaken in our lab expanded this picture, establishing that a priming mechanism involving the recruitment of the F-BAR proteins FBP17 and CIP4 downstream of Cdc42 prompts conversion of $PI(3,4,5)P_3$ into $PI(3,4)P_2$ at the leading edge by SHIP1/2 phosphatases. This is followed by recruitment of lamellipodin and finally endophilin. This "priming complex" disassembles within a few seconds in the absence of FEME cargo stimulation (Chan Wah Hak *et al.*, 2018). The discovery of this constant priming process helped explain a key feature of FEME: its prompt response to ligand stimulation.

Boucrot *et al.* (2015) also observed an interesting phenotype that hinted at a separate mechanism of FEME regulation. Addition of extra serum to cell media was observed to strongly increase the rate of FEME whilst removal of serum caused a strong decrease in FEME. Further study showed that this increase in FEME rate upon serum stimulation was observable in multiple cell types. These observations suggested that growth factor signalling plays a role in the control of FEME.

Endophilin plays a central role in FEME events. Following its recruitment to the plasma membrane by lamellipodin, endophilin serves as an adaptor for cargoes and scission factors (dynamin). The endophilin N-BAR domain is capable of inducing and stabilising the high degree of curvature observed on FEME tubules, and the protein marks tubules after they detach from

the plasma membrane and are transported into the cell interior. Regulation of FEME could disrupt the intermolecular contacts that endophilin makes when performing each of these roles. I designed an AP-MS experiment to investigate how the endophilin interactome changes upon growth factor stimulation.

Anti-endophilin magnetic beads were used to purify native endophilin complexes from the lysates of resting or serum stimulated cells. Quantitative bottom-up mass spectrometry was used to identify proteins strongly associated with endophilin complexes in a particular condition. Differential expression analysis identified an interesting candidate cargo adaptor, WW Domain Containing E3 Ubiquitin Protein Ligase 1 (WWP1), which appeared to associate with immunocomplexes isolated from the stimulated lysates. Due to their key role in growth factor signalling, the behaviour of any kinases differentially expressed was also noted. In total, 27 kinases were found to be differentially expressed and, intriguingly, Glycogen synthase kinase 3β (GSK3β) was enriched in resting rather than stimulated cells.

The behaviour of GSK3β was interesting as, during the wider investigation of FEME regulation in our lab, a panel of kinases had been screened using small molecule inhibitors. This screen had identified CDK5, p38 and GSK3 as potential inhibitors of FEME (see Figure 4.1, Ferreira *et al.*, submitted). GSK3 is an unusual kinase in that it recognises pre-phosphorylated residues, which occupy a docking site in GSK3, to guide the positioning of its kinase domain. This confers specificity for target serine/threonine residues 4 amino acids N-terminal of the "priming phosphate". GSK3 is regulated by phosphorylations targeting the unstructured N-terminus of the protein (S9 in GSK3β and S21 in GSK3α), allowing it to act as a pseudosubstrate for its own docking site, thereby inhibiting substrate recognition by the kinase (Cross *et al.*, 1995; Frame, Cohen and Biondi, 2001). These phosphorylations can be induced by growth factor signalling, including through phosphorylation by AKT and p90 S6 kinase (S6K) (Cross *et al.*, 1995; Eldar-Finkelman *et al.*, 1995). Lastly, GSK3β has been shown to participate in the regulation of endocytic events, inhibiting dynamin activity in CME but also potentiating activity-dependent bulk

endocytosis (ADBE) in synapses (Clayton *et al.*, 2010; Smillie and Cousin, 2012; Reis *et al.*, 2015). These factors prompted further study of GSK3 in the context of FEME regulation.

I investigated how GSK3 was being recruited to the leading edge to inhibit FEME. I performed a pulldown experiment and showed that GSK3β, but not GSK3α, was capable of binding directly to the EndophilinA2 SH3 domain in serum-starved but not resting cells. This discovery was built upon and has proven a fruitful route to understand FEME regulation.

Further work conducted in our lab has shown that GSK3β strongly colocalises with endophilin at the leading edge of cells, but that the majority of the GSK3β at the leading edge is in active form. Experiments also showed that GSK3β activity sets the upper and lower bounds within which FEME can be modulated. Indeed, FEME competence (the proportion of cells in a culture which display FEME at any one time) is strongly correlated with GSK3β activity. Finally, inhibition of GSK3 and/or CDK5 enhanced FEME at the levels of cargo sorting, dynamin and dynein recruitment to FEME carriers (Ferreira *et al.,* submitted). Interestingly, one of the cargoes shown to be under GSK3 control, ROBO1, triggers an intracellular cascade which has been shown to inactive GSK3β (Jiang *et al.*, 2005; Byun *et al.*, 2012).

The local recruitment of GSK3β by endophilin, whilst initially surprising, makes sense in the context of the need for spatial control of GSK3's activity. For a kinase, GSK3 has many targets, so restricting its activity to a specific region of the cell is crucial. EndophilinA2's inability to bind to GSK3α is likely explained by the absence of a suitable PRM in GSK3α for engagement by the Endophilin SH3 domain. GSK3β, by contrast, does possess such a PRM (amino acids 307-311).

My contribution and these further studies have built a picture of FEME regulation in which active GSK3β is recruited to the leading edge by endophilin to locally inhibit FEME. GSK3β inhibition can be quickly relieved by growth factor signalling. Given GSK3β's reliance on priming kinases, this implicates other kinases in FEME inhibition. CDK5 has been shown to prime GSK3

phosphorylation of dynamin and the PlexinA1 adaptor CRMP4 (Ferreira *et al.*, submitted), but is likely not the only priming kinase for GSK3β.

Several avenues for future investigation of FEME regulation exist. Firstly, it is likely that endophilin itself is a target for kinases during FEME regulation. Endophilin is phosphorylated at its N-terminal helix 0 (T14) by ROCK and at two positions in helix 1 (T73 and S75) by LRRK2. Both of these modifications are capable of inhibiting endophilin's membrane binding capacity and thus inhibiting the pathway (Kaneko *et al.*, 2005; Ambroso, Hegde and Langen, 2014; Arranz *et al.*, 2015). An attempt was made to detect and quantify changes in phosphopeptides in both AP-MS experiments reported here, but the phosphopeptides detected were of low confidence. This was expected as a relatively small amount of protein was analysed and no effort to enrich the sample for phosphopeptides was made. Furthermore, there is no guarantee that a majority of the endophilin captured by IP would be modified by GSK3β. Indeed, many cells do not exhibit FEME in cell extracts from which IPs were conducted (no FEME is observed in ~50% of cells in Resting and ~30% of cells in +10% FBS conditions). The proportion of captured endophilin monomers which bear phosphorylations relevant to FEME is therefore expected to be small despite enrichment by IP.

To assess changes in the phosphorylation of endophilin upon serum withdrawal or stimulation, a TiO_2 phosphopeptide enrichment step could be added to the workflow following tryptic digest. This step is widely used to exclude non-phoshorylated peptides from a sample and thus improve workflow sensitivity for phosphopeptides (Thingholm *et al.*, 2006). The use of larger amounts of starting material is essential to ensure that adequate amounts of phosphopeptide will be present following enrichment for detection by MS, given their low expected prevalence in the sample.

A second direction for future study is related to the question of how the inhibitory phosphorylations imposed by GSK3β and CDK5 are removed so quickly. It is possible that a single phosphatase rapidly removes the inhibitory signals upon ligand stimulation. A similar scenario is observed in neurons, where many endocytic proteins are inhibited by

phosphorylations (Tomizawa *et al.*, 2003; Liang *et al.*, 2007). An influx of Ca^{2+} upon axon depolarisation activates calcineurin phosphatase, which removes these phosphate groups and permits swift compensatory endocytosis. It has not been established if Ca^{2+} controls FEME in non-neuronal cells and it is not clear which phosphatase could be specifically activated downstream of the many FEME cargo receptors (calcium-dependent or otherwise). This raises an alternative possibility, whereby the inhibitory phosphorylations applied to FEME proteins are constantly being removed by a set of constitutive phosphatases. In this schema, FEME is held off through the constant activity of a kinase or group of kinases, inactivation of one of all of which tips the system toward favouring FEME. This notion makes sense in light of the fact that many phosphatases are constitutively active and have a wide range of substrates. It would be interesting to investigate which hypothesis is correct.

Further investigation of the candidate FEME adaptors/ cargoes identified in this work (WWP1, JUP and DYSF) would proceed by first testing if these candidates can bind to the endophilin SH3 domain. Promising results can be followed with antibody-based confirmation of *in vivo* associations with endophilin through immunoprecipitation, fluorescent microscopy and ultimately antibody feeding assays.

In this work I have been privileged to investigate a newly discovered endocytic pathway. What are the prospects for improving human health that spring from this work and how might FEME contribute to new therapies? Aside from the knowledge that Shiga and cholera toxins can gain entry to cell via FEME, it is possible that some viruses also exploit the pathway. Enterovirus 71 (EV71) causes infections via the oral-faecal route and mainly affects children. A subset of infections cause severe disease which may feature neurological defects and or cardiopulmonary failure (Solomon *et al.*, 2010). The virus has recently been described to enter cells via an endophilin-dependent, dynamin-dependent route that was independent of (Chen *et al.*, 2019). While the dynamics of the entry process were different (viral entry took a longer timeframe), knowledge of FEME characteristics helped direct study of the process and will undoubtedly inform the study of emerging pathogens in the future.

In the field of cancer therapeutics, a new strategy for enhancing monoclonal antibody treatment relying on inhibition of endocytosis has shown great promise. Chew *et al.* (2020) tested a novel treatment paradigm: the use of endocytosis inhibitors to potentiate antibody dependent cellular cytotoxicity (ADCC). ADCC is an important mechanism of tumour clearance during monoclonal antibody therapy, whereby opsonised tumour cells are recognised by natural killer cells of the innate immune system, which destroy the target cells through the release of granzymes and perforin (Wang *et al.*, 2015). An important factor in the success of a monoclonal antibody's ability to induce ADCC is the residence time of target receptors at the cell surface (Scott, Wolchok and Old, 2012). Using a squamous cell carcinoma cell line, Chew *et al*. (2020) showed that dynamin inhibition in the context of monoclonal antibody treatment caused a greater potentiation of ADCC than clathrin inhibitors alone. They attributed this to CME and FEME inhibition-dependent accumulation of EGFR at the surface of the cells. Finally, using a tumour *ex-vivo* model, the authors showed that a single dose of the dynamin inhibitor prochlorperazine (PCZ), redistributed EGFR to the surface of tumour cells, indicating that such potentiation of ADCC could work in humans.

Whilst FEME has not, so far, constituted a "magic bullet" opportunity for the development of novel therapies, the examples above demonstrate how an understanding of FEME has helped guide experiment designs and explain observations. This is likely to continue as the pathway is further elucidated.

6.3. Quantitative LC-MS profiling of spinal cord samples in a mouse model of motor neuron degeneration

A network of molecular chaperones ensures the correct folding of proteins in cells. This network is particularly important in neurons, where failures in proteostasis are associated with a number of neurological disorders (Hartl, Bracher and Hayer-Hartl, 2011). HSJ1 (DNAJB2 in humans) is a neuronally enriched HSP40 cochaperone that acts as an adaptor for the HSP70 chaperone

(Cheetham, Brion and Anderton, 1992). Interestingly, HSJ1 possesses two ubiquitin interacting motifs (UIM) which allows it to function as a shuttle linking the proteasome, ubiquitination and chaperone machinery of the cell (Westhoff *et al*., 2005). Loss-of-function mutations in the locus encoding HSJ1 have been causally linked to distal hereditary motor neuropathy (dHMN) and Charcot-Marie-Tooth syndrome type 2 (CMT2) in humans (Gess *et al.*, 2014; Tazir *et al.*, 2014). Both conditions feature the progressive dysfunction of motor neurons, with concomitant loss of sensory neurons in CMT2 only. Interestingly, HSJ1 overexpression is neuroprotective a number of cell models of neurodegenerative disease (Westhoff *et al*., 2005; Borrell-Pagès *et al*., 2006; Howarth *et al*., 2007; Labbadia *et al*., 2012, Novoselov *et al*., 2013). This investigation aimed to define the motor neuron clients of HSJ1 to explain the protein's neuroprotective effects.

A HSJ1 KO mouse was used as a model system for this study. Motor neurons innervating the lower limbs are located in the lumbar spinal cord. Previous work in the Cheetham group established that motor neuron numbers in the lumbar spinal cord of HSJ1 KO mice are reduced by 11.6% relative to their wild-type counterparts, and that this loss occurs between postnatal day 15 (P15) and postnatal day 20 (P20). Also, the elevated levels of total ubiquitinated proteins (1.9-fold increase versus WT) and the autophagy marker protein LC3-II (1.3-fold higher versus WT) were observed in lumbar spinal cord extracts at least as early as P15. Finally, elevated levels of phosphorylated eIF2α and phosphorylated IRE-1, were detectable in P10 lumbar spinal cord extracts from HSJ1 KO mice indicating an active unfolded protein response (UPR). Together, these markers indicated a defect in proteostasis in the cells of the lumbar spinal cord as early as P10 involving high levels of misfolded proteins, an increase in autophagy and induction of the UPR. This defect appears to cause the death of ~11% of neurons in the lumbar spinal cord by P20. Whilst HSJ1 KO mice are behaviourally normal, investigating the mechanism by which HSJ1 loss induces a stress response in the lumbar spinal cord could shed light on motor neuron disease in humans.

I hypothesised that HSJ1 client proteins would fail to fold correctly at a higher rate in HSJ1 KO animals, and would be degraded at the proteasome or lost in the insoluble fraction. Furthermore,

if HSJ1 is important for shuttling ubiquitinated clients to the proteasome, these would be expected to accumulate in KO animals. Therefore, by comparing the proteome complement of WT and KO mouse lumbar spinal cords at P10 using quantitative proteomics, I could identify any proteins which are differentially expressed as potential HSJ1 clients for further validation. In addition, sufficient coverage of the lumbar spinal cord proteome could independently confirm the observed induction of the UPR and increased autophagy in KO mice.

My analysis obtained quantitative information for 1538 proteins. The dataset displayed highest similarity (in protein abundances) to a dataset from adult mouse spinal cords, and an analysis of marker proteins showed that the sample originated from a mixed population of neuronal and glial cells. Finally, HSJ1 itself was only observed in WT samples. Together, these results confirmed that the sample preparation and MS analysis workflow had worked well and argue for the validity of any conclusions drawn. Unexpectedly, no differentially expressed proteins were identified and no significant changes in the proportion of GO terms in either condition were observed, indicating that HSJ1 loss does not induce major changes in protein abundance at the depth of coverage achieved and at this timepoint.

Turning to my qualitative data, I found 88 proteins unique to WT mouse samples and 76 unique in KO mice. Of these, three proteins unique in WT samples stood out as interesting candidates: Ubiquilin-1 (Ubqln1), Ganglioside Induced Differentiation Associated Protein 1 (Gdap1) and Ataxin-10 (Atxn10).

Gdap1 represents a particularly interesting candidate for future study, as mutations in this protein have been linked to CMT2 (Cuesta *et al.*, 2002; Tazir *et al.*, 2014). This disease closely resembles the pattern of motor neuron loss without demyelination observed in HSJ1-linked CMT2, raising the possibility that a failure to correctly fold Gdap1 underlies HSJ1-linked CMT2 disease mechanistically. Gdap1 is a 367-amino acid protein which is highly expressed in neurons (Cuesta *et al.*, 2002). It is localised to the outer mitochondrial membrane, where its exact function is unclear. Study of Gdap1 in cell and mouse models have proposed a role for Gdap1 in

mitochondrial fission (Niemann *et al.*, 2009) and positioning (Pla-Martín *et al.*, 2013), resistance to reactive oxygen species (ROS) (Noack *et al.*, 2012; Niemann *et al.*, 2014; Huber *et al.*, 2016) and Calcium homeostasis (Pla-Martín *et al.*, 2013; Barneo-Muñoz *et al.*, 2015; González-Sánchez *et al.*, 2017). Indeed, studies of Gdap1 deficiency in a neuroblastoma cell line and in a mouse model demonstrated that loss of Gdap1 resulted in lowered ER calcium levels (Barneo-Muñoz *et al.*, 2015; González-Sánchez *et al.*, 2017). As depletion of ER calcium is a well-documented UPR initiator (Shore, Papa and Oakes, 2011), Gdap1 misfolding in the absence of HSJ1 may underlie activation of the UPR in P10 lumbar spinal cord extracts.

To test the candidates identified above, a prudent next step would be to confirm the phenotypes described using an orthogonal measure of protein abundance such as western blotting or immunohistochemistry on spinal cord extracts or sections. If the loss of Gdap1 is confirmed, characterisation of mitochondrial distribution and ER calcium levels in motor neurons would be interesting.

The main limitation of my approach in this study was the relatively low depth of proteome coverage achieved. This is likely attributable to insufficient sample fractionation. During sample preparation, gels were cut into only six fractions, limiting the degree to which peptides could be separated before ionisation by nanoESI and thus compromising workflow sensitivity. Protocols for achieving deeper proteome coverage involve extensive sample fractionation. For example, the adult spinal cord sample dataset used for comparison in this work utilised a 40-fraction HPLC step, quantifying 7211 proteins (Hasan *et al*., 2019). Further fractionation at the peptide level could also be employed, perhaps using high pH RPLC fractionation or SCX after protein level fractionation.

The experiment design used here is suitable for identifying global changes in the proteomes of lumbar spinal cords but cannot directly define the HSJ1 interactome. Any candidates displaying differential expression may reflect adaptations to the loss of HSJ1. To find proteins which directly interact with or form a complex with HSJ1, AP-MS experiments could be used (using HSJ1 as a

bait). Indeed, the availability of HSJ1 KO samples presents the opportunity to perform a powerful controlled AP-MS experiment.

A further limitation of the model system is that only a subset of the cells from which proteins are extracted are likely to be affected by HSJ1 loss (the neurons). Proteins from the surrounding glia and muscle are also released, obscuring the neuronal proteome. A solution to this would be to use patient-derived induced pluripotent stem cells (iPSCs) to generate a homogeneous culture for profiling. Quantitative proteomic comparisons of case and control iPSC cultures could help narrow down the list of potential clients identified from orthologous techniques. It is hoped that a better understanding of how HSJ1 contributes to proteostasis will shed light on the disease mechanisms of motor neuron degeneration.

In summary, quantitative proteomics was applied to discover novel cargoes of the recently described endocytic pathway, FEME. Plasma membrane extracts from mammalian cells were profiled by MS following genetic perturbation of different endocytic routes, identifying several candidate FEME cargoes/ cargo sorting adaptors for future study. An investigation of the mechanisms which regulate FEME identified GS3Kβ as a binding partner for endophilin A2, helping to explain how GS3Kβ activity exerts such tight control over FEME events. Quantitative proteomics experiments were also conducted to help elucidate the causes of neuron death in a murine model of motor neuron degeneration. Spinal cord extracts from mice lacking the HSP40 co-chaperone HSJ1 were compared to their wild-type counterparts. Whilst quantitative measurements failed to identify differentially expressed proteins at the depth of coverage achieved, analysis of our qualitative highlighted several proteins as candidate HSJ1 clients including Gdap1. The datasets presented here serve as a valuable resource for future studies of FEME and HSJ1-asociated neurodegeneration.

7. Bibliography

Ackerley, S., James, P. A., Kalli, A., French, S., Davies, K. E. and Talbot, K. (2006) 'A mutation in the small heat-shock protein HSPB1 leading to distal hereditary motor neuronopathy disrupts neurofilament assembly and the axonal transport of specific cellular cargoes', *Human Molecular Genetics*, 15(2), pp. 347–354. doi: 10.1093/hmg/ddi452.

Ahadi, E. and Konermann, L. (2012) 'Modeling the Behavior of Coarse-Grained Polymer Chains in Charged Water Droplets: Implications for the Mechanism of Electrospray Ionization', *The Journal of Physical Chemistry B*, 116(1), pp. 104–112. doi: 10.1021/jp209344z.

Ambroso, M. R., Hegde, B. G. and Langen, R. (2014) 'Endophilin A1 induces different membrane shapes using a conformational switch that is regulated by phosphorylation', *Proceedings of the National Academy of Sciences of the United States of America*, 111(19), pp. 6982–6987. doi: 10.1073/pnas.1402233111.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y. and Oda, T. (2006) 'TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis', *Biochemical and Biophysical Research Communications*, 351(3), pp. 602–611. doi: 10.1016/j.bbrc.2006.10.093.

Arranz, A. M., Delbroek, L., van Kolen, K., Guimarães, M. R., Mandemakers, W., Daneels, G., Matta, S., Calafate, S., Shaban, H., Baatsen, P., de Bock, P. J., Gevaert, K., Berghe, P. Vanden, Verstreken, P., de Strooper, B. and Moechars, D. (2015) 'LRRK2 functions in synaptic vesicle endocytosis through a kinase-dependent mechanism', *Journal of Cell Science*, 128(3), pp. 541–552. doi: 10.1242/jcs.158196.

Bailey, A. P., Koster, G., Guillermier, C., Hirst, E. M. A., MacRae, J. I., Lechene, C. P., Postle, A. D. and Gould, A. P. (2015) 'Antioxidant Role for Lipid Droplets in a Stem Cell Niche of Drosophila', *Cell*, 163(2), pp. 340–353. doi: 10.1016/j.cell.2015.09.020.

Bansal, D., Miyake, K., Vogel, S. S., Groh, S., Chen, C.-C., Williamson, R., McNeil, P. L. and Campbell, K. P. (2003) 'Defective membrane repair in dysferlin-deficient muscular dystrophy.', *Nature*, 423(6936), pp. 168–72. doi: 10.1038/nature01573.

Barneo-Muñoz, M., Juárez, P., Civera-Tregón, A., Yndriago, L., Pla-Martin, D., Zenker, J., Cuevas-Martín, C., Estela, A., Sánchez-Aragó, M., Forteza-Vila, J., Cuezva, J. M., Chrast, R. and Palau, F. (2015) 'Lack of GDAP1 Induces Neuronal Calcium and Mitochondrial Defects in a Knockout Mouse Model of Charcot-Marie-Tooth Neuropathy', *PLoS Genetics*, 11(4), pp. 1–27. doi: 10.1371/journal.pgen.1005115.

Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A. P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R. L., Yoon, C., … Wollscheid, B. (2015) 'A mass spectrometric-derived cell surface protein atlas', *PLoS ONE*, 10(4), pp. 1–22. doi: 10.1371/journal.pone.0121314.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995.pdf', *Journal of the Royal Statistical Society Series B (Methological)*, 57(1), pp. 289–300. doi: 10.2307/2346101.

Biemann, K. (1992) 'Mass Spectrometry of Peptides and Proteins', *Annual Review of Biochemistry*, 61(1), pp. 977–1010. doi: 10.1146/annurev.bi.61.070192.004553.

Bitsikas, V., Corrêa, I. R. and Nichols, B. J. (2014) 'Clathrin-independent pathways do not contribute significantly to endocytic flux.', *eLife*, 3, p. e03970. doi: 10.7554/eLife.03970.

Blagoev, B., Ong, S. E., Kratchmarova, I. and Mann, M. (2004) 'Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics', *Nature Biotechnology*, 22(9), pp. 1139–1145. doi: 10.1038/nbt1005.

Borrell-Pagès, M., Canals, J. M., Cordelières, F. P., Parker, J. A., Pineda, J. R., Grange, G., Bryson, E. A., Guillermier, M., Hirsch, E., Hantraye, P., Cheetham, M. E., Néri, C., Alberch, J., Brouillet, E., Saudou, F. and Humbert, S. (2006) 'Cystamine and cysteamine increase brain levels of BDNF in Huntington disease via HSJ1b and transglutaminase', *Journal of Clinical Investigation*, 116(5), pp. 1410–1424. doi: 10.1172/JCI27607.

Boucrot, E., Saffarian, S., Zhang, R. and Kirchhausen, T. (2010) 'Roles of AP-2 in clathrin-mediated endocytosis', *PLoS ONE*, 5(5). doi: 10.1371/journal.pone.0010597.

Boucrot, E., Ferreira, A. P. A., Almeida-Souza, L., Debard, S., Vallis, Y., Howard, G., Bertot, L., Sauvonnet, N. and McMahon, H. T. (2015) 'Endophilin marks and controls a clathrin-independent endocytic pathway', *Nature*, 517(7535), pp. 460–465. doi: 10.1038/nature14067.

Brakke, M. K. (1953) 'Zonal separations by density-gradient centrifugation', *Archives of*

Biochemistry and Biophysics, 45(2), pp. 275–290. doi: 10.1016/S0003-9861(53)80005-6.

Brown, J. P., Couillard-Després, S., Cooper-Kuhn, C. M., Winkler, J., Aigner, L. and Kuhn, H. G. (2003) 'Transient Expression of Doublecortin during Adult Neurogenesis', *Journal of Comparative Neurology*, 467(1), pp. 1–10. doi: 10.1002/cne.10874.

Brunet, A., Datta, S. R. and Greenberg, M. E. (2001) 'Transcription-dependent and -independent control of neuronal survival by the PI3K-Akt signaling pathway', *Current Opinion in Neurobiology*, pp. 297–305. doi: 10.1016/S0959-4388(00)00211-7.

Brunet, A., Bonni, A., Zigmond, M. J., Lin, M. Z., Juo, P., Hu, L. S., Anderson, M. J., Arden, K. C., Blenis, J. and Greenberg, M. E. (1999) 'Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor', *Cell*, 96(6), pp. 857–868. doi: 10.1016/S0092-8674(00)80595-4.

Byun, J., Kim, B. T., Kim, Y. T., Jiao, Z., Hur, E.-M. and Zhou, F.-Q. (2012) 'Slit2 Inactivates GSK3β to Signal Neurite Outgrowth Inhibition', *PLoS ONE*. Edited by H. Ulrich, 7(12), p. e51895. doi: 10.1371/journal.pone.0051895.

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A., Thompson, W. J. and Barres, B. A. (2008) 'A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(1), pp. 264–78. doi: 10.1523/JNEUROSCI.4178-07.2008.

Castle, J. D. (2003) 'Purification of Organelles from Mammalian Cells', *Current Protocols in Immunology*, 56(1), pp. 8.1B.1-8.1B.57. doi: 10.1002/0471142735.im0801bs56.

Chan Wah Hak, L., Khan, S., Di Meglio, I., Law, A.-L., Lucken-Ardjomande Häsler, S., Quintaneiro, L. M., Ferreira, A. P. A., Krause, M., McMahon, H. T. and Boucrot, E. (2018) 'FBP17 and CIP4 recruit SHIP2 and lamellipodin to prime the plasma membrane for fast endophilin-mediated endocytosis', *Nature Cell Biology*, 20(9), pp. 1023–1031. doi: 10.1038/s41556-018-0146-8.

Chapple, J. P. and Cheetham, M. E. (2003) 'The chaperone environment at the cytoplasmic face of the endoplasmic reticulum can modulate rhodopsin processing and inclusion formation', *Journal of Biological Chemistry*, 278(21), pp. 19087–19094. doi: 10.1074/jbc.M212349200.

Chapple, J. P., van der Spuy, J., Poopalasundaram, S. and Cheetham, M. E. (2004) 'Neuronal

DnaJ proteins HSJ1a and HSJ1b: a role in linking the Hsp70 chaperone machine to the ubiquitin-proteasome system?', *Biochemical Society transactions*, 32(Pt 4), pp. 640–2. doi: 10.1042/BST0320640.

Cheetham, M. E., Brion, J. P. and Anderton, B. H. (1992) 'Human homologues of the bacterial heat-shock protein DnaJ are preferentially expressed in neurons.', *The Biochemical journal*, 284 (Pt 2), pp. 469–76.

Chen, S. L., Liu, Y.-G., Zhou, Y.-T., Zhao, P., Ren, H., Xiao, M., Zhu, Y.-Z. and Qi, Z.-T. (2019) 'Endophilin-A2-mediated endocytic pathway is critical for enterovirus 71 entry into caco-2 cells', *Emerging Microbes & Infections*, 8(1), pp. 773–786. doi: 10.1080/22221751.2019.1618686.

Chen, Z., Chang, K., Capraro, B. R., Zhu, C., Hsu, C.-J. and Baumgart, T. (2014) 'Intradimer/Intermolecular Interactions Suggest Autoinhibition Mechanism in Endophilin A1'. doi: 10.1021/ja411607b.

Chen, Z., Zhu, C., Kuo, C. J., Robustelli, J. and Baumgart, T. (2016) 'The N-Terminal Amphipathic Helix of Endophilin Does Not Contribute to Its Molecular Curvature Generation Capacity', *Journal of the American Chemical Society*, 138(44), pp. 14616–14622. doi: 10.1021/jacs.6b06820.

Chew, H. Y., De Lima, P. O., Gonzalez Cruz, J. L., Banushi, B., Echejoh, G., Hu, L., Joseph, S. R., Lum, B., Rae, J., O'Donnell, J. S., … Simpson, F. (2020) 'Endocytosis Inhibition in Humans to Improve Responses to ADCC-Mediating Antibodies', *Cell*, 180(5), pp. 895-914.e27. doi: 10.1016/j.cell.2020.02.019.

Chudakov, D. M., Matz, M. V., Lukyanov, S. and Lukyanov, K. A. (2010) 'Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues', *Physiological Reviews*, 90(3), pp. 1103–1163. doi: 10.1152/physrev.00038.2009.

Clayton, E. L., Sue, N., Smillie, K. J., O'Leary, T., Bache, N., Cheung, G., Cole, A. R., Wyllie, D. J., Sutherland, C., Robinson, P. J. and Cousin, M. A. (2010) 'Dynamin i phosphorylation by GSK3 controls activity-dependent bulk endocytosis of synaptic vesicles', *Nature Neuroscience*, 13(7), pp. 845–851. doi: 10.1038/nn.2571.

Cohen, P. and Frame, S. (2001) 'The renaissance of GSK3', *Nature Reviews Molecular Cell Biology*, pp. 769–776. doi: 10.1038/35096075.

Covey, T. R., Thomson, B. A. and Schneider, B. B. (2009) 'Atmospheric pressure ion sources',

Mass Spectrometry Reviews, 28(6), pp. 870–897. doi: 10.1002/mas.20246.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. (2011) 'Andromeda: A peptide search engine integrated into the MaxQuant environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N. and Mann, M. (2014) 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.', *Molecular & cellular proteomics : MCP*, 13(9), pp. 2513–26. doi: 10.1074/mcp.M113.031591.

Craig, R. and Beavis, R. C. (2004) 'TANDEM: matching proteins with tandem mass spectra', *BIOINFORMATICS APPLICATIONS NOTE*, 20(9), pp. 1466–1467. doi: 10.1093/bioinformatics/bth092.

Cross, D. A. E., Alessi, D. R., Cohen, P., Andjelkovich, M. and Hemmings, B. A. (1995) 'Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B', *Nature*, 378(6559), pp. 785–789. doi: 10.1038/378785a0.

Cuesta, A., Pedrola, L., Sevilla, T., García-Planells, J., Chumillas, M. J., Mayordomo, F., LeGuern, E., Marín, I., Vílchez, J. J. and Palau, F. (2002) 'The gene encoding ganglioside-induced differentiation-associated protein 1 is mutated in axonal Charcot-Marie-Tooth type 4A disease', *Nature Genetics*, 30(1), pp. 22–25. doi: 10.1038/ng798.

Dator, R. P., Gaston, K. W. and Limbach, P. A. (2014) 'Multiple enzymatic digestions and ion mobility separation improve quantification of bacterial ribosomal proteins by data independent acquisition liquid chromatography-mass spectrometry', *Analytical Chemistry*, 86(9), pp. 4264–4270. doi: 10.1021/ac404020j.

Datta, S. R., Dudek, H., Xu, T., Masters, S., Haian, F., Gotoh, Y. and Greenberg, M. E. (1997) 'Akt phosphorylation of BAD couples survival signals to the cell- intrinsic death machinery', *Cell*, 91(2), pp. 231–241. doi: 10.1016/S0092-8674(00)80405-5.

Davis, D. B., Doherty, K. R., Delmonte, A. J. and McNally, E. M. (2002) 'Calcium-sensitive phospholipid binding properties of normal and mutant ferlin C2 domains.', *The Journal of biological chemistry*, 277(25), pp. 22883–8. doi: 10.1074/jbc.M201858200.

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. and Sanchez, J. C. (2008) 'Relative quantification of proteins in human cerebrospinal fluids by MS/MS

using 6-plex isobaric tags', *Analytical Chemistry*, 80(8), pp. 2921–2931. doi: 10.1021/ac702422x.

deBlaquiere, J. and Burgess, A. W. (1999) 'Affinity purification of plasma membranes', *Journal of Biomolecular Techniques*, 10(2), pp. 64–71.

Deblaquiere, J. and Burgess, A. W. (1999) 'Affinity Purification of Plasma Membranes', *Journal of Biomolecular Techniques*, 10(2), pp. 64–71.

Deborde, S., Perret, E., Gravotta, D., Deora, A., Salvarezza, S., Schreiner, R. and Rodriguez-Boulan, E. (2008) 'Clathrin is a key regulator of basolateral polarity.', *Nature*, 452(7188), pp. 719–23. doi: 10.1038/nature06828.

Deeb, S. J., Cox, J., Schmidt-Supprian, M. and Mann, M. (2014) 'N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes.', *Molecular & cellular proteomics : MCP*, 13(1), pp. 240–51. doi: 10.1074/mcp.M113.033977.

Doherty, G. J. and McMahon, H. T. (2009) 'Mechanisms of endocytosis.', *Annual review of biochemistry*, 78, pp. 857–902. doi: 10.1146/annurev.biochem.78.081307.110540.

Dormeyer, W., Van Hoof, D., Braam, S. R., Heck, A. J. R., Mummery, C. L. and Krijgsveld, J. (2008) 'Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells', *Journal of Proteome Research*, 7(7), pp. 2936–2951. doi: 10.1021/pr800056j.

Dunlop, E. A., Dodd, K. M., Seymour, L. A. and Tee, A. R. (2009) 'Mammalian target of rapamycin complex 1-mediated phosphorylation of eukaryotic initiation factor 4E-binding protein 1 requires multiple protein-protein interactions for substrate recognition', *Cellular Signalling*, 21(7), pp. 1073–1084. doi: 10.1016/j.cellsig.2009.02.024.

Duve, C. (1975) 'Exploring cells with a centrifuge', *Science*, 189(4198), pp. 186–194. doi: 10.1126/science.1138375.

Eldar-Finkelman, H., Seger, R., Vandenheede, J. R. and Krebs, E. G. (1995) 'Inactivation of glycogen synthase kinase-3 by epidermal growth factor is mediated by mitogen-activated protein kinase/p90 ribosomal protein S6 kinase signaling pathway in NIH/3T3 cells', *Journal of Biological Chemistry*, pp. 987–990. doi: 10.1074/jbc.270.3.987.

Elkin, S. R., Oswald, N. W., Reed, D. K., Mettlen, M., MacMillan, J. B. and Schmid, S. L. (2016) 'Ikarugamycin: A Natural Product Inhibitor of Clathrin-Mediated Endocytosis', *Traffic*, 17(10), pp. 1139–1149. doi: 10.1111/tra.12425.

Endo, H., Nito, C., Kamada, H., Nishi, T. and Chan, P. H. (2006) 'Activation of the Akt/GSK3beta

signaling pathway mediates survival of vulnerable hippocampal neurons after transient global cerebral ischemia in rats.', *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 26(12), pp. 1479–89. doi: 10.1038/sj.jcbfm.9600303.

Evgrafov, O. V., Mersiyanova, I., Irobi, J., Van Den Bosch, L. V., Dierick, I., Leung, C. L., Schagina, O., Verpoorten, N., Van Impe, K., Fedotov, V., … Timmerman, V. (2004) 'Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy', *Nature Genetics*, 36(6), pp. 602–606. doi: 10.1038/ng1354.

Eyster, C. A., Higginson, J. D., Huebner, R., Porat-Shliom, N., Weigert, R., Wu, W. W., Shen, R. F. and Donaldson, J. G. (2009) 'Discovery of new cargo proteins that enter cells through clathrin-independent endocytosis', *Traffic*, 10(5), pp. 590–599. doi: 10.1111/j.1600-0854.2009.00894.x.

Feist, P. and Hummon, A. B. (2015) 'Proteomic challenges: Sample preparation techniques for Microgram-Quantity protein analysis from biological samples', *International Journal of Molecular Sciences*, 16(2), pp. 3537–3563. doi: 10.3390/ijms16023537.

Feng, S.-M., Muraoka-Cook, R. S., Hunter, D., Sandahl, M. A., Caskey, L. S., Miyazawa, K., Atfi, A. and Earp, H. S. (2009) 'The E3 Ubiquitin Ligase WWP1 Selectively Targets HER4 and Its Proteolytically Derived Signaling Isoforms for Degradation', *Molecular and Cellular Biology*, 29(3), pp. 892–906. doi: 10.1128/mcb.00595-08.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989) 'Electrospray ionization for mass spectrometry of large biomolecules.', *Science (New York, N.Y.)*, 246(4926), pp. 64–71.

Ford, M. G. J., Pearse, B. M. F., Higgins, M. K., Vallis, Y., Owen, D. J., Gibson, A., Hopkins, C. R., Evans, P. R. and McMahon, H. T. (2001) 'Simultaneous binding of PtdIns (4,5) P 2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes', *Science*, 291(5506), pp. 1051–1055. doi: 10.1126/science.291.5506.1051.

Frame, S., Cohen, P. and Biondi, R. M. (2001) 'A common phosphate binding site explains the unique substrate specificity of GSK3 and its inactivation by phosphorylation', *Molecular Cell*, 7(6), pp. 1321–1327. doi: 10.1016/S1097-2765(01)00253-2.

Francis, F., Koulakoff, A., Boucher, D., Chafey, P., Schaar, B., Vinet, M. C., Friocourt, G., McDonnell, N., Reiner, O., Kahn, A., McConnell, S. K., Berwald-Netter, Y., Denoulet, P. and

Chelly, J. (1999) ‘Doublecortin is a developmentally regulated, microtubule-associated protein expressed in migrating and differentiating neurons’, *Neuron*, 23(2), pp. 247–256. doi: 10.1016/S0896-6273(00)80777-1.

Fu, X., Yang, Y., Xu, C., Niu, Y., Chen, T., Zhou, Q. and Liu, J.-J. (2011) ‘Retrolinkin cooperates with endophilin A1 to mediate BDNF-TrkB early endocytic trafficking and signaling from early endosomes’, *Molecular Biology of the Cell*, 22(19), pp. 3684–3698. doi: 10.1091/mbc.e11-04-0308.

Fukuyama, H., Ndiaye, S., Hoffmann, J., Rossier, J., Liuu, S., Vinh, J. and Verdier, Y. (2012) ‘On-bead tryptic proteolysis: An attractive procedure for LC-MS/MS analysis of the Drosophila caspase 8 protein complex during immune response against bacteria’, *Journal of Proteomics*. doi: 10.1016/j.jprot.2012.03.003.

Gaidarov, I., Santini, F., Warren, R. A. and Keen, J. H. (1999) ‘Spatial control of coated-pit dynamics in living cells.’, *Nature cell biology*, 1(1), pp. 1–7. doi: 10.1038/8971.

Gallop, J. L., Jao, C. C., Kent, H. M., Butler, P. J. G., Evans, P. R., Langen, R. and McMahon, H. T. (2006) ‘Mechanism of endophilin N-BAR domain-mediated membrane curvature’, *The EMBO Journal*, 25(12), pp. 2898–2910. doi: 10.1038/sj.emboj.7601174.

Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. and Bryant, S. H. (2004) ‘Open mass spectrometry search algorithm’, *Journal of Proteome Research*, 3(5), pp. 958–964. doi: 10.1021/pr0499491.

Genet, G., Boyé, K., Mathivet, T., Ola, R., Zhang, F., Dubrac, A., Li, J., Genet, N., Henrique Geraldo, L., Benedetti, L., Künzel, S., Pibouin-Fragner, L., Thomas, J.-L. and Eichmann, A. (2019) ‘Endophilin-A2 dependent VEGFR2 endocytosis promotes sprouting angiogenesis’, *Nature Communications*, 10(1), p. 2350. doi: 10.1038/s41467-019-10359-x.

Geromanos, S. J., Vissers, J. P. C., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V, Bateman, R. H. and Langridge, J. I. (2009) ‘The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS.’, *Proteomics*, 9(6), pp. 1683–95. doi: 10.1002/pmic.200800562.

Geromanos, S. J., Hughes, C., Ciavarini, S., Vissers, J. P. C. and Langridge, J. I. (2012) ‘Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples’, *Analytical and Bioanalytical Chemistry*, 404(4), pp. 1127–1139. doi: 10.1007/s00216-

012-6197-y.

Gess, B., Auer-Grumbach, M., Schirmacher, A., Strom, T., Zitzelsberger, M., Rudnik-Schöneborn, S., Röhr, D., Halfter, H., Young, P. and Senderek, J. (2014) 'HSJ1-related hereditary neuropathies Novel mutations and extended clinical spectrum', *Neurology*, 83(19), pp. 1726–1732. doi: 10.1212/WNL.0000000000000966.

González-Sánchez, P., Pla-Martín, D., Martínez-Valero, P., Rueda, C. B., Calpena, E., Del Arco, A., Palau, F. and Satrústegui, J. (2017) 'CMT-linked loss-of-function mutations in GDAP1 impair store-operated Ca 2+ entry-stimulated respiration', *Scientific Reports*, 7. doi: 10.1038/srep42993.

Gordon, C. P., Venn-Brown, B., Robertson, M. J., Young, K. A., Chau, N., Mariana, A., Whiting, A., Chircop, M., Robinson, P. J. and McCluskey, A. (2013) 'Development of second-generation indole-based dynamin GTPase inhibitors', *Journal of Medicinal Chemistry*, 56(1), pp. 46–59. doi: 10.1021/jm300844m.

Green, N. M. (1975) 'Avidin', *Advances in Protein Chemistry*, 29, pp. 85–133. doi: 10.1016/S0065-3233(08)60411-8.

Grewal, R. P., Tayag, E., Figueroa, K. P., Zu, L., Durazo, A., Nunez, C. and Pulst, S. M. (1998) 'Clinical and genetic analysis of a distinct autosomal dominant spinocerebellar ataxia', *Neurology*, 51(5), pp. 1423–1426. doi: 10.1212/WNL.51.5.1423.

Guichet, A., Wucherpfennig, T., Dudu, V., Etter, S., Wilsch-Bräuniger, M., Hellwig, A., González-Gaitán, M., Huttner, W. B. and Schmidt, A. A. (2002) 'Essential role of endophilin A in synaptic vesicle budding at the Drosophila neuromuscular junction.', *The EMBO journal*, 21(7), pp. 1661–72. doi: 10.1093/emboj/21.7.1661.

Hanada, M., Feng, J. and Hemmings, B. A. (2004) 'Structure, regulation and function of PKB/AKT - A major therapeutic target', in *Biochimica et Biophysica Acta - Proteins and Proteomics*, pp. 3–16. doi: 10.1016/j.bbapap.2003.11.009.

Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011) 'Molecular chaperones in protein folding and proteostasis', *Nature*, pp. 324–332. doi: 10.1038/nature10317.

Hasan, M., Min, H., Rahaman, K. A., Muresan, A. R., Kim, H., Han, D. and Kwon, O. S. (2019) 'Quantitative Proteome Analysis of Brain Subregions and Spinal Cord from Experimental Autoimmune Encephalomyelitis Mice by TMT-Based Mass Spectrometry', *Proteomics*, 19(5), p. 1800355. doi: 10.1002/pmic.201800355.

Heir, R., Ablasou, C., Dumontier, E., Elliott, M., Fagotto-Kaufmann, C. and Bedford, F. K. (2006) 'The UBL domain of PLIC-1 regulates aggresome formation', *EMBO Reports*, 7(12), pp. 1252–1258. doi: 10.1038/sj.embor.7400823.

Hill, T. A., Gordon, C. P., McGeachie, A. B., Venn-Brown, B., Odell, L. R., Chau, N., Quan, A., Mariana, A., Sakoff, J. A., Chircop, M., Robinson, P. J. and McCluskey, A. (2009) 'Inhibition of dynamin mediated endocytosis by the Dynoles - Synthesis and functional activity of a family of indoles', *Journal of Medicinal Chemistry*, 52(12), pp. 3762–3773. doi: 10.1021/jm900036m.

Hinz, L., Ahles, A., Ruprecht, B., Kü Ster, B. and Engelhardt, S. (2017) 'Two serines in the distal C-terminus of the human ß 1-adrenoceptor determine ß-arrestin2 recruitment'. doi: 10.1371/journal.pone.0176450.

Hjelmeland, L. M. (1980) *A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis.*, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.77.11.6368.

Hornburg, D., Drepper, C., Butter, F., Meissner, F., Sendtner, M. and Mann, M. (2014) 'Deep proteomic evaluation of primary and cell line motoneuron disease models delineates major differences in neuronal characteristics', *Molecular and Cellular Proteomics*, 13(12), pp. 3410–3420. doi: 10.1074/mcp.M113.037291.

Howarth, J. L., Kelly, S., Keasey, M. P., Glover, C. P. J., Lee, Y. B., Mitrophanous, K., Chapple, J. P., Gallo, J. M., Cheetham, M. E. and Uney, J. B. (2007) 'Hsp40 molecules that target to the ubiquitin-proteasome system decrease inclusion formation in models of polyglutamine disease', *Molecular Therapy*, 15(6), pp. 1100–1105. doi: 10.1038/sj.mt.6300163.

Howes, M. T., Mayor, S. and Parton, R. G. (2010) 'Molecules, mechanisms, and cellular roles of clathrin-independent endocytosis', *Current Opinion in Cell Biology*, 22(4), pp. 519–527. doi: 10.1016/j.ceb.2010.04.001.

Huang, D. W., Sherman, B. T., Lempicki, R. A. and Huang DW, Sherman BT, L. R. A. (2009) 'DAVID Functional Annotation Bioinformatics Microarray Analysis', *Nature Protocols*, 4(1), pp. 44–57. doi: 4(1):44-57.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009) 'Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists', *Nucleic Acids Research*, 37(1), pp. 1–13. doi: 10.1093/nar/gkn923.

Huber, N., Bieniossek, C., Wagner, K. M., Elsässer, H. P., Suter, U., Berger, I. and Niemann, A. (2016) 'Glutathione-conjugating and membrane-remodeling activity of GDAP1 relies on amphipathic C-terminal domain', *Scientific Reports*, 6. doi: 10.1038/srep36930.

Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A. and Mann, M. (2010) 'Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions', *Journal of Cell Biology*, 189(4), pp. 739–754. doi: 10.1083/jcb.200911091.

Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M. and Krijgsveld, J. (2014) 'Ultrasensitive proteome analysis using paramagnetic bead technology', *Molecular Systems Biology*, 10(10). doi: 10.15252/msb.20145625.

Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E. and Engelhard, V. H. (1992) 'Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry.', *Science (New York, N.Y.)*, 255(5049), pp. 1261–3.

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M. P., Parzen, H., … Wade Harper, J. (2017) 'Architecture of the human interactome defines protein communities and disease networks', *Nature*, 545(7655), pp. 505–509. doi: 10.1038/nature22366.

Iavarone, A. T. and Williams, E. R. (2003) 'Collisionally activated dissociation of supercharged proteins formed by electrospray ionization.', *Analytical chemistry*, 75(17), pp. 4525–33.

Iribarne, J. V. (1976) 'On the evaporation of small ions from charged droplets', *The Journal of Chemical Physics*, 64(6), p. 2287. doi: 10.1063/1.432536.

Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., Michalik, A., De Vriendt, E., Jacobs, A., Van Gerwen, V., … Timmerman, V. (2004) 'Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy', *Nature Genetics*, 36(6), pp. 597–601. doi: 10.1038/ng1328.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005) 'Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein', *Molecular and Cellular Proteomics*, 4(9), pp. 1265–1272. doi: 10.1074/mcp.M500061-MCP200.

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M. L., Keenan, R. J. and Hegde, R. S. (2016) 'Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation', *Molecular*

Cell, 63(1), pp. 21–33. doi: 10.1016/j.molcel.2016.05.020.

Jiang, A., Lehti, K., Wang, X., Weiss, S. J., Keski-Oja, J. and Pei, D. (2001) 'Regulation of membrane-type matrix metalloproteinase 1 activity by dynamin-mediated endocytosis', *Proceedings of the National Academy of Sciences of the United States of America*, 98(24), pp. 13693–13698. doi: 10.1073/pnas.241293698.

Jiang, H., Guo, W., Liang, X. and Rao, Y. (2005) 'Both the establishment and the maintenance of neuronal polarity require active mechanisms: Critical roles of GSK-3β and its upstream regulators', *Cell*, 120(1), pp. 123–135. doi: 10.1016/j.cell.2004.12.033.

Johnsson, N. and Varshavsky, A. (1994) *Split ubiquitin as a sensor of protein interactions in vivo*, *Proc. Nati. Acad. Sci. USA*.

Kaksonen, M. and Roux, A. (2018) 'Mechanisms of clathrin-mediated endocytosis', *Nature Reviews Molecular Cell Biology*, pp. 313–326. doi: 10.1038/nrm.2017.132.

Kalmar, B., Innes, A., Wanisch, K., Kolaszynska, A. K., Pandraud, A., Kelly, G., Abramov, A. Y., Reilly, M. M., Schiavo, G. and Greensmith, L. (2017) 'Mitochondrial deficits and abnormal mitochondrial retrograde axonal transport play a role in the pathogenesis of mutant Hsp27-induced Charcot Marie Tooth Disease', *Human Molecular Genetics*, 26(17), pp. 3313–3326. doi: 10.1093/hmg/ddx216.

Kaltashov, I. A. and Mohimen, A. (2005) 'Estimates of protein surface areas in solution by electrospray ionization mass spectrometry.', *Analytical chemistry*, 77(16), pp. 5370–9. doi: 10.1021/ac050511+.

Kaneko, T., Maeda, A., Takefuji, M., Aoyama, H., Nakayama, M., Kawabata, S., Kawano, Y., Iwamatsu, A., Amano, M. and Kaibuchi, K. (2005) 'Rho mediates endocytosis of epidermal growth factor receptor through phosphorylation of endophilin A1 by Rho-kinase', *Genes to Cells*, 10(10), pp. 973–987. doi: 10.1111/j.1365-2443.2005.00895.x.

Kaneko, T. (2008) 'The SH3 domain- a family of versatile peptide- and protein-recognition module', *Frontiers in Bioscience*, Volume(13), p. 4938. doi: 10.2741/3053.

Kanellopoulos, A. H., Koenig, J., Huang, H., Pyrski, M., Millet, Q., Lolignier, S., Morohashi, T., Gossage, S. J., Jay, M., Linley, J. E., Baskozos, G., Kessler, B. M., Cox, J. J., Dolphin, A. C., Zufall, F., Wood, J. N. and Zhao, J. (2018) ' Mapping protein interactions of sodium channel Na V 1.7 using epitope-tagged gene-targeted mice ', *The EMBO Journal*, 37(3), pp. 427–445. doi:

10.15252/embj.201796692.

Karas, M. and Hillenkamp, F. (1988) ‘Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons’, *Analytical Chemistry*, 60(20), pp. 2299–2301. doi: 10.1021/ac00171a028.

Keren, B., Jacquette, A., Depienne, C., Leite, P., Durr, A., Carpentier, W., Benyahia, B., Ponsot, G., Soubrier, F., Brice, A. and Héron, D. (2010) ‘Evidence against haploinsuffiency of human ataxin 10 as a cause of spinocerebellar ataxia type 10’, *Neurogenetics*, pp. 273–274. doi: 10.1007/s10048-009-0227-8.

Kim, D. I., Birendra, K. C., Zhu, W., Motamedchaboki, K., Doye, V. and Roux, K. J. (2014) ‘Probing nuclear pore complex architecture with proximity-dependent biotinylation’, *Proceedings of the National Academy of Sciences of the United States of America*, 111(24). doi: 10.1073/pnas.1406459111.

Kim, J. Y., Woo, S. Y., Hong, Y. Bin, Choi, Heesun, Kim, J., Choi, Hyunjung, Mook-Jung, I., Ha, N., Kyung, J., Koo, S. K., Jung, S. C. and Choi, B. O. (2016) ‘HDAC6 Inhibitors Rescued the Defective Axonal Mitochondrial Movement in Motor Neurons Derived from the Induced Pluripotent Stem Cells of Peripheral Neuropathy Patients with HSPB1 Mutation’, *Stem Cells International*, 2016, pp. 1–14. doi: 10.1155/2016/9475981.

Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., … Pandey, A. (2014) ‘A draft map of the human proteome’, *Nature*, 509(7502), pp. 575–581. doi: 10.1038/nature13302.

Kleijnen, M. F., Shih, A. H., Zhou, P., Kumar, S., Soccio, R. E., Kedersha, N. L., Gill, G. and Howley, P. M. (2000) ‘The hPLIC proteins may provide a link between the ubiquitination machinery and the proteasome.’, *Molecular cell*, 6(2), pp. 409–19. doi: 10.1016/s1097-2765(00)00040-x.

Konermann, L., Rodriguez, A. D. and Liu, J. (2012) ‘On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization’, *Analytical Chemistry*, 84(15), pp. 6798–6804. doi: 10.1021/ac301298g.

Konermann, L., Ahadi, E., Rodriguez, A. D. and Vahidi, S. (2013) ‘Unraveling the mechanism of electrospray ionization.’, *Analytical chemistry*, 85(1), pp. 2–9. doi: 10.1021/ac302789c.

Kononenko, N. L., Puchkov, D., Classen, G. A., Walter, A. M., Pechstein, A., Sawade, L.,

Kaempf, N., Trimbuch, T., Lorenz, D., Rosenmund, C., Maritzen, T. and Haucke, V. (2014) 'Clathrin/AP-2 Mediate Synaptic Vesicle Reformation from Endosome-like Vacuoles but Are Not Essential for Membrane Retrieval at Central Synapses', *Neuron*, 82(5), pp. 981–988. doi: 10.1016/j.neuron.2014.05.007.

Konzer, A., Ruhs, A., Braun, H., Jungblut, B., Braun, T. and Krüger, M. (2013) 'Stable isotope labeling in zebrafish allows in vivo monitoring of cardiac morphogenesis', *Molecular and Cellular Proteomics*, 12(6), pp. 1502–1512. doi: 10.1074/mcp.M111.015594.

Kosaka, T. and Ikeda, K. (1983) 'Reversible blockage of membrane retrieval and endocytosis in the garland cell of the temperature-sensitive mutant of Drosophila melanogaster, shibirets1.', *The Journal of Cell Biology*, 97(2), pp. 499–507. doi: 10.1083/jcb.97.2.499.

Kristensen, A. R., Gsponer, J. and Foster, L. J. (2012) 'A high-throughput approach for measuring temporal changes in the interactome', *Nature Methods*, 9(9), pp. 907–909. doi: 10.1038/nmeth.2131.

Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. and Mann, M. (2014) 'Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells.', *Nature methods*, 11(3), pp. 319–24. doi: 10.1038/nmeth.2834.

Kurochkina, N. and Guha, U. (2013) 'SH3 domains: Modules of protein-protein interactions', *Biophysical Reviews*, 5(1), pp. 29–39. doi: 10.1007/s12551-012-0081-z.

Labbadia, J., Novoselov, S. S., Bett, J. S., Weiss, A., Paganetti, P., Bates, G. P. and Cheetham, M. E. (2012) 'Suppression of protein aggregation by chaperone modification of high molecular weight complexes', *Brain*, 135(4), pp. 1180–1186. doi: 10.1093/brain/aws022.

Lai, X. (2013) 'Reproducible method to enrich membrane proteins with high purity and high yield for an LC-MS/MS approach in quantitative membrane proteomics', *Electrophoresis*, 34(6), pp. 809–17. doi: 10.1002/elps.201200503.

Laplante, M. and Sabatini, D. M. (2012) 'MTOR signaling in growth control and disease', *Cell*, pp. 274–293. doi: 10.1016/j.cell.2012.03.017.

Larance, M., Bailly, A. P., Pourkarimi, E., Hay, R. T., Buchanan, G., Coulthurst, S., Xirodimas, D. P., Gartner, A. and Lamond, A. I. (2011) 'Stable-isotope labeling with amino acids in nematodes', *Nature Methods*, 8(10), pp. 849–851. doi: 10.1038/nmeth.1679.

Lau, K.-F., Miller, C. C. J., Anderton, B. H. and Shaw, P.-C. (1999) 'Expression analysis of

glycogen synthase kinase-3 in human tissues', *Journal of Peptide Research*, 54(1), pp. 85–91. doi: 10.1034/j.1399-3011.1999.00083.x.

Lewandrowski, U., Wortelkamp, S., Lohrig, K., Zahedi, R. P., Wolters, D. A., Walter, U. and Sickmann, A. (2009) 'Platelet membrane proteomics: A novel repository for functional research', *Blood*, 114(1). doi: 10.1182/blood-2009-02-203828.

Li, G. Z., Vissers, J. P. C., Silva, J. C., Golick, D., Gorenstein, M. V. and Geromanos, S. J. (2009) 'Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures', *Proteomics*, 9(6), pp. 1696–1719. doi: 10.1002/pmic.200800564.

Li, Y., Zhou, Z., Alimandi, M. and Chen, C. (2009) 'WW domain containing E3 ubiquitin protein ligase 1 targets the full-length ErbB4 for ubiquitin-mediated degradation in breast cancer', *Oncogene*, 28(33), pp. 2948–2958. doi: 10.1038/onc.2009.162.

Liang, S., Wei, F. Y., Wu, Y. M., Tanabe, K., Abe, T., Oda, Y., Yoshida, Y., Yamada, H., Matsui, H., Tomizawa, K. and Takei, K. (2007) 'Major Cdk5-dependent phosphorylation sites of amphiphysin 1 are implicated in the regulation of the membrane binding and endocytosis', *Journal of Neurochemistry*, 102(5), pp. 1466–1476. doi: 10.1111/j.1471-4159.2007.04507.x.

Ling, S. C., Polymenidou, M. and Cleveland, D. W. (2013) 'Converging mechanisms in als and FTD: Disrupted RNA and protein homeostasis', *Neuron*, pp. 416–438. doi: 10.1016/j.neuron.2013.07.033.

Liu, H., Sadygov, R. G. and Yates, J. R. (2004) 'A model for random sampling and estimation of relative protein abundance in shotgun proteomics', *Analytical Chemistry*, 76(14), pp. 4193–4201. doi: 10.1021/ac0498563.

Luo, Y., McDonald, K. and Hanrahan, J. W. (2008) 'Trafficking of immature ΔF508-CFTR to the plasma membrane and its detection by biotinylation', *Biochemical Journal*, 419(1), pp. 211–221. doi: 10.1042/bj20081869.

Macgregor, K. A., Abdel-Hamid, M. K., Odell, L. R., Chau, N., Whiting, A., Robinson, P. J. and McCluskey, A. (2014) 'Development of quinone analogues as dynamin GTPase inhibitors', *European Journal of Medicinal Chemistry*, 85, pp. 191–206. doi: 10.1016/j.ejmech.2014.06.070.

MacHacek, M., Hodgson, L., Welch, C., Elliott, H., Pertz, O., Nalbant, P., Abell, A., Johnson, G. L., Hahn, K. M. and Danuser, G. (2009) 'Coordination of Rho GTPase activities during cell

protrusion', *Nature*, 461(7260), pp. 99–103. doi: 10.1038/nature08242.

Macia, E., Ehrlich, M., Massol, R., Boucrot, E., Brunner, C. and Kirchhausen, T. (2006) 'Dynasore, a Cell-Permeable Inhibitor of Dynamin', *Developmental Cell*, 10(6), pp. 839–850. doi: 10.1016/j.devcel.2006.04.002.

Malumbres, M. (2014) 'Cyclin-dependent kinases', *Genome Biology*, 15(6), p. 122. doi: 10.1186/gb4184.

Manning, B. D. and Cantley, L. C. (2007) 'AKT/PKB Signaling: Navigating Downstream', *Cell*, pp. 1261–1274. doi: 10.1016/j.cell.2007.06.009.

März, P., Probst, A., Lang, S., Schwager, M., Rose-John, S., Otten, U. and Özbek, S. (2004) 'Ataxin-10, the spinocerebellar ataxia type 10 neurodegenerative disorder protein, is essential for survival of cerebellar neurons', *Journal of Biological Chemistry*, 279(34), pp. 35542–35550. doi: 10.1074/jbc.M405865200.

Masuda, M., Takeda, S., Sone, M., Ohki, T., Mori, H., Kamioka, Y. and Mochizuki, N. (2006) 'Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms', *The EMBO Journal*, 25(12), pp. 2889–2897. doi: 10.1038/sj.emboj.7601176.

McAlister, G. C., Nusinow, D. P., Jedrychowski, M. P., Wühr, M., Huttlin, E. L., Erickson, B. K., Rad, R., Haas, W. and Gygi, S. P. (2014) 'MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes', *Analytical Chemistry*, 86(14), pp. 7150–7158. doi: 10.1021/ac502040v.

Mccluskey, A., Daniel, J. A., Hadzic, G., Chau, N., Clayton, E. L., Mariana, A., Whiting, A., Gorgani, N. N., Lloyd, J., Quan, A., … Robinson, P. J. (2013) 'Building a better dynasore: The dyngo compounds potently inhibit dynamin and endocytosis', *Traffic*, 14(12), pp. 1272–1289. doi: 10.1111/tra.12119.

Meinecke, M., Boucrot, E., Camdere, G., Hon, W.-C., Mittal, R. and McMahon, H. T. (2013) 'Cooperative recruitment of dynamin and BIN/amphiphysin/Rvs (BAR) domain-containing proteins leads to GTP-dependent membrane scission', *The Journal of biological chemistry*, 288(9), pp. 6651–61. doi: 10.1074/jbc.M112.444869.

Milosevic, I., Giovedi, S., Lou, X., Raimondi, A., Collesi, C., Shen, H., Paradise, S., O'Toole, E., Ferguson, S., Cremona, O. and De Camilli, P. (2011) 'Recruitment of endophilin to clathrin-coated pit necks is required for efficient vesicle uncoating after fission.', *Neuron*, 72(4), pp. 587–

601. doi: 10.1016/j.neuron.2011.08.029.

Motley, A., Bright, N. A., Seaman, M. N. J. and Robinson, M. S. (2003) 'Clathrin-mediated endocytosis in AP-2-depleted cells', *Journal of Cell Biology*, 162(5), pp. 909–918. doi: 10.1083/jcb.200305145.

Müller, J. B., Geyer, P. E., Colaço, A. R., Treit, P. V., Strauss, M. T., Oroshi, M., Doll, S., Virreira Winter, S., Bader, J. M., Köhler, N., Theis, F., Santos, A. and Mann, M. (2020) 'The proteome landscape of the kingdoms of life', *Nature*, 582(7813), pp. 592–596. doi: 10.1038/s41586-020-2402-x.

Nalbant, P., Hodgson, L., Kraynov, V., Toutchkine, A. and Hahn, K. M. (2004) 'Activation of endogenous Cdc42 visualized in living cells', *Science*, 305(5690), pp. 1615–1619. doi: 10.1126/science.1100367.

Nazio, F., Strappazzon, F., Antonioli, M., Bielli, P., Cianfanelli, V., Bordi, M., Gretzmeier, C., Dengjel, J., Piacentini, M., Fimia, G. M. and Cecconi, F. (2013) 'MTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6', *Nature Cell Biology*, 15(4), pp. 406–416. doi: 10.1038/ncb2708.

Neymeyer, V., Tephly, T. R. and Miller, M. W. (1997) 'Folate and 10-formyltetrahydrofolate dehydrogenase (FDH) expression in the central nervous system of the mature rat.', *Brain research*, 766(1–2), pp. 195–204. doi: 10.1016/s0006-8993(97)00528-3.

Niemann, A., Wagner, K. M., Ruegg, M. and Suter, U. (2009) 'GDAP1 mutations differ in their effects on mitochondrial dynamics and apoptosis depending on the mode of inheritance', *Neurobiology of Disease*, 36(3), pp. 509–520. doi: 10.1016/j.nbd.2009.09.011.

Niemann, A., Huber, N., Wagner, K. M., Somandin, C., Horn, M., Lebrun-Julien, F., Angst, B., Pereira, J. A., Halfter, H., Welzl, H., Feltri, M. L., Wrabetz, L., Young, P., Wessig, C., Toyka, K. V. and Suter, U. (2014) 'The Gdap1 knockout mouse mechanistically links redox control to Charcot–Marie–Tooth disease', *Brain*, 137(3), pp. 668–682. doi: 10.1093/brain/awt371.

Noack, R., Frede, S., Albrecht, P., Henke, N., Pfeiffer, A., Knoll, K., Dehmel, T., Meyer zu Hörste, G., Stettner, M., Kieseier, B. C., Summer, H., Golz, S., Kochanski, A., Wiedau-Pazos, M., Arnold, S., Lewerenz, J. and Methner, A. (2012) 'Charcot–Marie–Tooth disease CMT4A: GDAP1 increases cellular glutathione and the mitochondrial membrane potential', *Human Molecular Genetics*, 21(1), pp. 150–162. doi: 10.1093/hmg/ddr450.

Noda, T. and Ohsumi, Y. (1998) ‘Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast’, *Journal of Biological Chemistry*, 273(7), pp. 3963–3966. doi: 10.1074/jbc.273.7.3963.

Nováková, L., Matysová, L. and Solich, P. (2006) ‘Advantages of application of UPLC in pharmaceutical analysis’, *Talanta*, 68(3), pp. 908–918. doi: 10.1016/J.TALANTA.2005.06.035.

Novoselov, S. S., Mustill, W. J., Gray, A. L., Dick, J. R., Kanuga, N., Kalmar, B., Greensmith, L. and Cheetham, M. E. (2013) ‘Molecular Chaperone Mediated Late-Stage Neuroprotection in the SOD1G93A Mouse Model of Amyotrophic Lateral Sclerosis’, *PLoS ONE*. Edited by H. H. Kampinga, 8(8), p. e73944. doi: 10.1371/journal.pone.0073944.

Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A. and Ahn, N. G. (2005) ‘Comparison of label-free methods for quantifying human proteins by shotgun proteomics’, *Molecular and Cellular Proteomics*, 4(10), pp. 1487–1502. doi: 10.1074/mcp.M500084-MCP200.

Olzmann, J. A. and Carvalho, P. (2019) ‘Dynamics and functions of lipid droplets’, *Nature Reviews Molecular Cell Biology*, pp. 137–155. doi: 10.1038/s41580-018-0085-z.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002) ‘Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.’, *Molecular & cellular proteomics : MCP*, 1(5), pp. 376–86.

Papadopoulos, C., Orso, G., Mancuso, G., Herholz, M., Gumeni, S., Tadepalle, N., Jüngst, C., Tzschichholz, A., Schauss, A., Höning, S., Trifunovic, A., Daga, A. and Rugarli, E. I. (2015) ‘Spastin Binds to Lipid Droplets and Affects Lipid Metabolism’, *PLOS Genetics*, 11(4), p. e1005149. doi: 10.1371/journal.pgen.1005149.

Pappireddi, N., Martin, L. and Wühr, M. (2019) ‘A Review on Quantitative Multiplexed Proteomics’, *ChemBioChem*, 20(10), pp. 1210–1224. doi: 10.1002/cbic.201800650.

Park, M., Reddy, G. R., Wallukat, G., Xiang, Y. K. and Steinberg, S. F. (2017) ‘β1-adrenergic receptor O-glycosylation regulates N-terminal cleavage and signaling responses in cardiomyocytes’, *Scientific Reports*, 7(1), p. 7890. doi: 10.1038/s41598-017-06607-z.

Pasterkamp, R. J. and Giger, R. J. (2009) ‘Semaphorin function in neural plasticity and disease’, *Current Opinion in Neurobiology*, pp. 263–274. doi: 10.1016/j.conb.2009.06.001.

Patel, P. and Woodgett, J. R. (2017) ‘Glycogen Synthase Kinase 3: A Kinase for All Pathways?’,

in *Current Topics in Developmental Biology*, pp. 277–302. doi: 10.1016/bs.ctdb.2016.11.011.

Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. and Scrivens, J. H. (2009) 'A comparison of labeling and label-free mass spectrometry-based proteomics approaches', *Journal of Proteome Research*, 8(7), pp. 3752–3759. doi: 10.1021/pr900080y.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999) 'Probability-based protein identification by searching sequence databases using mass spectrometry data', *Electrophoresis*, 20(18), pp. 3551–3567. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Petrelli, A., Gilestro, G. F., Lanzardo, S., Comoglio, P. M., Migone, N. and Giordano, S. (2002) 'The endophilin–CIN85–Cbl complex mediates ligand-dependent downregulation of c-Met', *Nature*, 416(6877), pp. 187–190. doi: 10.1038/416187a.

Petschnigg, J., Groisman, B., Kotlyar, M., Taipale, M., Zheng, Y., Kurat, C. F., Sayad, A., Sierra, J. R., Usaj, M. M., Snider, J., Nachman, A., Krykbaeva, I., Tsao, M.-S., Moffat, J., Pawson, T., Lindquist, S., Jurisica, I. and Stagljar, I. (2014) 'The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells', *Nature Methods*, 11(5), pp. 585–592. doi: 10.1038/nmeth.2895.

Pla-Martín, D., Rueda, C. B., Estela, A., Sánchez-Piris, M., González-Sánchez, P., Traba, J., de la Fuente, S., Scorrano, L., Renau-Piqueras, J., Alvarez, J., Satrústegui, J. and Palau, F. (2013) 'Silencing of the Charcot-Marie-Tooth disease-associated gene GDAP1 induces abnormal mitochondrial distribution and affects Ca2+ homeostasis by reducing store-operated Ca2+ entry', *Neurobiology of Disease*, 55, pp. 140–151. doi: 10.1016/j.nbd.2013.03.010.

Des Portes, V., Pinard, J. M., Billuart, P., Vinet, M. C., Koulakoff, A., Carrié, A., Gelot, A., Dupuis, E., Motte, J., Berwald-Netter, Y., Catala, M., Kahn, A., Beldjord, C. and Chelly, J. (1998) 'A novel CNS gene required for neuronal migration and involved in X- linked subcortical laminar heterotopia and lissencephaly syndrome', *Cell*, 92(1), pp. 51–61. doi: 10.1016/S0092-8674(00)80898-3.

Raiborg, C., Bache, K. G., Mehlum, A., Stang, E. and Stenmark, H. (2001) 'Hrs recruits clathrin to early endosomes.', *The EMBO journal*, 20(17), pp. 5008–21. doi: 10.1093/emboj/20.17.5008.

Ralser, M., Nonhoff, U., Albrecht, M., Lengauer, T., Wanker, E. E., Lehrach, H. and Krobitsch, S.

(2005) ‘Ataxin-2 and huntingtin interact with endophilin-A complexes to function in plastin-associated pathways’, *Human Molecular Genetics*, 14(19), pp. 2893–2909. doi: 10.1093/hmg/ddi321.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) ‘Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips’, *Nature Protocols*, 2(8), pp. 1896–1906. doi: 10.1038/nprot.2007.261.

Rasmussen, A., Matsuura, T., Ruano, L., Yescas, P., Ochoa, A., Ashizawa, T. and Alonso, E. (2001) ‘Clinical and genetic analysis of four Mexican families with spinocerebellar ataxia type 10’, *Annals of Neurology*, 50(2), pp. 234–239. doi: 10.1002/ana.1081.

Reaves, B. and Banting, G. (1994) ‘Overexpression of TGN38/41 leads to mislocalisation of γ-adaptin’, *FEBS Letters*, 351(3), pp. 448–456. doi: 10.1016/0014-5793(94)00813-2.

Reed, N. A., Cai, D., Blasius, T. L., Jih, G. T., Meyhofer, E., Gaertig, J. and Verhey, K. J. (2006) ‘Microtubule Acetylation Promotes Kinesin-1 Binding and Transport’, *Current Biology*, 16(21), pp. 2166–2172. doi: 10.1016/j.cub.2006.09.014.

Reis, C. R., Chen, P., Srinivasan, S., Aguet, F., Mettlen, M. and Schmid, S. L. (2015) ‘ Crosstalk between Akt/ GSK 3β signaling and dynamin-1 regulates clathrin-mediated endocytosis ’, *The EMBO Journal*, 34(16), pp. 2132–2146. doi: 10.15252/embj.201591518.

Remacle, A., Murphy, G. and Roghi, C. (2003) ‘Membrane type I-matrix metalloproteinase (MT1-MMP) is internalised by two different pathways and is recycled to the cell surface’, *Journal of Cell Science*, 116(19), pp. 3905–3916. doi: 10.1242/jcs.00710.

Renard, H.-F., Simunovic, M., Lemière, J., Boucrot, E., Garcia-Castillo, M. D., Arumugam, S., Chambon, V., Lamaze, C., Wunder, C., Kenworthy, A. K., Schmidt, A. a., McMahon, H. T., Sykes, C., Bassereau, P. and Johannes, L. (2015) ‘Endophilin-A2 functions in membrane scission in clathrin-independent endocytosis.’, *Nature*, 517(7535), pp. 493–6. doi: 10.1038/nature14064.

Renvoisé, B., Malone, B., Falgairolle, M., Munasinghe, J., Stadler, J., Sibilla, C., Park, S. H. and Blackstone, C. (2016) ‘Reep1 null mice reveal a converging role for hereditary spastic paraplegia proteins in lipid droplet regulation’, *Human Molecular Genetics*, 25(23), pp. 5111–5125. doi: 10.1093/hmg/ddw315.

Ridley, A. J. (2011) ‘Life at the leading edge’, *Cell*, 145(7), pp. 1012–1022. doi: 10.1016/j.cell.2011.06.010.

Ringstad, N., Nemoto, Y. and De Camilli, P. (1997) 'The SH3p4/Sh3p8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), pp. 8569–74.

Roepstorff, P. and Fohlman, J. (1984) 'Proposal for a common nomenclature for sequence ions in mass spectra of peptides', *Biological Mass Spectrometry*, 11(11), pp. 601–601. doi: 10.1002/bms.1200111109.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004) 'Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents', *Molecular and Cellular Proteomics*, 3(12), pp. 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

Royle, S. J., Bright, N. A. and Lagnado, L. (2005) 'Clathrin is required for the function of the mitotic spindle.', *Nature*, 434(7037), pp. 1152–7. doi: 10.1038/nature03502.

Safren, N., El Ayadi, A., Chang, L., Terrillion, C. E., Gould, T. D., Boehning, D. F. and Monteiro, M. J. (2014) 'Ubiquilin-1 Overexpression Increases the Lifespan and Delays Accumulation of Huntingtin Aggregates in the R6/2 Mouse Model of Huntington's Disease', *PLoS ONE*. Edited by D. Blum, 9(1), p. e87513. doi: 10.1371/journal.pone.0087513.

Saito, K., Yamashiro, K., Ichikawa, Y., Erlmann, P., Kontani, K., Malhotra, V. and Katada, T. (2011) 'cTAGE5 mediates collagen secretion through interaction with TANGO1 at endoplasmic reticulum exit sites', *Molecular Biology of the Cell*, 22(13), pp. 2301–2308. doi: 10.1091/mbc.E11-02-0143.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., Imamura, T., Ogawa, M., Masai, H. and Miyawaki, A. (2008) 'Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression', *Cell*, 132(3), pp. 487–498. doi: 10.1016/J.CELL.2007.12.033.

Sancak, Y., Peterson, T. R., Shaul, Y. D., Lindquist, R. A., Thoreen, C. C., Bar-Peled, L. and Sabatini, D. M. (2008) 'The rag GTPases bind raptor and mediate amino acid signaling to mTORC1', *Science*, 320(5882), pp. 1496–1501. doi: 10.1126/science.1157535.

Sancak, Y., Bar-Peled, L., Zoncu, R., Markhard, A. L., Nada, S. and Sabatini, D. M. (2010)

‘Ragulator-rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids’, *Cell*, 141(2), pp. 290–303. doi: 10.1016/j.cell.2010.02.024.

Santos, A. J. M., Nogueira, C., Ortega-Bellido, M. and Malhotra, V. (2016) ‘TAN GO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum’, *Journal of Cell Biology*, 213(3), pp. 343–354. doi: 10.1083/jcb.201603072.

Saraon, P., Grozavu, I., Lim, S. H., Snider, J., Yao, Z. and Stagljar, I. (2017) ‘Detecting Membrane Protein-protein Interactions Using the Mammalian Membrane Two-hybrid (MaMTH) Assay’, *Current Protocols in Chemical Biology*, 9(1), pp. 38–54. doi: 10.1002/cpch.15.

Savage, M. J., Goldberg, D. J. and Schacher, S. (1987) ‘Absolute specificity for retrograde fast axonal transport displayed by lipid droplets originating in the axon of an identified Aplysia neuron in vitro’, *Brain Research*, 406(1–2), pp. 215–223. doi: 10.1016/0006-8993(87)90785-2.

Schmitt, F., Hussain, G., Dupuis, L., Loeffler, J. P. and Henriques, A. (2014) ‘A plural role for lipids in motor neuron diseases: Energy, signaling and structure’, *Frontiers in Cellular Neuroscience*. doi: 10.3389/fncel.2014.00025.

Schuske, K. R., Richmond, J. E., Matthies, D. S., Davis, W. S., Runz, S., Rube, D. A., van der Bliek, A. M. and Jorgensen, E. M. (2003) ‘Endophilin is required for synaptic vesicle endocytosis by localizing synaptojanin.’, *Neuron*, 40(4), pp. 749–62.

Schwanhüusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) ‘Global quantification of mammalian gene expression control’, *Nature*, 473(7347), pp. 337–342. doi: 10.1038/nature10098.

Scott, A. M., Wolchok, J. D. and Old, L. J. (2012) ‘Antibody therapy of cancer’, *Nature Reviews Cancer*, 12(4), pp. 278–287. doi: 10.1038/nrc3236.

Servant, G., Weiner, O. D., Herzmark, P., Balla, T., Sedat, J. W. and Bourne, H. R. (2000) ‘Polarization of chemoattractant receptor signaling during neutrophil chemotaxis’, *Science*, 287(5455), pp. 1037–1040. doi: 10.1126/science.287.5455.1037.

Sharma, R., Dill, B. D., Chourey, K., Shah, M., VerBerkmoes, N. C. and Hettich, R. L. (2012) ‘Coupling a detergent lysis/cleanup methodology with intact protein fractionation for enhanced proteome characterization.’, *Journal of proteome research*, 11(12), pp. 6008–18. doi: 10.1021/pr300709k.

Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V. and Mann, M. (2007) ‘In-gel digestion for

mass spectrometric characterization of proteins and proteomes', *Nature Protocols*, 1(6), pp. 2856–2860. doi: 10.1038/nprot.2006.468.

Shliaha, P. V, Bond, N. J., Gatto, L. and Lilley, K. S. (2013) 'Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies'. doi: 10.1021/pr300775k.

Shore, G. C., Papa, F. R. and Oakes, S. A. (2011) 'Signaling cell death from the endoplasmic reticulum stress response', *Current Opinion in Cell Biology*, pp. 143–149. doi: 10.1016/j.ceb.2010.11.003.

Sielaff, M., Kuharev, J., Bohn, T., Hahlbrock, J., Bopp, T., Tenzer, S. and Distler, U. (2017) 'Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range', *Journal of Proteome Research*, 16(11), pp. 4060–4072. doi: 10.1021/acs.jproteome.7b00433.

Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P. and Geromanos, S. (2005) 'Quantitative proteomic analysis by accurate mass retention time pairs', *Analytical Chemistry*, 77(7), pp. 2187–2200. doi: 10.1021/ac048455k.

Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C. and Geromanos, S. J. (2006) 'Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition', *Molecular and Cellular Proteomics*, 5(1), pp. 144–156. doi: 10.1074/mcp.M500230-MCP200.

Simunovic, M., Evergren, E., Golushko, I., Prévost, C., Renard, H.-F., Johannes, L., McMahon, H. T., Lorman, V., Voth, G. A. and Bassereau, P. (2016) 'How curvature-generating proteins build scaffolds on membrane nanotubes', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1606943113.

Smillie, K. J. and Cousin, M. A. (2012) 'Akt/PKB Controls the Activity-Dependent Bulk Endocytosis of Synaptic Vesicles', *Traffic*, 13(7), pp. 1004–1011. doi: 10.1111/j.1600-0854.2012.01365.x.

Solomon, T., Lewthwaite, P., Perera, D., Cardosa, M. J., McMinn, P. and Ooi, M. H. (2010) 'Virology, epidemiology, pathogenesis, and control of enterovirus 71', *The Lancet Infectious Diseases*, 10(11), pp. 778–790. doi: 10.1016/S1473-3099(10)70194-8.

Sonnett, M., Yeung, E. and Wühr, M. (2018) 'Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster', *Analytical Chemistry*, 90(8), pp. 5032–

5039. doi: 10.1021/acs.analchem.7b04713.

Soto, C. (2003) 'Unfolding the role of protein misfolding in neurodegenerative diseases', *Nature Reviews Neuroscience*, 4(1), pp. 49–60. doi: 10.1038/nrn1007.

Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y. and Dikic, I. (2002) 'Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors.', *Nature*, 416(6877), pp. 183–7. doi: 10.1038/416183a.

Sowell, R. A., Hersberger, K. E., Kaufman, T. C. and Clemmer, D. E. (2007) 'Examining the proteome of Drosophila across organism lifespan', *Journal of Proteome Research*, 6(9), pp. 3637–3647. doi: 10.1021/pr070224h.

Speicher, K. D., Kolbas, O., Harper, S. and Speicher, D. W. (2000) 'Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies.', *Journal of Biomolecular Techniques*, 11(2), pp. 74–86.

Spence, E. F., Dube, S., Uezu, A., Locke, M., Soderblom, E. J. and Soderling, S. H. (2019) 'In vivo proximity proteomics of nascent synapses reveals a novel regulator of cytoskeleton-mediated synaptic maturation', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-08288-w.

Steinberg, F., Gallon, M., Winfield, M., Thomas, E. C., Bell, A. J., Heesom, K. J., Tavaré, J. M. and Cullen, P. J. (2013) 'A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose and metal ion transport', *Nature Cell Biology*, 15(5), pp. 461–471. doi: 10.1038/ncb2721.

Srinivasan, S., Wang, F., Glavas, S., Ott, A., Hofmann, F., Aktories, K., Kalman, D. and Bourne, H. R. (2003) 'Rac and Cdc42 play distinct roles in regulating PI(3,4,5)P3 and polarity during neutrophil chemotaxis', *Journal of Cell Biology*, 160(3), pp. 375–385. doi: 10.1083/jcb.200208179.

Sundborger, A., Soderblom, C., Vorontsova, O., Evergren, E., Hinshaw, J. E. and Shupliakov, O. (2011) 'An endophilin-dynamin complex promotes budding of clathrin-coated vesicles during synaptic vesicle recycling.', *Journal of Cell Science*, 124(Pt 1), pp. 133–43. doi: 10.1242/jcs.072686.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J. and Mering, C. von (2019) 'STRING v11:

protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.', *Nucleic acids research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Tan, R. H., Ke, Y. D., Ittner, L. M. and Halliday, G. M. (2017) 'ALS/FTLD: experimental models and reality', *Acta Neuropathologica*, pp. 177–196. doi: 10.1007/s00401-016-1666-6.

Tan, T. C., Valova, V. A., Malladi, C. S., Graham, M. E., Berven, L. A., Jupp, O. J., Hansra, G., McClure, S. J., Sarcevic, B., Boadle, R., Larsen, M. R., Cousin, M. A. and Robinson, P. J. (2003) 'Cdk5 is essential for synaptic vesicle endocytosis', *Nature Cell Biology*, 5(8), pp. 701–710. doi: 10.1038/ncb1020.

Tanabe, T., Maeda, M., Saito, K. and Katada, T. (2016) 'Dual function of cTAGE5 in collagen export from the endoplasmic reticulum', *Molecular Biology of the Cell*, 27(13), pp. 2008–2013. doi: 10.1091/mbc.E16-03-0180.

Tang, Y., Hu, L. A., Miller, W. E., Ringstad, N., Hall, R. A., Pitcher, J. A., DeCamilli, P. and Lefkowitz, R. J. (1999) 'Identification of the endophilins (SH3p4/p8/p13) as novel binding partners for the beta 1-adrenergic receptor', *Proceedings of the National Academy of Sciences*, 96(22), pp. 12559–12564. doi: 10.1073/pnas.96.22.12559.

Taylor, G. (1964) 'Disintegration of Water Drops in an Electric Field', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 280(1382), pp. 383–397. doi: 10.1098/rspa.1964.0151.

Taylor, M. J., Perrais, D. and Merrifield, C. J. (2011) 'A high precision survey of the molecular dynamics of mammalian clathrin-mediated endocytosis.', *PLoS Biology*, 9(3), p. e1000604. doi: 10.1371/journal.pbio.1000604.

Tazir, M., Hamadouche, T., Nouioua, S., Mathis, S. and Vallat, J. M. (2014) 'Hereditary motor and sensory neuropathies or Charcot-Marie-Tooth diseases: An update', *Journal of the Neurological Sciences*, 347(1–2), pp. 14–22. doi: 10.1016/j.jns.2014.10.013.

Teive, H. A. G. and Ashizawa, T. (2013) 'Spinocerebellar ataxia type 10: From amerindians to latin americans', *Current Neurology and Neuroscience Reports*, pp. 1–3. doi: 10.1007/s11910-013-0393-9.

The Global Proteome Machine (2016) *cRAP protein sequences*. Available at: https://www.thegpm.org/crap/ (Accessed: 14 October 2016).

Thingholm, T. E., Jørgensen, T. J. D., Jensen, O. N. and Larsen, M. R. (2006) 'Highly selective enrichment of phosphorylated peptides using titanium dioxide', *Nature Protocols*, 1(4), pp. 1929–1935. doi: 10.1038/nprot.2006.185.

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H. A., Alm, T., Asplund, A., Björk, L., Breckels, L. M., … Lundberg, E. (2017) 'A subcellular map of the human proteome', *Science*, 356(6340). doi: 10.1126/science.aal3321.

Tomizawa, K., Sunada, S., Lu, Y. F., Oda, Y., Kinuta, M., Ohshima, T., Saito, T., Wei, F. Y., Matsushita, M., Li, S. T., Tsutsui, K., Hisanaga, S. I., Mikoshiba, K., Takei, K. and Matsui, H. (2003) 'Cophosphorylation of amphiphysin I and dynamin I by Cdk5 regulates clathrin-mediated endocytosis of synaptic vesicles', *Journal of Cell Biology*, 163(4), pp. 813–824. doi: 10.1083/jcb.200308110.

Tong, X., Ao, Y., Faas, G. C., Nwaobi, S. E., Xu, J., Haustein, M. D., Anderson, M. A., Mody, I., Olsen, M. L., Sofroniew, M. V. and Khakh, B. S. (2014) 'Astrocyte Kir4.1 ion channel deficits contribute to neuronal dysfunction in Huntington's disease model mice', *Nature Neuroscience*, 17(5), pp. 694–703. doi: 10.1038/nn.3691.

Tracey, T. J., Steyn, F. J., Wolvetang, E. J. and Ngo, S. T. (2018) 'Neuronal lipid metabolism: Multiple pathways driving functional outcomes in health and disease', *Frontiers in Molecular Neuroscience*, 11(January), pp. 1–25. doi: 10.3389/fnmol.2018.00010.

Turriziani, B., Garcia-Munoz, A., Pilkington, R., Raso, C., Kolch, W. and von Kriegsheim, A. (2014) 'On-beads digestion in conjunction with data-dependent mass spectrometry: A shortcut to quantitative and dynamic interaction proteomics', *Biology*, 3(2), pp. 320–332. doi: 10.3390/biology3020320.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M. and Cox, J. (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nature Methods*. doi: 10.1038/nmeth.3901.

Uezu, A., Kanak, D. J., Bradshaw, T. W. A., Soderblom, E. J., Catavero, C. M., Burette, A. C., Weinberg, R. J. and Soderling, S. H. (2016) 'Identification of an elaborate complex mediating postsynaptic inhibition', *Science*, 353(6304), pp. 1123–1129. doi: 10.1126/science.aag0821.

Vázquez, F. X., Unger, V. M. and Voth, G. A. (2013) 'Autoinhibition of Endophilin in Solution via Interdomain Interactions', *Biophysical Journal*, 104(2), pp. 396–403. doi:

10.1016/j.bpj.2012.12.009.

Vehlow, A., Soong, D., Vizcay-Barrena, G., Bodo, C., Law, A.-L., Perera, U. and Krause, M. (2013) 'Endophilin, Lamellipodin, and Mena cooperate to regulate F-actin-dependent EGF-receptor endocytosis', *EMBO Journal*, 32, pp. 2722–2734. doi: 10.1038/emboj.2013.212.

Verstreken, P., Kjaerulff, O., Lloyd, T. E., Atkinson, R., Zhou, Y., Meinertzhagen, I. A. and Bellen, H. J. (2002) 'Endophilin mutations block clathrin-mediated endocytosis but not neurotransmitter release.', *Cell*, 109(1), pp. 101–12.

von Kleist, L., Stahlschmidt, W., Bulut, H., Gromova, K., Puchkov, D., Robertson, M. J., MacGregor, K. A., Tomilin, N., Pechstein, A., Chau, N., … Haucke, V. (2011) 'Role of the Clathrin Terminal Domain in Regulating Coated Pit Dynamics Revealed by Small Molecule Inhibition', *Cell*, 146(3), pp. 471–484. doi: 10.1016/j.cell.2011.06.025.

Wakamiya, M., Matsuura, T., Liu, Y., Schuster, G. C., Gao, R., Xu, W., Sarkar, P. S., Lin, X. and Ashizawa, T. (2006) 'The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10', *Neurology*, 67(4), pp. 607–613. doi: 10.1212/01.wnl.0000231140.26253.eb.

Wang, H., Lim, P. J., Yin, C., Rieckher, M., Vogel, B. E. and Monteiro, M. J. (2006) 'Suppression of polyglutamine-induced toxicity in cell and animal models of Huntington's disease by ubiquilin.', *Human molecular genetics*, 15(6), pp. 1025–41. doi: 10.1093/hmg/ddl017.

Wang, H. and Monteiro, M. J. (2007) 'Ubiquilin interacts and enhances the degradation of expanded-polyglutamine proteins', *Biochemical and Biophysical Research Communications*, 360(2), pp. 423–427. doi: 10.1016/j.bbrc.2007.06.097.

Wang, Q., Kaan, H. Y. K., Hooda, R. N., Goh, S. L. and Sondermann, H. (2008) 'Structure and Plasticity of Endophilin and Sorting Nexin 9', *Structure*, 16(10), pp. 1574–1587. doi: 10.1016/J.STR.2008.07.016.

Wang, Q., Navarro, M. V. A. S., Peng, G., Molinelli, E., Goh, S. L., Judson, B. L., Rajashankar, K. R. and Sondermann, H. (2009) 'Molecular mechanism of membrane constriction and tubulation mediated by the F-BAR protein Pacsin/Syndapin.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp. 12700–5. doi: 10.1073/pnas.0902974106.

Wang, W., Erbe, A. K., Hank, J. A., Morris, Z. S. and Sondel, P. M. (2015) 'NK cell-mediated antibody-dependent cellular cytotoxicity in cancer immunotherapy', *Frontiers in Immunology*,

6(JUL). doi: 10.3389/fimmu.2015.00368.

Weekes, M. P., Antrobus, R., Lill, J. R., Duncan, L. M., Hör, S. and Lehner, P. J. (2010) 'Comparative analysis of techniques to purify plasma membrane proteins', *Journal of Biomolecular Techniques*, 21(3), pp. 108–115.

Weiss, W., Weiland, F. and Görg, A. (2009) 'Protein detection and quantitation technologies for gel-based proteome analysis.', in *Methods in molecular biology (Clifton, N.J.)*, pp. 59–82. doi: 10.1007/978-1-60761-157-8_4.

Westhoff, B., Chapple, J. P., Van Der Spuy, J., Höhfeld, J. and Cheetham, M. E. (2005) 'HSJ1 is a neuronal shuttling factor for the sorting of chaperone clients to the proteasome', *Current Biology*, 15(11), pp. 1058–1064. doi: 10.1016/j.cub.2005.04.058.

White, M., Xia, G., Gao, R., Wakamiya, M., Sarkar, P. S., McFarland, K. and Ashizawa, T. (2012) 'Transgenic mice with SCA10 pentanucleotide repeats show motor phenotype and susceptibility to seizure: A toxic RNA gain-of-function model', *Journal of Neuroscience Research*, 90(3), pp. 706–714. doi: 10.1002/jnr.22786.

White, M. C., Gao, R., Xu, W., Mandal, S. M., Lim, J. G., Hazra, T. K., Wakamiya, M., Edwards, S. F., Raskin, S., Teive, H. A. G., Zoghbi, H. Y., Sarkar, P. S. and Ashizawa, T. (2010) 'Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCδ to mitochondria in spinocerebellar ataxia 10', *PLoS Genetics*, 6(6), pp. 1–12. doi: 10.1371/journal.pgen.1000984.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., … Kuster, B. (2014) 'Mass-spectrometry-based draft of the human proteome', *Nature*, 509(7502), pp. 582–587. doi: 10.1038/nature13319.

Willforss, J., Chawade, A. and Levander, F. (2019) 'NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis', *Journal of Proteome Research*, 18(2), pp. 732–740. doi: 10.1021/acs.jproteome.8b00523.

Wilm, M. and Mann, M. (1996) 'Analytical properties of the nanoelectrospray ion source.', *Analytical chemistry*, 68(1), pp. 1–8.

Wilson, R. C. and Doudna, J. A. (2013) 'Molecular Mechanisms of RNA Interference', *Annual Review of Biophysics*, 42(1), pp. 217–239. doi: 10.1146/annurev-biophys-083012-130404.

Wiśniewski, J. R., Zougman, A., Nagaraj, N. and Mann, M. (2009) 'Universal sample preparation

method for proteome analysis', *Nature Methods*, 6(5), pp. 359–362. doi: 10.1038/nmeth.1322.

Wiśniewski, J. R., Ostasiewicz, P., Duś, K., Zielińska, D. F., Gnad, F. and Mann, M. (2012) 'Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma', *Molecular Systems Biology*, 8. doi: 10.1038/msb.2012.44.

Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R. and Watts, J. D. (2009) 'Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins', *Nature Biotechnology*, 27(4), pp. 378–386. doi: 10.1038/nbt.1532.

Xu, X., Mei, H., Wang, S., Zhou, Q., Wang, G., Broske, L., Pena, A. and Korfmacher, W. A. (2005) 'A study of common discovery dosing formulation components and their potential for causing time-dependent matrix effects in high-performance liquid chromatography tandem mass spectrometry assays', *Rapid Communications in Mass Spectrometry*, 19(18), pp. 2643–2650. doi: 10.1002/rcm.2102.

Zhang, D., Raasi, S. and Fushman, D. (2008) 'Affinity Makes the Difference: Nonselective Interaction of the UBA Domain of Ubiquilin-1 with Monomeric Ubiquitin and Polyubiquitin Chains', *Journal of Molecular Biology*, 377(1), pp. 162–180. doi: 10.1016/j.jmb.2007.12.029.

Zhou, H., Li, X. M., Meinkoth, J. and Pittman, R. N. (2000) 'Akt regulates cell survival and apoptosis at a postmitochondrial level', *Journal of Cell Biology*, 151(3), pp. 483–494. doi: 10.1083/jcb.151.3.483.

Zubarev, R. A. (2013) 'The challenge of the proteome dynamic range and its implications for in-depth proteomics', *Proteomics*, 13(5), pp. 723–726. doi: 10.1002/pmic.201200451.

8. Appendix

8.1. Yield of plasma membrane protein from surface biotinylation

	Absorbance 750nm					
BSA(mg/ml)	1	2	3	Average (Technical measures)	Average (BSA duplicates)	Standard deviation (BSA duplicates)
1.0	0.229	0.23	0.229	0.229	0.220	0.013
1.0	0.211	0.207	0.214	0.211		0.004
0.2	0.088	0.086	0.088	0.087	0.084	0.005
0.2	0.08	0.081	0.079	0.080		0.001
0.5	0.137	0.14	0.135	0.137	0.130	0.010
0.5	0.123	0.125	0.122	0.123		0.002
Diluted Sample						
6X PM 1 (0.75X)	0.169	0.174	0.169	0.171		0.003
6X PM 1 (0.4X)	0.126	0.125	0.127	0.126		0.001

Table 8.1 Quantitation of concentrated plasma membrane (PM) fraction from 6 x 100 mm dishes of RPE1 cells at 2 dilutions with reference to BSA standards.

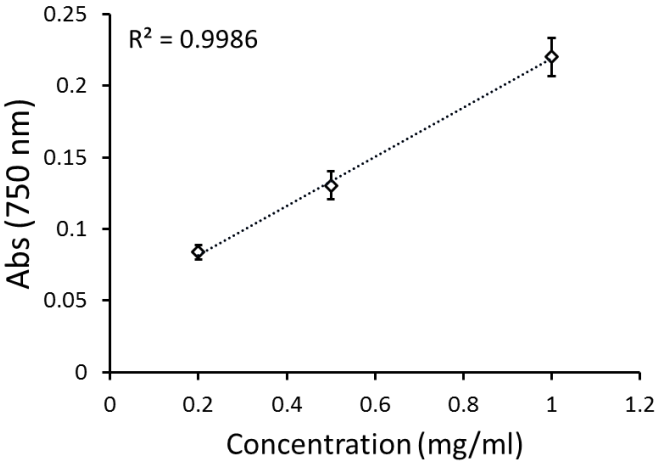

Figure 8.1 Standard curve from measurements in Table 8.1. Error bars represent mean ± SD.

Sample	Corrected Sample Concentration (mg/ml)
6X Pmx 1	0.96
6X Pmx 2	1.14
Mean + SD	**1.05 ± 0.1**

Table 8.2 Corrected sample concentrations and final estimate of protein concentration.

8.2. Differentially expressed proteins (endogenous Endophilin A2 IP)

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Dynein assembly factor 5_ axonemal	DNAAF5	8.545	0.0311
Ubiquitin carboxyl-terminal hydrolase BAP1	BAP1	8.061	0.0320
USP6 N-terminal-like protein	USP6NL	8.003	0.0155
Uncharacterized protein C14orf80	C14orf80	5.698	0.0321
DNA polymerase delta subunit 4	POLD4	4.649	0.0088
SRSF protein kinase 2	SRPK2	4.528	0.0389
NEDD4-like E3 ubiquitin-protein ligase WWP1	WWP1	3.936	0.0340
Diacylglycerol kinase zeta	DGKZ	3.458	0.0340
Death domain-associated protein 6	DAXX	3.440	0.0168
PAB-dependent poly(A)-specific ribonuclease subunit PAN3	PAN3	3.072	0.0496
General transcription factor IIF subunit 2	GTF2F2	3.032	0.0193
Kinesin-like protein KIF18B	KIF18B	2.966	0.0188
Zinc finger CCCH domain-containing protein 7B	ZC3H7B	2.738	0.0067
MICAL-like protein 1	MICALL1	2.718	0.0257
E3 ubiquitin-protein ligase TRIM33	TRIM33	2.615	0.0494
DTW domain-containing protein 2	DTWD2	2.474	0.0322
Mediator of RNA polymerase II transcription subunit 20	MED20	2.317	0.0427
cTAGE family member 6	CTAGE6	2.317	0.0001
Protein YIF1B	YIF1B	2.291	0.0130
Ribosomal RNA processing protein 1 homolog B	RRP1B	2.234	0.00004
Four and a half LIM domains protein 1	FHL1	2.191	0.0011
Probable E3 ubiquitin-protein ligase DTX3	DTX3	2.176	0.0167
Serine/threonine-protein kinase NIM1	NIM1K	2.170	0.0073
Rho GTPase-activating protein 17	ARHGAP17	2.168	0.0040
28S ribosomal protein S2_ mitochondrial	MRPS2	2.167	0.0069
T-complex protein 1 subunit alpha	TCP1	2.084	0.0221
Protein timeless homolog	TIMELESS	2.046	0.0402
Liprin-beta-1	PPFIBP1	2.039	0.0275
39S ribosomal protein L15_ mitochondrial	MRPL15	1.943	0.0137
Aminoacyl tRNA synthase complex-interacting multifunctional protein 2	AIMP2	1.942	0.0255
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1-related	HMG20B	1.880	0.0490
Protein RCC2	RCC2	1.875	0.0249
B-cell CLL/lymphoma 7 protein family member C	BCL7C	1.871	0.0394

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Tropomyosin alpha-4 chain	TPM4	1.839	0.0221
Actin-related protein 6	ACTR6	1.837	0.0101
Disabled homolog 2-interacting protein	DAB2IP	1.831	0.0015
General transcription factor 3C polypeptide 5	GTF3C5	1.811	0.0219
Anaphase-promoting complex subunit 16	ANAPC16	1.808	0.0408
F-actin-capping protein subunit alpha-1	CAPZA1	1.775	0.0351
Beta-crystallin A3	CRYBA1	1.758	0.0218
Protein Smaug homolog 1	SAMD4A	1.726	0.0314
Pleckstrin homology domain-containing family N member 1	PLEKHN1	1.724	0.0328
Endonuclease III-like protein 1	NTHL1	1.720	0.0290
Fructose-bisphosphate aldolase B	ALDOB	1.714	0.0377
WD repeat-containing protein 11	WDR11	1.701	0.0060
Phosphatidylglycerophosphatase and protein-tyrosine phosphatase 1	PTPMT1	1.698	0.0046
Headcase protein homolog	HECA	1.698	0.0241
Protein diaphanous homolog 1	DIAPH1	1.696	0.0176
Elongation factor 1-beta	EEF1B2	1.686	0.0255
CDGSH iron-sulfur domain-containing protein 2	CISD2	1.684	0.0348
BUB3-interacting and GLEBS motif-containing protein ZNF207	ZNF207	1.673	0.0491
Putative uncharacterized protein C20orf78	C20orf78	1.665	0.0255
Forkhead box protein C2	FOXC2	1.652	0.0288
DNA repair and recombination protein RAD54B	RAD54B	1.636	0.0321
Keratin_ type II cuticular Hb4	KRT84	1.629	0.0437
Centromere protein V	CENPV	1.594	0.0497
Fatty acid desaturase 3	FADS3	1.590	0.0333
Importin subunit alpha-1	KPNA2	1.587	0.0171
Importin subunit alpha-5	KPNA1	1.581	0.0320
Glial fibrillary acidic protein	GFAP	1.560	0.0390
ATP-dependent RNA helicase DDX55	DDX55	1.558	0.0236
Ankyrin repeat domain-containing protein 34A	ANKRD34A	1.554	0.0091
Serine/threonine-protein kinase RIO2	RIOK2	1.552	0.0499
Importin-9	IPO9	1.551	0.0031
E3 ubiquitin-protein ligase CBL-B	CBLB	1.551	0.0460
Histone deacetylase 1	HDAC1	1.550	0.0215
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	NDUFB10	1.547	0.0030
Sister chromatid cohesion protein PDS5 homolog A	PDS5A	1.544	0.0005
Partitioning defective 3 homolog B	PARD3B	1.539	0.0141
Eukaryotic translation elongation factor 1 epsilon-1	EEF1E1	1.538	0.0111
Long-chain-fatty-acid--CoA ligase 4	ACSL4	1.528	0.0250
ATP-dependent zinc metalloprotease YME1L1	YME1L1	1.525	0.0368
Putative heat shock 70 kDa protein 7	HSPA7	1.525	0.0160

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
TFIIH basal transcription factor complex helicase XPB subunit	ERCC3	1.522	0.0497
Nucleotide exchange factor SIL1	SIL1	1.521	0.0354
Importin subunit alpha-4	KPNA3	1.520	0.0016
E3 ubiquitin-protein ligase CHIP	STUB1	1.516	0.0314
Antigen peptide transporter 1	TAP1	1.507	0.0397
Monoacylglycerol lipase ABHD12	ABHD12	1.504	0.0131
Serine/threonine-protein kinase OSR1	OXSR1	1.497	0.0063
Elongation factor Tu_ mitochondrial	TUFM	1.496	0.0437
Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	SAMHD1	1.494	0.0496
Coiled-coil domain-containing protein 47	CCDC47	1.492	0.0029
Coiled-coil domain-containing protein 17	CCDC17	1.486	0.0383
Cyclin-dependent kinase 17	CDK17	1.486	0.0023
SH3 and PX domain-containing protein 2B	SH3PXD2B	1.484	0.0168
Glycerol-3-phosphate dehydrogenase_ mitochondrial	GPD2	1.482	0.0321
Transcription factor Dp-1	TFDP1	1.481	0.0038
Sulfide:quinone oxidoreductase_ mitochondrial	SQOR	1.476	0.0052
TOX high mobility group box family member 4	TOX4	1.473	0.0373
Importin subunit alpha-7	KPNA6	1.464	0.0033
F-box/LRR-repeat protein 6	FBXL6	1.460	0.0479
E3 ubiquitin-protein ligase RNF25	RNF25	1.457	0.0246
Polyadenylate-binding protein-interacting protein 1	PAIP1	1.454	0.0462
Protein TBRG4	TBRG4	1.452	0.0266
Histone deacetylase 2	HDAC2	1.451	0.0260
60 kDa heat shock protein_ mitochondrial	HSPD1	1.451	0.0438
Testican-1	SPOCK1	1.449	0.0392
Regulator of nonsense transcripts 3B	UPF3B	1.443	0.0381
Endothelial differentiation-related factor 1	EDF1	1.438	0.0031
Protein argonaute-4	AGO4	1.428	0.0050
WAS/WASL-interacting protein family member 1	WIPF1	1.427	0.0035
Ankyrin repeat and zinc finger domain-containing protein 1	ANKZF1	1.426	0.0421
Ankyrin repeat and LEM domain-containing protein 2	ANKLE2	1.416	0.0068
Putative RNA polymerase II subunit B1 CTD phosphatase RPAP2	RPAP2	1.412	0.0259
Zinc finger protein-like 1	ZFPL1	1.407	0.0159
Lamin-B receptor	LBR	1.405	0.0331
DNA topoisomerase 3-alpha	TOP3A	1.399	0.0431
PHD finger protein 23	PHF23	1.390	0.0192
Vitronectin	VTN	1.381	0.0190

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Plasminogen activator inhibitor 1	SERPINE1	1.381	0.0116
Probable C-mannosyltransferase DPY19L1	DPY19L1	1.376	0.0292
Catenin delta-1	CTNND1	1.371	0.0067
THAP domain-containing protein 11	THAP11	1.365	0.0049
U4/U6 small nuclear ribonucleoprotein Prp31	PRPF31	1.364	0.0182
T-complex protein 1 subunit gamma	CCT3	1.359	0.0360
Peroxidasin homolog	PXDN	1.358	0.0351
SH3KBP1-binding protein 1	SHKBP1	1.354	0.0380
Ras GTPase-activating protein-binding protein 1	G3BP1	1.353	0.0022
A disintegrin and metalloproteinase with thrombospondin motifs 1	ADAMTS1	1.351	0.0453
Trifunctional enzyme subunit alpha_ mitochondrial	HADHA	1.350	0.0138
T-complex protein 1 subunit delta	CCT4	1.347	0.0098
Metastasis-associated protein MTA2	MTA2	1.346	0.0123
Mapk-regulated corepressor-interacting protein 1	MCRIP1	1.342	0.0451
Protein ECT2	ECT2	1.340	0.0419
Unconventional myosin-If	MYO1F	1.338	0.0160
Transportin-1	TNPO1	1.336	0.0062
Proteasome subunit beta type-6	PSMB6	1.334	0.0244
Katanin p60 ATPase-containing subunit A-like 1	KATNAL1	1.334	0.0002
Nuclear pore complex protein Nup50	NUP50	1.328	0.0314
GTP-binding protein 1	GTPBP1	1.318	0.0005
Inner nuclear membrane protein Man1	LEMD3	1.315	0.0083
Ornithine aminotransferase_ mitochondrial	OAT	1.314	0.0202
GRIP and coiled-coil domain-containing protein 2	GCC2	1.313	0.0037
Twinfilin-1	TWF1	1.310	0.0297
Ras GTPase-activating protein nGAP	RASAL2	1.307	0.0319
Methyl-CpG-binding domain protein 2	MBD2	1.298	0.0482
Probable ATP-dependent RNA helicase DDX28	DDX28	1.296	0.0444
DEP domain-containing protein 1B	DEPDC1B	1.293	0.0164
Keratin_ type I cuticular Ha6	KRT36	0.764	0.0038
Rho guanine nucleotide exchange factor 17	ARHGEF17	0.763	0.0074
Cofilin-2	CFL2	0.756	0.0417
Replication factor C subunit 1	RFC1	0.741	0.0332
Serine protease HTRA3	HTRA3	0.740	0.0146
Hemoglobin subunit epsilon	HBE1	0.735	0.0204
Protein LLP homolog	LLPH	0.734	0.0227
Keratin_ type I cytoskeletal 20	KRT20	0.729	0.0112
Keratin_ type II cytoskeletal 2 epidermal	KRT2	0.720	0.0138
Unconventional prefoldin RPB5 interactor 1	URI1	0.720	0.0134
Protein bicaudal C homolog 1	BICC1	0.720	0.0443
Trypsin-3	PRSS3	0.714	0.0234
Thrombospondin-1	THBS1	0.712	0.0376
Cell division cycle 5-like protein	CDC5L	0.710	0.0056

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Cep170-like protein	CEP170P1	0.708	0.0444
ATP-binding cassette sub-family F member 1	ABCF1	0.701	0.0226
Deoxynucleotidyltransferase terminal-interacting protein 2	DNTTIP2	0.690	0.0131
Enhancer of rudimentary homolog	ERH	0.687	0.0149
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1	SMARCB1	0.686	0.0229
DNA-directed RNA polymerase II subunit RPB1	POLR2A	0.685	0.0204
YLP motif-containing protein 1	YLPM1	0.684	0.0288
Keratin_ type I cuticular Ha1	KRT31	0.680	0.0001
AT-rich interactive domain-containing protein 5B	ARID5B	0.672	0.0387
Poly [ADP-ribose] polymerase 4	PARP4	0.669	0.0466
Immunoglobulin heavy constant gamma 1	IGHG1	0.663	0.0108
Keratin_ type II cytoskeletal 4	KRT4	0.659	0.0188
CREB-regulated transcription coactivator 3	CRTC3	0.650	0.0474
Microtubule-associated protein 2	MAP2	0.645	0.0165
Desmoglein-1	DSG1	0.644	0.0049
Keratin_ type I cytoskeletal 10	KRT10	0.638	0.0065
Myotubularin-related protein 5	SBF1	0.636	0.0449
Heat shock 70 kDa protein 1A	HSPA1A	0.634	0.0446
Nuclear mitotic apparatus protein 1	NUMA1	0.631	0.0163
Homeobox protein Hox-D10	HOXD10	0.623	0.0025
Intraflagellar transport protein 74 homolog	IFT74	0.622	0.0339
Chromodomain-helicase-DNA-binding protein 1	CHD1	0.619	0.0028
PHD finger protein 3	PHF3	0.606	0.0240
Thyroid hormone receptor-associated protein 3	THRAP3	0.603	0.0126
Regulation of nuclear pre-mRNA domain-containing protein 2	RPRD2	0.598	0.0046
Plasma membrane calcium-transporting ATPase 4	ATP2B4	0.598	0.0265
AT-rich interactive domain-containing protein 1A	ARID1A	0.598	0.0475
Oxysterol-binding protein-related protein 3	OSBPL3	0.596	0.0370
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	0.595	0.0138
Actin-related protein 3	ACTR3	0.592	0.0039
Zinc finger protein 629	ZNF629	0.587	0.0315
Septin-7	SEPT7	0.578	0.0135
Centrosomal protein kizuna	KIZ	0.576	0.0011
E3 ubiquitin-protein ligase RBBP6	RBBP6	0.574	0.0462
Next to BRCA1 gene 2 protein	NBR2	0.569	0.0331
Myosin regulatory light polypeptide 9	MYL9	0.567	0.0220
Splicing factor_ suppressor of white-apricot homolog	SFSWAP	0.566	0.0128
Ral GTPase-activating protein subunit beta	RALGAPB	0.564	0.0209

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Probable ATP-dependent RNA helicase DDX46	DDX46	0.564	0.0027
Filaggrin-2	FLG2	0.562	0.0271
Nck-associated protein 5-like	NCKAP5L	0.561	0.0299
Prolactin-inducible protein	PIP	0.560	0.0118
Torsin-1A-interacting protein 1	TOR1AIP1	0.560	0.0391
Centrosomal protein of 89 kDa	CEP89	0.559	0.0207
Protein S100-A8	S100A8	0.553	0.0008
CD166 antigen	ALCAM	0.550	0.0130
Uncharacterized protein KIAA0754	KIAA0754	0.545	0.0123
DNA topoisomerase 2-beta	TOP2B	0.537	0.0427
Microtubule cross-linking factor 1	MTCL1	0.534	0.0364
Bromodomain adjacent to zinc finger domain protein 1A	BAZ1A	0.533	0.0211
Dermcidin	DCD	0.533	0.0120
Death-inducer obliterator 1	DIDO1	0.530	0.0234
N-acylneuraminate cytidylyltransferase	CMAS	0.529	0.0108
Methionine aminopeptidase 1	METAP1	0.527	0.0071
Protein polybromo-1	PBRM1	0.521	0.0191
Calumenin	CALU	0.520	0.0469
FERM domain-containing protein 4A	FRMD4A	0.519	0.0025
Caspase-14	CASP14	0.515	0.0143
Exocyst complex component 6B	EXOC6B	0.512	0.0340
E3 SUMO-protein ligase PIAS1	PIAS1	0.509	0.0294
Proline-rich protein 12	PRR12	0.506	0.0211
Sorting nexin-33	SNX33	0.500	0.0014
Clusterin	CLU	0.498	0.0151
Flap endonuclease GEN homolog 1	GEN1	0.497	0.0335
Alpha-actinin-4	ACTN4	0.490	0.0454
Collagen and calcium-binding EGF domain-containing protein 1	CCBE1	0.487	0.0424
Reticulocalbin-2	RCN2	0.484	0.0405
Splicing factor U2AF 35 kDa subunit-like protein	U2AF1L5	0.483	0.0118
Macoilin	TMEM57	0.482	0.0338
Serine/threonine-protein kinase PRP4 homolog	PRPF4B	0.479	0.0366
Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas	GNAS	0.474	0.0109
Pleckstrin homology domain-containing family A member 2	PLEKHA2	0.473	0.0451
60S ribosomal protein L35a	RPL35A	0.471	0.0423
DNA ligase 3	LIG3	0.471	0.0023
AT-rich interactive domain-containing protein 1B	ARID1B	0.470	0.0394
Sorting nexin-16	SNX16	0.468	0.0075
Exportin-2	CSE1L	0.467	0.0063

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
Uveal autoantigen with coiled-coil domains and ankyrin repeats	UACA	0.461	0.0175
Lysozyme C	LYZ	0.461	0.0095
Pleckstrin homology-like domain family B member 1	PHLDB1	0.456	0.0107
Serpin H1	SERPINH1	0.456	0.0498
RNA-binding protein 10	RBM10	0.455	0.0074
Casein kinase II subunit alpha	CSNK2A1	0.426	0.0169
Myosin-9	MYH9	0.426	0.0283
General transcription factor 3C polypeptide 1	GTF3C1	0.423	0.0151
Serpin B3	SERPINB3	0.420	0.0052
La-related protein 6	LARP6	0.413	0.0426
Aldehyde dehydrogenase_ dimeric NADP-preferring	ALDH3A1	0.408	0.0138
Keratin_ type I cytoskeletal 14	KRT14	0.403	0.0228
Phosphorylated adapter RNA export protein	PHAX	0.401	0.0089
Peripherin	PRPH	0.400	0.0072
E3 ubiquitin-protein ligase TRIM56	TRIM56	0.400	0.0002
Calcium signal-modulating cyclophilin ligand	CAMLG	0.397	0.0435
Serpin B4	SERPINB4	0.383	0.0007
28S ribosomal protein S31_ mitochondrial	MRPS31	0.367	0.0040
Keratin_ type I cytoskeletal 17	KRT17	0.363	0.0053
Keratin_ type I cytoskeletal 24	KRT24	0.354	0.0451
Serine/threonine-protein kinase Nek1	NEK1	0.347	0.0111
Poly [ADP-ribose] polymerase 14	PARP14	0.338	0.0078
Tyrosine-protein phosphatase non-receptor type 2	PTPN2	0.336	0.0181
Dual specificity protein phosphatase 16	DUSP16	0.333	0.0160
UNC119-binding protein C5orf30	C5orf30	0.330	0.0368
Keratin_ type II cytoskeletal 71	KRT71	0.330	0.0204
Eukaryotic translation initiation factor 2 subunit 2	EIF2S2	0.327	0.0331
LIM domain-binding protein 2	LDB2	0.323	0.0428
Sperm flagellar protein 2	SPEF2	0.321	0.0221
Hsp90 co-chaperone Cdc37	CDC37	0.312	0.0114
Keratin_ type I cytoskeletal 12	KRT12	0.297	0.0113
Heat shock 70 kDa protein 1-like	HSPA1L	0.289	0.0420
Replication factor C subunit 5	RFC5	0.262	0.0221
DNA topoisomerase 3-beta-1	TOP3B	0.257	0.0262
Hippocalcin-like protein 1	HPCAL1	0.256	0.0154
Keratin_ type I cytoskeletal 26	KRT26	0.250	0.0155
Keratin_ type I cytoskeletal 40	KRT40	0.249	0.0416
Histone-arginine methyltransferase CARM1	CARM1	0.247	0.0399
Kinesin-like protein KIF14	KIF14	0.240	0.0019
Mitochondrial import receptor subunit TOM40B	TOMM40L	0.206	0.0349

Protein Name	Gene Name	Fold change (Stim/Resting)	p-value
FAST kinase domain-containing protein 2_ mitochondrial	FASTKD2	0.201	0.0138
Isoleucine--tRNA ligase_ cytoplasmic	IARS	0.190	0.0433
Partitioning defective 3 homolog	PARD3	0.182	0.0354
AP-1 complex subunit mu-1	AP1M1	0.181	0.0368
Protein FAM84B	FAM84B	0.175	0.0068
DNA (cytosine-5)-methyltransferase 1	DNMT1	0.169	0.0404
Chromodomain-helicase-DNA-binding protein 2	CHD2	0.160	0.0496
Putative Ras-related protein Rab-1C	RAB1C	0.090	0.0129
Actin-binding LIM protein 3	ABLIM3	0.074	0.0214
Ribosome production factor 1	RPF1	0.007	0.0011

Table 8.3 List of proteins differentially expressed between endogenous IPs of endophilin from Stimulated or Resting RPE1 cell extracts. Proteins enriched in the Resting condition are shaded in grey.

8.3. List of proteins uniquely identified in HSJ1$^{+/+}$ and HSJ1$^{-/-}$ lumbar spinal cords

Gene name	Log2 Protein Abundance	Uniprot Accession	Description	Unique in
Dock11	3.362	A2AF47	Dedicator of cytokinesis protein 11	WT
Slc8a1	3.237	G3X9J1	Slc8a1 protein	WT
Gdap1	3.224	O88741	Ganglioside-induced differentiation-associated protein 1	WT
Ubr4	3.174	A2AN08-2	E3 ubiquitin-protein ligase UBR4 (Fragment)	WT
Cit	2.609	D3YU89	Citron Rho-interacting kinase	WT
Itsn1	2.586	D3Z6P4	Intersectin 1 isoform 12 (Fragment)	WT
Cdc42bpb	2.531	A0A1Y7VM95	Serine/threonine-protein kinase MRCK beta (Fragment)	WT
Phldb1	2.459	D3Z0X5	Isoform 2 of Pleckstrin homology-like domain family B member 1	WT
Igsf3	2.444	A0A0A6YX40	Immunoglobulin superfamily member 3	WT
Cep170	2.373	A0A0A6YVZ3	Centrosomal protein of 170 kDa (Fragment)	WT
Itsn2	1.953	B2RR82	Intersectin-2	WT
Agap2	1.939	Q3UHD9	Arf-GAP with GTPase_ ANK repeat and PH domain-containing protein 2	WT
Gm5580	1.918	A0A0N4SVP8	Predicted pseudogene 5580	WT
Taok2	1.831	Q6ZQ29	Isoform 2 of Serine/threonine-protein kinase TAO2	WT
Nav1	1.817	A0A087WSM9	Isoform 2 of Neuron navigator 1	WT
Sh3gl1	1.805	Q62419	Endophilin-A2	WT
Serpinb6a	1.800	E9PYY0	Serine (or cysteine) peptidase inhibitor_ clade B_ member 6a (Fragment)	WT
Gm14446	1.786	D3Z6F0	Interferon-induced protein with tetratricpeptide repeats 1B-like 1	WT
Arhgap44	1.732	F6T1Y2	Isoform 2 of Rho GTPase-activating protein 44	WT
Plxna2	1.728	P70207	Plexin-A2	WT
Grm5	1.725	E9QMC2	Isoform 2 of Metabotropic glutamate receptor 5	WT
Mast1	1.699	E9Q6Q5	Microtubule-associated serine/threonine-protein kinase 1	WT
Tuba1a	1.671	A0A2R8VHF3	Tubulin alpha chain (Fragment) OS	WT
Usp9x	1.633	E9PWA9	Probable ubiquitin carboxyl-terminal hydrolase FAF-X (Fragment)	WT
Zc2hc1a	1.625	Q8BJH1	Isoform 2 of Zinc finger C2HC domain-containing protein 1A	WT
Gpd1l	1.622	Q3ULJ0	Glycerol-3-phosphate dehydrogenase 1-like protein	WT
Sugt1	1.610	Q9CX34	Protein SGT1 homolog	WT
Pde9a	1.559	D3YTQ4	High affinity cGMP-specific 3'_5'-cyclic phosphodiesterase 9A (Fragment)	WT

Gene name	Log2 Protein Abundance	Uniprot Accession	Description	Unique in
Kif21b	1.527	E9Q0A4	Kinesin-like protein KIF21B	WT
Slc25a25	1.516	A2ASZ8	Calcium-binding mitochondrial carrier protein SCaMC-2	WT
Nagk	1.511	D3YXG2	N-acetyl-D-glucosamine kinase	WT
Sarm1	1.506	Q6PDS3	Isoform 2 of Sterile alpha and TIR motif-containing protein 1	WT
Rptor	1.488	A2ACM0	Isoform 2 of Regulatory-associated protein of mTOR	WT
Plxnb1	1.480	Q8CJH3	Plexin-B1	WT
Cntnap4	1.475	D3YWB9	Contactin-associated protein-like 4	WT
Slc12a6	1.398	A2AGJ9	Isoform 2 of Solute carrier family 12 member 6	WT
Atxn2l	1.385	A0A0U1RPL0	Ataxin-2-like protein (Fragment)	WT
Nrp2	1.383	O35375	Isoform A0 of Neuropilin-2	WT
Rasgrf1	1.381	P27671	Ras-specific guanine nucleotide-releasing factor 1	WT
Lrrc16a	1.335	D3Z030	F-actin-uncapping protein LRRC16A (Fragment)	WT
Gpr158	1.306	Q8C419	Probable G-protein coupled receptor 158	WT
Fkbp15	1.300	Q6P9Q6	FK506-binding protein 15	WT
Plcg1	1.294	A2A4A6	1-phosphatidylinositol 4_5-bisphosphate phosphodiesterase gamma-1	WT
Slc25a4	1.257	P48962	ADP/ATP translocase 1	WT
Btbd17	1.253	Q9DB72	BTB/POZ domain-containing protein 17	WT
Wdr44	1.235	Q6NVE8	Isoform 2 of WD repeat-containing protein 44	WT
Dock4	1.231	A0A1Y7VLY2	Dedicator of cytokinesis protein 4	WT
Magi2	1.216	A0A0G2JE00	Isoform 2 of Membrane-associated guanylate kinase_ WW and PDZ domain-containing protein 2	WT
Gm15800	1.215	Q6GQX8	AU042671 protein (Fragment)	WT
Atat1	1.212	Q8K341	Alpha-tubulin N-acetyltransferase 1	WT
Slmap	1.210	D3Z7V3	Isoform 2 of Sarcolemmal membrane-associated protein	WT
Cacnb4	1.179	J3QK20	Isoform 1 of Voltage-dependent L-type calcium channel subunit beta-4	WT
Atxn10	1.172	P28658	Ataxin-10	WT
Pten	1.141	O08586	Phosphatidylinositol 3_4_5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	WT
Slc12a7	1.121	Q9WVL3	Isoform 2 of Solute carrier family 12 member 7	WT
Ywhaq	1.105	F6VW30	14-3-3 protein theta (Fragment)	WT
Ppfia1	1.080	B2RXQ2	Ppfia1 protein	WT
Dhcr24	1.045	Q8VCH6	Delta(24)-sterol reductase	WT
Dnajb2	1.022	Q9QYI5	DnaJ homolog subfamily B member 2	WT
Alad	1.015	P10518	Delta-aminolevulinic acid dehydratase	WT
Ubr3	1.012	A2AV05	E3 ubiquitin-protein ligase UBR3 (Fragment)	WT
Ythdf2	1.004	Q91YT7	YTH domain-containing family protein 2	WT
Jmy	0.991	A0A0R4J0V4	Isoform 2 of Junction-mediating and -regulatory protein	WT
Irgq	0.969	Q8VIM9	Immunity-related GTPase family Q protein	WT

Gene name	Log2 Protein Abundance	Uniprot Accession	Description	Unique in
Sptbn4	0.947	E9PZC2	Spectrin beta_ non-erythrocytic 4	WT
Ppfia4	0.944	A0A087WPJ3	Liprin-alpha 4	WT
Lingo1	0.918	A9DA50	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1	WT
Nsun2	0.875	H3BKN0	Isoform 2 of tRNA (cytosine(34)-C(5))-methyltransferase	WT
Gm9044	0.859	A0A1W2P6Q8	Predicted gene 9040	WT
Armc2	0.859	Q3URY6	Armadillo repeat-containing protein 2	WT
Ythdf3	0.850	Q8BYK6	Isoform 2 of YTH domain-containing family protein 3	WT
Rraga	0.799	Q80X95	Ras-related GTP-binding protein A	WT
Polr1c	0.787	G3UX92	DNA-directed RNA polymerases I and III subunit RPAC1	WT
Lgi2	0.776	Q8K4Z0	Isoform 2 of Leucine-rich repeat LGI family member 2	WT
Asrgl1	0.771	Q8C0M9	Isoaspartyl peptidase/L-asparaginase	WT
Tom1l2	0.763	F6RBX1	Isoform 2 of TOM1-like protein 2	WT
Traf3	0.728	D3Z343	TNF receptor-associated factor 3 (Fragment)	WT
Enpp6	0.727	A0A1B0GS66	Ectonucleotide pyrophosphatase/phosphodiesterase family member 6	WT
Plxna3	0.723	P70208	Isoform 2 of Plexin-A3	WT
Ubqln1	0.711	Q8R317	Isoform 2 of Ubiquilin-1	WT
Kifc3	0.684	A0A0R4J220	Isoform 2 of Kinesin-like protein KIFC3	WT
Steap2	0.630	D3YUM6	MCG2960_ isoform CRA_a	WT
Luc7l3	0.553	Q5SUF2	Isoform 2 of Luc7-like protein 3	WT
Tmem30a	0.487	D3YVV1	Cell cycle control protein 50A	WT
Syngr3	0.448	Q8R191	Synaptogyrin-3	WT
Akt1	0.399	D3YXX3	RAC-alpha serine/threonine-protein kinase (Fragment)	WT
Trim2	0.288	E9QKC6	Tripartite motif-containing protein 2	WT
Ythdf1	0.035	A2AWN8	YTH domain family 1_ isoform CRA_a	WT
Gm13125	4.419	B1ARV6	PRAME family member 20	KO
Ndufa9	3.410	A0A0R3P9C8	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9_ mitochondrial	KO
Rpl5	3.137	D3YYV8	60S ribosomal protein L5 (Fragment)	KO
Hist1h1d	3.107	P43277	Histone H1.3	KO
Nnt	3.101	E9Q8F4	NAD(P) transhydrogenase_ mitochondrial	KO
Tgm2	**2.983**	**P21981**	**Protein-glutamine gamma-glutamyltransferase 2**	**KO**
Cpt1b	2.972	Q924X2	Carnitine O-palmitoyltransferase 1_ muscle isoform	KO
Plin1	2.971	A0A0U1RNP7	Isoform 2 of Perilipin-1	KO
Rpl7	2.930	F6XI62	60S ribosomal protein L7 (Fragment)	KO
Cd36	**2.910**	**Q08857**	**Platelet glycoprotein 4**	**KO**
Try10	2.841	Q792Z1	MCG140784	KO

Gene name	Log2 Protein Abundance	Uniprot Accession	Description	Unique in
Serpina1d	2.785	Q00897	Alpha-1-antitrypsin 1-4	KO
Ighg2b	2.690	A0A075B5P3	Immunoglobulin heavy constant gamma 2B (Fragment)	KO
Pdhb	2.654	Q9D051	Pyruvate dehydrogenase E1 component subunit beta_ mitochondrial	KO
Ewsr1	**2.557**	**Q5SUS9**	**RNA-binding protein EWS**	**KO**
Mylk	2.485	B1B1A8	Isoform 2 of Myosin light chain kinase_ smooth muscle	KO
Rpl13	2.439	P47963	60S ribosomal protein L13	KO
Ppfibp1	2.381	Q8C8U0	Isoform 2 of Liprin-beta-1	KO
Spata5	2.354	A0A0A0MQ80	Spermatogenesis-associated protein 5	KO
Mvp	2.249	E9Q3X0	Major vault protein	KO
Abcb6	**2.196**	**Q9DC29**	**ATP-binding cassette sub-family B member 6_ mitochondrial**	**KO**
Padi2	2.093	Q08642	Protein-arginine deiminase type-2	KO
Ldha	2.080	A0A1B0GQX5	L-lactate dehydrogenase (Fragment)	KO
Capn1	**2.057**	**O35350**	**Calpain-1 catalytic subunit**	**KO**
F13a1	2.028	Q8BH61	Coagulation factor XIII A chain	KO
Ywhae	1.952	D6REF3	14-3-3 protein epsilon	KO
Acat3	1.909	F2Z459	Acetyl-Coenzyme A acetyltransferase 3	KO
Plg	**1.900**	**P20918**	**Plasminogen**	**KO**
Rpl15	1.900	B8JKK2	60S ribosomal protein L15	KO
Fgg	**1.860**	**Q3UER8**	**Fibrinogen gamma chain**	**KO**
Acss1	1.850	Q99NB1	Acetyl-coenzyme A synthetase 2-like_ mitochondrial	KO
Hist1h2br	1.849	Q8CBB6	Histone H2B	KO
Golga5	1.828	Q9QYE6	Golgin subfamily A member 5	KO
Pck1	1.811	Q9Z2V4	Phosphoenolpyruvate carboxykinase_ cytosolic [GTP]	KO
Mut	1.810	P16332	Methylmalonyl-CoA mutase_ mitochondrial	KO
Sclt1	1.629	G5E861	Isoform 2 of Sodium channel and clathrin linker 1	KO
Bcat2	1.559	A0A1B0GST1	Branched-chain-amino-acid aminotransferase	KO
Acsl5	1.505	Q8JZR0	Long-chain-fatty-acid--CoA ligase 5	KO
Gfm1	1.469	Q8K0D5	Elongation factor G_ mitochondrial	KO
Lipe	**1.393**	**E9Q4M2**	**Hormone-sensitive lipase**	**KO**
Ugdh	1.379	O70475	UDP-glucose 6-dehydrogenase	KO
Aldh1a1	1.373	P24549	Retinal dehydrogenase 1	KO
Itih2	1.369	G3X977	Inter-alpha trypsin inhibitor_ heavy chain 2	KO
Rbm25	1.356	B2RY56	RNA-binding protein 25 (Fragment)	KO
Colgalt1	1.340	Q8K297	Procollagen galactosyltransferase 1	KO
Lbr	1.338	A0A0A6YW01	Lamin-B receptor (Fragment)	KO
Ighg2c	1.311	A0A0A6YY53	Immunoglobulin heavy constant gamma 2C (Fragment)	KO
Hibch	1.277	E0CX19	3-hydroxyisobutyryl-CoA hydrolase_ mitochondrial	KO
Isyna1	1.273	Q9JHU9	Inositol-3-phosphate synthase 1	KO

Gene name	Log2 Protein Abundance	Uniprot Accession	Description	Unique in
Nfia	1.266	B1AUB6	Isoform 1 of Nuclear factor 1 A-type	KO
Jakmip1	1.254	A0A1D5RLA4	Janus kinase and microtubule-interacting protein 1	KO
Aarsd1	1.192	Q3THG9	Alanyl-tRNA editing protein Aarsd1	KO
Fam129a	**1.189**	**E9PYV4**	**Protein Niban**	**KO**
Mapk9	1.186	Q9WTU6	Isoform Alpha-1 of Mitogen-activated protein kinase 9	KO
Ctrb1	1.068	Q9CR35	Chymotrypsinogen B	KO
Ilvbl	1.061	A0A1W2P727	Acetolactate synthase-like protein	KO
Ppia	1.038	P17742	Peptidyl-prolyl cis-trans isomerase A	KO
Tial1	1.013	A0A0U1RPD3	Nucleolysin TIAR (Fragment)	KO
Nans	1.007	Q99J77	Sialic acid synthase	KO
Cpq	0.924	A0A2I3BPG0	Carboxypeptidase Q (Fragment)	KO
Pbxip1	0.897	D3YUE1	Pre-B-cell leukemia transcription factor-interacting protein 1 (Fragment)	KO
Cul4a	0.884	Q3TCH7	Cullin-4A	KO
Snx6	0.880	A0A1W2P701	Sorting nexin-6	KO
Ppm1g	0.861	A0A0J9YVG0	Protein phosphatase 1G	KO
Asph	**0.842**	**A2AL85**	**Aspartyl/asparaginyl beta-hydroxylase**	**KO**
Eif2b5	0.830	Q8CHW4	Translation initiation factor eIF-2B subunit epsilon	KO
Ptpn6	**0.773**	**P29351**	**Isoform 2 of Tyrosine-protein phosphatase non-receptor type 6**	**KO**
Napb	0.754	D6RHL2	Beta-soluble NSF attachment protein	KO
Uap1l1	0.734	Q3TW96	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	KO
Sord	0.666	Q64442	Sorbitol dehydrogenase	KO
Abhd5	0.580	Q9DBL9	1-acylglycerol-3-phosphate O-acyltransferase ABHD5	KO
Rpa1	0.438	Q5SWN2	Replication protein A 70 kDa DNA-binding subunit	KO
Eln	0.407	P54320	Elastin	KO
L2hgdh	0.400	Q91YP0	L-2-hydroxyglutarate dehydrogenase_ mitochondrial	KO
Ccdc47	0.269	Q9D024	Coiled-coil domain-containing protein 47	KO
Gpi1	0.126	A0A0U1RP97	Glucose phosphate isomerase 1 (Fragment)	KO

Table 8.4 List of proteins uniquely detected in HSJ1 wild-type (WT) or HSJ1 knockout (KO) lumbar spinal cord extracts. Proteins detected in ≥2 biological samples from one condition and in none of the samples from the other condition were denoted unique in WT or KO as appropriate. Log_2 Protein abundance reflects normalised log_2 Hi3 protein abundance. Proteins in bold possess a GO Cellular compartment (GOCC) Plasma membrane annotation.

8.4. Co-occurrence of neuronal and astrocytic lineage markers in published datasets

Dataset	Neuronal markers	Lumbar	Primary MN	Adult SC	No match
Neuron Specific	Rbfox3 Nefl Nefm Nefh Gad1 Gad2 Th Dlg4 Syp Vamp1 Vamp2 Vapa Vapb Map2 Tubb3	Map2 Nefh Nefl Nefm Syp Tubb3	Map2 Nefh Nefl Nefm Syp Tubb3 Dlg4 Gad1 Gad2 Rbfox3 Vamp2 Vapa Vapb	Map2 Nefh Nefl Nefm Syp Tubb3 Th Vamp1	
Motor neuron specific	Chat En1 Evx1 Evx2 Fgf1 Mnx1 Isl1 Isl2 Lhx3 Nkx6 Ngfr Reg2 Sim1 Zeb1 Zeb2	Zeb2	Zeb2 Ngfr Zeb1	Zeb2 Chat Fgf1	En1 Evx1 Evx2 Mnx1 Isl1 Isl2 Lhx3 Nkx6 Reg2 Sim1
Cholinergic	Ache Chat Slc5a7 Slc18a3		Ache	Ache Chat Slc5a7	Slc18a3
Dopaminergic	Dbh Slc6a3 Slc6a2 Prkn Th			Dbh Slc6a2 Th	Slc6a3 Prkn
Serotonergic	Slc6a4 Tph2			Slc6a4 Tph2	
GABAergic	Slc6a1 Slca13 Slc6a11 Gad1 Gad2 Slc32a1	Slc6a11 Slc32a1 Gad1	Slc32a1 Gad1 Gad2 Slc6a1	Slc6a11 Slc32a1 Gad1 Gad2 Slc6a1	Slca13

Dataset	Neuronal markers	Lumbar	Primary MN	Adult SC	No match
Glutaminergic	Slc1a1 Glul Slc17a7 Slc17a6 Slc17a8		Glul Slc17a6 Slc1a1	Glul Slc17a6 Slc17a7	Slc17a8
Astrocytic	Gfap S100b Aqp4 Sox9 Aldh1l1	Aldh1l1 Gfap	Aldh1l1 Gfap Sox9	Aldh1l1 Aqp4 S100b	
Oligodendrocyte	Olig1 Mbp Olig2 Mog Olig3 Sox10 Cldn11 Osp	Mbp		Mbp Cldn11 Mog Sox10	Olig1 Olig2 Olig3 Osp

Table 8.5 List of cell lineage markers used in this study. Overlap with the indicated datasets is also shown.