Integration of Complex Phosphoproteomics and Network-Based Bioinformatics to Study Receptor Tyrosine Kinase Signalling

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2021

Joanne Watson

School of Biological Sciences

Table of Contents

List of Tables 4

List of Figures 6

List of Abbreviations 9

Abstract 11

Declaration 12

Copyright statement 13

Acknowledgements 14

Introduction to thesis 15

1 Introduction 17

1.1 Cells as systems 17
1.2 Receptor Tyrosine Kinase Signalling 19
1.3 Regulation of RTK signalling 38
1.4 Studying signalling processes with phosphoproteomics 60
1.5 Bioinformatics analysis of phosphoproteomics datasets 74
1.6 Aims and Objectives 85

2 Using Multilayer Heterogeneous Networks to Infer Functions of Phosphorylated Sites 87

2.1 Abstract 87
2.2 Introduction 87
2.3 Experimental Procedures 90
2.4 Results 95
2.5 Discussion 114
2.6 Conclusions 118

3 Spatially Resolved Phosphoproteomics Reveals FGFR2b Recycling-driven Regulation of Autophagy and Cell Survival 119

3.1 Abstract 119
3.2 Introduction 119
3.3 Results 122
3.4 Discussion 145
3.5 Methods 148

4 SubcellulaRVis: an app for the visualization of groups of proteins in subcellular compartments 163
4.1 Abstract 163
4.2 Author Summary 164
4.3 Introduction 164
4.4 Design and Implementation 166
4.5 Results 168
4.6 Availability and Future Directions 172
5 Ligand identity and concentration determine RTK signalling architecture 174
5.1 Introduction 174
5.2 Methods 176
5.3 Results 181
5.4 Discussion 195
6 Discussion 199
6.1 Summary of findings 199
6.2 Interpreting, predicting, and validating phosphorylation-mediated function 200
6.3 Scaling proteomics analysis methods to phosphoproteomics 202
6.4 Capturing dynamic signalling processes using phosphoproteomics 204
6.5 Biological insights on the regulation of FGFR and EGFR 207
6.6 Future perspectives 208
6.7 Conclusions 210
References 211
Appendix 1 253
Appendix 2 261

Word Count: 65682

List of Tables

Table 2-1 Multilayer heterogeneous network, constructed from MAPK/ERK validation data. GOBP terms were included as functional annotations. 97

Table 2-2 Multilayer heterogeneous network, constructed from Francavilla et al. EGF-regulated phosphorylated sites. GOBP terms were included as functional annotations.... 102

Table 2-3 Multilayer heterogeneous network, constructed from Francavilla et al. TGF-α-regulated phosphorylated sites. GOBP terms were included as functional annotations.... 103

Table 2-4. Multilayer heterogeneous network, constructed from Ruprecht et al. lapatinib-regulated phosphorylated sites in parental and lapatinib-resistant cell lines. KEGG pathways were included as functional annotations. 109

Table 2-5. Multilayer heterogeneous network, constructed from Ruprecht et al. lapatinib-regulated phosphorylated sites in lapatinib-resistant cell line. KEGG pathways were included as functional annotations. 109

Table 2-6. Multilayer heterogeneous network constructed using regulated sites from the Santra et al. dataset. GOBP terms were included as functional annotations. 112

Table 4-1. Results of GOCC enrichment from EnrichR on a list of 139 GPI-anchored proteins 166

Table 4-2. Number of child terms for each general parent term selected. 167

Table 4-3. Enrichment results using SubcellulaRVis for gene list of 139 GPI-anchored proteins. Break in table indicates non-significant enrichment. 170

Table 4-4. SubcellulaRVis enrichment results for each of the protein lists associated to cellular compartments by Go et al. (2021). The final column represents the overlap between the user-supplied gene list and genes annotated to the highest enriched SubcellulaRVis compartment. 171

Table 4-5. SubcellulaRVis enrichment results for each of the protein lists associated to cellular compartments by Nightingale et al. (2019). The final column represents the overlap between the user-supplied gene list and genes annotated to the highest enriched SubcellulaRVis compartment. 172

Table 5-1. List of reagents and consumables. 176

Table 5-2. R packages used for clustering. 181

Table 5-3. Number of regulated sites (following quantile filtering) in each stimulant condition. 185

Table 5-4. Number of clusters used for each stimulant condition. 189

List of Figures

Figure 1-1. Inputs, processing, and outputs of cellular systems. 18

Figure 1-2. Canonical signalling pathways downstream of receptor tyrosine kinase activation. 22

Figure 1-3. Schematic of EGFR (**A**) and FGFR (**B**) structure, phosphorylated sites following ligand activation, and recruited proteins. 32

Figure 1-4**.** Regulation of receptor tyrosine kinase signalling by ligands through A) functional selectivity and B) ligand abundance. 38

Figure 1-5. Trafficking routes of internalised receptor tyrosine kinases, resulting in recycling back to the plasma membrane or degradation in the lysosome. 44

Figure 1-6. Conceptualisation of dynamic signalling processes. 54

Figure 1-7. mTOR signalling is mediated from intracellular compartments. 57

Figure 1-8. Mass-spectrometry based (phospho)proteomics. 61

Figure 2-1. Overview of the multilayer heterogeneous network construction from phosphoproteomics data. 90

Figure 2-2. Details of validation data. 98

Figure 2-3. Site- and gene-centric analysis of validation data. 100

Figure 2-4. Distribution of annotations and classifications of sites extracted from PhosphoSitePlus. 102

Figure 2-5. Site-centric and gene-centric analysis of EGF-regulated sites from Francavilla et al. (2016) dataset. 106

Figure 2-6. Site-centric and gene-centric analysis of TGF-α-regulated sites from Francavilla et al. (2016) dataset. 107

Figure 2-7. Cluster containing sites annotated to “cell growth, inhibited” in PhosphoSitePlus, spiked into Francavilla et al. networks. 108

Figure 2-8. Site-centric and gene-centric analysis of Ruprecht et al. datasets. 111

Figure 2-9. Site-centric and gene-centric analysis of Santra et al. datasets. 114

Figure 3-1. Analysis of FGFR2b trafficking reveals FGFR2b activation is not affected by receptor sub-cellular localization. 123

Figure 3-2. Immunoblot analysis reveals FGFR2b activation is not affected by receptor sub-cellular localization. 124

Figure 3-3. Phosphoproteomics analysis identifies FGFR2b internalization- and recycling-dependent signalling pathways. 125

Figure 3-4. Characterisation of FGFR2b internalization- and recycling- signatures in HeLa_FGFR2b and T47D. 127

Figure 3-5. Design of spatially resolved phosphoproteome (SRP) experiment to elucidate FGFR2b and RAB11a proximal interactors. 129

Figure 3-6. APEX2 tagged-FGFR2b does not have an altered trafficking response downstream of FGF7 or FGF10 stimulation. 130

Figure 3-7. APEX2 tagged-FGFR2b and RAB11a identifies compartment-specific signalling partners upon FGF10 stimulation. 131

Figure 3-8. Workflow of the spatially resolved proteomics and phosphoproteomics experiments. 132

Figure 3-9. Spatially resolved proteomics and phosphoproteomics reveal FGFR2b-dependent regulation of mTOR signalling and autophagy. 134

Figure 3-10. FGFR2b regulates mTOR signalling and autophagy from the REs. 136

Figure 3-11. FGFR2b regulates autophagy from the REs 138

Figure 3-12. Phosphorylated ULK1 is found at RAB11-positive REs. 139

Figure 3-13. Phosphorylated ULK1 colocalises with FGFR2b at the RE. 140

Figure 3-14. Phosphorylated ULK1 recruitment at the REs depends on FGFR2b recycling. 141

Figure 3-15. mTOR activation is dependent on FGFR2b recycling. 142

Figure 3-16. Inhibiting the recycling of FGFR2b causes dysregulated autophagy and alters the balance of proliferation and cell death. 144

Figure 3-17. Model of FGFR2b global and proximal signalling partners during recycling to the plasma membrane. 145

Figure 4-1. SubcellulaRVis workflow 168

Figure 4-2. Visualization of enrichment results using SubcellulaRVis on a gene list of GPI anchored proteins (The UniProt Consortium, 2021). 169

Figure 4-3. Visualization of enrichment results using SubcellulaRVis, using the trafficking-specific endosomal view, on a gene list of GPI anchored proteins (The UniProt Consortium, 2021). 173

Figure 5-1. Set up and quality assessment of phosphoproteomics investigation of concentration-dependent regulation of intracellular EGF and FGF1 signalling. 183

Figure 5-2. Validation of phosphoproteomics approach. 185

Figure 5-3. Analysis pipeline of phosphoproteome and 36-hour proteome data. 186

Figure 5-4. Similarity between different clustering methods and measures to determine optimal number of clusters. 188

Figure 5-5. Number of GOBP terms resulting from RWHN applied to short-term phosphoproteome. 189

Figure 5-6. Gene targets of differentially phosphorylated transcription factors have altered protein abundance 36 hours following stimulation with high or low concentrations of EGF and FGF1. 191

Figure 5-7. Overlap of terms found in phosphoproteome and proteome analysis. 192

Figure 5-8. Networks based on kinase-substrate relationships (from PhosphoSitePlus) and other protein interactions (from STRING database). 194

Figure 5-9. Phosphorylation profiles of RAB7A-S72 and CCZ1-S266 downstream of EGF (**A**) and RAB12-S21 downstream of FGF1 (**B**). 195

List of Abbreviations

AMPK	AMP activated protein kinase
ANOVA	Analysis of variance
APEX	Ascorbate peroxidase
AP-MS	Affinity proteomics mass spectrometry
ATP	Adenosine 5’-triphosphate
DDA	Data dependent acquisition
DIA	Data independent acquisition
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ERK	Extracellular regulated kinase
FDR	False discovery rate
FGF	Fibroblast Growth Factor
FGFR	Fibroblast growth factor receptor
FRS2	FGFR substrate 2
GAB	GRB2-associated binder
GO	Gene Ontology
GOBP	Gene Ontology Biological Process
GOBP	Gene Ontology biological processes
GOCC	Gene Ontology Cellular Compartment
GOMF	Gene Ontology Molecular Function
HS	Heparan sulphate
IEG	Intermediate early genes
IGF	Insulin growth factor
JNK	c-Jun amino (N)-terminal kinases
KDE	Kernel density estimation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LFQ	Label free quantification
m/z	Mass-to-charge ratio
MAPK	Mitogen activated protein kinase
MAPKK	Mitogen activated protein kinase kinase
MAPKKK	Mitogen activated protein kinase kinase kinase
MS	Mass Spectrometry
mTOR	mechanistic Target of Rapamycin
mTORC1/2	mechanistic Target of Rapamycin complex 1/2
ORA	Overrepresentation analysis
PCA	principal component analysis
PI3K	Phosphoinositide 3-kinase
PIP2	Phosphatidylinositol-3,4-bisphosphate
PIP3	Phosphatidylinositol-3,4,5-trisphosphate
PKC	Protein kinase C
PLC	Phospholipase C

PPI	Protein-protein interaction
PPI	protein–protein interaction
PSP	PhosphoSitePlus
PTM	Post translational modification
RSK	p90 ribosomal S6 kinase 1-4
RTK	Receptor Tyrosine Kinase
RWHN	random walk on the heterogeneous network
RWR	random walk with restart
S	Serine
SHP2	SH2 domain–containing phosphatase 2
SILAC	Stable isotope labelling by amino acids in cell culture
SOS1	Son of sevenless 1
SPRY	Sprouty
T	Threonine
TGF-a	Transforming growth factor-α
TMT	Tandem mass tagging
TSC	Tuberous Sclerosis Complex
Y	Tyrosine

Abstract

In multi-cellular tissues and organisms, extracellular chemical stimuli are processed by cells via a network of interactions between protein molecules, that sequentially modify each other through post-translational modifications such as phosphorylation to transmit information throughout the cell. Mass spectrometry-based quantitative phosphoproteomics, which allows systematic identification of proteins modified by phosphorylation, has become an essential approach to study intracellular signalling.

The work in this thesis relates to the bioinformatics analysis of phosphoproteomics data. The modification of proteins downstream of different receptor tyrosine kinase (RTK) signalling cascades is known to change their function. An ongoing issue in the field of phosphoproteomics is the lack of phosphorylated site-centric analysis methods as alternatives to widely used gene- or protein-centric methods. As a solution to this, I propose a method to relate phosphorylated sites to biological functions using multilayer networks. By applying a random walk on heterogeneous network (RWHN) algorithm, phosphorylated sites are related to functional descriptors. This method outperformed the standard enrichment analyses.

Network-based bioinformatics approaches were further used to analyses complex phosphoproteomics datasets describing phosphoproteomic changes downstream of epidermal growth factor receptor (EGFR) and fibroblast growth factor receptor (FGFR) signalling cascades. We used a bespoke proximity-dependent biotinylation approach to generate a spatially resolved phosphoproteomics dataset describing global and recycling endosome-proximal phosphorylation events of activated FGFR2b. From this, we revealed that FGFR2b locally suppresses autophagy through the mTOR pathway from the recycling endosome, downstream of FGF10 activation. As a result of this work, I developed an application to visualise enrichment of GOCC terms. I also investigated the role of ligand concentration in regulating phosphorylation events and long-term proteome-level changes downstream of FGF1 and EGF stimulation. The analysis method utilised links short-term phosphorylation dynamics to long-term changes in protein abundance, downstream of different ligands and ligand concentrations, via modified transcription factors, to identify differential responses.

The work in this thesis demonstrates the valuable role of integrating network-based methods in the analysis of phosphoproteomics datasets in order to better understand RTK signalling cascades.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning

Copyright statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420, in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

First, I would like to thank my supervisors, Dr Jean-Marc Schwartz and Dr Chiara Francavilla for their supervision and guidance of my PhD work. They were always generous with their advice and support, and constantly encouraged me to improve and take opportunities. These are qualities that I was especially grateful for when the pandemic hit, and I would specifically like to thank them for all the extra support they gave during this time. I am a more confident researcher from having worked with them.

I would like to acknowledge the people I have had the opportunity to work with during my PhD. Thank you to Pat Caswell for his advice on the direction of project and his generosity in letting me learn from him and his team during the first year of the project. A special thanks goes to all the current and previous members of the Francavilla and Schwartz groups: Jamie Soul, Paul Fullwood, Heyuan Mitchell, Harriet Ferguson, Joe Parsons, Michael Smith, Jen Haworth, and the many talented students who have passed through the labs. The amount of help I received from you all would be impossible to describe succinctly here; I am very grateful for it all. It has been a pleasure to work with you all.

I thank my funding body, the Biotechnology and Biological Sciences Research Council, for their financial support of this work, through which I had the opportunity to attend conferences and gain work-experience abroad.

Finally, thank you to my mum, dad and sister, who gave me everything I needed to succeed and always made me feel loved and capable. To my partner, Gabriel, whose support truly carried me whilst I was doing this work: thank you for not letting me give up.

Introduction to thesis

This thesis is being presented in the alternative format in accordance with the rules and regulations of the University of Manchester. As chapter 2 has already been published in a peer-reviewed journal and chapters 3 and 4 have been submitted and are under consideration, the presentation of the thesis as publication style chapters was appropriate. Chapter 5 contains on-going work that will be submitted for publication in the future. Each chapter includes a self-contained introduction, methodology, results and/or discussion section; chapters 2, 3 and 4 also contain an abstract.

Below are the details of each chapter, details of contributions to the work and, where appropriate, the journal of submission. An asterisk (*) indicates the corresponding author and a hash (#) indicates equal contribution of authors, where appropriate. Supplementary figures have been included as appendices.

Chapter 2: Using Multilayer Heterogeneous Networks to Infer Functions of Phosphorylated Sites.
Joanne Watson, Jean-Marc Schwartz, and Chiara Francavilla*. Journal of Proteome Research 20, no. 7 (July 2, 2021): 3532–48. https://doi.org/10.1021/acs.jproteome.1c00150.

JW designed the method, performed all analyses, and wrote the manuscript. *JMS* conceived the idea. *JMS* and *CF* supervised the work and edited the manuscript.

Please note that the figures have been modified from the original publication in order to meet the formatting requirements of the thesis.

Chapter 3. Spatially Resolved Phosphoproteomics Reveals FGFR2b Recycling-driven Suppression of Autophagy and Cell Survival.

Joanne Watson[#], Harriet R. Ferguson[#], Rosie M. Brady, Jennifer Ferguson, Paul Fullwood, Hanyi Mo, Katherine H. Bexley, David Knight, Jean-marc Schwartz, Michael P. Smith*, Chiara Francavilla* *Under revision at Nature Communications and available on BioRXiv.*

JW performed bioinformatics analysis of the proximal APEX2 experiment, provided ideas for further experimentation, and interpreted results. *MPS* and *RB* performed the dominant negative perturbation experiment; *MPS*, *RB* and *PF* prepared samples for mass spectrometry. HF analysed dominant negative perturbation dataset. *MPS* performed the proximal APEX experiment. *DK* and the BioMS facility at the University of Manchester performed all mass

spectrometry analysis. *MPS* and *CF* performed all non-proteomics experimental analysis. *HM* and *KB* provided data analysis support. *MPS* supervised *RB* and *KB*. *JMS* and *CF* supervised *JW* and HM. *CF* conceived the idea, supervised the people involved in the work and obtained funding for the project. *JW*, *HF*, *MPS* and *CF* co-wrote the manuscript.

Chapter 4. SubcellulaRVis: an app for the visualization of groups of proteins in subcellular compartments.
Joanne Watson*, Michael Smith, Chiara Francavilla* and Jean-Marc Schwartz*. *To be submitted to PLOS Computational Biology. Written in the style of the journal's Software Article requirements.*

JW designed and implemented the app and wrote the manuscript. All authors contributed to the idea. *MPS* provided user-testing. *CF* and *JMS* supervised the work and edited the manuscript.

Chapter 5. Ligand identity and concentration determine RTK signalling architecture.

Joanne Watson, Hanyi Mo, Siqiang Fan, Paul Fullwood, Jennifer Ferguson, David Knight, Jean-Marc Schwartz, Chiara Francavilla*.

JW performed the experimental work and bioinformatics analysis. *JW* and *PF* prepared samples for mass spectrometry. Ideas were contributed by *HM* and *SF* as part of their masters' and undergraduate dissertations, respectively. The *BioMS facility* at University of Manchester performed all mass spectrometry analysis. *CF* conceived the idea and designed the phosphoproteomics experiment. *JMS* and *CF* supervised the work.

Below are details of publications to which I have contributed but are not included in the thesis:

Watson J and Francavilla C. (2018) "Regulation of FGF10 Signaling in Development and Disease." *Frontiers in Genetics* 9; 500. https://doi.org/10.3389/fgene.2018.00500.

Smith MP, Ferguson HR, Ferguson J, Zindy E, Kowalczyk KM, Kedward T, Bates C, Parsons J, **Watson J**, Chandler S, Fullwood P, Warwood S, Knight S, Clarke RB, Francavilla C (2021) "Reciprocal priming between receptor tyrosine kinases at recycling endosomes orchestrates cellular signalling outputs." *The EMBO Journal*. 40: e107182.

1 Introduction

1.1 Cells as systems

Cells existing singularly or in the context of tissues, organs and multicellular organisms are constantly sensing changes in their environments. The process of sensing and responding to such changes is critical for cellular survival and homeostasis in multicellular organisms (Schmidt, Dall and Gils, 2010; Frieden and Gatenby, 2011). In mammalian cells, a highly tuned and complex system has evolved to receive, process, and respond to mechanical, physical, and chemical stimuli (Figure 1-1) (Gatenby, 2019). The intracellular processing of the stimuli is referred to as cellular signalling. Several layers of molecular regulation orchestrate the intracellular response to extracellular stimuli, however intracellular proteins are the primary processors. Once the signal has traversed the plasma membrane via cell surface receptors, it will be passed throughout the cell by the sequential post-translational modification (PTM) of proteins in a signalling cascade. PTM of proteins ultimately impacts protein function, and can cause their activation, changes in subcellular localisation or degradation (Day, Sosale and Lazzara, 2016). A key outcome of PTM-mediated signalling cascades is the activation of transcription factors, which alter gene expression within the cell. This may induce greater expression of target proteins or nucleic acid regulators (e.g. micro RNAs, miRNA) (Balázsi and Oltvai, 2005). Ultimately, the signalling cascades induce changes in cellular behaviour, through altered proliferation, differentiation, and migration, or by tipping the balance of cell fate towards survival or death.

'Omics approaches aim to define the molecular layers in a cellular system: genomics describes genes; transcriptomics describes RNA transcripts; proteomics describes proteins. In cellular signalling there is a clear interplay between these layers; regulation of genomes and gene expression is influenced by actions within the proteome, which will alter the transcriptome. The relationship between 'omic layers is particularly pertinent in the context of disease. For example, cancer, which can be considered a disease either of gene mutation or tissue disorganisation, is ultimately driven by intracellular protein action resulting from either type these changes (Soto and Sonnenschein, 2011; Yaffe, 2013, 2019). It has been argued that the candidates for personalised cancer medicine should be sought in proteomics profiling (Doll, Gnad and Mann, 2019).

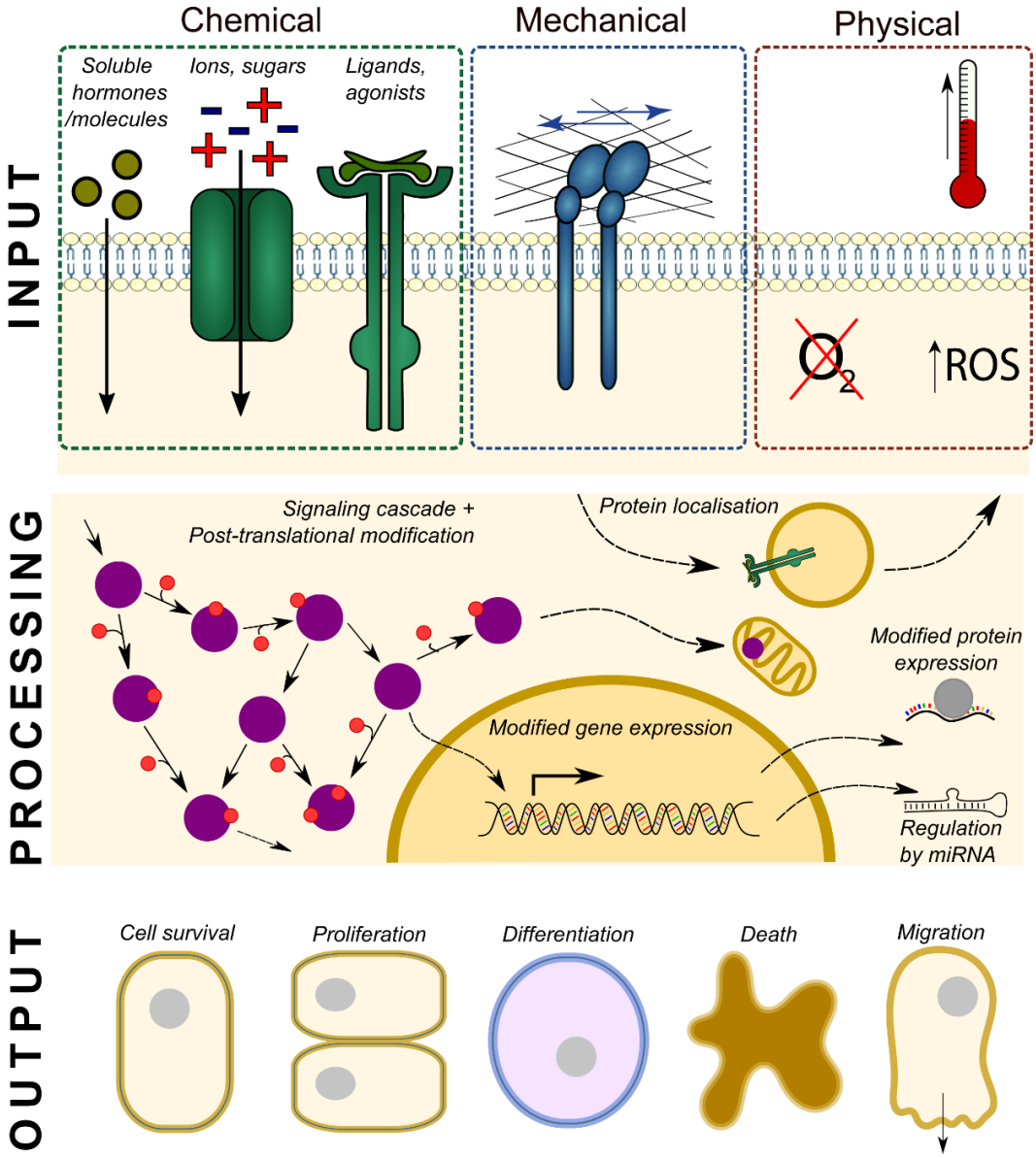


Figure 1-1. Inputs, processing, and outputs of cellular systems.
Cells are systems that receive inputs from their environment and process them intracellularly, resulting in output changes in cellular behaviour.

In the context of cellular signalling, however, protein action is driven by PTM induced through the initiating stimuli. PTMs are either proteolytic cleavages, resulting in changes to the polypeptide structure of the protein, or covalent additions to amino acid side chains (Pang and Wilkins, 2019). PTM has important effects on individual proteins, including (but not limited to) conformation changes; anchoring to membranes or inducing re-localisation; switching potential binding partners; and altering protein stability (Pang and Wilkins, 2019). Even minor modifications can alter whole protein structure and function (Tang and Kaneko, 2020).

By far the most common PTM in cellular signalling processes is the reversible phosphorylation of serine (S), threonine (T), and tyrosine (Y) residues (Day, Sosale and Lazzara, 2016). Around 75% of eukaryotic proteins will undergo phosphorylation during their lifespan, though the number of phosphorylation sites considered functional is still under speculation (Venerando, Cesaro and Pinna, 2017; Ochoa *et al.*, 2019). The modification is the product of phosphate transferred from adenosine 5'-triphosphate (ATP), catalysed by enzymes called kinases and reversed by phosphatases (Hunter, 1995). Kinases can be categorised into two groups: S/T or Y kinases, named based on the amino-acid residues they show specificity for. Histidine kinases and phosphorylation are widespread in prokaryotes, and more common in eukaryotes than previously thought, but their role in eukaryotic cellular signalling is minor or unknown (Adam and Hunter, 2018); for this reason, only to S, T and Y phosphorylation is referred to in this thesis. Work in the last 15 years has shown that phosphorylation primarily occurs in intrinsically disordered regions of proteins (Collins *et al.*, 2008; Palmeri *et al.*, 2014). The current hypothesis is that intrinsically disordered regions allow for enzymes and binding proteins to gain access to target sequences or phosphate groups, to regulate local protein conformation and activity (van der Lee *et al.*, 2014). Ultimately, phosphorylation is a dynamic and transient modification, and is reversed through the action of phosphatases (Yadav *et al.*, 2017). The tight regulation of kinases and phosphatases dictates their competitive effects on target protein function and structure.

1.2 Receptor Tyrosine Kinase Signalling

Receptor Tyrosine Kinases (RTKs) are a superfamily of cell-surface receptors comprised of 58 known members in humans, which belong to 20 subfamilies (Wintheiser and Silberstein, 2021). The work in this thesis pertains to two of the RTK subfamilies: epidermal growth factor (EGF) receptor (EGFR)/ErbB and fibroblast growth factor (FGF) receptor (FGFR). In this section, I will describe commonalities between the structure, function and downstream signalling cascades of the RTK subfamilies before describing the specifics of EGFR and FGFR in sections 1.2.2 and 1.2.3, respectively.

1.2.1 Molecular background of Receptor Tyrosine Kinases

1.2.1.1 Receptor tyrosine kinases share a common structure & activation mechanism.

Broadly, RTKs are structurally characterised by an extracellular region, or ectodomain, containing the ligand binding domain; a single pass transmembrane region; and a cytoplasmic tail that contains the kinase domains. The ectodomains, which are a modular structure, are the most divergent between the different RTKs and are used to classify the

subfamilies (Lawrence and Ward, 2015). Some module types are commonly found in RTK ectodomains, such as the immunoglobulin-like (Ig-like) (including the Ig domain, fibronectin Type III domain and cadherin domains), which is found in the ectodomains of 10 of the 20 RTK subfamilies (Bork, Holm and Sander, 1994; Lemmon and Schlessinger, 2011). Others are more restricted to specific RTK subfamilies, such as laminin/EGF-like cysteine-rich domains which are only found in ErbB and insulin RTK subfamilies (Bork *et al.*, 1996). The ectodomains differ not only in the type of modules contained within them but also in the number of modules and totally size of the ectodomain, which ranges 40 to 1830 residues across RTK superfamily (Lawrence and Ward, 2015).

In the classic model of RTK activation, monomeric RTKs form oligomers, usually dimers, upon ligand binding. The oligomerisation and structural change from ligand binding induce kinase activity, leading to autophosphorylation and transphosphorylation of the RTK cytoplasmic tail. This may be partially facilitated by co-binding partners, such as heparan sulphate (HS) for the fibroblast growth FGFR family (Guimond *et al.*, 1993; Pellegrini *et al.*, 2000; Kreuger *et al.*, 2005). Insulin receptor (IR) is the exception to the dynamic oligomerisation of the RTK family, since it exists as a stable homodimer covalently linked by disulphide bonds; maximal ligand binding induces a conformational change, moving the IR out of the autoinhibited state (Sparrow *et al.*, 1997; Uchikawa *et al.*, 2019). Each successive phosphorylation event has a modest effect on catalytic activity, with the first event having the most significant impact (Favelyukis *et al.*, 2001). This is demonstrated by FGFR1, which has been shown to require sequential phosphorylation of Y653 followed by Y654 (both in the activation loop of the kinase domain) for maximal kinase activity of the receptor (Furdui *et al.*, 2006). It should be noted that there is evidence that ligand binding does not necessarily precede dimerization. In models based on EGFR, it has been suggested that dimerization doesn't necessarily encourage ligand binding, but part of the role of ligand binding could be in encouraging dimerization (Bessman *et al.*, 2014). Indeed, dimerization does not appear to be critical for auto- and transphosphorylation, rather the ligand binding is (Freed, Alvarado and Lemmon, 2015); instead, dimerization essentially primes the receptor for activation (Chung *et al.*, 2010). Moreover, the affinity of the ligand for the ligand-binding region will determine the stability of the receptor and have downstream effects on the length of the immediate response to the ligand binding. This has been demonstrated by studies on EGFR, describing how different ligands that show differential affinity for the EGFR result in changes in dimer conformation and the length of time the receptor is dimerised (Freed *et al.*, 2017).

Phosphorylated tyrosine residues on activated RTKs act as recruitment or docking sites for adaptor proteins. The recruited proteins fall into two main categories: those containing src homology 2 (SH2) or phosphotyrosine binding (PTB) domains, or specialised docking proteins. Canonical examples of the former include the adaptor proteins Grb2 and Shc1, alongside well-studied signalling proteins phosphoinositide 3-kinase (PI3K, specifically the p85 subunit), phospholipase C- γ (PLC-γ) and Ras-GTPase (guanosine triphosphatase)-activating protein (Ras-GAP) (Schlessinger and Lemmon, 2003; Pawson, 2004). Docking proteins, on the other hand, often only interact with specific RTKs. For example, FRS2 is mainly recruited to FGFR or nerve growth factor (NGF) receptors, whilst insulin receptor substrate 1-4 (IRS1-4) is recruited to insulin growth factor (IGF) receptors (Schlessinger, 2000). These docking proteins contain tyrosine phosphorylation sites which are substrates for their associated RTK; they act as platforms for the recruitment of more signalling and adaptor proteins (Lemmon and Schlessinger, 2011). Since there are many phosphorylated residues on the receptor which can recruit proteins, it serves as a platform around which a signalling complex can form. Work in EGFR has demonstrated that turnover of phosphorylation, through potent phosphatase activity, at the receptor following activation is in the order of a few seconds (Kleiman et al. 2011). Thus, the signalling complex at the receptor is highly dynamic and dependent on the length of the immediate-early response to the ligand, which will be determined by oligomer stability, ligand binding affinity, and, potentially, ligand abundance.

1.2.1.2 Receptor tyrosine kinases activate a common set of signalling pathways.

Signalling cascades downstream of RTKs are often mediated by a common set of signalling molecules, which, when first discovered, were considered to compose several distinct, linear pathways (Figure 1-2). As more work has been done, the interconnectivity of these signalling modules has been better understood, however the modules are still regularly discussed and considered as separate entities. The best studied of these pathways included those involving mitogen activated protein kinase (MAPK) family members, AKT/PI3K, and PLCγ/ protein kinase C (PKC), which will be discussed in more detail.

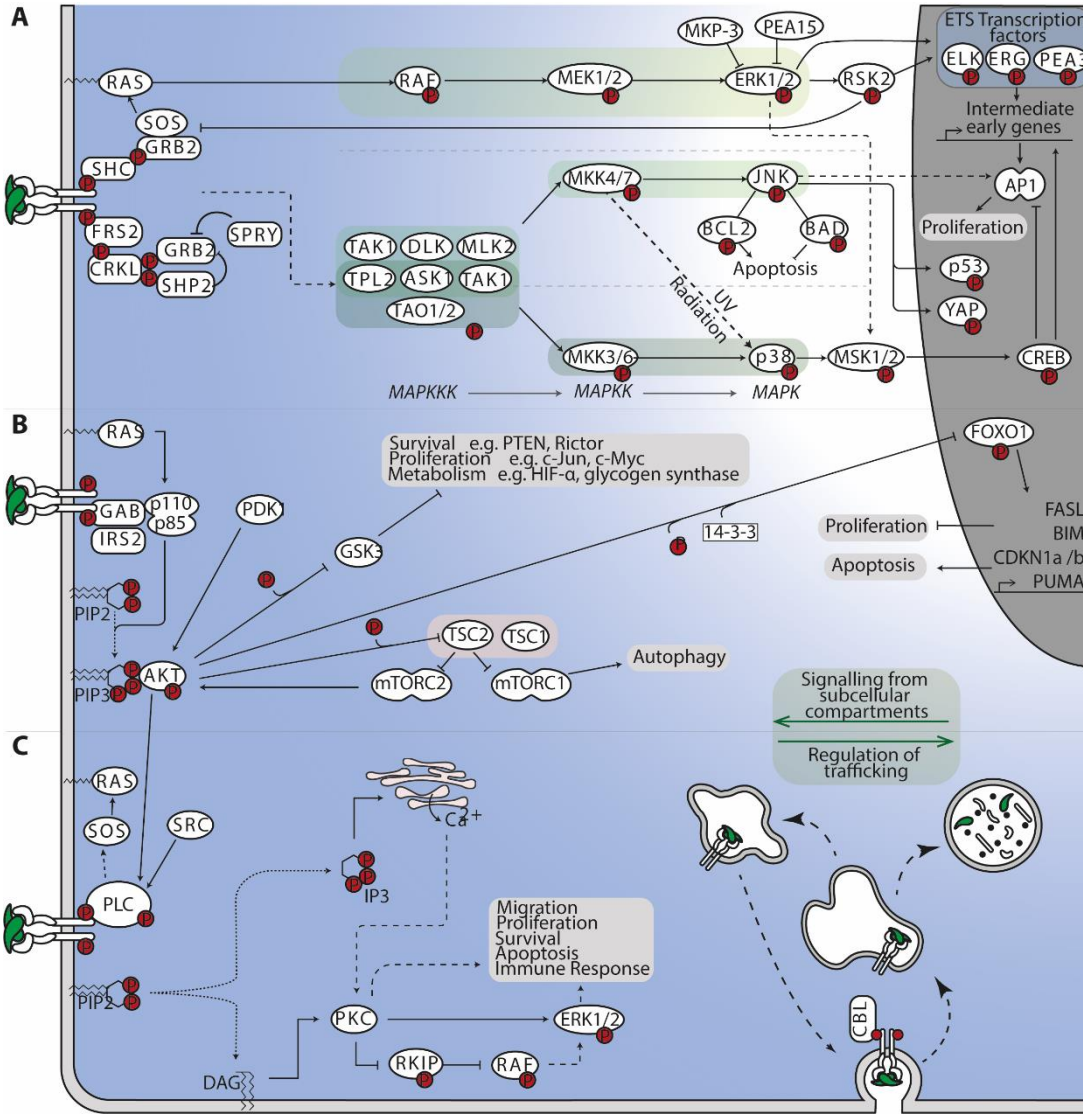


Figure 1-2. Canonical signalling pathways downstream of receptor tyrosine kinase activation.
A. *MAPK pathways.* ***B.*** *PI3K/AKT, and* ***C.*** *PLC.*

1.2.1.2.1 MAPK Signalling pathways

The MAPK family is highly conserved throughout eukaryotes, and can be characterised into seven groups in mammals (Widmann *et al.*, 1999). The groups containing the canonical MAPKs are: extracellular regulated kinase (ERK) 1/2, c-Jun amino (N)-terminal kinases 1/2/3 (JNK1/2/3), p38 isoforms (α, β, γ, and δ), and ERK5 (Cargnello and Roux, 2011). The latter is the least studied, with the others having been extensively investigated in numerous contexts. ERK, JNK and p38 all mediate aspects of proliferation, survival, and growth, however family specific roles include migration by ERK and inflammation and apoptosis induction by JNK and p38 (Lavoie, Gagnon and Therrien, 2020). Accordingly, ERK1/2 (and ERK5) are the main family activated downstream of RTKs, with p38 and JNK more frequently associated with activation downstream of cellular stress (e.g. UV, hypoxia, nutrient deprivation) and cytokine signalling,

though with known crosstalk to the ERK pathway (Beyer and MacBeath, 2012). The signalling modules up- and downstream of the MAPKs ERK1/2, p38 and JNK share a common, three-tiered structure (Figure 1-2A), where the MAPK is phosphorylated and activated by a specific set of kinases (MAPKK), and the MAPKKs activated but a further upstream level of kinases (MAPKKKs).

The best characterised activation of the ERK signalling module is largely initiated from the plasma membrane. In the context of RTKs, the first stage is the recruitment of son of sevenless 1 (SOS1) to the activated receptor by Grb2; SOS1 is a guanine exchange factor that acts on members of the membrane-bound RAS family, such as KRAS, stimulating their GTPase activity (Kessler *et al.*, 2021). RAS family members activate MAPKKKs, including RAF family members such as RAF-1 and BRAF. RAF phosphorylates the MAPKKs MEK1/2 (also known as MAP2K1, MAP2K3), which subsequently activate ERK1/2. It has been shown that the two isoforms of ERK and MEK have functional redundancy, and therefore are generally considered as indistinct (Frémin *et al.*, 2015; Aoidi, Maltais and Charron, 2016). There is evidence that MEK activation of ERK requires a processive, sequential phosphorylation (Shi *et al.*, 2015), though this diverges from early evidence suggesting that the order of phosphorylation is not important for ERK activation (Burack and Sturgill, 1997; Aoki *et al.*, 2011).

In its inactive, unphosphorylated state, ERK is found in the cytoplasm, where it is bound to the inhibitors PEA-15 and MKP-3 (Formstecher *et al.*, 2001; Karlsson *et al.*, 2004). Due to the nuclear-export sequence on PEA-15, ERK is maintained in the cytoplasm. However upon phosphorylation, it can be imported to the nucleus (Chen, Sarnecki and Blenis, 1992), where many of its substrates reside. The localisation of ERK highlights the contested role of dimerization in ERK activity. Some evidence suggests that ERK binds its nuclear targets as a monomer and its cytoplasmic targets as a dimer (Casar, Pinto and Crespo, 2008; Herrero *et al.*, 2015). However, a dimerization-dependent, autocatalytically phosphorylated residue, T190, was demonstrated to be required for nuclear import (Lorenz *et al.*, 2009; Lai and Pelech, 2016). It has long been accepted that active ERK performs its roles as a dimer, but in light of the contrasting evidence the current hypothesis is that upon T190 phosphorylation the ERK dimer dissociates and the activated ERK monomers are imported into the nucleus, but this remains to be proven (Lavoie, Gagnon and Therrien, 2020).

ERK substrates have a short P-XS/T-P consensus motif, where X represents any residue (Gonzalez, Raden and Davis, 1991). The frequency of the target motif makes the proximal

docking sites crucial to maintaining true ERK-substrate specificity. These docking sites are known as D- and F- motifs; though not found at every ERK site, they increase substrate affinity of ERK (Lee *et al.*, 2004). Over 2500 putative ERK substrates, on more than 1000 proteins, have been identified, largely from high-throughput phosphoproteomic studies (Ünal, Uhlitz and Blüthgen, 2017). These substrates include negative feedback regulators of the pathway, for example through inhibitory phosphorylation at the level of RTK, adaptors and core RAS-ERK signalling module.

The activation of ERK signalling by mitogenic growth factors such as EGF is associated with the rapid induction of a transcriptional programme. ERK phosphorylates transcription factors from the ETS family, such as ELK, ERG and PEA3, which bind to the promotors of the so-called intermediate-early genes (IEG) (Verger *et al.*, 2001; Hollenhorst *et al.*, 2011; Hollenhorst, McIntosh and Graves, 2011; Padul *et al.*, 2015). Also activated by ERK is the ETS transcriptional repressor ERF, which in resting cells suppresses IEG transcription but dissociates and is exported to the cytoplasm upon ERK phosphorylation (Sgouras *et al.*, 1995; le Gallic *et al.*, 1999). A key component of IEG upregulation is the phosphorylation of p90 ribosomal S6 kinase 1-4 (RSK1-4) by cytoplasmic ERK. RSK1-4 belong to the MAPK activated protein kinase (MAPKAPK) family and are thought to be the evolutionary product of a gene fusion event, giving these proteins two kinase domains from two distinct protein kinases (Jones *et al.*, 1988; Fisher and Blenis, 1996). They therefore share homology not only with other MAPKAPK members but also with the AGC family of kinases, which includes PKC, AKT and S6Ks (Bjørbaek, Zhao and Moller, 1995). A major consequence of RSK activation by ERK is the expanded repertoire of cellular processes mediated downstream of ERK signalling, such as actin cytoskeleton regulation (Woo *et al.*, 2004; Tanimura *et al.*, 2016) and glucose metabolism (Houles *et al.*, 2018). One way in which RSK enhances ERK-mediated response is through the phosphorylation and activation of the transcription factor serum response factor (SRF), which works in concert with the ERK-phosphorylated ELK1, ELK3 and ELK4 to regulate recruitment of the transcription machinery to IEG promotors (Rivera *et al.*, 1993; Mylona *et al.*, 2016; Uhlitz *et al.*, 2017). RSK1 and RSK2 are also responsible for negative feedback of upstream signalling proteins such as SOS1 (Lavoie, Gagnon and Therrien, 2020).

The best characterised IEGs belong to the EGR, FOS, JUN and MYC families of transcription factors, many of which are themselves ERK targets. For example, phosphorylation of FOS and JUN family members by ERK increases their stability and DNA-binding activity. Phosphorylated FOS and JUN form transcriptionally active AP-1 complexes, which induce the

expression of the G1-phase cyclin D1, driving cell cycle progression (Albanese *et al.*, 1995; Lavoie *et al.*, 1996; Shaulian and Karin, 2001; Ünal, Uhlitz and Blüthgen, 2017). The IEG transcriptional programme also includes the expression of negative regulators such as phosphatases, from the DUSP family as well as PTPN7, and the RAS inhibitor SPRY4 (Ünal, Uhlitz and Blüthgen, 2017).

JNK is activated by the MAPKKs MKK4 and MKK7, which are phosphorylated by a diverse range of MAPKKKs (Krishna and Narang, 2008). Different MAPKKKs are activated depending on the stimuli being responded to, for example TAK1 in response to inflammatory cytokines and toll like receptors (Wan *et al.*, 2006). Like ERK, once activated JNK can relocate to the nucleus, where its known targets are found; though a proportion of activated JNK remains in the nucleus, and it has roles in extra-nuclear processes (e.g. endoplasmic reticulum stress, response to growth factor withdrawal), few high-confidence JNK cytoplasmic targets have been identified (Mizukami *et al.*, 1997; Cargnello and Roux, 2011). These include inhibition of the pro-apoptotic Bcl-2 homology domain 3 (BH3)-only Bcl-associated death protein (BAD) (Deng *et al.*, 2011), and the activation of antiapoptotic BCL-2, which subsequently induces autophagy in response to nutrient deprivation (Wei *et al.*, 2008). One common target between JNK and ERK signalling is c-Jun; JNK activity also induces AP-1 complex formation and upregulation of cell cycle genes such as cyclin D1 (Sabapathy *et al.*, 2004). However, numerous other transcription factor targets of JNK have been identified which regulate processes relating to cell survival and stress, such as p53 and Yes-activated protein (YAP) (Tomlinson *et al.*, 2010).

The final canonical MAPK family, p38, bears more resemblance in terms of function and interactors to JNK than ERK. The MAPKKs responsible for activating p38 are MKK3 and MKK6, however in response to UV radiation the JNK MAPKK, MKK4, can also phosphorylate p38 (Dérijard *et al.*, 1995; Enslen, Raingeaud and Davis, 1998; Brancho *et al.*, 2003). Most MAPKKKs are also shared between the pathways, including TAK1. An important target of p38 is the mitogen- and stress-activated kinases (MSK1, MSK2), which are MAPKAPKs like RSK. MSK1/2 are also activated by ERK; evidence suggests that MSK can be activated by p38 or ERK signalling modules regardless of stress or mitogenic stimulus, indicating a point of crosstalk between the two modules (Deak *et al.*, 1998). The best described role of MSK is the activation by phosphorylation of the transcription factor cAMP response element-binding protein (CREB), which participates in the transcriptional activation of IEGs such as c-Fos, JunB and Egr1 (Lonze and Ginty, 2002). Other key p38 targets are those relating to cell cycle

progression. Unlike ERK and JNK, p38 downregulates cyclins and upregulates CDK inhibitors; for example, the actions of p38 induce the G1/S checkpoint, through the stimulation of cyclin D1 degradation (Thoms, Dunlop and Stark, 2007). Through the activation of HBP1, cyclin D1 gene transcription is inhibited and p53 is activated, inducing proliferation arrest (Stramucci, Pranteda and Bossi, 2018; Bollaert, Serra and Demoulin, 2019). The ability to activate p53 points to p38's role in responding to cellular stress, which is further highlighted from evidence in breast cancer cells, where p38 was shown to coordinate the DNA damage response. Pharmacological inhibition of p38 increased replication errors, DNA damage and chromosome instability (Cánovas *et al.*, 2018) . Contrastingly, in some hematopoietic cells and cancer cell lines, p38 can positively regulate proliferation, indicating it is under mitogenic regulation alongside cell stress (Wagner and Nebreda, 2009; Faust *et al.*, 2012).

1.2.1.2.2 PI3K/AKT Signalling Pathway

PI3K family is comprised of four classes. They are responsible for catalysing the phosphorylation of the 3-OH group of inositol membrane lipids, regulating many aspects of cell function (Vanhaesebroeck, Stephens and Hawkins, 2012; Vanhaesebroeck *et al.*, 2021). Activated class I PI3Ks are well-characterised effectors of activated RTKs; they phosphorylate phosphatidylinositol-3,4-bisphosphate (PIP2) to generate the second messenger phosphatidylinositol-3,4,5-trisphosphate (PIP3) (Fruman *et al.*, 2017) (Figure 1-2B). The enzyme is a constitutive dimer, comprised of a p110 catalytic subunit and a p85 regulatory subunit. There are four p110 isoforms, encoded by *PIK3CA*, *PIK3CB*, *PIK3CG*, and *PIK3CD*. The primary role of the regulatory subunit is in regulating the activity and subcellular localisation of the complex. The class I complex will be referred to as PI3K hereon (Vanhaesebroeck, Stephens and Hawkins, 2012).

The PI3K is recruited via the regulatory subunit to activated RTKs, either by directly binding to phosphorylated tyrosine residues on the receptor or via adaptor proteins such as insulin receptor substrate 1 (IRS1) or GRB2-associated binder (GAB) (Burke and Williams, 2015). Binding the receptor achieves several key changes. The first is a conformational change; in the inactive state, SH2 domains on the regulatory subunit form inhibitory contacts with the helical, kinase, and C2 domains of the catalytic subunit. Upon recruitment via the SH2 domains to the receptor, these contacts are relieved (Backer, 2011). Once in the active conformation, the complex is positioned with the catalytic subunit towards the cytoplasmic face of the plasma membrane, where its PIP2 substrates are. In the case of the p110α catalytic subunit, once recruited the complex can also receive input from members of the membrane-anchored Ras superfamily (Burke and Williams, 2015). The actions of PI3K can be

reversed by phosphatase and tensin homolog (PTEN), which reverses PIP3 phosphorylation. PTEN is an important regulator of PI3K activity and a tumour-suppressor.

PIP3 is crucial for recruiting pleckstrin-homolgy (PH) domain effector proteins to the signalling complex at the plasma membrane. The best known of these effectors is AKT, also known as protein kinase B. Once recruited to the plasma membrane, AKT can be phosphorylated and activated by 3-phosphoinositide-dependent protein kinase-1 (PDK1) (Higuchi *et al.*, 2008). Maximal AKT activity is achieved through phosphorylation by the mTORC2 complex, which is associated with cell survival and metabolism. mTORC2 is recruited to PIP3 at the plasma membrane by the PH-domain in SIN1 (Sarbassov *et al.*, 2005; Gaubitz *et al.*, 2016). Active AKT has greater than 100 known substrates (Sugiyama, Fairn and Antonescu, 2019), with a consensus sequence of R-X-R-X-X-S/T-φ (where X is any amino acid and φ denotes a preference for large hydrophobic residues) (Manning and Toker, 2017). Phosphorylation by AKT is generally, though not exclusively, associated with inhibition or downregulation of the target protein.

The best-characterised substrates are found on the Forkhead Box O Family (FOXO) transcription factors, Glycogen Synthase Kinase 3 (GSK3) and Tuberous Sclerosis Complex 2 (TSC2). All of these release suppressions on cell proliferation and survival. FOXO induces translation of the pro-apoptotic proteins FASL, BIM and PUMA, as well as cyclin inhibitors such as CDKN1a and CDKN1b (Webb and Brunet, 2014; Manning and Toker, 2017), among others. AKT phosphorylates FOXO on three sites, creating a binding motif that is recognised by 14-3-3 proteins. Upon 14-3-3 binding, FOXO is exported from the nucleus and its dampening effects on target genes released. The second AKT target, GSK3, is a promiscuous protein kinase (Hermida, Dinesh Kumar and Leslie, 2017). In the absence of exogenous signals, GSK3 is active and, like AKT, it generally downregulates its targets, which are involved in cell proliferation (e.g. c-Jun, c-Myc), survival (e.g. PTEN, Rictor) and metabolism (e.g. glycogen synthase, HIF-α) (Kaidanovich-Beilin and Woodgett, 2011). Finally, phosphorylation of TSC2 relieves the inhibition on the mTORC1 complex. TSC2 operates in complex with TSC1, acting as a specific GTPase activating protein for Rheb, inactivating it. As Rheb is an essential activator of the mTORC1 complex, mTORC1 activity is repressed in the presence of active TSC1/2. Though mTORC1 is mediated by numerous signals, largely relating to nutrient availability, this is its main regulation downstream of growth factor signalling (Saxton and Sabatini, 2017).

1.2.1.2.3 Phospholipase C-γ / Protein Kinase C Signalling Pathway

PLC family is comprised of thirteen members that can be subdivided into six classes; the γ class (PLCγ) is activated downstream of RTKs (Bill and Vines, 2020). PLCγ is recruited to the activated RTK by its N-terminal SH2 domain, whereupon it is phosphorylated at the conserved residue Y783. Compelling evidence from Huang et al. (2016) suggests that one partner in the RTK heterodimer serves as the recruiter, with the other acting as the kinase for the key Y783 site. The phosphorylation of Y783 relieves the autoinhibition by the C-terminal SH2 domain on the catalytic active site, giving PLCγ access to substrates on the inner face of the plasma membrane (Nakamura and Fukami, 2017). There is some evidence of pathway crosstalk converging on or from PLCγ , for example potential phosphorylation by Src (K. Sato *et al.*, 2003), recruitment of and phosphorylation by AKT (Wang, Wu and Wang, 2006) and the enhancement of SOS1-driven activation of RAS (Kim *et al.*, 2000).

All members of the PLC family catalyse the hydrolysis of phosphatidylinositols including PIP2 (Figure 1-2C). The cleavage event yields two products: diacylglycerol (DAG) and inositol 1,4,5 trisphosphate (IP_3). DAG is responsible for the activation of PKC, whilst IP_3 diffuses through the cytosol to induce release of Ca^{2+} from intracellular stores such as the endoplasmic reticulum. Thus, PLC mediates two means of regulation. PKC is a critical regulator of numerous downstream cellular processes. The family of enzymes has 15 members, characterised into classical, novel or atypical classes based on their mode of activation: requiring DAG and Ca^{2+} (PKCα, β, γ), DAG alone (PKCδ, ε, η, θ), or phosphatidyl-serine (PKCι, ζ), respectively. The classical and novel PKCs are recruited to the plasma membrane by DAG, inducing a conformational change which activates them (Black and Black, 2013). Unlike other kinases activated in the initial stages of RTK signalling, such as AKT, PKC doesn't undergo further phosphorylation for its activation, as it is constitutively phosphorylated; thus, changes in its phosphorylation do not indicate modulation of its activity (Rosse *et al.*, 2010; Black and Black, 2013). The range of processes regulated by PKC downstream of PLCγ are very broad due to context- and isoform-specificity. Cell migration, proliferation, survival, apoptosis, immune response are just a sample of the processes described to be regulated by this pathway (Larsson, 2006; Reyland, 2009; Rosse *et al.*, 2010). One potential mechanism for mediating such a large number of processes is through crosstalk with the MAPK/ERK pathway, through direct phosphorylation of ERK and activation of RAF-1 via inhibitory phosphorylation of the RAF-1 inhibitor RKIP (Corbit *et al.*, 2000, 2003).

1.2.2 Epidermal Growth Factor Receptor

1.2.2.1 EGFR Structure & Activation

EGFR, also known as ErbB1 or HER1, is widely considered the "model" RTK, given it was the first RTK to be characterised and how well-studied the system is (Chen *et al.*, 2016). EGFR is one of four members of the ErbB family and can form heterodimers with other ErbB family members (Roskoski, 2014), altering their structural conformation and the downstream signalling cascade (Ward and Leahy, 2015). The preferential heterodimerisation partner of EGFR is ErbB2 (also known as HER2), which is commonly overexpressed in human diseases, including breast and ovarian cancers (Moasser, 2007). EGFR has four domains within its ectodomain, two of which are capable of ligand binding (Figure 1-3A). The intracellular region of EGFR is composed of a tyrosine kinase domain and a long cytoplasmic tail (Figure 1-3A). The ErbB receptors have a common structure but differ most markedly in their ectodomains; the exception is ErbB3 which is kinase defunct (Lemmon, Schlessinger and Ferguson, 2014; Kovacs *et al.*, 2015). The reliance of ErbB3 on other family members to initiate downstream kinase signalling highlights the importance of heterodimer formation for this family (Kovacs *et al.*, 2015).

EGFR homodimers and ErbB2-EGFR heterodimers are differentially activated by 7 ligands: EGF, transforming growth factor-α (TGF-α), heparin-binding EGF-like growth factor (HBEGF), amphiregulin (AREG), betacellulin (BTC), epiregulin (EREG), and epigen (EPGN) (Schneider and Wolf, 2009; Singh, Carpenter and Coffey, 2016). The most potent activators of the receptor, in terms of downstream signalling and receptor stability, are EGF and TGF-α (Harris, Chung and Coffey, 2003; Freed *et al.*, 2017). From when it was first isolated in 1961, EGF has been known for its potent mitogenic properties (Cohen, 1962). Following synthesis, ErbB family ligands are presented on the cell surface as transmembrane proteins, before being cleaved from their transmembrane domain by matrix metalloproteases and released into the extracellular space (Dempsey, Garton and Raines, 2002; Sahin *et al.*, 2004).

1.2.2.2 Role of EGFR in Physiology and Disease

The physiological roles of EGFR begin during the pre-implantation stage of embryogenesis; EGFR alongside EGF and TGF-α are expressed from the 4- to 8- cell stage, where they are associated with promoting development and expansion of blastocysts (Kane, Morgan and Coonan, 1997). There is an indication of stage-specific and organogenesis roles for EGFR from different mice strains that have EGFR deletions, most of which do not survive past birth but have different causes of gestational death (depending on genetic background), including lung immaturity and poor placental development (Miettinen *et al.*, 1995; Threadgill *et al.*, 1995).

EGFR-dependent processes have been described for the development of most major organs (Chen *et al.*, 2016).

Due to the mitogenic and pro-survival processes it regulates, the most appreciated role of EGFR in disease has been in tumorigenesis. Mutations of EGFR itself are found in >10% of glioblastoma, lung adenocarcinoma, stomach, colorectal and head/neck cancers, as recorded in The Cancer Genome Atlas (TGCA, accessed 30th July 2020). These include the EGFRvIII mutant, associated with glioblastoma, which is the result of an EGFR gene amplification and rearrangement (Grandal *et al.*, 2007). EGFRvIII has a truncated ectodomain that is incapable of binding ligands, is constitutively phosphorylated and active, and is unable to be internalised, maintaining the active receptor at the plasma membrane (Grandal *et al.*, 2007; Gan, Cvrljevic and Johns, 2013). EGFR point mutations are prevalent, for example in non-small-cell lung cancer 10% of patients have a mutation in the kinase activation loop (L858R), putting the receptor in a constitutively active state (Lynch *et al.*, 2004). EGFR is also over-expressed in approximately half of triple-negative breast cancer and inflammatory breast cancer cases (Masuda *et al.*, 2012). ErbB2, the preferential heterodimerisation partner of EGFR, is also commonly amplified in cancers, most notably in breast and ovarian cancer but also gastrointestinal and bladder cancers (TGCA, accessed 30th July 2020). Overexpression of ErbB2 was associated with spontaneous dimerization and receptor activation in the absence of ligand, leading to increased mitogenic signalling (Chung *et al.*, 2010).

Given that EGFR and ErbB2 are frequent heterodimers, treatments such as lapatinib for ErbB2 overexpressing breast cancer and dacomitinib for non-small cell lung cancer, act as a tyrosine kinase inhibitor of both (Esteban-Villarrubia *et al.*, 2020). Regarding the treatment of EGFR and its family members, there has been frequent discussion in the literature regarding eventual resistance to therapies targeting them. This is due to acquisition of different mutations; for example, approximately 50% of non-small cell lung cancer cases with mutant EGFR will acquire another EGFR mutation, T790M, following treatment with tyrosine kinase inhibitors (Stewart *et al.*, 2015). Another putative reason these targeted therapies may not be successful in the long term is due to nuclear localisation of EGFR (Chen *et al.*, 2016). An important therapeutic insight was gained regarding the spatiotemporal regulation of EGFR when it was reported that nuclear localisation of EGFR has been associated with poor prognosis in numerous cancers (Min, Lee and Baek, 2016). Knowledge of where EGFR

may be internalised could have implications for the way EGFR is pharmaceutically targeted in the future. This is discussed in section 1.3.2.3.

1.2.2.3 Main Signalling Pathways downstream of EGFR

Signalling downstream of EGFR is considered the standard template to which other RTK-initiated cascades can be compared to (Lemmon, Schlessinger and Ferguson, 2014). Thus, the study bias of EGFR should be kept in mind whilst reviewing the literature for EGFR, since the mechanisms may not be unique but rather first discovered in this system.

Following activation, EGFR recruits several adaptor proteins to the phosphorylated sites (Figure 1-3A). Grb2 is recruited following phosphorylation by EGFR, or by Bcr/Abl in acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL) (Li *et al.*, 2001). A comprehensive targeted proteomics study showed that Grb2 also cooperates with Shc1 to regulate recruitment to EGFR and its downstream activity; less than 1 minute after stimulation, Shc1 binds Grb2, which was shown to be essential for the propagation of early ERK and AKT activation (Zheng *et al.*, 2013). AKT phosphorylation of Shc1 acts as a negative feedback mechanism by providing a binding site for the tyrosine phosphatase PTPN12, which dephosphorylates EGFR and causes the dissociation of Shc1 (Zheng *et al.*, 2013). GAB1 is another adaptor utilised by EGFR; GAB1 is recruited to the activated receptor by Grb2, and is associated with the subsequent recruitment of the phosphatase SH2 domain–containing phosphatase 2 (SHP2) (Gu and Neel, 2003). However, a mechanism for maintaining GAB1-SHP2 complexes distal from the activated receptor, in the early-intermediate phase of signalling, has been described and shown to be dependent on cycles of Src-family kinase phosphorylation of GAB1 (Furcht, Buonato and Lazzara, 2015).

The E3 ubiquitin ligase Cbl is a well-studied interactor of EGFR. Initially it was considered a negative regulator; it is now better understood as a mediator of EGFR internalisation and its subsequent trafficking route (Ettenberg *et al.*, 1999; Stern, Place and Lill, 2008; Roepstorff *et al.*, 2009). Cbl recruitment to EGFR is Grb2 dependent (Fukazawa *et al.*, 1996). A recent study associated the Cbl isoform Cbl-b with internalised EGFR, suggesting that the isoforms are physically as well as functionally linked to the subcellular location of EGFR. This study showed an interaction between both isoforms and Grb2 and another EGFR adaptor protein, Nck2 (Liu *et al.*, 2021). Cbl is negatively regulated through phosphorylation by Sprouty 2 (SPRY2), inhibiting it's activity and prolonging EGFR activation at the cell surface (Fong *et al.*, 2003).

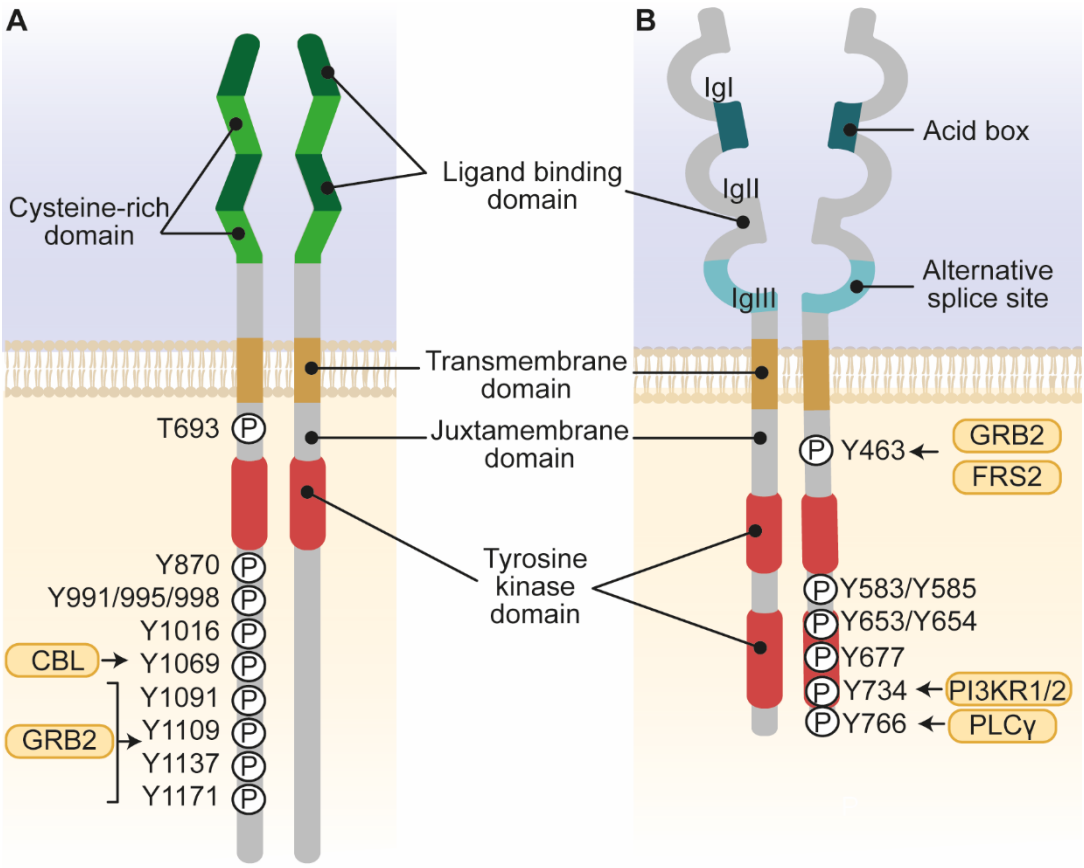


*Figure 1-3. Schematic of EGFR (**A**) and FGFR (**B**) structure, phosphorylated sites following ligand activation, and recruited proteins.*
Adapted from Ferguson et al. (2021); compiled using Lemmon et al. (2014), Lundby et al. (2019), Francavilla et al. (2016) and Ferguson et al. (2021).

1.2.3 Fibroblast Growth Factor Receptor

1.2.3.1 FGFR structure & ligand activation

The FGF family of 22 ligands (FGF1-22) comprises small signalling proteins, the majority of which are secreted and act in a paracrine manner, binding to cell surface FGFRs. The exceptions are endocrine FGF15, FGF19, FGF21 and FGF23 and the non-secreted, intracellular FGF11-14. The paracrine, or canonical, FGFs are expressed in nearly all tissues, with myriad functions from development to homeostasis. Their associated receptors compose the FGFR family, which are the product of four genes (*FGFR1-4*). FGFR1-3 are alternatively spliced, resulting in two isoforms each and seven FGFR receptors in total: FGFR1b-c, FGFR2 b-c, FGFR3b-c, FGFR4 (Ornitz and Itoh, 2015). The seven receptors have a shared structure: the intracellular region contains two tyrosine kinase domains with an intermediate linker, whilst the extracellular region contains three Ig-like domains (Figure 1-3B). The IG-like domains II and III, and the linker between them, confer receptor specificity to subsets of FGF ligands (Plotnikov *et al.*, 2000; Yeh *et al.*, 2003). The alternatively spliced

forms of *FGFR1-3* differ in the expression of IG-like III domain, which has a IIIb and IIIc form; the isoforms are denoted 'b' and 'c' accordingly (Miki *et al.*, 1992; Werner *et al.*, 1992; Chellaiah *et al.*, 1994). The alternative splicing event has substantial effects on ligand specificity. This is most notable for FGFR2, where the b and c isoforms show binding affinity for a distinct set of FGFs; FGFR2b has specificity for FGF3, FGF7, FGF10 and FGF22 whilst FGFR2c has specific for all other FGF isoforms (Ornitz and Itoh, 2015; Watson and Francavilla, 2018). The exception is FGF1, which is the only FGF to show binding affinity to all FGFR splice variants (Ornitz *et al.*, 1996; Xiuqin Zhang *et al.*, 2006). The isoforms of *FGFR1-2* also show variation in tissue expression, with the 'b' form expressed in epithelial cells and the 'c' form in mesenchymal cells (Ornitz and Itoh, 2015; Ranieri *et al.*, 2018). There is some evidence that, like EGFR, FGFR dimers form in the absence of ligand, and even undergo some autophosphorylation. However, ligand bound states are activated and differentially regulated by FGF ligands (Sarabipour and Hristova, 2016).

Canonical FGFs also utilise a co-receptor, HS, which is a complex carbohydrate chain that is found bound to membrane-anchored proteins or proteins diffuse in the extracellular matrix. The role of HS has been showed to be non-essential but highly beneficial in stabilising FGFs, particularly those that exist as unstable proteins (Buchtova *et al.*, 2015), such as FGF1, which has a short half-life and denatures in the range of physiological temperatures (Zakrzewska et al. 2009). FGF2 variants with increased thermal stability were shown to activate downstream pathways more potently at lower concentrations with less dependence on HS (Koledova et al. 2019). Furthermore, the binding affinities of FGF7 and FGF10 for HS has been shown to underlie their respective roles in regulating branching morphogenesis of salivary and lacrimal gland epithelium buds; mutating FGF10 in the HS-binding pocket caused the buds to branch rather than elongate, as was seen for upon FGF7 stimulation (Makarenkova *et al.*, 2009).

The wide range of FGFRs and FGF ligands give rise to a complex system, where multiple ligands are specific to the same receptor (e.g. FGFR2b has binding affinity for FGF1, -3,-7, -10 and -22), however there is little redundancy, as the receptor:ligand pairs result in distinct cellular outputs (Ornitz *et al.*, 1996; Xiuqin Zhang *et al.*, 2006; Francavilla *et al.*, 2013; Zinkle and Mohammadi, 2019). A further layer of regulation by the FGFR family is the cell-specific expression seen in tissues and tumours. Inducing FGFR2b expression in epithelial cells has been shown to promote expression of FGF4 and FGF7 that can activate FGFR2c on fibroblasts, causing thyroid cancer progression (Guo *et al.*, 2012). Cancer associated fibroblasts from mouse mammary tumours were shown to have high expression of FGF2,

increasing cell proliferation within the tumours and bypassing typical hormone receptor signalling, allowing for resistance to endocrine therapy (Giulianelli *et al.*, 2008; Shee *et al.*, 2018)

1.2.3.2 Role of FGFR in Physiology and Disease

It would be difficult to overstate the diversity of cellular outputs and the crucial role FGF-FGFR signalling plays in development, homeostasis and human pathology. FGF signalling has been shown to mediate: mesoderm formation, axes specification and lineage specification in vertebrate development (reviewed by Dorey and Amaya 2010); wound healing, particularly reepithelization and angiogenesis (Müller, Meyer and Werner, 2012); cardioprotection following ischemic injury (Kardami *et al.*, 2007; Liao *et al.*, 2007) and lung epithelial repair (Braun *et al.*, 2004). Knock-out mice models have shed light on the often-redundant roles of FGF ligands in development. *Fgf1* knockout mice are viable and show no phenotype, indicating that their role can be performed by the other paracrine FGFs (Miller *et al.*, 2000). This correlates with their ability to bind and activate all FGFR isoforms. Knockout of *Fgf9, Fgf10* or *Fgf18* induce death shortly after birth, due to impairments in the development of multiple organs (Ohuchi *et al.*, 2000; Colvin *et al.*, 2001; Ohbayashi *et al.*, 2002). Others induce viable litters, but with defects in specific areas. For example, knockout of *Fgf7* results in impaired hair and kidney development (Guo, Degenstein and Fuchs, 1996; Qiao *et al.*, 1999).

As these are processes defined by cell proliferation, survival, and migration, it is expected that the deregulation of FGF signalling often results in an oncogenic phenotype; accordingly, signalling in the correct context is tumour suppressive. The oncogenic role of FGF signalling was emphasized by a screen of 1000 somatic mutations corresponding to 518 kinase genes in 210 diverse human cancers, in which FGF signalling participants were the most represented (Greenman *et al.*, 2007). The aberration of FGFR signalling differs between tumour types, however activating mutations (e.g. FGFR2 mutations in 10-12% endometrial carcinomas (Gatius *et al.*, 2011)), gene amplification (e.g. FGFR2 amplification in gastric cancers (Takeda *et al.*, 2007; Kunii *et al.*, 2008; Hur *et al.*, 2020)) and chromosomal translocation (e.g. translocation that links FGFR3 to the immunoglobulin heavy chain *IGH* locus in 15% of multiple myelomas (Chesi *et al.*, 1997; Avet-Loiseau *et al.*, 1998)) have been described for FGFRs in multiple cancer types (Turner and Grose, 2010). Ligand-dependent tumorigenesis has also been described, for example the frequent amplification of FGF1 in ovarian cancer, which is associated with poor survival (Birrer *et al.*, 2007). Germline single

nucleotide polymorphisms of FGF-FGFR family members have also been described, for instance of FGFR2 in oestrogen receptor-positive breast cancer (Garcia-Closas *et al.*, 2008).

Considering this, FGFR is the target of multiple pharmaceuticals for the treatment of cancer. Many of these are pan- or specific-FGFR inhibitors, though other therapies include FGFR-blocking antibodies and FGF ligand traps (Babina and Turner, 2017). Several generations of FGFR inhibitors have been developed, but their use for the most part in clinic has been limited, either by too high toxicity or poor response (Schuler, 2020). The most recent generation is an irreversible, pan FGFR inhibitor that is structurally distinct from the previous generations of therapeutics; futibatinib, the first of this class of medicines, is currently in phase II clinical trials (Bahleda *et al.*, 2020). Recent work from Fernández-Nogueira et al. (2020) has suggested that FGFR2 expressing fibroblasts in the tumour microenvironment may be enabling HER2+ cancers obtaining resistance to HER2-targeted therapy. This suggests that targeting FGFR may aid in the treatment of cancers even if there is not a genetic abnormality in FGFR itself.

1.2.3.3 Signalling Specificity downstream of FGFR

Upon ligand binding, receptor activation is induced by sequential autophosphorylation of five tyrosine residues in the kinase domain: Y463, Y583, Y585, Y653 and Y654 (Figure 1-3B). If this stepwise activation is altered, it may lead to an oncogenic signalling phenotype (Lew *et al.*, 2009). Following activation, FGFR can recruit the adaptor protein FGFR substrate 2 (FRS2), which is a relatively unique component of the FGFR "signalosome". FRS2 is constitutively bound to the juxtamembrane region of FGFR and anchored to the cell membrane by a myristoyl-group (Kouhara *et al.*, 1997). Its activation is dependent on FGFR phosphorylation on Y463 and, in some contexts, interaction with other adaptor proteins, such as CRKL (Larsson *et al.*, 1999; Collins *et al.*, 2018). FRS2 acts as the main platform from which the MAPK and PI3K pathways are initiated from, through recruitment of SHP2, GRB2 and GAB1 (Kouhara *et al.*, 1997; Ong *et al.*, 2000; Gotoh, 2008). Src is also recruited to FRS2; disturbance of this interaction at the activated receptor results in perturbation of the PI3K/AKT and ERK pathways (Sandilands *et al.*, 2007). Moreover, Src recruited to FRS2 was shown to be critical, in coordination with the Src substrate Eps8, for internalisation of the receptor through clathrin coated pits (Auciello *et al.*, 2013). Knockdown of Eps8 resulted in fewer clathrin coated pits, retention of the receptor at the plasma membrane and dampening of ERK signalling (Auciello *et al.*, 2013). The importance of receptor internalisation as a mode of spatial regulation is discussed in detail in section 1.3.2.3.

Negative regulators of FGFR signalling are also recruited to FRS2. A ternary complex of FRS2-Grb2 and Cbl was shown to be important in ubiquitination of FGFR2 following stimulation (Wong *et al.*, 2002). Cbl was separately shown to attenuate FGFR2 induced PI3K-AKT signalling (Dufour *et al.*, 2008). One form of negative feedback on FRS2 itself occurs from the MAPK pathways. ERK phosphorylates the C-terminal of the FGFR (at site S777) and on FRS2 (Zakrzewska *et al.*, 2013). Similarly, p38 is able to phosphorylate FRS2 and can compensate for inhibited ERK1/2 feedback through excessive activation (Zakrzewska *et al.*, 2019). There is evidence that the negative feedback on signalling downstream of FGF2 by ERK and p38, alongside suppression of the JNK pathway, elicits a cardioprotective response (House *et al.*, 2005; Liao *et al.*, 2007; Sørensen *et al.*, 2008).

Proteins are also recruited directly to the receptor, for example PI3K is recruited to phosphorylated Y734 and PLC-γ is recruited to phosphorylated Y766 (Mohammadi *et al.*, 1991; Peters *et al.*, 1992; Francavilla *et al.*, 2013). Intriguingly, preventing phosphorylation at PLCγ-Y766 inhibits internalisation of the receptor (Sorokin *et al.*, 1994). Mutation at this site was also shown to prevent activation of PLC-γ as well as PI3K, resulting in changes in FGFR1 mediated cytoskeletal reorganisation but not its mitogenic response (corroborated by the lack of impact on the MAPK pathway) (Cross *et al.*, 2000). In cells expressing wild-type FGFR, PLC-γ recruitment and association was negatively regulated by Grb14, which was also recruited to Y766 (Browaeys-Poly *et al.*, 2010).

SPRY family proteins 1-4 are negative regulators of RAS-ERK and PI3K/AKT signalling downstream of FGFR. The most studied isoform is SPRY2, which has been shown to inhibit RAS-ERK signalling at the level of RAF and Grb2 (Yusoff *et al.*, 2002; Lao *et al.*, 2007). Following FGFR activation, SPRY1/2 translocates to the plasma membrane and is dephosphorylated at its basally phosphorylated S112 and S115 sites by PP2A (Tefft *et al.*, 2002; Lao *et al.*, 2007). Once proximal to the activated receptor, it can bind Grb2, antagonising its recruitment to the docking protein FRS2 or dephosphorylation by SHP2 (Hanafusa *et al.*, 2002; Lao *et al.*, 2007). This is opposite to the function of SPRY2 downstream of EGF:EGFR (Fong *et al.*, 2003). Expression of SPRY4 in several cancer models was associated with suppression of FGFR-driven tumorigenesis, emphasising its importance as a tumour suppressor (Qiu *et al.*, 2019; Lu *et al.*, 2020). Knockout of SPRY proteins is also associated with numerous perturbations in organogenesis (Guy *et al.*, 2009).

Another negative regulator of RAS-ERK signalling downstream of FGFR is Similar Expression to FGF (Sef), which was originally identified in zebrafish as an FGFR inhibitor (Fürthauer *et*

al., 2002; Tsang *et al.*, 2002). In humans, the gene was originally designated as a member of the interleukin-17 receptor family due to the sequence similarity of the intracellular region (Yang *et al.*, 2003). Contrasting methods of action have been described for Sef, in which it exerts its activity at either the receptor or within the RAS-ERK cascade itself. Sef was shown to directly interact with FGFR1 and FGFR2, resulting in regulation of phosphorylation of RAF, MEK and ERK, and inhibiting differentiation in PC12 cells, respectively (Kovalenko *et al.*, 2003; Xiong *et al.*, 2003). However, work by Torii et al. (2004) demonstrated that Sef is found in the Golgi apparatus until stimulation with FGF, when a portion of the Sef population is trafficked to the plasma membrane. Activated MEK and ERK were shown to co-localise with Sef at the plasma membrane and Golgi apparatus; Sef activity blocked nuclear translocation of ERK and inhibited dissociation of the MEK-ERK complex. Cytoplasmic targets of ERK such as RSK2 were not affected, though the expression of IEGs was (Torii *et al.*, 2004). RAS-ERK signalling is also downregulated by Dual Specificity Phosphatase 6 (DUSP6) downstream of FGFR *in vivo* (Li *et al.*, 2007; Ekerot *et al.*, 2008). Inhibition by DUSP6 was shown to be crucial for limb bud and neural development in chicken and mouse embryos (Dickinson *et al.*, 2002; Eblaghie *et al.*, 2003).

An interesting aspect of spatial FGFR regulation has been described, whereupon stimulation leads to translocation of the ligand:receptor pair to the nucleus. There is evidence that the translocation may be dependent on the phosphorylation of the ligand, as described for FGF1 and FGF2, potentially via p38 feedback (Małecki *et al.*, 2004; Więdłocha *et al.*, 2005; Sørensen *et al.*, 2008). Nuclear localisation of FGFR may have anti-apoptotic function (Kostas *et al.*, 2018). FGFR1:FGF10 have been shown to translocate to the nucleus in invading breast cancer cells and regulate the expression of migration-related genes, thus contributing to the invasive phenotype (Chioni and Grose, 2012). The nuclear translocation of FGFR1 has been indicated to determine cellular behaviour in other cell models too, with neurite outgrowth occurring when FGFR1 is at the plasma membrane but not when internalised (Csanaky, Hess and Klimaschewski, 2019). The mechanism of the translocation is likely to be a different method to EGFR/ERBB2 nuclear translocation. EGFR/ERBB2 were shown to translocate to the nucleus via a retrograde plasma membrane-Golgi apparatus-endoplasmic reticulum route, whilst FGFR1 may be translocated indirectly via importin β (Reilly and Maher, 2001; Wang *et al.*, 2012).

1.3 Regulation of RTK signalling

1.3.1 Regulation from the ligand perspective

Ligand-dependent regulation of RTK signalling takes two main forms: functional selectivity and changes in ligand abundance. Functional selectivity describes the phenomenon of different ligands showing affinity for the same receptor yet eliciting distinct downstream responses (Figure 1-4A); this is not well understood for RTKs (Karl *et al.*, 2020). Examples of functional selectivity can be drawn from across the cell-surface receptor superfamily. Biased agonism is well described for GPCRs (Mailman, 2007) , and it has been shown that biased agonists of β-arrestin associated pathways or G-protein associated pathways are associated with different behaviours (Smith et al., 2018). The toll-like receptor (TLR) family member TLR4 is activated by the ligands LPS and PAL, which induce pro- and anti-inflammatory effects respectively (Fischer *et al.*, 2019) Using single cell RNA-seq, the authors identified distinct transcriptional states following TLR4 activation with either ligand in a macrophage cell line. Differentially expressed genes in downstream of each ligand were antagonistic to genes that were differentially expressed by the other.

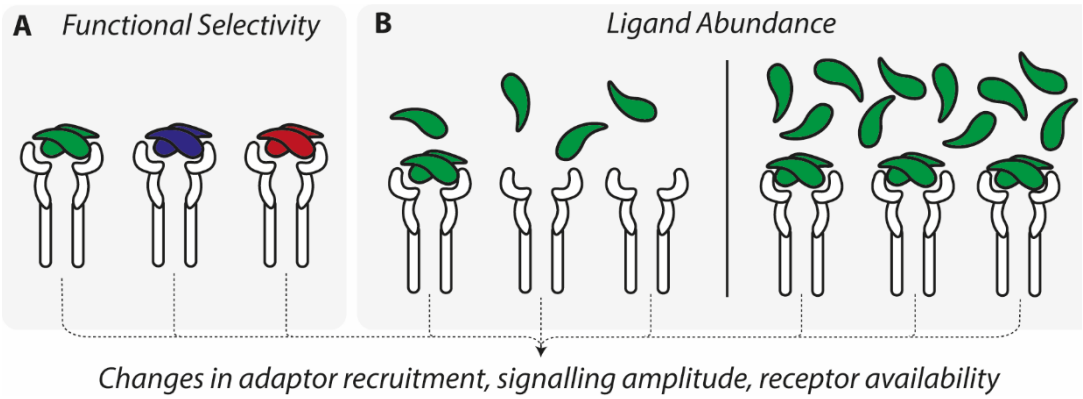


Figure 1-4. Regulation of receptor tyrosine kinase signalling by ligands through A) functional selectivity and B) ligand abundance.

Only recently has the importance of functional selectivity in RTK signalling started to be appreciated (Karl *et al.*, 2020). The best-studied example in the RTK family is EGFR, which, as previously discussed, is differentially activated by seven ligands that alter receptor conformation and subsequent downstream response (Harris, Chung and Coffey, 2003; Freed *et al.*, 2017). The ligands that have the strongest affinity, EGF and TGF-α, were shown to have distinct profiles of response at the level of receptor modification (phosphorylation/ubiquitination), recruitment of interactors to the activated receptor, and on the trafficking of the receptor itself (Roepstorff *et al.*, 2009; Francavilla *et al.*, 2016). The differential effects between EGF and another EGFR ligand, HRG, was shown to induce

different expression patterns of the IEG c-Fos (Nakakuki *et al.*, 2010). An interesting model for studying functional selectivity is FGFR2b, which is differentially activated by FGF1, FGF3, FGF7 and FGF10. FGF7 and FGF10 show the strongest affinity for FGFR2b, and FGF7 does not bind any others in the FGFR family (Ornitz *et al.*, 1996; Xiuqin Zhang *et al.*, 2006). FGF10 stimulation was shown to induce rapid phosphorylation of Y734 on FGFR2b allowing recruitment of PI3K and the receptor internalisation adaptor, SH3 domain-binding protein 4 (SH3BP4). Y734-FGFR2b phosphorylation and subsequent protein recruitment was not seen downstream of FGF7. Differences in the signalling cascade and receptor trafficking routes (discussed in greater detail in section 1.3.2.3) were also observed, and when considered together were contributing to the molecular regulation of cell behaviour: FGF7 induced a proliferative phenotype whilst FGF10 induced a migratory phenotype.

Ligand availability and abundance also impacts downstream signalling (Figure 1-4B). A high concentration of extracellular ligand increases the number of activated receptors, whilst reducing the number of receptors accessible to form new receptor:ligand pairs (I.e. increasing receptor occupancy). Numerous ligand dose-dependent effects have been described for RTKs. The opposing ERK and AKT pulsatile responses downstream of several growth factors, including EGF, were found to be dose-dependent, directly influencing the activity of FOXo family transcription factors (Gross and Rotwein, 2016; Sampattavanich *et al.*, 2018). Varying concentrations of EGF in cells at steady state induced fluctuating profiles of ERK pulsing activity, known as frequency modulation. Entry into the cell cycle was found to have a non-linear dependency on pulse profiles (Albeck, Mills and Brugge, 2013). Another example of EGF dose-dependent effects is in breast cancer cells; cell motility was found to be highest at 50 ng/ mL EGF, whilst lower (25 ng/ mL) or higher (100 ng/mL) concentrations induced more limited motility compared to unstimulated cells (Wang *et al.*, 2004). Low concentrations of FGF2 were shown to induce p38-dependent survival and differentiation whilst inhibiting proliferation, but high concentrations do the near opposite (Garcia-Maya et al. 2006); follow up work associated the biphasic response to binding competition with the HS co-factors (Kanodia *et al.*, 2014). A concentration dependent switch has also been investigated based PDGF concentration, with low (1ng/mL) concentration inducing migration, and maximal activation of signalling associated with migratory phenotypes, and high (>5ng/ml) concentration inducing migration and signalling associated with cell cycle induction (Donatis *et al.*, 2008).

It should be noted that these experiments often use a range of concentrations that are beyond the level of receptor saturation. Levels of EGF in blood plasma and serum are often sub-saturating, in the range of 0.4-20ng/mL, but can surpass 50 ng/ml and be as high as 500 ng/ml in tissues and organs (Konturek *et al.*, 1989; Sigismund *et al.*, 2005; Koseska and Bastiaens, 2020). It is common for experimentalists to use concentrations of 50 ng/ml, which is thought to be receptor saturating in most cell types. One of the most cited high-throughput studies on EGF signalling used a concentration of 150 ng/ml (Olsen *et al.*, 2006). Notably, once an EGF concentration "threshold" of approximately 1ng/ml is passed, EGFR phosphorylation and ubiquitination is sufficient to recruit adaptors (I.e. GRB2 and Cbl) for the progression of the signalling cascade and regulation (Capuani *et al.*, 2015). The physiological levels of FGFs have not been studied, and it remains complicated to do so given their high dynamic range. However, the local concentration of FGFs is thought to be in the range of 100 ng/mL based on FGF binding to its low-affinity receptor HS-proteoglycans in the extracellular matrix (Sun *et al.*, 2016).

1.3.2 Regulation from the receptor perspective

1.3.2.1 Structural changes in RTKs modulate activity

Receptors are intricate information-transfer devices, that directly regulate downstream signalling via structural changes, be that from physical constraints of ligand binding, oligomerisation and protein recruitment, or post-translational modification.

As previously mentioned, RTK di- or oligomerisation is a critical component of RTK activation, that may or may not precede ligand binding. Regardless, ligand binding is thought to achieve two key structural changes: 1) the release of the kinase domain from an inhibitory state, and 2) confer ligand identity through intracellular domain conformation (Olsen *et al.*, 2004; Xuewu Zhang *et al.*, 2006; Diwanji, Thaker and Jura, 2019). The latter has been described for ligands of both EGFR and FGFR, with implications on downstream signalling cascades if perturbed (Plotnikov *et al.*, 2000; Freed, Alvarado and Lemmon, 2015). However, there is controversy around whether ligand binding results in structural changes to the kinase domain or only to the transmembrane and extracellular domains (Karl *et al.*, 2020). Some work suggests that the extracellular structural changes induced upon ligand binding are not propagated to the intracellular structure, due to flexible linkers between the domains (Lu *et al.*, 2010, 2012; Mi *et al.*, 2011). This would imply that the RTK kinase activity is not differentially regulated by the structural changes induced by the ligand, however that would contrast with the known changes in phosphorylation patterns downstream of different RTK

ligands (Francavilla *et al.*, 2013). One possible resolution to these contrasting positions is that while ligand identity does not influence kinase activity, it does influence the stability of the dimerised RTK, and thus the duration (rather than the amplitude) of activity (Freed *et al.*, 2017; Huang *et al.*, 2017). It is important that these open questions on RTK structural regulation are addressed in order to aid design and deployment of pharmaceutical interventions against RTKs (Roskoski, 2020).

As kinases that auto- and transphosphorylate, the impact of modification on RTK structure is equally critical. Many of the stimulatory phosphorylation sites on RTKs are in the kinase domain and C-terminal tail, however they are also found in the juxtamembrane region of PDGFR, EPH and insulin receptors (Wybenga-Groot *et al.*, 2001; Hubbard, 2004), as well as the kinase insert domain (KID) regions of PDGFR, VEGFR and FGFR (Chen *et al.*, 2007; Locascio and Donoghue, 2013). It has been shown that phosphorylation stabilises EGFR in a ligand-specific active conformation; a common mutation in cancer, L834R, stabilises the receptor in this conformation independent of ligand binding, causing constitutive activation (Shan *et al.*, 2012; Brewer *et al.*, 2013). Phosphorylation of the juxtamembrane region has been shown to relieve autoinhibition, complementing these other conformational changes (Hubbard, 2004).

Phosphorylation sites are now understood to commonly occur in intrinsically disordered regions (Iakoucheva *et al.*, 2004; Collins *et al.*, 2008). Intrinsically disordered regions are defined as polypeptide sections that are unlikely to form a classic protein 3D structure but are still functional, unlike misfolded proteins (van der Lee *et al.*, 2014). Though disordered, phosphorylation does change the properties of intrinsically disordered regions to allow for access of kinases, phosphatases and binding proteins to the modification site (van der Lee *et al.*, 2014). Indeed, the issue of access is hypothesised as the key advantage of intrinsically disordered regions. Recent work has shown that this is exaggerated upon multi-site phosphorylation (Jin and Gräter, 2021). Further, predictions by Palmeri et al. (2014) suggest that phosphorylation site conservation is not encoded in sequence similarity, but rather structural similarity, highlighting the plasticity of these regions. As understanding of intrinsically disordered regions increases, so too does the appreciation of this novel mechanism for RTK sensitivity to ligand abundance and identity (Simon *et al.*, 2014). These concepts are being used in bioengineering in the hope of improving the sensitivity and responsiveness of biosensors, genetic logic gates and "smart" materials (Simon *et al.*, 2014).

Irreversible modifications are also part of the structural regulation at the receptor level. This may be due to processes related to cell death, for example the shedding of the extracellular domain by matrix metalloproteases following endoplasmic reticulum stress or hypoxia, and cleavage of intracellular domains by caspases during apoptosis initiation (Kreitman, Noronha and Yarden, 2018). Proteolysis, frequently mediated by ubiquitination at different sites and within different cellular contexts , is another example of irreversible modification (Miller, Sullivan and Lauffenburger, 2017; Kreitman, Noronha and Yarden, 2018). Downstream of activation, RTKs can be directed towards proteolysis in the lysosome upon ubiquitination by E3 ubiquitin ligases such as Cbl which are recruited to the activated receptor. Cellular stress also induces the addition of ubiquitin chains, and directs the receptor to degradation in the proteasome; this is mediated by heat shock protein 90 and other chaperones (Ancot *et al.*, 2009; Kreitman, Noronha and Yarden, 2018).

1.3.2.2 Expression levels of RTKs

As discussed in section 1.3.1, the abundance of ligand has an impact on cell signalling. However, the transfer of this external information into the cell necessitates a minimum expression level of the associate RTK. A high level of RTK will less readily become saturated, but the total number of RTKs activated will be higher (Lam, Pickering and Mac Gabhann, 2019). Receptor abundance has been shown to influence tissue development, for example, unequal distribution of EGFR in daughter cells in forebrain development was shown to lead to distinct cell fates (Sun, Goderie and Temple, 2005). A similar effect was seen in the expression of VEGFR in the development of retinal vasculature (Nakayama *et al.*, 2013). As different cell types will express different levels of RTKs, the physiological response to changes in ligand abundance is controlled and the effects are likely to be consistent.

RTK overexpression or depletion is often disease-causing. The broadest example is cancer, from which myriad examples of RTK expression deregulation can be drawn, several of which are described in sections 1.2.2.2 and 1.2.3.2. It has also been reported that changes in EGFR expression are a risk factor for Alzheimer's disease (Hiramatsu *et al.*, 1988; Sibilia *et al.*, 1998; Chen *et al.*, 2016). The change in expression of co-factors, such as HS for FGFRs, may also be a consideration; HS proteoglycans from breast cancer cells were shown to form FGF2:FGFR complexes more readily than HS from "normal" cells (Qiao *et al.*, 2003).

1.3.2.3 Internalisation and intracellular trafficking of RTK

1.3.2.3.1 Endocytosis of active RTKs occurs moments after activation

Structural changes occur within seconds of receptor activation (Bessman *et al.*, 2014; Lemmon, Schlessinger and Ferguson, 2014; Ferguson, Smith and Francavilla, 2021), however receptor internalisation via endocytosis occurs concomitantly in the early stages of activation. Thus, receptor internalisation and trafficking, termed receptor-mediated endocytosis, is a second method of receptor-level regulation of intracellular signalling. Internalisation and trafficking to different subcellular compartments introduces the receptor to different binding partners, leading to localised interactomes (Lobingier *et al.*, 2017; Sigismund *et al.*, 2021). When internalised, receptors are trafficked through the endosomal system, which is traditionally considered a series of membrane bound organelles that mediate incoming and outgoing cellular cargo and receptors, as well as homeostatic turnover of intracellular molecules (Hsu, Bai and Li, 2012). Briefly, activated RTKs are endocytosed from the plasma membrane to the early endosome (also called the sorting endosome); from here they can be recycled back to the plasma membrane, either directly or via the recycling endosome, sent for degradation via the late endosome (also called multi vesicular bodies) and lysosome, or passed through other intracellular trafficking routes, for example to the Golgi apparatus or endoplasmic reticulum (Figure 1-5) (Goh and Sorkin, 2013; Cullen and Steinberg, 2018). When reviewing the characteristics and functions of the endosomal system, it is quickly apparent that there are many gaps in the field, not least of which how to define distinct endosomal compartments. Identifying endosomal compartments has been typically difficult due to the dynamics of the system, through which many proteins and lipids flow as the compartments mature or respond to environmental stimuli. For this reason, Rab GTPases are considered the key markers of distinct endosomal compartments. Rab GTPases are responsible for orchestrating the molecular events in the endosomal system. Some of these events include assembly of protein complexes, called coats, that surround the vesicles budding from a membrane, tethering to cytoskeletal motors for vesicle motility, providing identity to vesicles so they fuse with the correct compartment, and tethering to those destination compartments by recruiting factors such as Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) (Stenmark, 2009). Discrepancies in Rab activity or localisation are associated with tumorigenesis (reviewed by Tzeng and Wang (2016)).

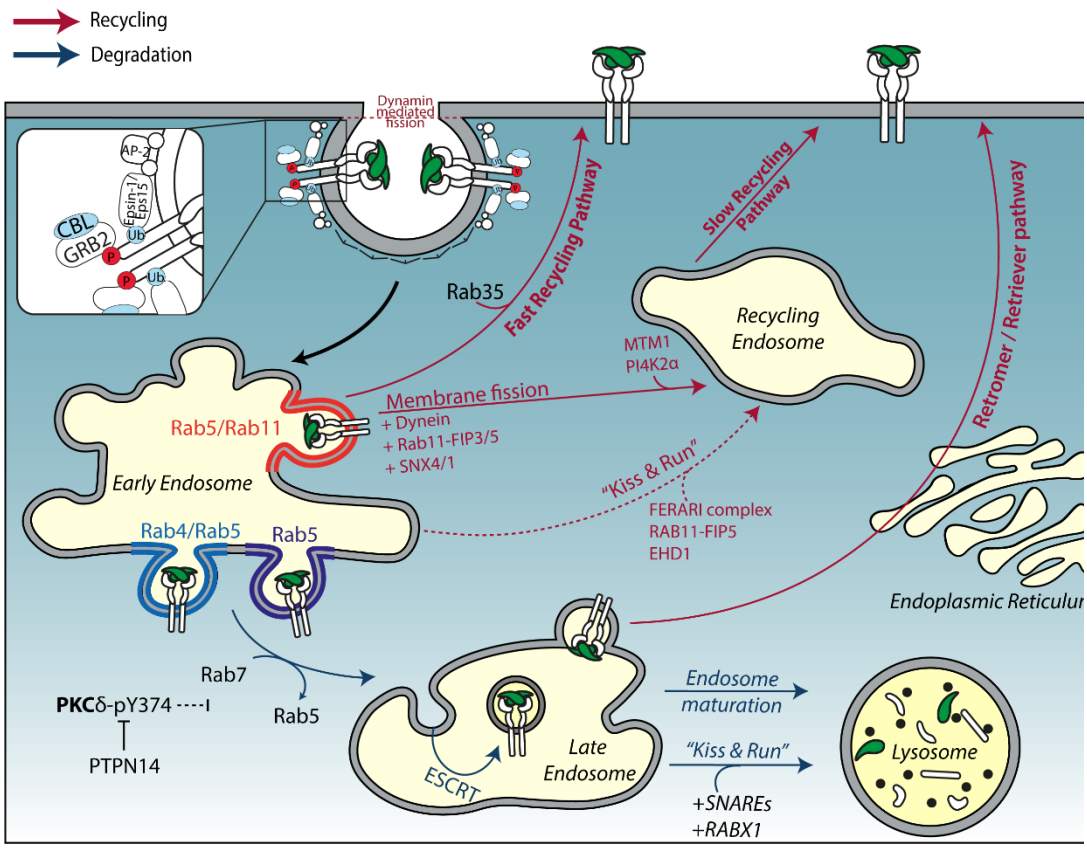


Figure 1-5. Trafficking routes of internalised receptor tyrosine kinases, resulting in recycling back to the plasma membrane or degradation in the lysosome.

Some argue that receptor-mediated endocytosis and endosomal trafficking events are in fact the master regulators of cellular signalling (Sigismund *et al.*, 2012). Indeed, seminal work in the field revealed that blocking endocytosis (through the dominant-negative mutation of dynamin, a key membrane fission enzyme) has been shown to induce considerable changes in signalling response, with some pathways augmented and others dampened (Vieira, Lamaze and Schmid, 1996). EGFR remains the best-studied RTK model of receptor-mediated endocytosis and intracellular trafficking; all molecular details described here relate to EGFR, unless otherwise stated.

Constitutive receptor-mediated endocytosis is typically quite slow (Hopkins, Miller and Beardmore, 1985), however when endocytosis is ligand-induced various factors are recruited to the activated receptor, dramatically increasing the efficiency of the process. receptor-mediated endocytosis occurs either by clathrin-mediated endocytosis or clathrin-independent endocytosis. In clathrin-mediated endocytosis, the E3 ubiquitin ligase Cbl is recruited by GRB2 following receptor and direct interaction with phosphorylated Y1045 (Levkowitz *et al.*, 1999; Jiang and Sorkin, 2003), where it mono- and poly-ubiquitinates lysine residues on the receptor (Huang *et al.*, 2006). Ubiquitination of the receptor is relatively-well

established as a prerequisite to degradation in the lysosome, however the role of ubiquitination in other trafficking routes, for example in the recruitment of clathrin adaptor proteins, remains controversial (Sorkin and Goh, 2008; Tomas, Futter and Eden, 2014). The most accepted hypothesis states that Cbl ubiquitinates EGFR creating recruitment sites for ubiquitin-binding proteins of the clathrin coat, such as epsin-1, Eps15 and AP-2, which act as a bridge between the receptor and clathrin proteins (Kazazic *et al.*, 2009; Traub, 2009; Tomas, Futter and Eden, 2014). However, numerous studies have shown that clathrin-mediated endocytosis of EGFR is not dependent on the ubiquitination by Cbl (Huang et al. 2006) and AP-2 recruitment (Sorkin and Carpenter, 1993; Johannessen *et al.*, 2006; Madshus and Stang, 2009), but disrupting either of these had effects on protein turnover and lysosomal degradation of the receptor. By whatever means it is recruited, clathrin polymers coat the budding vesicle, causing membrane deformation from their curved structure (Dannhauser and Ungewickell, 2012; Stachowiak *et al.*, 2012). The fission GTPase dynamin regulates the separation of the budding vesicle from the plasma membrane, forming the mature clathrin-coated vesicle (Mettlen *et al.*, 2018). Early experiments found that clathrin-mediated endocytosis pathways were easily saturated when cells were stimulated with greater than physiological levels of EGF (>1-2 ng/mL) (Wiley, 1988). Subsequent work highlighted that when clathrin-mediated endocytosis pathways were saturated, EGFR would be internalised by clathrin-independent endocytosis pathways (Sigismund *et al.*, 2008). Moreover, levels of receptor trafficked via clathrin-independent endocytosis tracked high levels of ubiquitination (which influence trafficking to the lysosome) and increasing EGF concentration (Sigismund *et al.*, 2013). Clathrin-independent endocytosis has a much less defined role in receptor-mediated endocytosis, but appears to rely on some of the same players as clathrin-mediated endocytosis, namely Eps15 and epsin-1 (Sigismund *et al.*, 2005).

1.3.2.3.2 Sorting internalised receptors at the early endosome

The initial destination for all endocytosed vesicles is the early endosome. The early endosome is defined by PIP3 enrichment in the membrane bilayer (Corvera, D'Arrigo and Stenmark, 1999), a luminal pH of ~6.2 (maintained or further acidified by the V-ATPase proton pump) (Murphy, Powers and Cantor, 1984), and association with proteins containing a FYVE-domain (Gaullier *et al.*, 1998), such as EEA1 and rabenosyn-5. Intriguingly, regardless of the quantity of internalised receptors, the mean amount of receptor per early endosome is constant; rather than concentrating more internalised receptor in each early endosome, the cell responds by increasing the number of early endosome (Villaseñor *et al.*, 2015). This was suggested to be important for ensuring signalling fidelity and specificity.

The mechanism which dictates the onward route of the internalised receptor from EEs is not well understood; the current understanding indicates that the early endosome is a dynamic and heterogenous organelle. Some evidence suggests that distinct membrane domains form in the early endosome, characterised by regions enriched for Rab5, Rab4/Rab5 and Rab4/Rab11 (Sönnichsen *et al.*, 2000). The receptor trafficking route may be dictated by which membrane domain it is localised to: Rab5 domains mature into late endosome, characterised by replacement of Rab5 with Rab7 (Rink *et al.*, 2005); Rab4/Rab5 domains likely direct receptor cargo towards lysosomal degradation by a similar mechanism (Nag *et al.*, 2018); Rab4/Rab11 domains regulate trafficking to the recycling endosome (Takahashi *et al.*, 2012). The replacement of Rab5 by Rab7 has been investigated as a particularly crucial point mediating the proportion of internalised RTKs that are pushed towards recycling vs. degradation. PKCδ phosphorylated at Y374 has been shown to stabilise Rab5, preventing replacement by Rab7 and thereby tilting RTK fate towards recycling (Lonic *et al.*, 2021); the tumour suppressor PTPN14 negatively regulates PKC in this context (Belle *et al.*, 2015). Taken together, there is evidence for a sorting role for Rab4 and Rab5 at the early endosome, whilst Rab11 and recruitment of Rab7 mediate onward endocytic route of the internalised receptor (Redpath *et al.*, 2020). Alongside membrane heterogeneity, the morphology of the early endosome appears to be important; evidence in breast cancer cell lines suggests that differing levels of Rab4 alter early endosome size and number, with concomitant changes in cellular signalling and the trafficking of cargo (Tubbesing *et al.*, 2020).

1.3.2.3.3 Receptors targeted for degradation are trafficked to the lysosome

Receptors localised to Rab7-associated membranes are directed towards the late endosome-lysosomal degradative pathway. LEs have a lower pH than EEs (Griffiths, 1989) and their membranes are enriched in phosphatidylinositol (PI) (3,5) bisphosphate (PI(3,5,)P2) (Shaw *et al.*, 2003). They are generally considered more closely related to lysosomes than early endosome, though the growing consensus is that LEs are generated from multivesicular bodies, which are derived from EEs (Scott, Vacca and Gruenberg, 2014; Naslavsky and Caplan, 2018). Multivesicular bodies are the early endosome structures that undergo the Rab5-to-Rab7 conversion and are specifically defined by intralumenal vesicles, containing soluble internalised cargo. Intraluminal vesicles are formed from the invagination of the endosomal membrane; receptors bound for degradation are thought to be designated by ubiquitinated residues which are recognised by endosomal sorting complexes required for transport (ESCRT) complex members (Pryor and Luzio, 2009; Christ *et al.*, 2017). The mechanism of transfer of material intended for degradation to the lysosome from the late

endosome/multivesicular bodies is still unknown. A popular hypothesis is that late endosome/multivesicular bodies further mature in lysosomal compartments (Huotari and Helenius, 2011). This is supported by the molecular similarity of late endosome and lysosomal membranes; it is typical to only be able to distinguish them by their morphology using microscopy (Naslavsky and Caplan, 2018). A further distinction is intralumenal pH, which is far more acidic than endosomal compartments; it has been argued that a maturation model does not answer how intralumenal late endosome and lysosome content could be distinct (Pryor and Luzio, 2009). Therefore, a "kiss and run" model has been gaining traction, defined by transient tethering of late endosome and lysosome membranes, mediated by SNAREs and RABX1 (Pryor and Luzio, 2009; Saffi and Botelho, 2019; Laiouar *et al.*, 2020). Intracellular positioning of the lysosome is dependent on nutrient conditions; in starvation, lysosomes will cluster at the perinucleus, whilst mTORC1 activation in nourished cells correlates with localisation at peripheral lysosomes (Korolchuk *et al.*, 2011).

1.3.2.3.4 Retrieval from the degradative pathway via retromer and retriever

Internalised receptors in Rab5/Rab7 endosomal compartments, or in Rab5/Rab7 enriched early endosome membrane domains, can still escape the degradative pathway in order to be recycled, directly or via the trans-Golgi network (Redpath *et al.*, 2020). Escape from the degradative pathway is mediated by the retromer and retriever/CCC (together termed commander) complexes, which recognise specific sorting sequences in membrane bound cargo (Cullen and Steinberg, 2018). Retromer is thought to be composed of two subcomplexes: VPS5:VPS17, which are sorting nexins (SNX) that contain BAR domains (SNX-BAR), and the evolutionary conserved VPS26:VPS29:VPS35, termed the cargo-selective complex (CSC). CSC was originally characterised in yeast, however subsequent studies in higher eukaryotes suggest a more transient role for SNX-BAR (Chen, Healy and Collins, 2019). It has been speculated that CSC enrichment in endosomal membranes captures cargo requiring export from late endosome/ multivesicular body (Cullen and Steinberg, 2018; Norris and Grant, 2020). Briefly, retromer mediates cargo removal by first linking the cargo to the actin cytoskeleton through association of VPS35 with actin-binding Wiskott–Aldrich syndrome and SCAR homologue (WASH) complex. The proximity of WASH to the membrane initiates actin restructuring and membrane deformation by the proteins in retromer containing a SNX-BAR domain, forming membrane tubules that bud off from the endosomal compartment (Burd and Cullen, 2014; Cullen and Steinberg, 2018).

The retriever complex was discovered more recently and is highly similar to retromer. It is composed of DSCR3, C16orf62 and VPS29, and was found to function with CCDC93, CCDC22

and COMMD (the CCC complex), which together are termed the commander complex (McNally *et al.*, 2017; Chen, Healy and Collins, 2019). Similarities between commander and retromer include a close association with WASH, the subunit Vps29 and predicted structural similarities (though little sequence similarity) (McNally *et al.*, 2017; Chen, Healy and Collins, 2019). The molecular and functional details of retromer and commander mediated cargo recycling is still being uncovered, and indeed how these complexes recognise unique cargos to be rescued from degradative pathways and their route to recycling. The broad hypothesis is that retromer and retriever coated compartments are destined for the trans-Golgi network, from which cargo will be trafficked back to the plasma membrane. Recent evidence demonstrated that EGFR utilises a retromer-dependent pathway of recycling in unstimulated conditions, that is disrupted upon EGF activation (Ioannou *et al.*, 2017).

1.3.2.3.5 Recycling to the plasma membrane via the recycling endosome

It is thought that only a small population of internalised receptors utilise the retromer/retriever mechanism of recycling, though more work into this area is needed. The bulk of internalised material that is recycled from the early endosome back to the plasma membrane is diverted from the early endosome rather than the late endosome. Two routes back to the plasma membrane from the early endosome have been described: “fast” recycling, that recycles directly from the early endosome, or “slow”, which transports cargo for recycling via the recycling endosome. Fast recycling is regulated by Rab35 (Kouranti *et al.*, 2006; Patino-Lopez *et al.*, 2008).

An interesting characteristic of the recycling endosome is its positioning depending on cell function. In most cell types, the recycling endosome is localised in the perinuclear region, where it appears closely associated with the microtubule organising complex (MTOC) (Horgan *et al.*, 2010; Hehnly *et al.*, 2012), however in neurones and T-cells the recycling endosome is dispersed throughout the cell, likely due to the high levels of endo- and exocytosis (Joensuu *et al.*, 2017; Compeer *et al.*, 2018). Rab11 is generally considered as the marker of the recycling endosome, and it is also defined by its tubular structure (Ullrich *et al.*, 1996; Maxfield and McGraw, 2004; Takahashi *et al.*, 2012). Like the early endosome, the recycling endosome shows membrane heterogeneity, with cargo internalised by clathrin-mediated endocytosis and clathrin-independent endocytosis segregated in the recycling endosome; the functional role of this segregation has yet to be fully investigated (Xie *et al.*, 2016).

Cargo targeted for slow recycling via the recycling endosome can be found in Rab4/11-enriched early endosome membrane domains, where Rab11 is in an inactive state (Redpath *et al.*, 2019). Once activated, by PI(3)P enrichment and PI3KCA, Rab11 displaces Rab5 (Franco *et al.*, 2014; Campa *et al.*, 2018). The population of PI(3)P is critical, as Rab11 dissociates if the PI(3)P population is not maintained (Campa *et al.*, 2018). As with late endosome-lysosome material transfer, there are two proposed models for early endosome-recycling endosome transfer: Rab11-positive vesicle budding from the early endosome, or a "kiss and run" model defined by membrane contact between the early endosome and recycling endosome. The former requires the recruitment of fission molecules such as the microtubule motor protein dynein, recruited by the Rab11-interacting protein Rab11-FIP3 (Horgan *et al.*, 2010), which is thought to coordinate fission with the Rab-FIP5 recruited components SNX4/1 (Traer *et al.*, 2007; Campa *et al.*, 2018). Once the vesicle has budded from the early endosome membrane, PI(3)P is rapidly converted to PI(4)P by MTM1 and phosphatidylinositol 4-kinase-2α (PI4K2α), which is enriched in the recycling endosome membrane (Ketel *et al.*, 2016). The more recently described kiss-and-run model for recycling endosome-early endosome cargo transfer posits the recruitment of the factors for endosome recycling and Rab interaction (FERARI) complex to the early endosome membrane by Rab5 (Solinger *et al.*, 2020). Rab11-FIP5 acts as a bridge between FERARI and Rab11-positive compartments; the contact between the membranes themselves has been suggested to be provided by EHD1 (Solinger *et al.*, 2020), which is recruited to phosphatidyl-serine positive compartments (Uchida *et al.*, 2011; Lee *et al.*, 2015). These two different models for cargo transfer would both depend on the movement of endocytic compartments, on microtubule tracks, to different regions of the cell, due to the apical and perinuclear localisation of the early endosome and recycling endosome, respectively (Akhmanova and Hammer, 2010; Takahashi *et al.*, 2012; Semenova *et al.*, 2014).

It is known that Rab11 depletion leads to accumulation of cargo in plasma membrane-adjacent compartments, indicating it's crucial role in mediating exocytosis from the recycling endosome back to the plasma membrane (Takahashi *et al.*, 2012). This study also demonstrated that removal of Rab11 disturbed recruitment of the exocyst complex. The exocyst mediates return of cargo to the plasma membrane, and is recruited to membranes enriched for PI(4)P. The hydrolysis and phosphorylation of PI(3)P enriched in the recycling endosome membrane by the phosphatidylinositol 3-phosphatase MTM1 and PI3K2α, respectively, results in pools of PI(4)P to which the exocyst can be recruited to (Ketel *et al.*, 2016; Campa *et al.*, 2018).

1.3.2.3.6 Intracellular trafficking of RTKs is a ligand:receptor directed process

The trafficking route of ligand-activated RTKs appears dependent on the ligand:receptor pair, as well as the ligand-dependent regulation described in the section 1.3.1. In terms of functional selectivity, it has long been established that the seven EGFR ligands have a differential effect on receptor sorting (Roepstorff *et al.*, 2009). For example, EGF and TGF-α induce EGFR degradation or recycling, respectively (Decker, 1990; Ebner and Derynck, 1991). Early evidence suggested that this is partially a function of the pH sensitivity of the ligand, with TGF-α dissociating from the receptor at endosomal pH but EGF staying in the receptor:ligand pair (French *et al.*, 1995), though this has been disputed through evidence of TGF-α induced EGFR endosomal signalling (Francavilla *et al.*, 2016). A similar switch has been defined for FGFR2b, with FGF7 inducing degradation and FGF10 inducing recycling of the receptor (Francavilla *et al.*, 2013). Ligands alone have been shown not to encode the trafficking route, as FGF1 stimulated FGFR1 and FGFR3 were shown to have different internalisation dynamics, with FGF1 inducing slower FGFR3 internalisation and utilising clathrin-independent endocytosis alongside clathrin-mediated endocytosis, unlike FGFR1 which is only internalised by clathrin-mediated endocytosis (Haugsten *et al.*, 2011). The receptor:ligand relationship is not unique to RTKs; biased antagonists of GPCRs may present therapeutic value by blocking G protein signalling but not GPCR internalization (Hitchinson *et al.*, 2018; Lorenzen *et al.*, 2018)

As previously mentioned, ligand abundance has been shown to influence EGFR trafficking; at high EGF concentrations (> 10 ng/ml), EGFR tends to be internalised by clathrin-independent endocytosis, targeting the receptor towards degradation and therefore attenuating the signal. At lower concentrations, the clathrin-mediated endocytosis pathway is utilised, and signalling is prolonged. Although there is little evidence suggesting that, in general, clathrin-independent endocytosis is the internalisation pathway for RTKs targeted for degradation (especially considering the evidence of segregated clathrin-mediated endocytosis and clathrin-independent endocytosis cargo in the recycling endosome), in the case of EGFR clathrin-independent endocytosis is only utilised upon high levels of EGF:EGFR pairs and saturation of clathrin-mediated endocytosis machinery (Schmidt-Glenewinkel *et al.*, 2008; Sigismund *et al.*, 2013; Hemalatha and Mayor, 2019). The authors of the original work posit a ubiquitination threshold that controls the fate of the receptor and the regulation of downstream signalling, demonstrating the important role of PTM in the decisions of RTK fate (Sigismund *et al.*, 2005, 2013). Recent work has added nuance to the hypothesis of dose-dependent regulation of EGFR trafficking; in mitosis, the concentration dependent effect of

EGF:EGFR internalisation is lost, with even low doses of EGF (2 ng/mL) inducing internalisation by clathrin-independent endocytosis. The authors attributed the switch to clathrin-independent endocytosis to the mitosis-induced ubiquitination of EGFR (Wee and Wang, 2018).

The trafficking of FGF1:FGFR receptor pairs has also been studied as a function of ubiquitination. FGF:FGFR4 was shown to recycle to the plasma membrane following internalisation, whilst FGF1:FGFR1-3 were primarily degraded in the lysosome, with FGF1:FGFR1 showing the most efficient degradation (Haugsten *et al.*, 2005). It was later shown that recycling of FGF1:FGFR4 is Rab11 dependent; when Rab11 is depleted, signalling downstream of FGF1-FGFR4 was deregulated (Haugsten *et al.*, 2014). Total ubiquitination on FGFR4 was less than FGFR1-3 upon FGF1 stimulation, with the most influential modifications occurring on the C-terminal tail (Haugsten *et al.*, 2005; Sørensen *et al.*, 2006). Using proximity labelling, it has been shown that trafficking of FGFR4 upon FGF1 stimulation is via clathrin-mediated endocytosis, via the recycling endosome and trans-Golgi network (Haugsten *et al.*, 2016). Whether FGF1:FGFR4 internalisation and trafficking is modulated in a concentration dependent manner, as with EGF:EGFR, has not been investigated. The ligands of FGFR2b have also been shown to have a differential effect on receptor trafficking; FGF7 induces Cbl-mediated degradation and FGF10 induces recycling of FGFR2b (Belleudi *et al.*, 2007).

Though ubiquitination of RTKs is generally associated with degradative pathways, a sophisticated multi-proteomics screen by Francavilla et al. (2016) demonstrated that TGF-α:EGFR is recycled despite extensive ubiquitination of the receptor. Moreover, phosphorylation has been shown to have an equally deterministic role in intracellular trafficking. For example, phosphorylation of FGFR1-S789 by RSK2 has been shown to have a role in FGFR1 internalisation (Nadratowska-Wesolowska *et al.*, 2014).

1.3.2.3.7 Regulation of RTKs through endocytic trafficking

The endosomal processes described here represent a "canonical" view of the role internalised receptors play in signalling cascade regulation; the prevalent view has been that internalisation is the process by which RTK activity is attenuated. The activity attenuation is prolonged in the long-term by degradation or in the short-term by recycling. This view is backed by evidence of the majority of signalling processes being initiated and regulated from the plasma membrane. It has been demonstrated that a substantial portion of EGFR induced signalling, particularly through MAPK and AKT, is initiated from the plasma membrane (Sousa *et al.*, 2012). Depletion of ESCRT machinery, which prevented EGFR being trafficked to the

lysosome and retains it in the endosomes, was shown to increase the activity of EGFR and MAPK signalling, but didn't alter the transcriptional profile of target genes (Brankatschk *et al.*, 2012). Early work from Wells et al. (1990) found that the internalisation of EGFR preceded mitosis, however cells expressing an internalisation-defunct copy of EGFR entered mitosis at an eighth of the concentration of EGF than wild type cells; thus, internalisation acted as an attenuator of a transformed phenotype. These studies confirm the role of receptor internalisation as a means of physical or spatial downregulation through the reduction of receptor levels at the plasma membrane. This also alters sensitivity to ligand concentration (Villaseñor, Kalaidzidis and Zerial, 2016).

However, the nuanced role of continued signalling from activated & internalised RTKs has yet to be fully explored, due to the difficulty of isolating localised signals from specific sub cellular compartments (Lill and Sever, 2012; Sigismund *et al.*, 2012). During trafficking, spontaneous or continuous activation of the receptor (as occurs at the plasma membrane) is partially inhibited by high abundance of PTP1B on vesicular trafficking routes (Baumdick *et al.*, 2015). Vesicular dynamics determine a spatially distributed EGFR-PTP network that may act as a method of spatiotemporal regulation (Eden *et al.*, 2010; Stanoev *et al.*, 2018). However, the activity of phosphatases may be attenuated by the generation of reactive oxygen species, particularly hydrogen peroxide, by nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex found on endosome membranes; reactive oxygen species induce the reversible inactivation of tyrosine phosphatases, leading to potentially sustained signalling (Janssen-Heininger *et al.*, 2008; Oakley *et al.*, 2009; Miaczynska, 2013). Moreover, it has been shown that RTKs can continue to recruit signalling proteins from the early endosomes, such as Grb2 (Fortian and Sorkin, 2014), Vav2 (Ménard, Parker and Kermorgant, 2014) and Shc (Villaseñor *et al.*, 2015). An intriguing recent paper from my group pointed to a point of crosstalk between the FGFR and EGFR pathways, through the reciprocal priming of internalised FGFR2b and EGFR (Smith *et al.*, 2021); phosphorylation by ERK of T693 on EGFR, downstream of FGF10 stimulation and FGFR2b activation, was shown to drive cell cycle progression downstream of both receptors.

It is becoming clear that the RTK internalisation is not solely a means in reducing receptor activity. Instead, RTKs are able to sustain or induce particular signalling modules intracellularly (Sigismund *et al.*, 2021). The role of the downstream signalling proteins activated subcellularly is discussed in greater detail in section 1.3.3.2, however the activity of the receptors themselves, whether through priming, recruitment of scaffolds, or selective

activation or deactivation, provides a yet-to-be understood layer of nuance to these processes. There are indications of pathophysiological relevance, for example in cancer metastasis (Boissan *et al.*, 2014; Khan, Gril and Steeg, 2019; Khan and Steeg, 2021). Tellingly, pharmaceutical interventions for cancer treatment targeting RTKs at particular points of their endocytic trafficking routes are being considered, for example for IGF-1R, which undergoes oncogenic rewiring of its trafficking and thus has deregulated signalling (Crudden *et al.*, 2019).

1.3.3 Regulation from the perspective of intracellular signalling networks

If only the perspective of the ligand and receptor were taken into account, it would be reasonable to suggest that the ligand:RTK pair encodes the downstream cellular response. However these serve primarily as an input device, converting external stimuli into internal signals that are processed by the cell in a sensitive, non-linear manner (Brent, 2009). The internal component orchestrates a response to a multitude of extracellular signals; predeterminants of the response, such as abundance or expression of particular signalling proteins, can be cell specific. Therefore, the cellular context is important for the regulation of RTK signalling (Lam, Pickering and Mac Gabhann, 2019).

1.3.3.1 Dynamic activity of signalling proteins

The few signalling pathways downstream of RTKs, described in section 1.2.1.2, mediate a huge number of effects on cellular behaviour and transcriptional programming (through the regulation of transcription factors). This is commonly referred to as the hourglass effect, with many possible inputs being processed by a limited and conserved set of intracellular proteins and pathways, leading to a multitude of potential impacts on cell behaviour & fate (Figure 1-6A) (Oda *et al.*, 2005; Lam, Pickering and Mac Gabhann, 2019). Moreover, crosstalk between the pathways creates a highly modular, modifiable system. Since there are many ways to achieve the same functional outcome, keeping the pool of core signalling proteins small whilst being able to encode the signal through PTMs or changes in expression is considered a feature of a robust signalling system (Kitano, 2007). An example of this is in cell cycle progression, where altered phosphorylation stoichiometry on core proteins induces dynamic responses (Olsen *et al.*, 2010). Indeed, protein modification state is inherently linked to its function; proteins fluctuate through different conformational states, with modifications on one region of the protein being sensed by another region despite spatial distance (Tang and Kaneko, 2020). *En masse,* protein modification thus influences cell behaviour.

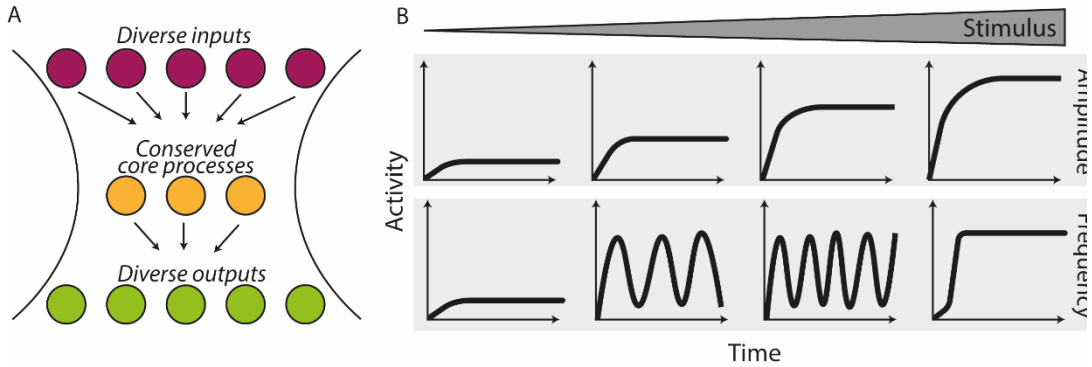


Figure 1-6. Conceptualisation of dynamic signalling processes.
A. *The hourglass effect of intracellular signalling: diverse inputs are processed by a limited, core set of proteins that result in diver outputs.* ***B.*** *Protein modification dynamics follow patterns of amplitude or frequency modulation.*

Intracellular protein modification dynamics, and thus activity, are strongly linked to the activating ligand:RTK pair. The terms frequency or amplitude-modulation are often used to describe the modification profiles of the signalling response (Figure 1-6B). A well-known example has been described in PC-12 cells; upon EGF stimulation, ERK activity is transient and results in cell cycle entry, whilst stimulation with NGF results in sustained ERK activity leading to differentiation (Marshall, 1995). These activity levels occur in pulses that are similar in amplitude; increasing EGF concentration leads to increased frequency of the pulses, rather than increased amplitude (Albeck, Mills and Brugge, 2013). Mathematical modelling using single-cell imaging data has suggested that the pulses are the result of negative and positive feedback loops within the signalling network, rather than driven from the activity of the receptor itself (Sparta *et al.*, 2015). Similar pulsatile behaviour is also exemplified by NF-κB in response to TNF-α (Ashall *et al.*, 2009). At the level of transcription factors, FOXO3, under the combinatorial control of AKT and ERK, was found to encode the identity and abundance of growth factors through oscillatory translocations between the nucleus and cytosol (Sampattavanich *et al.*, 2018). To take one example from this study, the identity of IGF1 was encoded by high levels of activity through the AKT pathway and less so through the ERK pathway; this was associated with early translocation of FOXO3 to the nucleus, followed by pulsatile translocation between the cytosol and nucleus (Sampattavanich *et al.*, 2018). A study specifically investigating the dynamics of AKT activity found that IGF1 induces sustained Akt phosphorylation whilst EGF induces transient phosphorylation (Gross and Rotwein, 2016). The latter correlates with the known EGF-ERK dynamics discussed previously.

The flow of signal through the network has several characteristics. First, it is combinatorial, having to account for multiple and complex signals that the cell needs to decode. This requires more than the correct combination of players to be present in the system, but also for signalling proteins to be activated (for example through phosphorylation) and in the correct subcellular location. Alterations of these sets of factors could inadvertently induce an alternative cellular response (Buchler, Gerland and Hwa, 2003; Behar and Hoffmann, 2010; Adlung *et al.*, 2017). Relatedly, the network needs to account for noise. It has been suggested that in biological networks noise is more than a problematic biproduct of living systems, but also plays a crucial role in processes such as probabilistic differentiation of cells across tissue populations during development (Eldar and Elowitz, 2010). Finally, it has become well accepted that PTMs, particularly phosphorylation, are a useful read-out for the study of dynamic signalling systems. Though many PTMs and biomolecules are required for transmission of intracellular signals, the role of phosphorylation is generally considered to be less redundant, since kinases are often mutated or have altered expression in diseases of cell signalling, such as cancer (Day, Sosale and Lazzara, 2016; Yaffe, 2019). High-throughput phosphoproteomics methods, described further in section 1.4, are opening the way to large-scale models of context-dependent signalling network dynamics and topology, allowing all these characteristics to be defined in more detail (Choudhary and Mann, 2010; Imamura *et al.*, 2010; Roux and Thibault, 2013).

1.3.3.2 Subcellular signalling

As described in section 1.3.2.3.7, the internalisation of RTKs does not necessarily impede their ability to signal; their subcellar location exposes them to a distinct interactome for eliciting further intracellular responses. Likewise, the subcellular localisation of key signalling proteins is an important regulator of their function and activity. In its simplest form, the role of localisation can be modelled in terms of distance. Considering, for example, a protein phosphorylated by an RTK at the plasma membrane, a gradient of phosphorylated to unphosphorylated protein will form with distance from the membrane, as those proteins become closer to cytosolic phosphatases (Brown and Kholodenko, 1999; Kholodenko, 2009). Thus, domains of restricted activity are created through the cytoplasm.

A layer of complexity can be added to this model, by the tethering or association of populations of signalling proteins to cellular compartments. One example is RAS, which is known to localise to the plasma membrane, Golgi apparatus, endoplasmic reticulum, and endosomal system (Matallanas *et al.*, 2006; Prior and Hancock, 2012; Santra *et al.*, 2019). In turn, pools of activated ERK are found in the same locations, which is thought to be tethered

to these membranes by scaffold proteins (Chiu *et al.*, 2002; Casar *et al.*, 2009; Lavoie, Gagnon and Therrien, 2020). The functional impact of RAS-ERK localisation remains to be elucidated, as does the role of scaffolds, whether they are nucleators of signalling complexes or allosteric regulators of RAS-ERK signalling. PKC has also been shown to have roles in the endosomal system, and in the regulation of EGFR fate following stimulation. When phosphorylated at Y374 by the tyrosine kinase FER, Lonic et al. (2021) showed that PKC colocalised with internalised EGFR. This prevented RAB5 release from the same endosomes, promoting the recycling rather than the degradation of EGFR.

AKT/PI3K localisation in and signalling from the endosomal system has received increased attention in recent years. Though the plasma membrane hosts the largest populations of PI3K and AKT, both are also found within the endosomal system. Although there is decades old evidence of PI3K recruitment to the endosomes following growth factor stimulation, it had seemingly fallen out of the narrative of the cell signalling community (Kelly, Ruderman and Chen, 1992; M. Sato *et al.*, 2003). The prevailing view stated that PI3K and AKT were always activated at the plasma membrane, and redistributed to other compartments as opposed to activation *in-situ* (Sugiyama, Fairn and Antonescu, 2019; Batrouni and Baskin, 2020). Recent evidence suggests there may be more nuance. A significant paper recently published by Thapa et al (2020) reported that PI3K was anchored to internal membrane vesicles aligned along microtubules by the microtubule associated protein 4 (MAP4). When RTKs are internalised following activation (a microtubule-dependent process), they are brought into proximity with PI3K leading to its internal activation. Work by Francavilla et al. (2013) suggested that PI3K can take a more active role in driving the internalisation of FGFR2b upon FGF10 activation; whether this is related to the PI3K-MAP4 interaction warrants further investigation. AKT, as a key PI3K effector, may also display spatial regulation. Of the three canonical AKT isoforms, AKT1 and AKT3 appear to follow the typical paradigm of activation at the plasma membrane, however recent evidence suggests that AKT2 can be selectively activated at the early endosomes, where it is recruited following PI3K activation (Liu *et al.*, 2018). Finally, PTEN may also be tethered to endosomal membranes, thereby negatively regulating PI3K activity, and AKT recruitment, in these cellular compartments (Naguib *et al.*, 2015). Together, this body of evidence describes a signalling module within the endosomal system. It remains to be seen whether the endosomal population of PI3K/AKT is entirely distinct from the population at the plasma membrane, or whether these modules work in concert.

These examples all describe canonical growth factor signalling pathways that are primarily activated at the plasma membrane but can be diverted to subcellular compartments. However, some elements of the intracellular signalling network are solely regulated at organelles. mechanistic Target of Rapamycin (mTOR) signalling at the lysosome is perhaps the best studied example (Figure 1-7A). mTOR is a S/T kinase that exerts its activity as part of one of two complexes, named mTORC1 and mTORC2. Both complexes contain three core components and several regulatory subunits. Along with mTOR, mTORC1 is composed of regulatory protein associated with mTOR (Raptor) (Hara *et al.*, 2002; Kim *et al.*, 2002), which is responsible for substrate recruitment (Nojima *et al.*, 2003) and correct localisation of the complex (Kwak *et al.*, 2016), and mammalian lethal with Sec13 protein 8 (mLST8, also known as GβL) (Kim *et al.*, 2003), which is thought to help stabilise the mTOR kinase activation loop (Yang *et al.*, 2013). mTORC2 also contains mTOR and mLST8, but contains rapamycin insensitive companion of mTOR (Rictor) that serves an analogous role to Raptor in mTORC1 (Sarbassov *et al.*, 2004). Both complexes contain an inhibitory subunit named DEP domain containing mTOR interacting protein (DEPTOR) (Peterson *et al.*, 2009), as well as some complex-specific regulatory components (Saxton and Sabatini, 2017). Generally, mTOR signalling is considered a master regulator of autophagy and cell growth, promoting cell survival and growth whilst inhibiting autophagy. mTORC1 and mTORC2 play distinct roles in these intersecting processes, through activation by different stimuli. Both are solely active at the lysosome, with some evidence suggesting they may also function at the RAB11-positive endosomes (Longatti *et al.*, 2012).

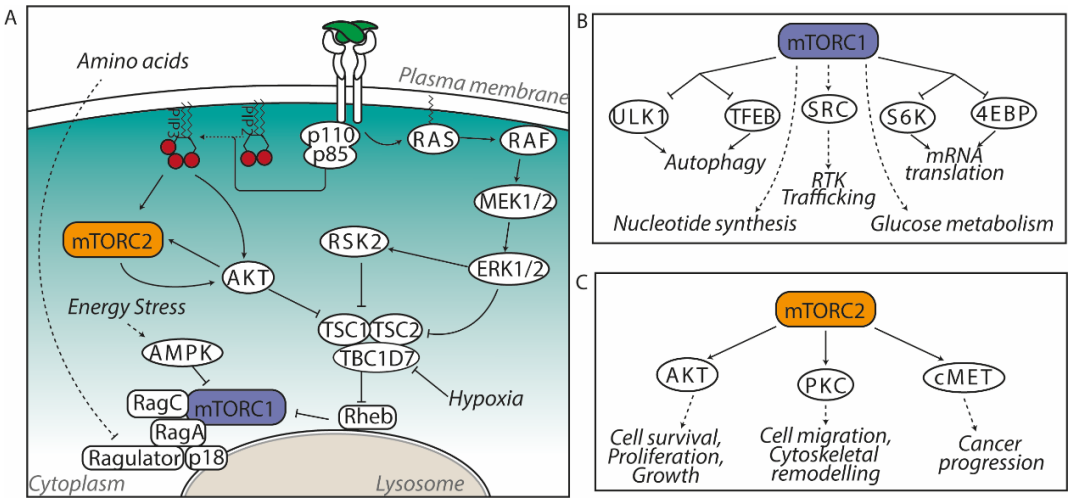


Figure 1-7. mTOR signalling is mediated from intracellular compartments.
A. *mTORC1 and mTORC2 activation by RTKs, hypoxia and nutrient stress.* ***B.*** *mTORC1 effectors and downstream outputs.* ***C.*** *mTORC2 effectors and downstream outputs. Adapted from Saxton and Sabitini (2017).*

mTORC1 is sensitive to a number of conditions that need to be met for its activity to be induced: normoxia, sufficient nutrition (I.e. amino acids, glucose), lack of cellular stress (e.g. low reactive oxygen species), and the presence of pro-growth signals (Saxton and Sabatini, 2017) . When these conditions are not met, mTORC1 is under the inhibition of its negative regulator, TSC complex (composed of TSC1, TSC2, and TBC1D7) (Dibble *et al.*, 2012), which is recruited to the lysosome by Ras GTPase-activating protein-binding proteins (G3BPs) (Prentzell *et al.*, 2021). There it acts as a GTPase activating protein for the crucial mTORC1 activator Rheb (Long *et al.*, 2005). Consequently, when TSC1 is activated cell growth is repressed, and autophagy is activated (Inoki *et al.*, 2003; Tee *et al.*, 2003). Growth factor signalling and permissive environmental cues induce the inhibition of TSC; in the case of growth factor signalling, convergence on TSC occurs via AKT or ERK/RSK, which both phosphorylate and inhibit TSC2 (Inoki *et al.*, 2002; Roux *et al.*, 2004; Ballif *et al.*, 2005; Menon *et al.*, 2014; Jacobs *et al.*, 2017). When the balance is tilted in favour of mTORC1 activation, the complex is recruited to the lysosome where it can interact with its essential activator, Rheb. Rheb itself must be recruited to the lysosomal membrane, which is prevented in amino-acid deprived conditions (Garrido, Brandt and Djouder, 2016). mTORC1 activation on the lysosomal membrane occurs at a lysosomal protein complex called the Ragulator, which is anchored to the lysosome by p18 (Sancak *et al.*, 2010). One of the key functions of the Ragulator is evidenced by its association with the lysosomal vacuolar ATPase (v-ATPase) (Zoncu *et al.*, 2011). v-ATPase allows the amino acid sensing subunits of the Ragulator, the GTPases RagA and RagC, to sense availability of these nutrients. In low amino-acid conditions, RagA/C bind GTP, preventing mTORC1 recruitment to the Ragulator and thus mTORC1 activation (Kim *et al.*, 2008; Sancak *et al.*, 2010).

Active mTORC1 mediates a huge number of downstream processes that are largely related to anabolism (Lamming and Bar-Peled, 2019). The best described downstream process is the inhibition of autophagy, which can be broadly defined as the bulk catabolism of cellular components, induced by stress conditions such as starvation (Bento *et al.*, 2016). Autophagy induces the formation of the autophagosome, a membrane-bound compartment that eventually fuses with a lysosome for the breakdown of the internal components (Bento *et al.*, 2016). A well-studied target of mTOR is ULK1, which initiates autophagy through autophagosome formation (Zachari and Ganley, 2017). mTORC1 is known to inhibit ULK1 by phosphorylation, as well as its interacting partner ATG13 (Hosokawa *et al.*, 2009; Jung *et al.*, 2009). The inhibition of ULK1 by mTORC1 is sufficient to prevent its activation by AMP activated protein kinase (AMPK), a nutrient sensor and autophagy promotor (Kim *et al.*,

2011). ULK1/ATG13 are also regulated through the cell cycle by CDK1 phosphorylation, when mTOR is also inhibited but autophagy must be controlled to allow for cell cycle progression (Li *et al.*, 2016, 2020; Odle *et al.*, 2020).

The autophagic transcriptional programme is also downregulated by mTORC1, which phosphorylates transcription factor EB (TFEB) creating a binding site for 14-3-3 proteins, preventing its nuclear translocation and thus its ability to transcribe genes critical for autophagy to proceed (Martina *et al.*, 2012; Roczniak-Ferguson *et al.*, 2012; Settembre *et al.*, 2012). The dampening of autophagy is a hallmark of mTORC1 signalling, however mTORC1 also mediates metabolic processes as distinct as the induction of protein translation from mRNA (via S6K1 and 4EBP) (Düvel *et al.*, 2010; Thoreen *et al.*, 2012), nucleotide synthesis (Ben-Sahra *et al.*, 2016), and promoting a shift from oxidative phosphorylation to glycolysis (Düvel *et al.*, 2010; Saxton and Sabatini, 2017).

The convergence of RTK activity on mTORC1 is also apparent through the effect on other cellular processes (Figure 1-7B). An intriguing study by Dauner et al. (2017) demonstrated a critical role for mTORC1 activity to maintain delivery of endocytosed material to the recycling endosome; when mTORC1 was inactive, material bound to the recycling endosome was delivered to the lysosome instead. Dauner et al. focused primarily on the trafficking in nutrient rich or starved conditions, but in a separate study the same effect was observed in EGFR trafficking, and was shown to be SRC-dependent (Colella *et al.*, 2020). This provides a point of regulation of mTORC1 on RTKs, complementing the better studied converse. Another example of the role of mTORC1 in RTK signalling regulation is evidence of mTORC1 interactors enhancing ERK activity (Martinez-Lopez *et al.*, 2013).

mTORC2 is far less understood than mTORC1, due to it being discovered more recently. Its activity is thought to be primarily reliant on cues from insulin as well as PI3K signalling downstream of GFs. A recent phosphoproteomic study showed that loss of RICTOR in brown adipocytes suppressed insulin signalling via AKT (Entwisle *et al.*, 2020). There is evidence of the PH-domain of the mTORC2 subunit SIN1 being the point of recruitment to the plasma membrane upon PIP3 generation by PI3K (Liu *et al.*, 2015). A putative positive feedback loop through AKT phosphorylation of SIN1 has also been described, which enhances mTORC2 kinase activity (G. Yang *et al.*, 2015). However, the spatial regulation of the complex has not been thoroughly investigated. mTORC2 regulation should be of interest to the RTK signalling community, due to its downstream effectors (Figure 1-7C). For example, mTORC2 phosphorylates AKT, promoting cell survival, proliferation and growth through the effectors

described in section 1.2.1.2.2 (Sarbassov *et al.*, 2005; Baffi *et al.*, 2021). mTORC2 also phosphorylates numerous members of the PKC family, regulating cell migration and cytoskeletal remodelling (Sarbassov *et al.*, 2004; Gan *et al.*, 2012; Li and Gao, 2014; Baffi *et al.*, 2021). It has also been shown in colorectal cancer cells that basal-PI3K/AKT/mTORC1 activation is insufficient to supress autophagy; the loss of autophagy-inhibition by mTORC1 was attributed to mTORC2-dependent phosphorylation of several RTKs, including c-MET (Lampada *et al.*, 2017).

The increased understanding of mTOR signalling has changed the perception in the signalling community of the lysosome. It is now understood to be more than the "cellular recycling centre" and instead a "metabolic signalling hub", where nutrients are stored and cell behaviour and metabolism is regulated from, primarily via mTORC1; furthermore, the localisation of mTORC1 and its regulators is critical for its function (Demetriades, Plescher and Teleman, 2016; Lamming and Bar-Peled, 2019). For example, recent work demonstrated that lysosomes proximal to focal adhesions were hubs of mTORC1 activity, where the lysosome coupled mTORC1 to EGFR and IGFR1 enriched in the focal adhesion, in order to sense amino acid availability (Rabanal-Ruiz *et al.*, 2021).

1.4 Studying signalling processes with phosphoproteomics

Prior to the advent of high-throughput technologies to profile proteins, experimental biologists were reliant on traditional low-throughput biochemical methods, such as western blotting, to identify proteins of interest and their modifications. The rapid technological advances in mass spectrometry-based (MS) proteomics in the 1990s and early 2000s provided experimentalists the ability to identify and measure many proteins simultaneously. The technology was well adopted by cell biologists investigating signalling cascades and other intracellular changes; though it was known that phosphorylation was crucial in these processes, the coverage and frequency of stimuli- or environment-induced phosphorylation was truly appreciated upon the inspection of phosphoproteomes (Venerando, Cesaro and Pinna, 2017). However, proteomics has yet to gain the accessibility of genomics or transcriptomics methods (Sidoli, Kulej and Garcia, 2017). Furthermore, phosphoproteomics methodology currently lags behind proteomics in terms of resolution and depth (Solari *et al.*, 2015).

Here, I summarise common protocols and practices for mass spectrometry-based proteomics and phosphoproteomics, how recent advances continue to improve and widen the uses of the technology, and the technological caveats and limitations. As the work in this

thesis pertains to bottom-up (or "shotgun") proteomics approaches, top-down approaches and targeted approaches such as selected reaction monitoring (SRM; also referred to as multiple reaction monitoring, MRM) will not be discussed here.

1.4.1 Mass spectrometry-based proteomics

1.4.1.1 Sample preparation and instrumentation

Proteins from cells and tissues are extracted, purified, and denatured, before being digested into peptides, commonly using trypsin which cleaves at lysine or arginine residues. This generates predictable peptide sequences, between 6 -40 amino acids long, that can be compared against pre-existing peptide libraries (Swaney, Wenger and Coon, 2010). Digestion performance can be improved by a pre-digestion step with Lys-C; this is particularly useful for phosphoproteomics samples, where phosphorylation groups close to the trypsin cleavage site have been reported to reduce digestion efficiency (Wiśniewski and Mann, 2012; Dickhut *et al.*, 2014). Digestion can be performed on protein samples fractionated in an SDS-PAGE gel (so-called in-gel digestion) or on samples in solution (Shevchenko *et al.*, 2006). Fractionation is useful for separating complex protein samples into less-complex, molecular weight-related groupings. Though not essential, by reducing the samples through fractionation proteins can be better resolved by the mass spectrometer (Manadas *et al.*, 2010).

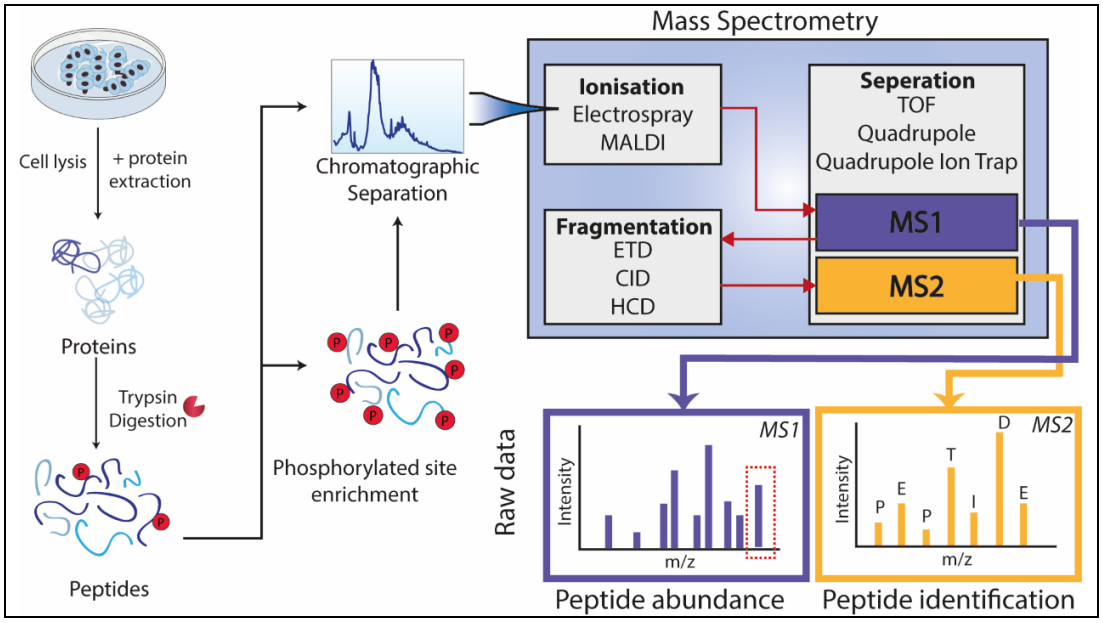


Figure 1-8. Mass-spectrometry based (phospho)proteomics.

As phosphorylated peptides comprise a small percentage of the total peptide population in a given sample, enrichment for phosphorylation gives a better chance of accurate detection by the mass spectrometer (Roux and Thibault, 2013; Low *et al.*, 2021). A consequence of this

is a significant increase in the required amount of starting material to still have a sufficient representation of the sample following enrichment; a standard proteome analysis requires 50 µg or less of protein, whereas for phosphoproteomes 3-5 mg is typically recommended for non-specialized laboratories (Roux and Thibault, 2013). Different enrichment methods can be employed to extract modified peptides from the digested sample. Titanium dioxide (TiO_2) beads can be used to extract phosphorylated peptides on S, T and Y residues. In an acidic environment, TiO_2 has a positively charged surface that selectively attracts the negatively charged phosphorylation groups in the peptide mix (Thingholm *et al.*, 2006; Dunn, Reid and Bruening, 2009). Immobilized Metal Ions Affinity Chromatography (IMAC) uses affinity chromatography to extract phosphorylated peptides from the sample. An IMAC column is composed of three parts: metal (III/IV) cations (e.g. Fe^{3+}, Ti^{4+}; binds the negatively charged phosphorylated peptides), a chelating group (e.g. iminodiacetic acid) and a supporting matrix (Posewitz and Tempst, 1999; Zhou *et al.*, 2013; Bekker-Jensen *et al.*, 2020; Low *et al.*, 2021). Alternatively, phosphotyrosine antibodies can be used to extract peptides with only pY residues (Rush *et al.*, 2005; van der Mijn *et al.*, 2015). As pY accounts for less than 1% of total phosphorylation (Sharma *et al.*, 2014), an extra enrichment with pY antibodies increases resolution. pY enrichment has been successfully used in dissecting signalling-induced phosphorylation patterns, for example downstream of EGF (Sharma *et al.*, 2014; Lundby *et al.*, 2019) and through the spleen tyrosine kinase (Syk) in breast cancer cells (Naldi *et al.*, 2017). Though the focus here is on phosphoproteomes, many other PTMs can be enriched using similar capture techniques.

The final step prior to detection by MS is separation of the peptide mix. Peptides digested in-gel are extracted from the gel and put back into solution before this step. Separation by liquid chromatography (LC) is most common; the LC column is often coupled, or “on-line”, with the mass spectrometer, in a system called LC-MS, allowing the sample, separated by hydrophobicity, to be fed directly from the chromatography column to the mass spectrometer. As with fractionation, the separation of the sample aids in reducing the complexity of the sample being injected into the mass spectrometer, improving accuracy and range of peptide identifications. Advances in LC, such as high-pressure LC (HPLC) or the use of longer LC columns, has been attributed to improvements in the sampling rate of MS, though these technologies remain restricted to specialist laboratories (Shishkova, Hebert and Coon, 2016). At the end of the chromatography column, the sample is fed into the mass spectrometer.

The invention of the mass spectrometer in 1913 revolutionised many scientific fields, not least cell and molecular biology, through the ability to explore molecular structures and identify unknown molecular entities (Maher, Jjunju and Taylor, 2015). The fundamental concepts of mass spectrometers themselves have not changed drastically since the inception of MS, despite continued and remarkable development of the instrumentation: compounds in a sample are ionised and then separated by their mass-to-charge ratio (m/z) before being detected. The output is presented to the user as a mass spectrum, which plots the relative abundance (or intensity) by the m/z (Steen and Mann, 2004). The methods utilised for each of these steps differ based on the physical principles applied to them. For example, early mass spectrometers utilised thermal ionisation; in the biological sciences, electrospray ionisation is most-frequently employed though other methods are implemented in specialised mass spectrometers, such as matrix-assisted laser desorption/ionization (MALDI) (Wilm, 2011; Dreisewerd, 2014).

There are three main separation methods utilised by the mass spectrometers used in biological proteomics research: time of flight (TOF), quadrupole, quadrupole ion traps and Fourier-transform analysers (Haag, 2016). TOF separates ions of different m/z based on velocity, as measured by the passage of ions accelerating through an electric-field-free flight tube. Quadrupole mass spectrometers filter for ions of a specific m/z by applying a particular sinusoidal potential. Quadrupole ion traps "catch" ions in an electric field, before ejecting them sequentially onto the detectors based on their m/z using a second electric field; these are a type of linear ion trap. Another type of ion trap frequently found in mass spectrometers for the life sciences, which is based on Fourier transformation principles, is the Orbitrap, which circulates ions around a central electrode and converts the oscillation signal to *m/z* through Fourier transformation (Hu *et al.*, 2005).

Modern mass spectrometers perform two stages of separation and detection, in a method known as tandem MS (MS/MS). First, the peptides are separated based on m/z and detected, providing mass spectra known as MS^1 which describe the abundance (or intensity) of each peptide in the sample. Next, peptides are fragmented, for example by collision with inert gases (collision induced dissociation, CID) or chemically with introduced ions (electron-transfer dissociation, EID) (Guthals and Bandeira, 2012; Earley *et al.*, 2013). There are two methods that dictate which peptides will be fragmented. In data-dependent acquisition (DDA), the top-*N* most abundant peptides are filtered for and fragmented, whilst in data-independent acquisition (DIA), peptides within a given m/z window are selected and

fragmented. An example of DIA is SWATH (Sequential Windowed Acquisition of All Theoretical Fragment ions), in which the entire MS^1 mass range is split into m/z windows between 2 – 200 in size; the instrument measures all fragment ions within the m/z window (Gillet *et al.*, 2012; Meyer and Schilling, 2017; Bekker-Jensen *et al.*, 2020). The merits and limitations of DDA and DIA are discussed in more detail in section 1.4.1.2. The peptides are fragmented at the amine bonds, which are the weakest bonds within the peptide, yielding ions designated as b-ions or y-ions. The fragment ions conveniently contain one amino acid side-chain, or R-chain, each, allowing them to be identified. The fragments are then separated by m/z again, allowing a mass spectrum to be detected for each; these are the MS^2 spectra, which are essential for the reconstruction of the peptide sequence and subsequent protein identification. Tandem mass spectrometers such as the Q-Exactive HF, Orbitrap Fusion Lumos and Orbitrap Exploris 480 employ a linear quadrupole ion trap for MS^1 followed by Orbitrap for MS^2 (Scheltema *et al.*, 2014; Levy, Washburn and Florens, 2018; Bekker-Jensen *et al.*, 2020). The Impact II utilises a quadrupole and TOF to the same end (Beck *et al.*, 2015).

1.4.1.2 Processing raw mass-spectrometry data

There are several common software suites used to process raw MS data. Popular options include the freely available MaxQuant suite (Cox and Mann, 2008; Tyanova, Temu and Cox, 2016) and Proteome Discoverer™ from ThermoFisher Scientific. Both apply different statistical and computational methods for the inferences of raw MS data; some of the fundamental concepts will be discussed here.

A point of divergence in MS spectra interpretation is between the analysis of DDA and DIA spectra. In both cases, the interpretation of raw MS^1 and MS^2 spectra is a computationally complex task, and the methods to do so differ between DDA and DIA. DDA commonly identifies peptides from their MS^2 spectra using database matching. The databases are precomputed and contain putative peptide sequences generated from *in silico* digestion of proteins by trypsin (Eng *et al.*, 2011). The MS^2 spectra are matched against the sequences in these databases and given a score based on similarity. Each spectrum will likely generate multiple potential matches, called peptide spectrum matches (PSMs); sets of PSMs are characterised using a statistical test such as a false discovery rate (FDR). Individual PSMs, on the other hand, are statistically characterised by q-values or posterior error probabilities (PEPs) rather than FDRs. In the target-decoy strategy, decoys (most frequently the reverse of the peptide sequences) can be added to the database; from these, an FDR threshold can be calculated for each putative sequence in the database, to compare the PSM FDRs to (Elias

and Gygi, 2007). The Andromeda search engine, implemented in MaxQuant (Cox *et al.*, 2011) and MASCOT, one of the search engines implemented in Proteome Discoverer (Koenig *et al.*, 2008), both implement a target-decoy strategy. Correct fragment ion identification is particularly important in phosphoproteomics where modified sites are identified based on one peptide, unlike proteomics where protein identification is ideally based on the cumulation of multiple identified peptides (Lee, Jones and Hubbard, 2015). In either case, complications arise when identified peptides match multiple parent proteins.

One of the limitations of DDA is that only the most abundant peptides get fragmented and thus sequenced, which causes several problems. First, since the selection of abundant peptides is a semi-stochastic process, there may not be enough unique peptides from a protein identified. Divergences in peptide selection can be the culprit for irreproducibility between replicates, since different peptides may be selected in each MS run. Methods such as "Match between runs" implemented in MaxQuant have been proposed to get around this (Tyanova, Temu and Cox, 2016; Lim, Paulo and Gygi, 2019). Secondly, selecting only the most abundant ions leads to low-abundance peptides and proteins not being identified. Though bias towards high-abundance proteins is generally considered acceptable for discovery proteomics, it is a more considerable problem when studying stochastic processes or interactomes (Hu, Noble and Wolf-Yadlin, 2016).

DIA is a more recent method that aims to improve the issues encountered with DDA. As previously mentioned, DIA fragments all the ions within an m/z window, rather than just the most abundant. Since DIA repeatedly cycles over all the fragment ions within the window, sufficient information can be captured to identify all the peptides. However, this results in highly complex and mixed MS^2 spectra; alongside the time it takes to do DIA MS runs, this remains a key limitation to the technology, especially when applied to phosphoproteomics. To deconvolve DIA MS^2 spectra, spectral libraries rather than sequence databases are used to search against. Spectral libraries contain pseudo-spectra of each peptide ion based on the elution profiles of peptide-fragment ions; these are frequently built on DDA of the sample being investigated (Yang *et al.*, 2020). The reliance on DDA is time-consuming and makes DIA less unbiased than it could be. Since the inception of DIA, there has been focus in the field on spectral library-free, or peptide-centric, methods; recently, a method was proposed based on regression models in order to extend DIA methods to phosphoproteomics (Bekker-Jensen *et al.*, 2020; Martinez-Val, Dorte Breinholdt Bekker-Jensen, *et al.*, 2021). A recent publication also reported a spectral library generated *in silico* using deep learning (Yang *et*

al., 2020). DIA has been shown to have higher reproducibility and accuracy than DDA, by addressing concerns of selecting peptides, but this currently comes at the cost of time and computational complexity (Barkovits *et al.*, 2020; Fernández-Costa *et al.*, 2020)

As previously alluded to, the level of complexity when interpreting the MS^2 spectra increases when assigning phosphorylation site localisation. There are several reasons for this. First, the lability of the phosphoryl-group increases the ratio of unmodified to modified peptides, giving different MS^1 spectra for the same peptide. Second, though the target-decoy strategy is successfully implemented for modified proteins, as demonstrated by its use in the Andromeda search engine (Cox *et al.*, 2011), when there are multiply phosphorylated residues or putative residues close together in the same peptide, mislocalisation is more likely to occur (Fu and Qian, 2014). Finally, the low stoichiometry of phosphorylated peptides, as demonstrated in seminal work by Olsen et al. (2010) relies on a highly-sensitive MS set-up. To account for these issues, numerous algorithms have been proposed for the correct localisation of modifications on peptides. Several of these tools were recently benchmarked by Locard-Paulet et al. (2020), who investigated the performance of MaxQuant/Andromeda, Proteome Discoverer and PeptideShaker. They found that the algorithms implemented by each required different thresholds, however when this was accounted for high-confidence localisations were consistent across platforms. Moreover, the correct identification of the peptide was critical. This was highlighted by Kim, Zhong, and Pandey (2016) using peptides derived from the tyrosine kinase domains of Src-family nonreceptor tyrosine kinases; a tryptic peptide with a tyrosine localisation in the same position is shared by Fyn, Lck, Src and Yes. The authors suggest using another digestion enzyme, such as LysC, to overcome this issue, but recognise that one cannot predict this will work in all cases. This further demonstrates the non-trivial role of database search in the case of DDA-based phosphoproteomics. DIA approaches have been used with success in phosphoproteomics, even in unenriched samples, but are not routinely used (Bauer *et al.*, 2014; Bekker-Jensen *et al.*, 2020).

1.4.1.3 Quantitative (phospho)proteomics

As the resolution of mass spectrometers has improved, methods to improve the quantification of peptides and proteins have improved alongside. These can be classified, broadly into three categories: label-free quantification (LFQ), metabolic labelling and isobaric labelling. LFQ is the simplest and most economical of these approaches in terms of sample preparation but the most challenging computationally. It remains an open question in the field how best to translate LFQ-MS data to measures of true abundance without undergoing

rigorous adherence to uniform sample preparation. Though DIA methods may well provide a solution for this, several computational methods have been proposed to overcome this issue when using DDA. Broadly, the most common methods take one of two approaches: spectral counting relies on counting the number of spectra belonging to a protein across multiple LC-MS runs, whilst peak-intensity uses the area under the peptide ion or fragment ion spectral peaks as a read-out for abundance (Matzke *et al.*, 2013) A recent evaluation study found that methods based on peak-intensities were best-performing: the MaxLFQ algorithm, implemented in MaxQuant, performed most strongly in terms of accuracy and precision, however SEQUEST in Proteome Explorer had fewer missing values for low abundance proteins (Cox *et al.*, 2014; Zhao *et al.*, 2020).

Metabolic labelling involves the incorporation of chemical markers through the growth media of the samples. An example of this is stable isotope labelling by amino acids in cell culture (SILAC), in which cells are cultured with growth medium containing amino acid isotopes (e.g. arginine [R] and lysine [K]). These are referred to as “heavy” (H, e.g. K8, R10), “medium” (M, e.g. K4, R6) or “light” (L, e.g. K0, R0). The alternate amino acids get incorporated in the proteins of the cells in culture; these proteins will have a different molecular weight and thus be separated by MS. Extracted proteins from cell lysates from the three conditions (L, M, H) can be combined as one sample for MS, and as the peptides will be separated in the MS^1 spectra based on the incorporated isotopic amino acids, this serves as a point of relative quantification between experimental conditions (Ong *et al.*, 2002; Ong and Mann, 2007). One of the limitations of SILAC is that is difficult to use for model organisms, though has been incorporated in mouse (Zanivan, Krueger and Mann, 2012), and it cannot be readily used for clinical or *ex vivo* samples since the metabolic labels need incorporating during growth. One solution proposed is super-SILAC or spike-in SILAC, in which SILAC-labelled cell-lines representing the cell types in the clinical sample are incorporated, or spiked-in, prior to analysis by MS, providing internal quantification (Geiger *et al.*, 2010, 2011). This has been successfully used to analyse ex vivo microglia (Pinho *et al.*, 2017) and the role of matrix stiffness in metastasis (Reid *et al.*, 2017).

Another consideration when designing SILAC experiments is the limited multiplexing, with only three conditions able to be compared to one another. Isobaric labelling provides an alternative labelling method that overcomes this limitation. The two main methods are called tandem mass tagging (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ) (Thompson *et al.*, 2003; Ross *et al.*, 2004). Both are based on similar principles. Once

protein samples have been digested into peptides, each experimental sample is treated with an isobaric label and then combined for a single MS run. For TMT, MS runs can be 6-plex or 10-plex (Werner *et al.*, 2014); recently, ThermoFisher Scientific released tags to perform 16-plex TMT. The tags are isobaric, meaning they have identical chemical properties and appear as one peak in MS^1. However, upon fragmentation, the tags dissociate and appear as distinct masses. Thus, the tag can be used to infer the relative abundance of peptide in each experimental sample (Rauniyar and Yates, 2014). Quantification based on the MS^2 spectra is one of the key ways in which isobaric labelling differs from LFQ and metabolic labelling methods, which are sometimes referred to as full-scan based, since the quantification is based on MS^1 spectra (Li *et al.*, 2012; Cox *et al.*, 2014). Since the isolation window for selecting fragment ions will sometimes contain more than one isobaric tag, quantification can be inaccurate – this is referred to as ratio compression and is a standing problem with isobaric analysis. One solution, which can be performed with hybrid mass spectrometers like the Q-Exactive HF, is to perform a further round of fragmentation and separation to generate MS^3 spectra. The most abundant second-generation fragment ions are then selected and quantification is calculated based on these; this is thought to improve quantitative precision (McAlister *et al.*, 2014). Taken together, isobaric labelling is generally considered to improve the sensitivity of the quantification and reduce the number of missing values (Rauniyar and Yates, 2014). However, a recent study investigating the precision and scalability of 10-plex TMT found that although these benefits were clear for single batches, integrating results with other TMT batches increased the proportion of missing values from <2% to 24% (Brenes *et al.*, 2019). There was also poor reproducibility between batches of the same type, with a particularly high incidence of false positives. Similar issues have been previously reported for iTRAQ tagging (Karp *et al.*, 2010). These issues can be counteracted with strict quality control and well-considered experimental design.

A benchmarking study comparing three quantification techniques and their suitability for large-scale phosphoproteomics yielded useful insights on the pros and cons of each of the methods discussed above. They found that MS^2-based TMT performed best in terms of reproducible identification, however LFQ and SILAC were superior in terms of accurate quantification; MS^3-based TMT yielded the least precise results.(Hogrebe *et al.*, 2018) These results are consistent with those discussed above from Brenes et al. (2019), who also found poor performance with MS^3 spectra on proteomics (rather than phosphoproteomics) data, but did not investigate using just MS^2. Despite the strong performance of MS^2-based TMT, the main obstacle to its wide use is the expense. LFQ remains the most accessible option for

experimentalists, as the sample prep does not require the addition of expensive labels. SILAC, though cheaper than isobaric labelling, requires more time than either LFQ or isobaric labelling, due to the need to incorporate the isotopic amino acids into the cells in culture; in a previously published protocol, the authors suggest that HeLa cells will fully incorporate SILAC amino acids after approximately five *passages* (Ong and Mann, 2007).

1.4.2 **Experimental strategies for analysing cell signalling with (phospho)proteomics**

1.4.2.1 Time-resolved phosphoproteomics

Signalling processes have tight temporal regulation. From the moment they are initiated, changes can be detected in the phosphoproteome of the cell; these modifications happen at a much faster rate than changes in protein levels (through translation or degradation). MS-based phosphoproteomics therefore provides a powerful tool to take systems-level snapshots of different periods of signalling processes. A caveat remains, in that the requirements for MS sample preparation, and the expertise in instrumentation and data analysis, is still somewhat of a barrier to performing high dimensional studies such as comprehensive time courses. However, as the technology itself (particularly in terms of resolution and quantitation) has improved, many studies untangling the temporal regulation of various signalling processes have been published.

MS was first applied for the investigation of phosphorylation-dependent growth factor signalling dynamics in the early 2000s (Blagoev *et al.*, 2003; Kratchmarova *et al.*, 2005; Zhang *et al.*, 2005). These studies utilised phosphorylated peptide enrichment with antibodies and were seminal in resolving temporal phosphorylation events following growth factor stimulation, which previously could only be studied with one or two proteins at a time. Through a combination of quantitation (with SILAC) and TiO_2 enrichment, changes in the phosphoproteome following stimulation with 150 ng/ml EGF for 1, 5, 10 and 20 minutes were more fully explored (Olsen *et al.*, 2006). This was the deepest study of its kind at the time, with 6600 phosphorylated sites identified, and was important in developing the methodology to analyse large-scale dynamic phosphoproteomes quantitatively.

Several studies have taken and built upon the quantitative phosphoproteomics approach to examine temporal regulation of RTK signalling cascades more deeply. The functional selectivity of FGFR2b to FGF7 and FGF10 was investigated using a similar SILAC approach as above, with samples collected 1, 8 and 40 minutes following stimulation (Francavilla *et al.*, 2013). The authors identified phosphorylation of FGFR2b at Y734 as a molecular switch from

FGF7-induced receptor degradation to FGF10-induced recycling, among other phosphorylation-dependent mechanisms regulating the functional selectivity of FGFR2b. The 1 minute time point included in the study highlights that very early changes can be investigated using these powerful methods. The temporal resolution available through quantitative phosphoproteomics was also demonstrated by a study investigating signalling downstream of insulin growth factor 1 receptor (IGF1R), FGFR1, and platelet derived growth factor receptor (PDGFR), 3 and 15 minutes following stimulation with their respective ligands: IGF1, FGF2 and PDGF (Batth *et al.*, 2018). This revealed differential phosphorylation throughout canonical signalling pathways; the authors particularly highlighted the negative regulation of PDGFR-induced tyrosine phosphorylation by the phosphatase Shp-2. The methodology used in these examples was further developed by the same group, who integrated the analysis of the phosphoproteome, ubiquitinome, receptor interactome and late proteome (to examine long-term proteome changes) of the EGFR system when stimulated with either EGF or TGF-α (Francavilla *et al.*, 2016).

As algorithms for LFQ have improved, their use in phosphoproteomics has increased. For example, LFQ was used to characterise ERK1/2 effectors within the phosphoproteome of serum-stimulated IEC-6 cells (Courcelles *et al.*, 2013). 2296 phosphorylated sites containing the ERK1/2 consensus motifs, which could be filtered to 232 based on the dynamics in the time course (I.e. Increasing upon serum stimulation). LFQ was also used to perform the most comprehensive characterisation to date of regulated phosphorylated sites following EGF stimulation in HeLa, with more than 38,000 phosphorylated sites identified with high confidence (Sharma *et al.*, 2014).

1.4.2.2 Mapping interactomes with MS-based proteomics

Proteins are nodes in a large interaction network. In a standard shotgun proteomics or phosphoproteomics experiment, reconstructing protein-protein interaction (PPI) networks can be difficult for two key reasons. First, proteins may not have been identified in the experiment, but that is not reason enough to assume they are not there, due to the limitations in MS-based profiling discussed above. Second, the identification of a protein does not indicate its interaction with other proteins, since the dataset will represent a global snapshot of the sample; this is particularly relevant when considering phosphoproteomics, as a protein in a particular modification state may not be able to interact with other proteins, or vice versa. Considering these concerns, methods to specifically probe interactions between proteins have been developed. These are affinity proteomics (AP-MS), proximity-

dependent labelling, and organellar mapping. They are broadly considered as different styles of spatial proteomics (Lundberg and Borner, 2019).

AP-MS typically involves enrichment of a bait protein (e.g. a protein of interest that one wishes to find the interactors of) via immunoprecipitation. An impressive example comes from the three iterations of the large-scale BioPlex interaction networks (Huttlin *et al.*, 2015, 2017, 2021). In these studies, HA-flag tagged proteins are pulled down by IP and their interactors identified with MS. In the most recent version, BioPlex 3.0, 10,128 HA-flag tagged proteins were purified from 293T cells, defining 118,162 interactions between 14,586 proteins (over half of the proteome). On a smaller scale, AP-MS was used to identify the interactors of 54 human phosphatases; 838 interactors were identified, and further investigation associated many with colorectal cancer (Yadav *et al.*, 2017). It is worth noting that, while a powerful and accessible method, AP-MS may not be well suited to identifying low-affinity interactions or interactions between, or with, membrane proteins, due to the biochemistry of the IP process (Dunham, Mullin and Gingras, 2012).

Proximity-dependent labelling is an alternative method of AP-MS that does not require protein-interactions to be maintained during the sample purification process. The most common strategies of proximity-dependent labelling involve covalent biotinylation, which is introduced by proteins (called “baits”) tagged with an enzyme capable of catalysing the reaction; proteins (called “preys”) will be biotinylated when proximal to the bait. The preys can then be purified with streptavidin beads. The two common strategies use different types of enzymes to perform the biotinylation. BioID utilises biotin protein ligases, commonly BirA (Roux *et al.*, 2012), BioID2 (Kim *et al.*, 2016) and TurboID (Branon *et al.*, 2018). These can activate biotin, generating a “cloud” of reactive biotinoyl-AMP intermediate which is directly transferred to lysine residues on prey proteins within a 10 nm radius (Kim *et al.*, 2014; Samavarchi-Tehrani, Samson and Gingras, 2020). Labelling typically takes 16−18 hours, making it best suited for experiments investigating static or gradually changing components (Liu *et al.*, 2020). One of the first applications of BioID was to characterise the components and interactors of the nuclear pore complex (Kim *et al.*, 2014). It has since been used in a variety of scenarios, for example mapping the composition of focal adhesions (Dong *et al.*, 2016), integrin interaction partners (Myllymäki *et al.*, 2021) and to characterise the HEK293 interactome in a similar manner to BioPlex (Go *et al.*, 2021) . Biotin peroxidases, such as ascorbate peroxidase (APEX) and the modified APEX2, are alternatives to BioID that provide more rapid labelling (1 minute) and larger labelling radius than BioID, of approximately 20

nm (Gingras, Abe and Raught, 2019). To put the labelling ranges into context, the diameter of a HeLa cell is in the region of 17 µm (Borle, 1969). APEX oxidises biotin-phenol to biotin-phenoxyl radicals, which then covalently bonds with tyrosine residues, predominantly (Hung *et al.*, 2014). The first use of APEX based proximity-dependent biotinylation was to characterise the mitochondria (Rhee *et al.*, 2013; Hung *et al.*, 2014). Due to its short labelling time, it has since been used to understand dynamic processes such as cellular signalling and intracellular trafficking. One study used APEX2 to characterise Rab interactomes; the early endosome RAB4a, RAB5a and RAB21 were shown to have many interactors in common but strong association with unique proteins in each case provided further evidence for the endosomal microdomain theory (Del Olmo *et al.*, 2019). APEX2 has also been used to identify FGF1 proteoglycan co-receptors at the cell surface (Zhen *et al.*, 2018). An interesting application of the APEX methodology was in studying the spatiotemporal regulation of GPCR upon activation (Lobingier *et al.*, 2017). By comparing APEX- GPCRs and APEX-tagged subcellular markers, such as FYVE for the endosomes, the researchers were able to capture time-resolved interactomes at different subcellular locations following receptor activation.

Since the labelling is performed in live cells, samples that have been labelled with either BioID or APEX2 can be (and many argue should be) coupled with the metabolic or chemical labelling approaches discussed in section 1.4.1.3 (Hung *et al.*, 2016). Sample preparation for MS following APEX and BioID is similar in both cases and can be performed in one of two ways; either biotinylated proteins are enriched prior to digestion using streptavidin beads, or enrichment is performed following digestion into peptides with biotin-antibody conjugated beads (Samavarchi-Tehrani, Samson and Gingras, 2020). The former approach, though likely to yield generally higher identification, will be less efficient at eluting highly-biotinylated proteins, whilst the latter approach will elute highly-biotinylated proteins and can also be used to identify sites of biotinylation. Another consideration in the experimental design is ways in which artefacts can be mitigated against; due to the nature of the labelling, "passers-by" are likely to be labelled alongside true interactors. Trinkle-Mulcahy (2019) suggest looking for mitochondrial proteins (many of which bind biotin directly) in the dataset, well-considered placement of the proximity tag, and appropriate expression levels of the bait protein to mitigate against noise. It can be noted that there are currently no examples in the literature for proximity-dependent labelling being used to analyse proximal phosphorylated proteins, through enrichment of biotin-labelled and phosphorylated peptides; advancements in this area would need careful experimental design to account for the existing limitations of these approaches.

1.4.2.3 Breadth & Depth of phosphoproteomics

The power of phosphoproteomics is best demonstrated by the extent to which samples can be profiled accurately. Several pivotal studies have demonstrated what is possible with MS-based phosphoproteomics. Considering progress within the last few decades, the highest number of high-confidence phosphorylated sites identified in a single experiment has expanded from 6600 (Olsen *et al.*, 2006), to > 20,000 (Olsen *et al.*, 2010) to the most recent high of > 38,000 (Sharma *et al.*, 2014). Based on this trajectory the total number of phosphorylated sites is predicted to be 100,000 (encompassing 90% of the proteome), indicating that sensitivity of the methodology and instrumentation has some way to go (Zubarev, 2013; Sharma *et al.*, 2014).

A recent perspective from Timp & Timp (2020) argued that without high-resolution mass spectrometers and dedicated resources for state-of-the art sample preparation and data capture, the breadth and depth of identification demonstrated in these studies remains limited to a few laboratories. Sidoli, Kulej, & Garcia (2017) also argued this point, suggesting that while competitive technology remains inaccessible, researchers will continue to select accessible genomics and transcriptomics technologies for their research. This is problematic, especially for cell signalling research, given that genome and transcriptome-level changes do not translate to the proteome or phosphoproteome level (Gygi *et al.*, 1999; Lundberg *et al.*, 2010; Yaffe, 2013, 2019). While profiling at the proteome and phosphoproteome level remains incomplete, good practice considers that analysis of these datasets should be performed with statistical rigour followed by experimental validation (Oberg and Mahoney, 2012; Handler *et al.*, 2018). The latter is the basis of the functional proteomics approach, first discussed in the early 2000s (Godovac-Zimmermann and Brown, 2001).

Functional phosphoproteomics describes a pipeline of discovery phosphoproteomics being used as a basis for hypotheses generation, followed by bioinformatics analyses with wet laboratory investigations. These investigations aim to prove the role of a phosphorylated site or protein, or a set of phosphorylated sites or proteins. Common strategies include mutagenesis, in which the residue which has the putative functional phosphorylation is mutated to a 'neutral' amino acid (i.e. serine/threonine to alanine, or tyrosine to phenylalanine). Alternatively, the residue can be mutated to a phosphomimetic amino acid, such as aspartic acid or glutamic acid, which may prevent errors in protein folding and generate a constitutively "phosphorylated" proteoform. Mutagenesis can be performed through a CRISPR approach, via "knock-in" of a point mutation (Bak, Gomez-Ospina and Porteus, 2018). Alternatively, expression of the endogenous proteins could be supressed

with a “knock-out” of the gene (Joung *et al.*, 2017); the mutated form could then be re-expressed through the transfection of an exogeneous genetic vector. This would be beneficial if a range of mutated proteoforms needed to be investigated. Endogenous protein expression can also be “knocked-down” through small interfering RNA (siRNA) and the mutated form re-expressed as described above. Rather than directly mutating or altering the phosphorylated site or protein of interest, genetic approaches could also be used to knock-out, knock-down or overexpress the upstream kinase. This can be predicted from bioinformatics tools such as NetworKIN or NetPhorest if experimental data describing the kinase-substrate relationship is unavailable (Horn *et al.*, 2014). An alternative method to prevent phosphorylation of a site of interest would be to use pharmacological kinase inhibitors, which are widely available for many commonly studied kinases.

The success of strategies aiming to disrupt phosphorylation are commonly investigated using immunoblot analysis if antibodies are available for the phosphorylated sites or proteins of interest. Phospho-specific antibodies are significantly more common for phosphotyrosine residues, given their higher immunogenicity (Roux and Thibault, 2013). Immunoblotting can then be used to assess how disrupting the phosphorylation impacts on other phosphorylation events, for example those known to be downstream of the protein of interest. The broader consequences on cellular behaviour can be investigated using functional assays. For example, cell proliferation can be measured by incorporation of the nucleotide analogues BrdU or EdU, or simply by counting cells before and after perturbation.

1.5 Bioinformatics analysis of phosphoproteomics datasets

1.5.1 Common approaches in phosphoproteomics data analysis

1.5.1.1 Preliminary statistical assessments

The statistical assessment of processed proteomics or phosphoproteomics data is a non-trivial task. Background noise, few or poorly correlated replicates (due to the cost and time of running large-scale proteomic screens) and missing values are among the special considerations that must be accounted for during analysis. Here I will discuss some of the statistical approaches for overcoming these issues and identifying differentially regulated proteins and sites.

1.5.1.1.1 Differentially abundant or regulated proteins and phosphorylated sites

To determine if a protein or phosphorylated site of interest is differentially regulated, the researcher must first define what an interesting difference is. In the case of proteins, is it the colocalization with a bait protein, or an increase in abundance? For phosphorylated sites, is

it a change from the baseline, or the pattern of dynamics over time? Many statistical methods have been suggested or applied to proteomics data which suit these different research questions.

Common statistical hypothesis tests of differences between groups, such as Student's t-test and analysis of variance (ANOVA), have been widely applied in proteomics. The former was expanded by Tusher, Tibshirani, and Chu (2001) in a method called Significance Analysis of Microarrays (SAM), which, as the name suggests, was initially for application to microarrays but has been applied to proteomics and phosphoproteomics, most notably by its inclusion in the Perseus analysis suite (Tyanova *et al.*, 2016). SAM differs to a standard t-test through the permutation of repeated measurements to estimate the FDR. One issue with the application of t-tests or ANOVA is the reliance on assumption or assessment by the researcher that the data are normally distributed (Ellison, Barwick and Farrant, 2009). Further to this is the impact of missing values, as discussed in section 1.5.1.1.3, which reduce the number of degrees of freedom and the accuracy of the standard error, reducing the power of the statistical tests (Ting *et al.*, 2009). However, this has been examined in the context of ANOVA applied to SILAC proteomics data, where it was found that even if the data violated the assumption of normality there was high correlation between the results of ANOVA and the equivalent non-parametric test, Kruskal-Wallis (Albaum *et al.*, 2011). It should be noted that where quantitative data are represented as ratios, as for metabolic quantitative methods such as SILAC, an arbitrary cut-off is often applied to determine differential regulation of sites. It has been previously suggested that hypothesis tests are not suitable for use on ratios, particularly those derived from a small number of replicates (I.e., duplicate or triplicate) (Cho *et al.*, 2007). However, no assessment of the appropriate methods for determining a cut-off currently exists in the literature.

An alternative to statistical hypothesis tests is methods based on linear models. Linear models for microarray data (Limma), which estimates differential expression based on gene-wise regression models, is a popular tool for performing these analyses (Smyth, 2004; Ritchie *et al.*, 2015). Though designed for analysing RNA microarray and RNA-seq data, it has been applied to proteomics data with success (Ting *et al.*, 2009; Schwämmle, León and Jensen, 2013; Kammers *et al.*, 2015). However, a mixed linear model (MLM) method was shown to outperform Limma when applied to TMT data and when certain normalisation methods are used (D'Angelo *et al.*, 2017). For proteomics analysis, an extension of the popular DESeq2 method for transcriptomics has been developed, named DEqMS (Differential Expression

analysis of quantitative Mass Spectrometry data). Like DESeq2, DEqMS accounts for the number of peptides or PSMs used to calculate protein abundance more accurately between experimental conditions (Love, Huber and Anders, 2014; Zhu *et al.*, 2020). However, this cannot be further extended to phosphoproteomics, which inherently relies on quantification of a single (phosphorylated) peptide.

An interesting aspect in the quantification of phosphorylated peptides is understanding the proportion of the protein population that has been phosphorylated at a given site; this is called the site stoichiometry or occupancy. A method to calculate occupancy using SILAC data of phosphorylated and corresponding non-phosphorylated peptides has been reported, however is held back by two main caveats: the assumption that both the phosphorylated and non-phosphorylated peptide have been correctly profiled, and that the phosphorylated peptide exists in a singly-phosphorylated state (Olsen *et al.*, 2010). This highlights another requirement in uncovering the differential regulation of phosphorylated sites: that changes in protein abundance are not falsely accentuating changes at phosphorylated sites.

An extra consideration needs to be taken with protein lists derived from AP-MS or proximity-dependent biotinylation based experiments, to address how likely the proteins and sites identified are true interactors of the bait protein. There have been several algorithms proposed to address this question. Significance analysis of interactome (SAINT) calculates the probability of true protein-protein interaction by comparing to negative controls using Bayesian statistics (Choi *et al.*, 2011). Further implementations of SAINT have developed on the original method: SAINTexpress is based on similar principles but is less computationally demanding (Teo *et al.*, 2014) and SAINTq has been optimised for DIA-MS data (Teo *et al.*, 2016). The Mass spectrometry interaction STatistics (MiST) method scores bait-prey interaction likelihood on the abundance of the prey protein, the uniqueness of the interaction when compared with other baits and the reproducibility across replicates. Comparative Proteomic Analysis Software Suite (CompPASS) scores interactions in a similar way to MiST but also takes into account previously validated PPIs (Sowa *et al.*, 2009).

1.5.1.1.2 Identifying patterns with dimensionality reduction & clustering

Given the huge number of phosphorylated sites that may be identified as differentially regulated in a single experiment, it is crucial to be able to determine those that are co-regulated. There are two main unsupervised pattern recognition methods used in phosphoproteomics: dimensionality reduction with principal component analysis (PCA) and clustering.

Dimensionality reduction is the solution proposed to distinguish the many potential features that will be found in a dataset containing thousands of data points. PCA does this by transforming the quantitative data from each sample or experimental condition so the more variant elements can be extracted; these are then used to find "principal components" that each account for a decreasing amount of variance in the data (David and Jacobs, 2014). The greater the size of each principal component, the more variation there is in the data (Meyer-Baese *et al.*, 2014). The principal components can be plotted to determine how related or divergent different samples, replicates or experimental conditions are to one another (Alter, Brown and Botstein, 2000).

If dimensionality reduction is used to see how the samples or experimental conditions are related to each other, clustering typically accounts for relations between the proteins or phosphorylated sites within them. Patterns of regulation are useful to identify functionally related players in the system being examined. Popular clustering methods used for phosphoproteomics include K-means (Hartigan and Wong, 1979), self-organising maps (SOM) (Kiang, 2001), Fuzzy C-means (Bezdek, 1973; Futschik and Carlisle, 2005; Schwämmle and Jensen, 2010) and hierarchical clustering (Gordon, 1987). Benchmarking studies have concluded that fuzzy C-means is a superior option for highly overlapping data, with hierarchical methods typically outperformed by the other methods (Mingoti and Lima, 2006; Albaum *et al.*, 2011). A phosphoproteomics specific method called CLUE has been developed; biologically meaningful clusters are generated by integrating k-means clustering and kinase-substrate annotations extracted from PhosphoSitePlus (P. Yang *et al.*, 2015).

1.5.1.1.3 Dealing with missing values

Missing values are an important consideration in proteomics and phosphoproteomics due to their frequency. The cause of missing values is generally divided into two main categories: missing completely at random (MCAR) or missing not at random (MNAR) (Rubin, 1976). MCAR generally accounts for technical causes of missingness, such as the stochastic nature of selection in DDA MS or a misidentified peptide. missing values from MCAR can be expected through the entire dataset in a uniform distribution (Karpievitch *et al.*, 2009; Lazar *et al.*, 2016). Values which are MNAR are related to the measurability of the protein or site or abundance below the mass spectrometer limit of detection. MNAR is thought to account for the majority of missing values in proteomics, and typically is considered to have a left skew (Karpievitch *et al.*, 2009).

Missing values can be replaced in the data with valid values using data imputation methods. Imputation of data is somewhat controversial. Some argue that since many of the missing values are MNAR, imputation makes an assumption of the abundance of the protein or phosphorylated peptide, and applied across the whole dataset this may generate misleading results (Schwämmle, Verano-Braga and Roepstorff, 2015). On the other hand, statistical approaches such as PCA and several clustering methods (e.g. K-Means) do not handle missing data, and therefore the data needs to be imputed or filtered to include only complete cases (that is, peptides or proteins with quantitative values for all samples). The latter in itself introduces bias to the dataset (Bartlett, Harel and Carpenter, 2015). Researchers may therefore consider a compromise between the two approaches, whereby a threshold is set for the number of acceptable missing values for each peptide or protein. The compromise is harder to reach when there are a high number of missing values; it has been shown that the higher the proportion of missing values, the poorer imputation methods tend to perform (Lazar *et al.*, 2016).

Missing values which are MNAR are best dealt with using imputation methods suited for left-censored data (Webb-Robertson *et al.*, 2015). These include imputing missing values with a fixed value, such as a value near the limit of detection or the lowest value in the dataset, or imputing from random draws of a truncated normal distribution (Tyanova *et al.*, 2016). The former has been shown to perform poorly across several benchmarking studies, however the latter performs well on datasets that are known to have the majority of values MNAR (Webb-Robertson *et al.*, 2015; Lazar *et al.*, 2016). However, the proportions of MNAR and MCAR are difficult to discern in a standard (phospho)proteomics dataset, and imputing from random draws of a truncated normal distribution performs poorly when imputed missing values are MCAR (Lazar *et al.*, 2016). Therefore, hybrid approaches have been proposed, such as k-Nearest Neighbours (kNN), maximum likelihood estimation (MLE) and random forest (RF). These methods generally perform well in benchmark studies, which cover LF and isobaric labelling-based datasets but are hindered by the computational time required, often in the order of hours (Webb-Robertson *et al.*, 2015; Kokla *et al.*, 2019; Bramer *et al.*, 2021; Jin *et al.*, 2021).

1.5.1.2 Functional analysis

Once sets of differentially regulated phosphorylated sites have been identified with statistical analyses, functional analyses can help reveal their roles. The more common methods are based on protein-level (rather than phosphorylated site-level or modified protein-level) functional annotations and interactions. The merits and disadvantages of

protein-level analyses for analysing phosphoproteomics data are discussed in section 1.5.2. However, protein-level functional analysis provides a broad overview of the potential role of the modified protein. The function of a protein can be considered in two ways: the function is the contribution of the protein to the biological program or system-level processes (e.g. cell cycle, metabolism); or it is the specific events that it coordinates (e.g. binding a protein, catalysing a reaction), which may contribute to a larger process or may be biological "noise" (Thomas, 2017). When performing functional analyses, the first step primarily considers the former definition of function. This is partly out of practicality; given the number of differentially regulated or expressed proteins that will be found in a proteomics experiment, the researcher analysing the data cannot do an individual literature search to understand the specific roles coordinated by each.

Analyses based on standard functional descriptors of proteins, such as those defined by the Gene Ontology (GO) are critical to building understanding based on the broad definition of function. GO provides a restricted vocabulary to describe the biological processes (BP), molecular functions (MF) and cellular compartments (CC, that proteins are located in) that proteins are related to (The Gene Ontology Consortium, 2018). GO is the largest and most-widely used resource of its kind (The Gene Ontology Consortium, 2018). However, there are several limitations to the GO resource: mis-annotations in GO (e.g. from automated curation of annotations inferring orthologous protein function across taxa) are relatively common. There is also a tendency towards annotating well-studied genes (the so-called rich-getting-richer effect) (Haynes, Tomczak and Khatri, 2018). Lastly, annotations frequently change between versions, ultimately impacting the interpretation of protein function in 'omics datasets (Tomczak *et al.*, 2018; Wei *et al.*, 2020). Though these problems are not unique to GO, corroboration with other databases containing protein-level annotations is prudent. For example, there are several databases containing pathway annotations, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY (Kanehisa and Goto, 2000) and Reactome (Fabregat *et al.*, 2018). In these databases, proteins are annotated to pathways related to (but not limited to) signalling, metabolism, and disease. An option that encompasses GO categories and pathways as well as co-expression profiles, genome location clusters and sequence similarity is the Molecular Signatures Database (MSigDB) (Liberzon *et al.*, 2011). With any of these sources, it is important to note the difficulty in defining molecular function, as many molecules participate in numerous context-dependent biological processes and pathways (Stoney *et al.*, 2015).

To statistically assess whether there is a substantial number of proteins involved in the same processes, enrichment tests are commonly used. These are typically based on a hypergeometric or Fisher Exact test, broadly called overrepresentation analysis (ORA). These statistically test the null hypothesis that the number of proteins annotated to a particular process in the set being analysed is proportional to the number annotated to all the proteins in the system being studied. Rejecting the null hypothesis would suggest that there are a higher proportion of proteins in the set annotated to a process than in "background" protein population. Web tools such as EnrichR (Kuleshov *et al.*, 2016) and WebGestalt (Liao *et al.*, 2019) can be used for performing ORA. However, several benchmarking studies have considered ORA a poor option for performing this type of analysis, highlighting that the low computational time and ease of use, rather than it's suitability, is the reason for its widespread use (Nguyen and Holmes, 2019; Maleki *et al.*, 2020). A recent study benchmarking enrichment tests on transcriptomics data suggested otherwise, highlighting that ORA was the best option where differential expression of sample or background proteins/transcripts is not factored into the null hypothesis (Geistlinger *et al.*, 2021). An alternative to ORA are the Functional Class Scoring methods, such as the popular gene set enrichment analysis (GSEA), which allow the quantitative values of the differentially expressed molecule to be regarded (Subramanian *et al.*, 2005; Maleki *et al.*, 2020). GSEA is based on a weighted Kolmogorov–Smirnov-like statistic, and is considered a powerful alternative to ORA, though the statistical methods have also been questioned (Damian and Gorfine, 2004; Mathur *et al.*, 2018). Methods accounting for quantitative values are useful for proteomics, but less so where phosphoproteomics quantitative values would have to be collapsed or summarised to the protein-level, as the annotations describe proteins and not phosphorylated sites or proteins in a particular modification state.

Enrichment analyses address the aspect of protein function from the perspective of whole processes or systems. Considering the second aspect of protein function defined above, a method to identify the specific events coordinated by proteins is by understanding how much interaction there is within the protein list. Graph databases that store known or predicted PPIs such as STRING, BioGRID and INTact can be used to construct networks of protein sets (Orchard *et al.*, 2014; Chatr-aryamontri *et al.*, 2017; Szklarczyk *et al.*, 2019). Module detection can then be used to find functionally related proteins within the network. Functional modules contain nodes that are highly interconnected and are useful in identifying components of biological processes (Hartwell *et al.*, 1999). The multifunctionality of many proteins is demonstrated by the overlap of modules, particularly those containing

hubs (Csermely *et al.*, 2012; Stoney *et al.*, 2015). There are many methods of detection of functional modules, reviewed extensively by Fortunato (2010) and Ji et al. (2014).

The use of PPI networks is somewhat hampered by three key aspects. First, the graph databases they are based on are unavoidably incomplete, therefore true interactions or associations may not be adequately highlighted (Kovács *et al.*, 2019). Secondly, study bias is also an issue in PPI-based network analyses. Study bias may derive from the biochemical methods used in the acquisition of PPI data, which extract or highlight proteins with certain properties more than others, and the tendencies of researchers to study previously studied proteins (Gillis, Ballouz and Pavlidis, 2014). Finally, the "guilt-by-association" principle that is assumed in the detection of network modules does not always hold; the functions of proteins are not encoded in the network nor network modules, and the incomplete information and study bias may feed into this (Gillis and Pavlidis, 2012).

1.5.2 Protein-centric vs. phosphorylation site-centric analysis for phosphoproteomics

1.5.2.1 Is protein-centric analysis suitable for phosphoproteomics?

Though protein-centric methods are widely used for analysing phosphoproteomics data, they require compromise. Primarily this is in the handling of the quantitative data associated to each phosphorylated site. The researcher must decide to disregard the quantitative values, or the quantitative values get collapsed or summarised to the protein level. The latter is only an issue if multiple phosphorylated sites are detected from the same protein. Alternatively, values can be encoded using clustering, as phosphorylated sites with similar patterns of regulation will be grouped and phosphorylated proteins in each cluster analysed. The importance of interrogating data at the level of the phosphorylated site was emphasised by Ochoa et al. (2019), who used machine learning approaches on phosphoproteomics datasets from the PRIDE repository to identify 59 features that indicate the myriad functional roles individual phosphorylated sites can be involved in.

Several repositories contain annotations of phosphorylated sites to known functions – for example, PhosphositePlus (Hornbeck *et al.*, 2015), dbPTM (Huang *et al.*, 2019), and the PTMsigDB repository (Krug *et al.*, 2019). However, their use in enrichment analyses like those using protein-level annotations is limited. Very few phosphorylated sites are annotated in these repositories; 4271 of the more than 230,000 human phosphorylated sites in PhosphositePlus and 8029 phosphorylation sites in PTMsigDB are annotated (Hornbeck *et al.*, 2015; Krug *et al.*, 2019). It is unknown how many phosphorylated sites have an impact

on function, however evolutionary studies have predicted some 45% may be (Landry, Levy and Michnick, 2009; Gray and Kumar, 2011). These repositories may also be focussed on a limited number of well-studied phosphorylated sites.

1.5.2.2 Proposed site-centric methods

To complement, or as alternatives to, the protein-centric methods, phosphorylated site-centric methods can be used.

Enrichment for particular amino-acid sequence motifs in the region of phosphorylated sites can be measured using tools such as motif-x (Chou and Schwartz, 2011). In the case of phosphoproteomics, this is useful for capturing kinase-substrate relationships, as the motifs around phosphorylated sites can be mapped to the target sequences of kinases. A number of kinase-substrate specific motif enrichment tools have been developed, including Kinase Substrate Enrichment Analysis (KSEA) (Casado *et al.*, 2013; Wiredja, Koyutürk and Chance, 2017) and Kinase Enrichment Analysis (KEA) (Lachmann and Ma'ayan, 2009). These methods rely on previously known kinase motifs, however other methods infer these relationships through supervised approaches, such as KSR-LIVE (Domanova *et al.*, 2016) and IKAP (Mischnik *et al.*, 2016). The Phosphomatics web platform provides a user-friendly interface for performing these types of analyses (Leeming *et al.*, 2020). Kinase-substrate analysis has also been used to enhance the clustering of temporal phosphoproteomics data, by using the enrichment of kinase signatures found in PhosphositePlus to justify data partitions (P. Yang *et al.*, 2015).

K-S analysis methods are useful for describing the cellular context, especially when used for network inference (as discussed in section 1.5.2.3), however they do not capture the functional role of the phosphorylated sites, nor the participation of modified proteins in pathways and cellular processes (Mantini *et al.*, 2021). To this end, the GSEA method was modified into PTM Signature Enrichment Analysis (PTM-SEA) (Krug *et al.*, 2019). The PTM-SEA method integrates a developed version of GSEA, named single sample GSEA (ssGSEA), and the PTMsigDB database, which contains a curated set of site-specific signatures for pathways, diseases, and processes. The authors highlight that, whilst the PTMsigDB is restricted to a small number of processes and functionally relevant sites, the use of PTM-SEA is limited to datasets where a clear phenotype is expected and, therefore, can be used more as a validation than a discovery tool.

1.5.2.3 Construction of phosphorylation-centric networks

As discussed in section 1.5.1.2, networks have long been used to represent protein interactions; PPI networks, and most large networks, are referred to as scale-free networks, meaning they do not follow a Poisson degree (the number of edges associated to a node) distribution, as would be expected from a random network (Vidal, Cusick and Barabási, 2011). This implies that most nodes show low connectivity, whilst there are few highly connected 'hub' nodes. There is a higher probability that new nodes will connect to these hubs, and the hubs often represent essential components of the network (Barabási and Albert, 1999; Jeong *et al.*, 2001). It is for this reason that hubs identified in PPI are popular choices as drug targets (Csermely *et al.*, 2012). However, representing and visualising phosphorylated sites as a network is not as simple as a PPI network. In this section, three methods of phosphorylated site-centric network inference will be discussed: projection onto a PPI network, kinase-substrate networks, and data-dependent inference approaches.

PPI networks can be constructed from large scale interactome or AP-MS studies or from graph databases (such as STRING or BioGRID), as discussed in sections 1.4.2.2 and 1.5.1.2 respectively. However, two issues need to be addressed before phosphoproteomics data or quantitation can be projected onto a PPI network: how to summarise or account for multiple phosphorylated sites on the same protein node and how to represent dynamic changes in the network due to the modification state of a protein node. The PHOTON (phosphoproteomics dissection using networks) method attempts to address these issues (Rudolph *et al.*, 2016). PHOTON uses interaction data from BioGRID and weights the edges of the PPI network according to confidence of the interaction. Then, each node is assigned a "signalling functionality score" based on the quantitation of the phosphorylated sites belonging to that node and its neighbours. Thus, the core assumption of PHOTON is that a phosphorylated site is functional if its interactors are differentially phosphorylated within a network module enriched for high scoring nodes. PHOTON addresses how multiple phosphorylated sites can be accounted for by a single protein node and goes some way to addressing how protein activity dynamics can also be integrated using quantitative phosphoproteomics. However, the issue with using PPI networks as the basis is that edges will be drawn between nodes regardless of whether the interaction is occurring; scoring the nodes only goes some way to solve this. As previously highlighted, these networks are also often incomplete. Temporal Pathway Synthesiser (TPS) aims to mitigate both the issue of missingness and context-dependence by constraining the PPI network based on time-resolved quantification at the phosphorylated sites of the protein nodes (Köksal *et al.*, 2018).

Another approach, CelNOptR, utilises interaction networks related to cellular signalling built from literature curation rather than using PPI networks from graph databases (Terfve *et al.*, 2012; Gjerga *et al.*, 2020). These literature-curated networks are converted to Boolean logic models, then quantitative phosphoproteomics data is used to train them.

Another solution to creating a network with edges that are dependent on the context of the phosphoproteomics experiment they are based on is by inferring the interactions from kinase-substrate analyses. Though all the kinase-substrate methods discussed in section 1.5.2.2 could be used for this purpose, only some have been applied specifically to network inference. Perhaps the most popular option is the KinomeXplorer platform, which constructs kinase-substrate networks based on predictions from two algorithms: NetworKIN and NetPhorest (Linding *et al.*, 2008; Miller *et al.*, 2008; Horn *et al.*, 2014). NetworKIN uses a context-specific PPI network and kinase motifs to predict the likelihood of a kinase-substrate interaction , whilst NetPhorest uses a phylogenetic tree based algorithm to describe kinase-motif-substrate relationships (Linding *et al.*, 2008; Miller *et al.*, 2008). PHONEMeS (PHOsphorylation NEtworks for Mass Spectrometry) is a method that, similarly to CelNOptR, relies on a prior-knowledge network. The network contains known kinase-substrate and phosphatase-substrate reactions and is used to build a Boolean logical model based on quantitative phosphoproteomics data (Terfve *et al.*, 2015; Gjerga *et al.*, 2021). Since the principles are so similar to CellNOptR, PHONEMeS is restricted by constraint on what has been included in the prior-knowledge network. PHONEMeS has been used to predict the involvement of AKT, JNK, PKC and AMPK in endothelin-ephrin signalling-induced cell migration (Schäfer *et al.*, 2019). A combination of the PPI and kinase-substrate methods has been proposed by Narushima et al. (2016), whose approach integrates signalling pathway maps from Pathway Commons with kinase-substrate relationships inferred from the phosphoproteomics data; the resulting network can be visualised with the PTMapper app. This is similar to the approach by Buffard et al (2019), who filter the edges within signalling pathway maps (selected based on enrichment) based on kinase-substrate relationships identified in the phosphoproteomics dataset.

A drawback of both kinase-substrate and PPI based network inference is the semi-reliance on *a priori* knowledge. Several data-dependent inference methods based on the quantitative phosphoproteomics data itself have been proposed. Imamura et al. (2010) inferred edges between phosphorylated sites based on the Pearson correlation (R) of time-resolved quantitation; by overlaying the inferred networks onto signalling pathway maps from

NetPath, the authors demonstrated their approach could capture dynamic changes induced by phosphorylation. A similar approach was taken in the implementation of the SELPHI (Systematic Extraction of Linked PHospho-Interactions) web app, which provides a platform for network building based on correlation (Petsalaki *et al.*, 2015). The method proposed by Locasale & Wolf-Yadlin (2009) is also based on correlation between phosphorylated sites, however they then constrain the network based on the entropy maximization principle. The entropy maximization principle is increasingly considered a powerful means of reducing statistical uncertainty when handling biological datasets that are high-dimensional but have less-than-complete coverage (De Martino and De Martino, 2018). A non-correlation based approach is Sorad, which uses a flexible non-parametric ODE model to recreate signalling dynamics from phosphoproteomics data (Äijö, Granberg and Lähdesmäki, 2013). However, Sorad may not be scalable to large phosphoproteomics datasets due to its computational requirements.

The examples of data-dependent inference described here are those that have been specifically designed for or applied to phosphoproteomics data. However, methods designed generally for time-resolved data may also be applicable for describing phosphoproteome dynamics. For example, a dynamic extension of the maximum entropy principle, maximum caliber, has been generalised for application to biological networks and could be applied to phosphoproteomics (Weistuch *et al.*, 2020).

1.6 Aims and Objectives

Phosphoproteomics is an invaluable tool for profiling the impact of external perturbations on intracellular signalling cascades. However, challenges remain in gaining biological insights from these datasets. Challenges derive from all steps of the experimental pipeline – from experimental design, sample preparation, mass spectrometry technology for sample analysis, peptide identification/sequencing and the final bioinformatics analysis. One of the key goals and greatest bottleneck in the bioinformatics analysis of phosphoproteomic datasets is the identification of functionally and biologically relevant phosphorylated sites. What can be defined as functionally relevant is dependent on the experimental context under study. For phosphoproteomics experiments aiming to understand RTK signalling cascades and their regulation, function can be as broad as modulating whole signalling pathways or altering cellular behaviour.

The work described in this thesis aims to address this bioinformatics challenge by developing and applying approaches to understand the function of phosphorylated sites in RTK signalling cascades. This aim was delivered through two key objectives:

1. To develop phosphorylated site-centric analysis methods.
2. To consider how complex phosphoproteomics datasets can be analysed with network-based analysis methods to reveal biological function of regulated phosphorylated sites in RTK signalling cascades.

Objective 1 is primarily addressed in chapter 2, which presents a method developed to infer the context-dependent function of phosphorylated sites from shotgun phosphoproteomics data. In chapter 3, the results of a project describing the contribution of internalised FGFR2b (located at the recycling endosome) to signalling cascades and cell fate are detailed. I analysed a high-dimension spatial phosphoproteomics datasets based on APEX proximity-dependent biotinylation; given that this approach has not previously been used for phosphoproteomics, analytic methods had to be considered or scaled from existing proteomics analytic methods. This addresses objective 2. Following from the work described in chapter 3, a need was identified for a compelling way to understand and visualise the subcellular locations the phosphorylated proteins identified by our analysis were found in. Consequently, in chapter 4, I describe the development of a web-based tool to visualise the enrichment of Gene Ontology Cellular Components, by projecting the enrichment results onto a schematic of a mammalian cell. Therefore, chapter 4 addresses both objectives. Finally, in chapter 5, a complex phosphoproteomics experiment that compares the dynamic response to high (100 ng/mL) and low (5 ng/mL) concentrations of EGF or FGF1 activating EGFR and FGFRs, respectively, is presented and a pipeline for data analysis is discussed. I showed how site-centric approaches (including the new method described in chapter 2) can be employed to understand the dose-dependent differential effects of these two critical growth factors. Chapter 5 address both objectives 1 and 2.

2 Using Multilayer Heterogeneous Networks to Infer Functions of Phosphorylated Sites

2.1 Abstract

Mass spectrometry-based quantitative phosphoproteomics has become an essential approach in the study of cellular processes such as signaling. Commonly used methods to analyze phosphoproteomics datasets depend on generic, gene-centric annotations such as Gene Ontology terms, which do not account for the function of a protein in a particular phosphorylation state. Analysis of phosphoproteomics data is hampered by a lack of phosphorylated site-specific annotations. We propose a method that combines shotgun phosphoproteomics data, protein–protein interactions, and functional annotations into a heterogeneous multilayer network. Phosphorylation sites are associated to potential functions using a random walk on the heterogeneous network (RWHN) algorithm. We validated our approach against a model of the MAPK/ ERK pathway and functional annotations from PhosphoSitePlus and were able to associate differentially regulated sites on the same proteins to their previously described specific functions. We further tested the algorithm on three previously published datasets and were able to reproduce their experimentally validated conclusions and to associate phosphorylation sites with known functions based on their regulatory patterns. Our approach provides a refinement of commonly used analysis methods and accurately predicts context-specific functions for sites with similar phosphorylation profiles.

2.2 Introduction

Phosphorylation is the most studied post-translational modification (PTM) due to its central role in cellular regulation. It is thought to be the principal PTM in the human proteome and an essential mediator of protein–protein interactions (PPIs) and protein functions (Sugiyama, Imamura and Ishihama, 2019). Transient changes occur at specifically regulated phosphorylation sites, of which there may be multiple on each protein. Regulation of phosphorylation is often dependent on perturbations such as the activity of extracellular ligands, drug treatment, or physical stimuli in the extracellular environment (Day, Sosale and Lazzara, 2016). By comparing changes in the phosphoproteome of cells under different experimental conditions through mass spectrometry-based phosphoproteomics, phosphorylated sites that are key players in cellular processes and functions can be uncovered in an unbiased, high-throughput manner (Vyse, Desmond and Huang, 2017).

Functional analysis of phosphoproteomics datasets is typically based on gene-centric enrichment of Gene Ontology (GO) terms or involvement in known pathways (Munk, Refsgaard and Olsen, 2016). However, this approach disregards information captured by phosphoproteomics data on changes at specific phosphorylated sites, by limiting the analysis to the protein level. The modification state of a protein is inherently coupled to its function; PTMs alter protein activity and the ability to interact with different sets of proteins. Furthermore, if a protein is phosphorylated on multiple sites, each with a different function and regulatory pattern, this information is not revealed by gene-centric analysis (Sacco, Perfetto and Cesareni, 2018). For instance, the well-studied signaling protein MAPK1 has 18 known phosphorylated sites recorded in the database PhosphoSitePlus (PSP); however, only six have been annotated to a downstream function (Hornbeck *et al.*, 2015). In the most recent release of PSP (v6.5.9.3), only 4271 of the more than 230,000 human phosphorylated sites recorded in the database are associated to 19 generic biological functions (*e.g.*, "cell cycle" and "transcription"), which are qualified with one of the following: "induced", "inhibited", "regulated", or "altered". Enrichment analyses that rely on generalizations based on protein-level or gene-centric descriptions exclude the details that are encoded in the phosphorylation signature. Analyses are thus hampered by the lack of phosphorylation site-specific functional annotations.

Several network-based methods have been proposed to move toward phosphorylation site-specific analyses of phosphoproteomics data. A significant focus has been on the inference of kinase–substrate networks, such as NetworKIN (Linding *et al.*, 2008), KSEA (Casado *et al.*, 2013), and IKAP (Mischnik *et al.*, 2016), among others. These methods are useful for reconstructing the architecture of signaling intracellular networks, which can be informative for identifying modules of modified proteins involved in cellular processes, but again remain hampered by the lack of site-specific functional annotation (Needham *et al.*, 2019). They may also be biased toward the most studied kinases and the exclusion of non-kinase proteins. Rudolph *et al.* (2016) proposed a method to address this issue named PHOTON, which identified differentially regulated proteins based on the level of phosphorylation of their binding partners in a high-confidence PPI network and then used logistic regression to identify the involvement of the phosphorylated proteins in known signaling pathways. PHOTON is not truly site-centric, however, as it relies on summarized quantitative values of phosphorylated sites. An alternative approach described by Krug *et al.* (2019) uses a modification to the gene-centric method gene set enrichment analysis (GSEA), which relies on a curated resource containing literature-derived phosphorylated site-specific signatures

to assign functions to sites. However, the use of such a resource is limited by the lack of phosphorylated site annotation to these signatures or specific cellular processes; although it is useful for identifying well-studied sites, predictions of functions of under-studied sites or sites in alternative contexts would not be captured. These do not fulfil the same role as popular gene-centric methods such as over-representation analysis (ORA) (Boyle *et al.*, 2004) in aiding prediction and hypothesis generation when performing exploratory analysis of shotgun phosphoproteomics data.

In recent years, heterogeneous or multilayer networks have been used to represent many types of omics datasets (Battiston *et al.*, 2017). These specialized networks are used to describe multiple types of associations with nodes representing different entities. To identify relationships between the different biological layers, random walk algorithms have been applied to these networks. The random walk on the heterogeneous network (RWHN) with the restart (RWR) method (Li and Patra, 2010) has been particularly popular. Jiang (Jiang, 2015) used RWR to prioritize disease candidate genes in a PPI–phenome network; similarly, Soul *et al.* (2015) applied it to a PPI–phenome network to identify disease mechanisms. Recent work has extended the method to multiple layers of biological information, for example, to infer disease-associated m6A RNA methylation *via* known gene–disease associations (Tang *et al.*, 2019). Similar methodology was used to associate phosphorylation sites recorded in the PSP database to diseases *via* kinase–substrate interactions (Xu and Wang, 2016).

Here, we propose an algorithm that uses RWHN to associate phosphorylated sites to context-specific function *via* a heterogeneous multilayer network using shotgun phosphoproteomics data. The network combines three layers of information: phosphorylated sites, protein interactions, and GO terms. Clustering of phosphoproteomics data is used to find common features within datasets and is generally followed by enrichment analyses. This is based on the assumption that common patterns of phosphorylation, based on temporal changes or those in response to a particular stimulus, treatment, or environmental context, are a likely indicator of involvement in common functions or processes (Behar and Hoffmann, 2010). We utilize this concept in our algorithm by connecting phosphorylated sites that have been clustered together and therefore share regulatory patterns, within the multilayer network. The algorithm is intended for use on large phosphoproteomics datasets, assessing perturbations or processes as opposed to an interpretation of the full phosphoproteomics network in the cell.

We first apply our algorithm to a small-scale, manually curated validation network. We also assess the ability of our method and the most commonly used alternative, ORA, to capture the functional descriptions recorded in PSP, which are based on experimental analysis. We then demonstrate the utility of our algorithm using three previously published shotgun phosphoproteomics datasets, describing early signaling events in HeLa cells upon EGF and TGF-α stimulation (Francavilla *et al.*, 2016). phosphorylation-mediated changes in breast cancer cells resistant to lapatinib treatment (Ruprecht *et al.*, 2017), and subcellular location-dependent signaling events downstream of HRAS (Santra *et al.*, 2019). We demonstrate that phosphorylated sites can be differentially assigned to functional annotations and this is driven by changes in the context-dependent modification of these sites. Our method is suitable for use with any phosphoproteomics dataset and could be generalized for data describing other PTMs.

2.3 Experimental Procedures

A multilayer heterogeneous network was constructed to relate biological functions to phosphorylation sites *via* a protein–protein interaction network (Figure 2-1). Three types of nodes are contained within the network: phosphorylation sites (called "sites" here on in for brevity), proteins, and functional annotations [either Gene Ontology biological process (GOBP) terms or KEGG pathways]. The edges describe five possible associations, which are either bipartite (*I.e.*, site to protein and protein to function) or between the same type of nodes.

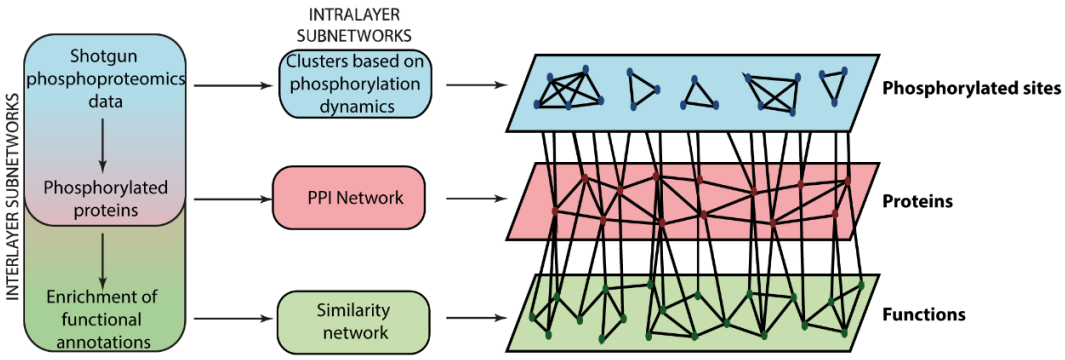


Figure 2-1. Overview of the multilayer heterogeneous network construction from phosphoproteomics data.

2.3.1 Dataset Accession and Processing

Sites annotated to have a role in biological processes were extracted from the "Regulatory sites" dataset, available for download from PhosphoSitePlus (version 6.5.9.3). The dataset

was filtered on the “ON_PROCESS” column to only include sites annotated to biological roles, rather than those with molecular or interaction roles.

The results of the phosphoproteomics experiment described in Francavilla *et al.* (2016) were taken from Supporting Information Table S1 of their manuscript. The authors had processed the raw mass spectrometry data using MaxQuant (Cox and Mann, 2008), filtered identified phosphorylated sites based on a localization probability of greater than 0.75, and normalized remaining data to the ratio of EGFR at 1 min after stimulation with EGF or TGF-α. Prior to multilayer heterogeneous network construction, the data were further filtered to remove sites with two or more missing ratios in their time series and containing at least one SILAC ratio higher than 2 or lower than 0.5 (as described in the methods of Francavilla *et al.* (2016)). Missing data were imputed with random draws from a truncated distribution, as previously described (Tyanova *et al.*, 2016), using the *impute.QRLIC* function from the *imputeLCMD* R package.

Data from the Ruprecht *et al.* (2017) study were taken from Supporting Information file named “Filtered and normalized phosphoproteome dataset”. The raw mass spectrometry data had been processed using MaxQuant, filtered to include only identified phosphorylated sites with a localization probability of greater than 0.75, and normalized. Sites that did not show a significant change (FDR <1%) either between the untreated parental and lapatinib-treated resistant (SILAC ratios H/L) or between the lapatinib-treated parental and lapatinib-treated resistant (SILAC ratios H/M) experimental conditions were filtered from the data, as described in the methods of Ruprecht *et al.* (2017) Missing data were imputed with the same method used for the Francavilla *et al.* dataset.

Data from the Santra *et al.* study (2019) were taken from Supporting Information Table S2. The raw mass spectrometry data had been processed using MaxQuant and filtered to include only identified phosphorylated sites with a localization probability of greater than 0.75; significant sites were identified using a two-sample *t*-test with FDR. The downloaded dataset was normalized using the *limma* R package function *normalizeBetweenArrays* using the quantile method (Ritchie *et al.*, 2015). Missing data were imputed with the same method used for the Francavilla *et al.* dataset. The data were then filtered to exclude sites that had not been identified as significant in the original analysis by Santra *et al.*

Before applying the algorithm, all datasets were clustered based on what was reported in the original publication, where available. The model dataset was clustered using the fuzzy C-

means (FCM) method (Schwämmle and Jensen, 2010), with the number of clusters selected based on the temporal trends we could visually identify in the data; generally, silhouette analysis is recommended for selecting the number of clusters using FCM for real experimental datasets (Campello and Hruschka, 2006). The sites extracted from PSP were clustered based on what process they were annotated to (as indicated in the "ON_PROCESS" column in the Regulatory_sites file available for download on the PSP website). The regulated phosphorylated sites from Francavilla *et al.* were clustered using FCM (Figure 2-5, Figure 2-6); since the original publication used a cluster number of six, the same was done here for better comparison. The regulated phosphorylated sites from Ruprecht *et al.* and Santra *et al.* were both clustered using the k-means method, with the cluster number being selected using the elbow method (Figures S4 and S5, respectively) (Zambelli, 2016).

2.3.2 Construction of the Multilayer Network

2.3.2.1 Phosphorylation Site—Phosphorylation Site Subnetwork

Edges were drawn between sites in the same cluster that had a Pearson correlation (R^2) between all the data points greater than or equal to 0.99 (Figure 2-1).

2.3.2.2 Protein—Protein Subnetwork

The protein–protein interaction network was extracted from STRING (Szklarczyk *et al.*, 2019). All interactors of the proteins included in the phosphoproteomics datasets with an experimental confidence score of greater than 0.4 were included (Figure 2-1).

2.3.2.3 Function—Function Subnetwork

We used either GOBP terms or KEGG pathways as functional annotators in this work. If GOBP terms were included in the multilayer network, the *GOSemSim* package from Bioconductor (Yu *et al.*, 2010) was used to calculate the semantic similarity of enriched GOBP terms. An edge was drawn between terms with a semantic similarity of greater than 0.7, as calculated using the Wang method included in the *goSim* function of *GOSemSim* (Yu *et al.*, 2010; Szklarczyk *et al.*, 2019). In the case of KEGG pathways, edges were drawn between pathways that had greater than 70% pairwise similarity in their functional annotation profiles, following the method described in Stoney *et al.* (2015) (Figure 2-1).

2.3.2.4 Phosphorylation Site—Protein Bipartite Edges

Edges were drawn between phosphorylated sites and their parent protein. Therefore, sites will only have one edge, but proteins will have edges to all the sites found on that protein that were differentially regulated in the dataset (Figure 2-1).

2.3.2.5 Protein—Function Bipartite Edge

We assumed that closely connected nodes in the protein–protein subnetwork would more likely be involved in similar biological processes. Therefore, we computed modules of the protein–protein subnetwork using the Louvain module detection method (Blondel *et al.*, 2008). Proteins from each module were analyzed for enrichment (hypergeometric test with Benjamini–Hochberg correction, FDR <0.05%) of functional annotations (from either the "GO_Biological_Process_2018" or "KEGG_2019_Human" libraries, included in enrichR and listed on the EnrichR website) using the *enrichR* (Kuleshov *et al.*, 2016) R interface. This increased the specificity of terms to be included in the network. When GO terms were included, high-frequency (annotated to more than 5% of genes) and semantically redundant terms (similarity >0.9) were filtered using the Bioconductor *GOSemSim* (Yu *et al.*, 2010) and *GO.db* packages (Carlson, 2019) (Figure 2-1).

2.3.3 Random Walk on the Heterogeneous Network

The heterogeneous network can be represented as an adjacency matrix as follows

$$A = \begin{bmatrix} A_{\mathrm{RR}} & A_{\mathrm{RP}} & 0 \\ A_{\mathrm{PR}} & A_{\mathrm{PP}} & A_{\mathrm{PF}} \\ 0 & A_{\mathrm{FP}} & A_{\mathrm{FF}} \end{bmatrix} \quad (1)$$

where A_{RR} represents site–site associations, A_{PP} represents protein–protein associations, A_{FF} represents function–function associations, A_{RP} represents site–protein associations with A_{PR} as the transpose, and A_{PF} represents protein–function associations with A_{FP} as the transpose.

As described in previous work (Li and Patra, 2010), a transition matrix (M) was calculated for use in the first stage of the algorithm

$$M = \begin{bmatrix} M_{\mathrm{RR}} & M_{\mathrm{RP}} & 0 \\ M_{\mathrm{RP}^{\mathrm{T}}} & M_{\mathrm{PP}} & M_{\mathrm{PF}} \\ 0 & M_{\mathrm{PF}^{\mathrm{T}}} & M_{\mathrm{FF}} \end{bmatrix} \quad (2)$$

The bipartite inter-subgraph transition matrices (M_{RP} and M_{PF}) were calculated as

$$(M_{\mathrm{RP}})_{i,j} = \begin{cases} \lambda (A_{\mathrm{RP}})_{i,j} / \sum_j (A_{\mathrm{RP}})_{i,j}, & \text{if } \sum_j (A_{\mathrm{RP}})_{i,j} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$(M_{\mathrm{PF}})_{i,j} = \begin{cases} \lambda (A_{\mathrm{PF}})_{i,j} / \sum_j (A_{\mathrm{PF}})_{i,j}, & \text{if } \sum_j (A_{\mathrm{PF}})_{i,j} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where λ is the transition probability (*I.e.*, the likelihood of the walker moving between two layers of the network).

The intra-subgraph transition matrices (M_{RR}, M_{PP}, and M_{FF}) were calculated as

$$(M_{RR})_{i,j} = \begin{cases} (A_{RR})_{i,j} / \sum_j (A_{RR})_{i,j}, & \text{if } \sum_j (A_{RR})_{i,j} = 0 \\ (1-\lambda)(A_{RR})_{i,j} / \sum_j (A_{RR})_{i,j}, & \text{otherwise} \end{cases} \quad (5)$$

$$(M_{PP})_{i,j} = \begin{cases} (A_{PP})_{i,j} / \sum_j (A_{PP})_{i,j}, & \text{if } \sum_j (A_{PP})_{i,j}, (A_{PF})_{i,j} = 0 \\ (1-\lambda)(A_{PP})_{i,j} / \sum_j (A_{PP})_{i,j}, & \text{otherwise} \end{cases} \quad (6)$$

$$(M_{FF})_{i,j} = \begin{cases} (A_{FF})_{i,j} / \sum_j (A_{FF})_{i,j}, & \text{if } \sum_j (A_{FF})_{i,j} = 0 \\ (1-\lambda)(A_{FF})_{i,j} / \sum_j (A_{FF})_{i,j}, & \text{otherwise} \end{cases} \quad (7)$$

RWHN is a ranking algorithm; nodes are ranked based on the probabilities of finding the random walker at a given node in the steady state, having started at a given seed node or a set of seed nodes. In this work, we set the seed nodes to be those sites belonging to a particular cluster. The probability of finding the random walker at each node for each step is calculated based on the iterative equation

$$\begin{cases} P_{s+1} = (1-r) \times M^T \times P_s + r \times P_0 \\ P_{s+1} - P_s \leq 10^{-6} \end{cases} \quad (8)$$

where r is the restart probability (set to 0.7 as described in Köhler *et al.*(2008)), P_0 is the initial probability vector, and P_s is the probability vector at step s. P_0 was calculated such that all seed nodes were given equal probabilities with their sum equal to 1. All other nodes in the site–site subnetwork were assigned an initial probability of 0. Nodes in other subnetworks were assigned equal probabilities with their sum equal to 1 and weighted with the tunable parameters η_P and η_F that are used to weigh the influence of each layer.

The output of the algorithm is a ranked list of all the nodes, based on the probability of finding the random walker at each node in the steady state. This list is filtered by retaining functional annotations and removing proteins and sites. We remove annotations that are ranked in the same position regardless of the seed node, in order to focus on specific

functions related to each cluster. The top 5% of functions in the probability distribution are retained.

We implemented the algorithm in R using packages available from Bioconductor and CRAN. The source code is available at github.com/jowatson2011/RWHN_phosphoproteomics, and an R package for general use of the tool can be found at github.com/JoWatson2011/phosphoRWHN. RWHN on the validation network took less than 30 s to run; however, the larger experimental datasets took several hours on a moderately powerful computer (32GB RAM, Intel i7-4770 Processor).

2.3.4 Over-representation Analysis

The names of modified proteins in each cluster were used as input for the *enrichR* R interface (Kuleshov *et al.*, 2016). The libraries "GO_Biological_Process_2018" or "KEGG_2019_Human" (included in enrichR and listed on the EnrichR website) were used depending on the one used for the RWHN analysis of the same data. The list of over-represented terms was filtered to only include those with an FDR <0.05.

2.4 Results

2.4.1 Overview of the Algorithm

To associate phosphorylated sites of unknown functions to potential cellular functions, we developed an algorithm to apply to shotgun phosphoproteomics data. First, in order to help experimental biologists to identify the context- or perturbation-specific roles of modified proteins changing in each given experiment, a multilayer heterogeneous network is constructed based on the regulation of phosphorylated sites. Regulated phosphorylated sites may be defined differently by the user depending on the quantitation method used in their phosphoproteomics experiment (*e.g.*, label-free quantification and SILAC) or their experimental question. The network represents three layers of biological entities and information: regulated phosphorylated sites, protein–protein interactions, and biological functions (Figure 2-1). We then apply a ranking algorithm, RWHN, which ranks nodes of each layer based on (I) the distance from the phosphorylated sites of interest, which are assigned as "seed" nodes, and (ii) the topology of the multilayer heterogeneous network (Li and Patra, 2010). Functions that are highly ranked can be considered more correlated with a set of seed nodes.

In the multilayer heterogeneous network, edges are drawn between sites based on similarity in the regulatory pattern of their phosphorylation, which we determine using k-means or

fuzzy C-means clustering. For the protein layer, a PPI network is constructed of the phosphorylated proteins and their interactors using the STRING database (Szklarczyk *et al.*, 2019). Interactors are included to account for non-phosphorylated proteins or those below the limit of detection of the experiment (Riley and Coon, 2016). To get a comprehensive and specific list of functional annotations, we identified closely connected groups of proteins in the PPI network using module detection and performed functional enrichment analysis on these modules. Edges between functional annotations are drawn based on functional similarity and overlap (Yu *et al.*, 2010; Stoney *et al.*, 2015).

The RWHN algorithm simulates a walker moving from a starting node(s) (called a seed) and then node to node through the multilayer network. Each step is influenced by the probability of transition to another layer (λ), the weighting of the protein and function layers (η_P and η_F), and the probability of restart [*I.e.*, teleportation back to the seed node(s), *r*]. The output of RWHN consists of a list of ranks for all the nodes in the network based on the likelihood of the walker reaching those nodes.

To optimize the algorithm, we ran RWHN with each of these parameters (λ, *r*, η_P, and η_F) tested over a range of values (0.1–0.9), changing one while setting all others to 0.5 (Appendix 1 Figure 1). Performance was decidedly stable over the range of η_P and η_F; however, altering λ and *r* resulted in differential ranking depending on the parameter value. Previous work has suggested an ideal λ and *r* of 0.7 (Köhler *et al.*, 2008; Li and Patra, 2010; Blatti and Sinha, 2016), while η_P and η_F were set to 0.7 and 0.3, respectively, to prioritize movement in the protein layer and reduce the number of terms having the same rank regardless of the seed.

The parameters used in the construction of the PPI network were also assessed. Proteins interacting with the phosphorylated proteins found in a given sample were extracted from STRING; those proteins whose interaction had a STRING experimental confidence score above a given threshold were included in the multilayer heterogeneous network. To investigate the impact of choosing different STRING confidence scores on the results, we tested a confidence score range of 0.1–0.9. First, we looked at whether connectivity of the graph was altered by the score. As expected, the greater the confidence score, the more components the PPIs were split into; above a score of 0.7, the PPIs had two or more components (Appendix 1 Figure 2B). We next investigated whether the RWHN results were affected by changing the STRING confidence score increasing components within the protein layer. Using weighted tau correlation (Shieh, 1998) to compare the rankings from each set of results, we found that the correlation was generally in the range of 0.8–1.0, with the lowest

value being 0.54 between networks constructed with 0.1 or 0.9 confidence threshold, which were the extremes of the range tested.

2.4.2 Validation 1: Model of the MAPK/ERK Pathway

MAPK/ERK signaling has been well studied, and the temporal phosphorylation status of component proteins in response to growth factor stimuli is relatively established (Fey *et al.*, 2012; Arkun and Yasemi, 2018) (Figure 2-2A). Based on this general understanding, we compiled a simple validation dataset of phosphorylation dynamics at particular sites of the main signaling proteins within the pathway (Figure 2-2B). Sites were chosen based on whether their phosphorylation is known to activate or inactivate protein activity, as recorded in the PSP database (Hornbeck *et al.*, 2015). We clustered the data into five clusters, referred to here as clusters 1–5, using the fuzzy c-means algorithm (Figure 2-2B). The features of the multilayer network that was constructed are summarized in Table 2-1. Enrichment of GOBP terms was used to form the functional annotation layer of the network.

Table 2-1 Multilayer heterogeneous network, constructed from MAPK/ERK validation data. GOBP terms were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	28	19
Protein-protein	932	313
Function-function	148	88
Site-protein	19	19 sites, 8 proteins
Protein-function	1796	261 proteins, 298 functions

We ran the RWHN algorithm over the multilayer heterogeneous network with seed nodes set to all the sites belonging to one of the clusters; this was then repeated for each cluster (Figure 2-3A). The highest ranked GOBP term was the same ("protein autophosphorylation") regardless of the seed nodes. However, it was possible to differentiate between the five clusters based on the ranking of terms below the first one.

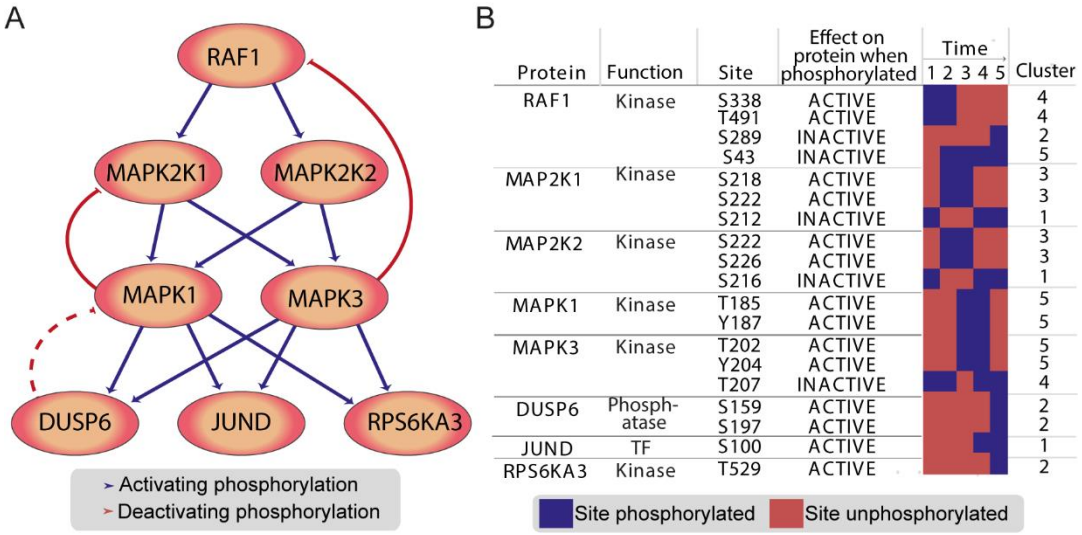


Figure 2-2. Details of validation data.
A. *Construction of validation data was based on a traditional representation of the MAPK/ERK pathway.* ***B****. Validation dataset was constructed based on the MAPK/ERK pathway, simulating phosphorylation dynamics over time after pathway activation. The sites were clustered using fuzzy c-means clusters. These data were used to construct a heterogeneous multilayer network.*

Where available, we could use the biological process annotations in PSP (v6.5.9.3) as a "benchmark" for the highly ranked terms associated to each cluster. RPS6KA3_Y529, annotated to "apoptosis, altered" in PSP was found in cluster 2, along with DUSP6 sites; this cluster highly ranks "positive regulation of programmed cell death". Of the four RAF1 sites included in this model dataset, two were annotated to "cell cycle regulation" in PSP: RAF1_S289, in cluster 2, and RAF1_S338, in cluster 5. These two clusters both rank "regulation of the mitotic cell cycle" in the top 5% of terms.

Largely due to the inherent step in our method in which the redundant terms are reduced in the multilayer heterogeneous network, fewer terms were found in the results of RWHN than in ORA (Figure 2-3A,B). We next wanted to assess whether the terms in the RWHN result were drawn from a particular level of the GO hierarchy; if they were, this would suggest a bias toward more specific or broader terms compared to ORA. From the GO hierarchy, we extracted the parent and ancestor terms of the terms found in each set of the results and compared them relative to the top of the GO hierarchy (Figure 2-3C). We found that RWHN terms were drawn from the same range of the hierarchy as the ORA terms, with slightly more coming from lower in the hierarchy, indicating that the results have more specificity than the ORA results. We note that there is also less redundancy in the RWHN results, despite the ORA results also being reduced for semantically similar/redundant terms. For example, the

terms “regulation of the ERK1 and ERK2 cascades” and “ERK1 and ERK2 cascades” both come up in the ORA results and are associated to different clusters.

To assess the robustness of the algorithm, we performed a random permutation control. RWHN was run 100 times with the seeds set as before and random permutations of the subnetworks and bipartite edges maintained. Kernel density estimation (KDE) was calculated to assess how often each GO term occurred at each rank in the 100 random permutations for each set of seed nodes. In each case, there was a subset of terms that were more likely to have a rank greater than 50; however, none strongly correlated with the actual rankings for each set of seed nodes (Appendix 1 Figure 2A). This confirms that the rankings are not random but primarily determined by the network topology. We next investigated the rate of potential false positives in the results. Previous implementations of RWHN, primarily using genomics data, have utilized an arbitrary threshold of the top *x* ranked terms or the top *x* % of ranked terms (based on the probability vector as calculated in eq 8). A threshold that is too stringent would not remove lower ranking terms from the results, while a too low threshold would increase the number of false positives. Using the results of RWHN run on the randomly permuted networks, we calculated the probability that the “true” ranking (*i.e.*, from RWHN run on the non-permuted network) was different from the random rankings ($p < 0.05$, Mann–Whitney *U* test with Benjamini–Hochberg adjustment (H. B. Mann and D. R. Whitney, 1947; Benjamini and Hochberg, 1995)). We found that most of the rankings differed from random (Appendix 1 Figure 2B). The terms that had $p > 0.05$ were checked against the RWHN results with different thresholds selected (top 1, 5, 10, and 15% of ranked terms); none were found above the 1% threshold for any of the clusters, and only one term (“peptidyl serine modification” in cluster 2) was found above the 5 and 10% thresholds (Appendix 1 Figure 2B). We chose to set the default threshold to the top 5% of terms to be more conservative while retaining useful annotations.

For a given set of seed nodes, the algorithm is capable of giving high ranks to functions with known associations to sites. It is also capable of predicting non-random, reasonable functions for sites of unknown functions in this context.

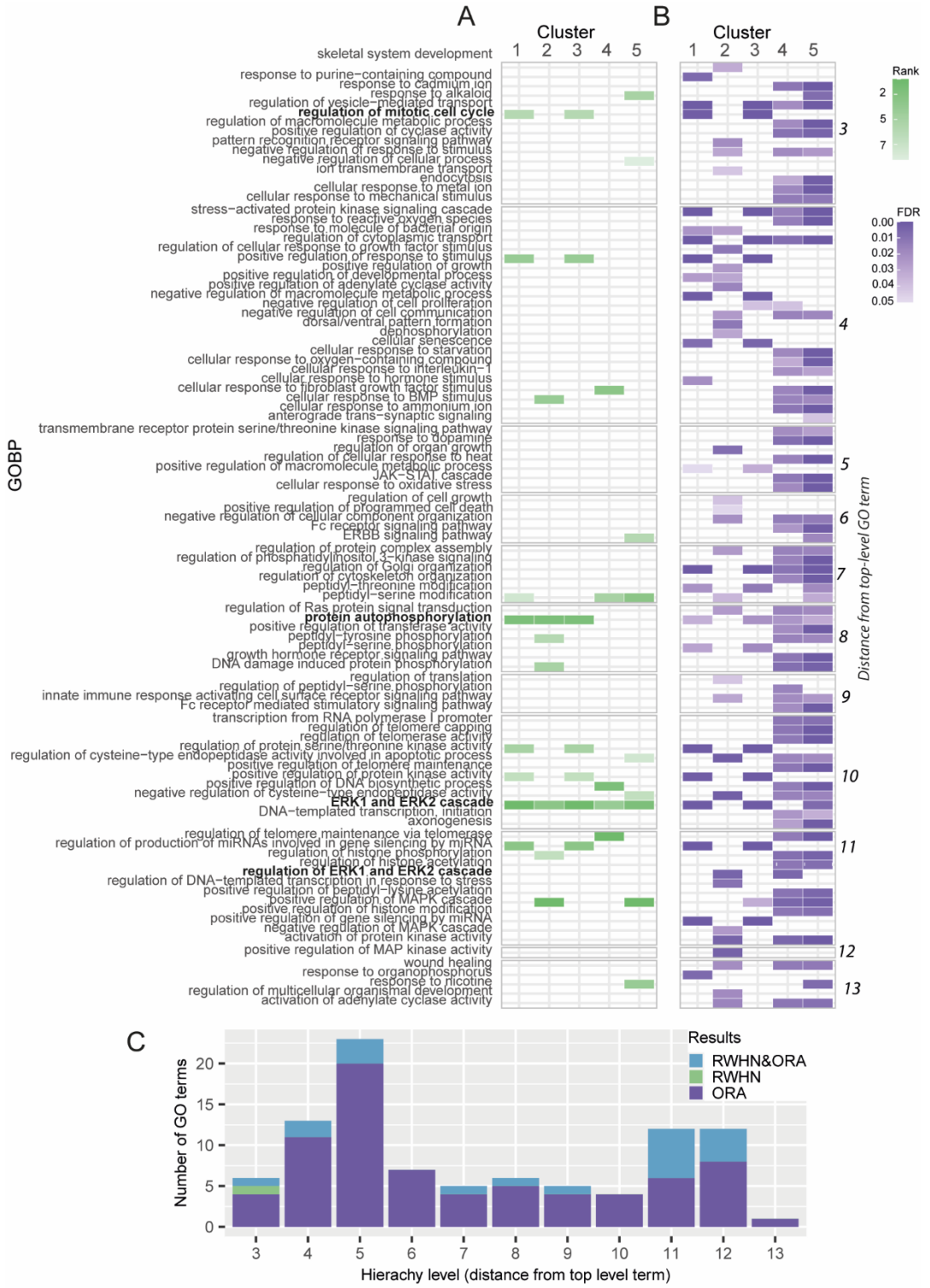


Figure 2-3. Site- and gene-centric analysis of validation data.
A. *RWHN results with sites annotated to functional roles in PSP; seed nodes are equivalent to the clusters.* ***B.*** *ORA results with modified proteins contained in toy data.* ***C.*** *Number of terms from ORA or RWHN that are from different levels of the GO hierarchy.*

2.4.3 Validation 2: Classification of PhosphoSitePlus-Annotated Sites

A comprehensive benchmark of dynamically regulated phosphorylated sites is not currently available, due to the lack of phosphorylated sites with a regulatory pattern universally defined independent of the experimental conditions. Therefore, we used the static classifications of phosphorylated site functions recorded in PSP to benchmark the predictive power of our algorithm. The PSP database has annotations to biological processes for ~4000 human phosphorylated sites. As these annotations are non-standard, each one was mapped to the most closely related GOBP term, where available (Appendix 1 Table 1). The majority of sites were annotated to signal transduction, gene expression, or cell differentiation (Figure 2-4A). Of these, 1496 sites were annotated to more than one function; as these annotations are manually assigned based on literature searches, the multiple functions may be due to the differences in experimental questions in the source datasets. The sites were clustered based on which function they were annotated to and analyzed with RWHN and ORA. The GOBP terms in the results of both methods were associated to the child terms of each mapped PSP GO term, as per the GO hierarchy. The terms in the results of both methods showed a similar distribution to the terms in PSP as a whole, the main exception being in the cell differentiation category (Figure 2-4B). The two methods showed a considerable difference in the number of terms in the output, with ORA resulting in several hundred more terms in the output than RWHN. To test whether the two sets of results were drawn from the same parts of the GO hierarchy (with those from the top of the hierarchy being broader and those from the bottom being more specific), we constructed a tree based on the GO hierarchy. The tree was then divided into levels, to capture terms that were of similar descriptiveness. Finally, we counted the number of terms that came up in the RWHN or ORA results (Figure 2-4C). Both sets of results draw from the same part of the hierarchy, with a skew toward the broader terms. We therefore concluded that RWHN resulted in similarly descriptive but more restricted and focused annotations than ORA. This was notable given that RWHN did not associate any terms to the smaller categories in PSP, such as exocytosis, endocytosis, cell growth, and autophagy (Figure 2-4B).

2.4.4 Experimental Case Study 1: Dissecting EGF- and TGF-α-Induced Dynamic Phosphorylation

Given that PSP contains static annotations of sites from various sources, we wanted to test our algorithm on experimental datasets with controlled perturbations. We reasoned that this would allow us to assess whether our algorithm could differentially associate sites to their context-dependent function. The phosphoproteomics data retrieved from Francavilla *et al.*(2016) describe the effect in HeLa cells of stimulation with EGF or TGF-α over a period of

90 min (with time points at 1, 8, 40, and 90 min). In this study, one of the findings was that EGF or TGF-α stimulation of EGFR induces receptor degradation or recycling, respectively. TGF-α was associated with a more potent mitogenic and migratory response than EGF over time (Francavilla *et al.*, 2016). Phosphorylated sites from Supporting Information Table S1 of the original publication were filtered based on regulation by EGF or by TGF-α and divided based on which stimuli the regulation was dependent on. The separated data were clustered into six clusters using the fuzzy C-means method, as per the original publication, with each cluster representing a distinct dynamic profile of phosphorylation (Appendix 1 Figure 3). A multilayer heterogeneous network was constructed for both sets of data, described in Table 2-2 and Table 2-3.

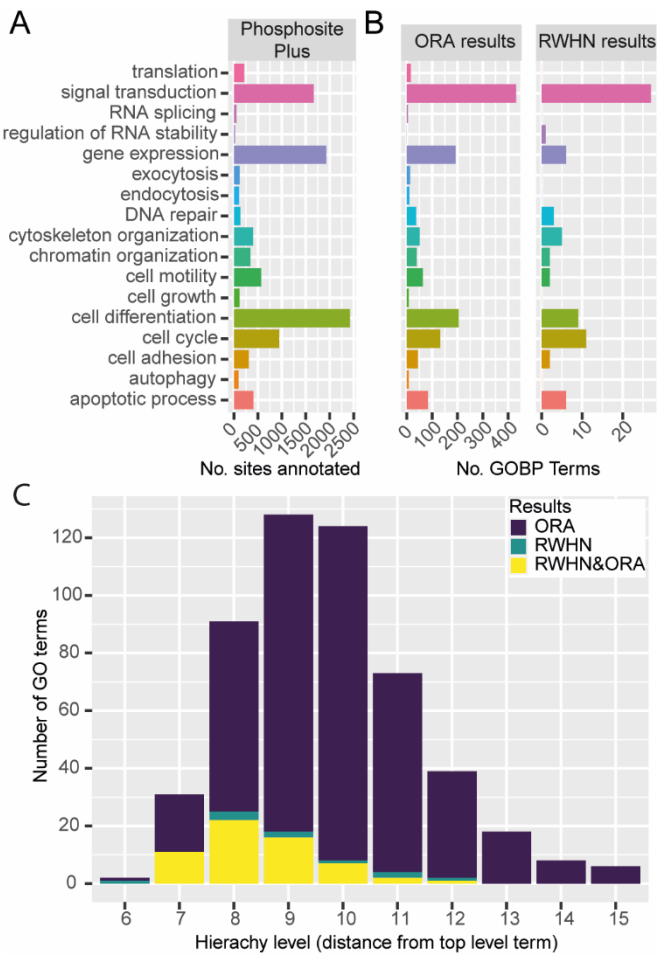


Figure 2-4. Distribution of annotations and classifications of sites extracted from PhosphoSitePlus. ***A.*** *Distributions of functionally annotated sites (to mapped GO terms).* ***B.*** *Distribution of terms in ORA results and RWHN results when analyzing functionally annotated sites from PSP.* ***C.*** *Number of terms from ORA or RWHN that are from different levels of the GO hierarchy.*

Table 2-2 Multilayer heterogeneous network, constructed from Francavilla et al. EGF-regulated phosphorylated sites. GOBP terms were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	4347	733

Protein-protein	16654	4072
Function-function	532	544
Site-protein	733	733 sites, 433 proteins
Protein-function	1223	115 proteins, 532 functions

Table 2-3 Multilayer heterogeneous network, constructed from Francavilla et al. TGF-α-regulated phosphorylated sites. GOBP terms were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	3085	675
Protein-protein	16556	3960
Function-function	332	496
Site-protein	675	675 sites, 403 proteins
Protein-function	1015	109 proteins, 486 functions

Several biologically relevant terms were differentially ranked in the EGF and TGF-α networks (Figure 2-5, Figure 2-6). Consistent with the original publication, the term "regulation of ERK1 and ERK2 cascades" was highly ranked using RWHN when seeds were set to sites in cluster 2 (representing EGF late responders) for the EGF network and in cluster 1 (representing TGF-α early responders) for the TGF-α network. Despite their central role in TGF-α signaling, when we performed standard ORA enrichment, terms related to MAPK/ERK were not found in any of the clusters (Figure 2-5). However, both ORA and our method highlighted the role of the proteins belonging to TGF-α cluster 1 in regulating EGFR/ERBB signaling, with terms referring to regulation of these pathways ranked highly (Figure 2-6). We used cluster 2 under the EGF conditions and cluster 1 under the TGF-α conditions to verify the robustness and accuracy of our approach. As the MAPK/ERK cascade is well studied in respect to EGF/TGF-α signaling, many of its associated sites have documented functions that we could use to benchmark our results. EGF cluster 2 contains sites of several known players of the MAPK/ERK cascades, including RSP6KA3_S375, SOS1 (T1119 and S1137), RAF1 (S289, S96, and S301), JUN_S243, and JUNB_259. All of these sites are predicted MAPK1/MAPK3 targets, with the exception of SOS1_T1119, which is a MAP2K1 target (Linding *et al.*, 2008). As this cluster represents those sites phosphorylated late in the time course, we would expect ERK1/2 feedback regulation to emerge. Besides confirming the term "regulation of ERK1 and ERK2 cascades" in EGF cluster 2, results from the RHWN algorithm add detail to this picture, with "establishment of protein localization to organelles" ranked highly in this cluster (Figure 2-5). This finding confirms the conclusion of Francavilla *et al.* of EGFR trafficking to the lysosome and subsequent degradation when stimulated by EGF (Francavilla *et al.*, 2016). EGF cluster 2 is

the only one which did not highly rank “regulation of clathrin-dependent endocytosis” compared to the five other clusters; as this cluster represents sites that are highly phosphorylated late in the time course, this coincides with the understanding of receptor endocytosis as a process regulated in the early stage. This cluster is one of the five that highly ranks “regulation of cell migration” (Figure 2-5); PSP also annotates several of the sites in this cluster to cell motility (PTPN12_S571, SLC9A1_S703, and FLNA_S2152).

TGF-α cluster 1, on the other hand, contains sites that are phosphorylated within the first 8 minutes upon TGF-α stimulation. Accordingly, this cluster contains different members of the MAPK/ERK cascade, such as the canonical activating sites of MAPK1, MAPK3, and MAPK14 (also known as p38) and several sites on EGFR (EGFR_S991, EGFR_S995, and EGFR_T693). A few of these sites, such as EGFR_T693, are annotated to “receptor internalization” in PSP. This cluster also contains several phosphorylated sites on known players in receptor trafficking, including RAB7A_Y183 and RAB11B_Y8, which are highlighted in the original publication (Francavilla *et al.*, 2016), along with CBL_Y674 and SH3BP4_T118. Our algorithm ranked terms such as “endosomal transport,” “establishment of protein localization to organelles,” and “rab protein signal transduction” high when seed nodes were set to the sites found in this cluster (Figure 2-6). This was only partially captured by the ORA (Figure 2-6), which showed enrichment of the term “regulation of vesicle mediated transport” but also “protein localization to the lysosome”, the opposite of the experimentally derived conclusion of Francavilla *et al.*

EGF cluster 1 representing phosphorylated cycling sites ranks both “regulation of cell migration” and “positive regulation of cell migration” in the top 5% of terms but did not have a similar association with trafficking terms as TGF-α cluster 1 in the RWHN output. Consistent with Francavilla *et al.*, we concluded that there was no association between trafficking and regulation of migration upon EGF stimulation(Francavilla *et al.*, 2016). Upon further investigation of this cluster, we found several sites that were shared with TGF-α cluster 1, such as the MAPK1/3/14, EGFR_S991, and EGFR_T693 sites. We also found the inhibitory RAF1 site S259 and SCRIB_1448. SCRIB has a putative role in cell migration (Dow *et al.*, 2008; Nola *et al.*, 2008).

TGF-α clusters 4 and 5 also highly rank terms related to intracellular trafficking. In the ORA analysis (Figure 2-6), cluster 4 was also enriched for “regulation of clathrin-dependent endocytosis”, corroborating this association. Of particular interest in cluster 4 are the EGFR (Y1110, Y1125, and Y1197), CBL (Y700), and CBLB (Y665, Y763, and Y889) sites. None of the

CBL/CBLB sites have roles described in PSP, but these proteins are known regulators of receptors in trafficking *via* clathrin (Pennock and Wang, 2008; Roepstorff *et al.*, 2009). Moreover, EGFR_Y1110 and EGFR_Y1197 are both annotated to receptor internalization in PSP and inducing cell motility. This is in line with "regulation of epithelial cell migration" highly ranked in this cluster (Figure 2-6). Taken together, these data indicate that TGF cluster 4 sites may have an initial role in regulating receptor internalization and migration, while the cluster 1 sites are regulating later parts of the process.

As a positive control for assignment of functions to clustered sites, sites annotated in PSP to the function "cell growth, inhibited" were added as a "spike-in" cluster to the EGF and TGF-α data, with sites already found in the data excluded from this new cluster. RWHN and ORA were then applied as done previously. We found that both RWHN and ORA could assign these sites to the function "negative regulation of cell growth" (Figure 2-7A). However, ORA assigned 134 and 122 other terms to the sites added to the EGF and TGF data, respectively, while RWHN assigned a refined list of 36 and 39, respectively, (Figure 2-7B). This result suggests that RWHN assigns more specific functions to each phosphorylated site. Indeed, for each of the investigated clusters, RWHN was able to assign sites belonging to those clusters to functions that closely resemble those that have been experimentally proven by Francavilla et al. or described in PSP. The ORA results, though not inaccurate, assigned more generic biological functions and did not assign specific terms. These results suggest that the RWHN algorithm can distinguish between sites of known functions and can assign functional terms to differentially regulated phosphorylation sites that could be further investigated experimentally.

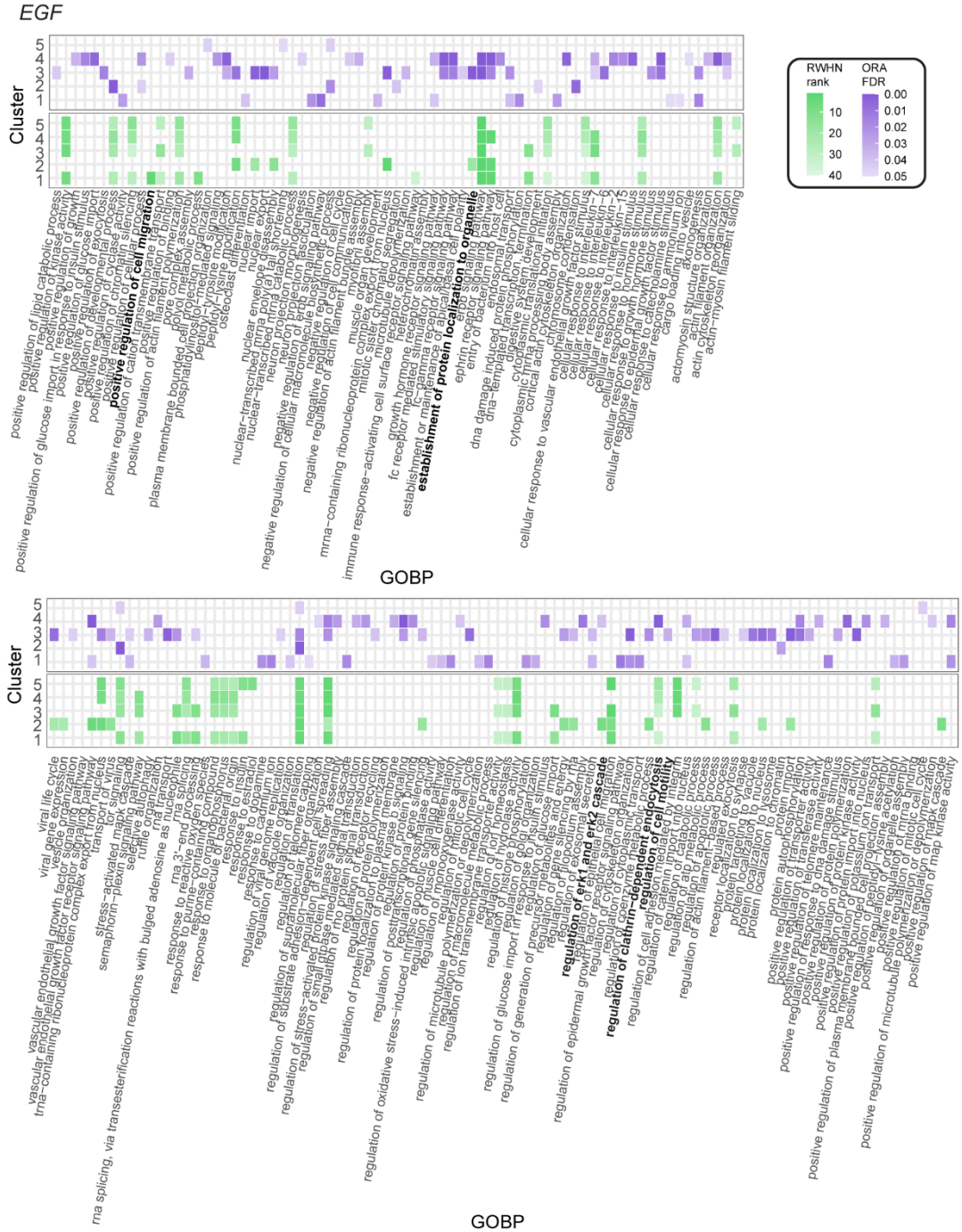


Figure 2-5. Site-centric and gene-centric analysis of EGF-regulated sites from Francavilla et al. (2016) dataset.

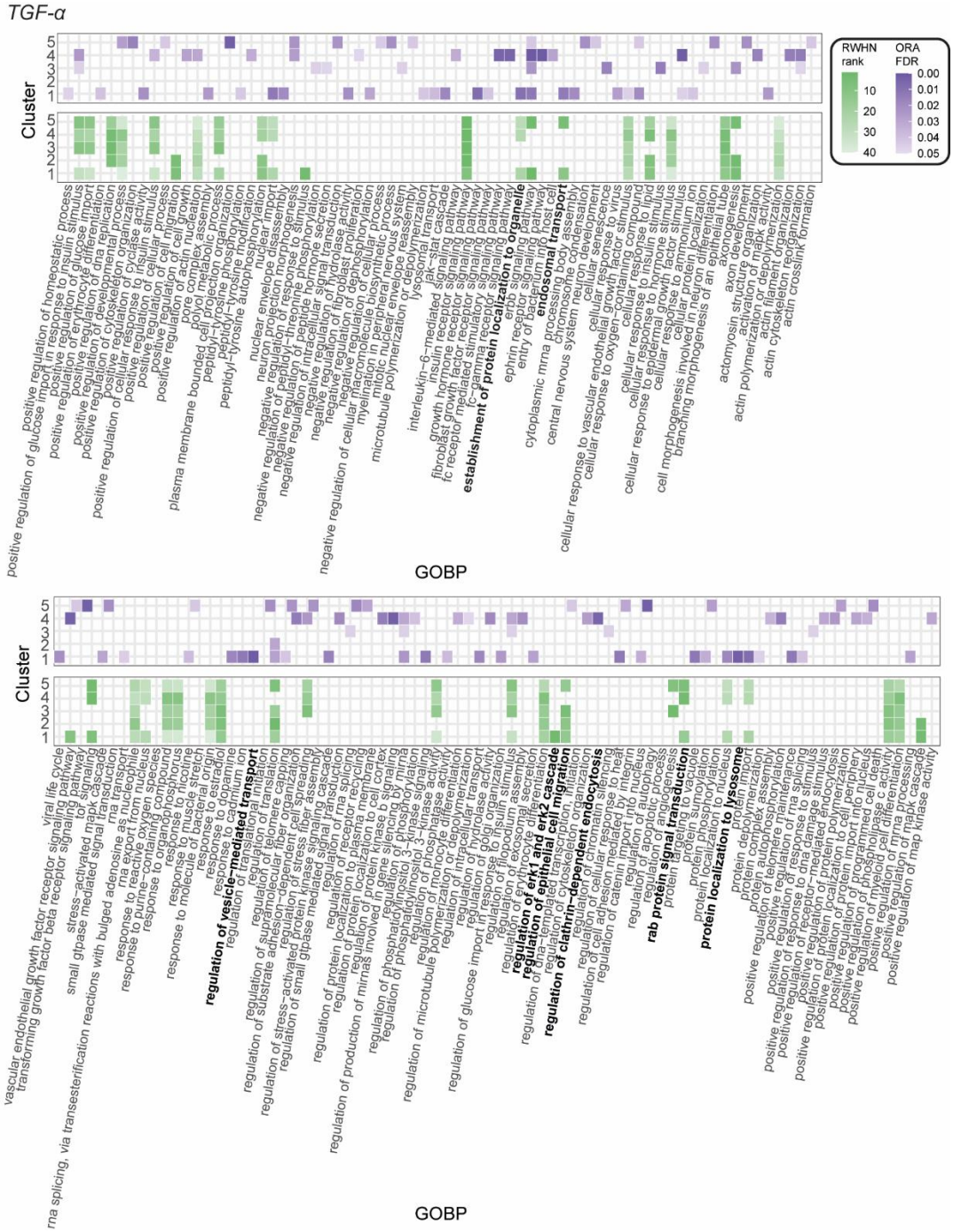
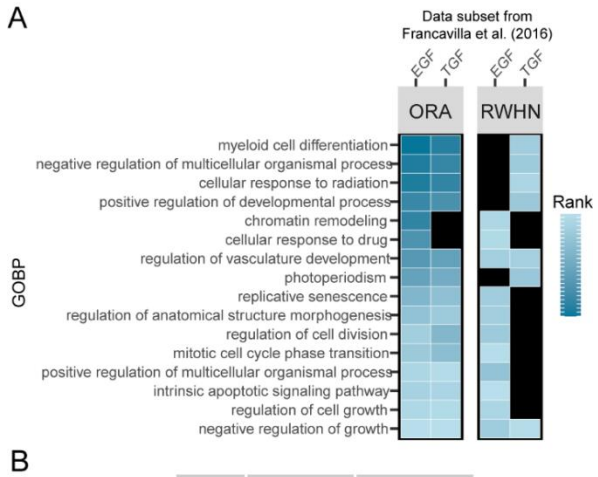


Figure 2-6. Site-centric and gene-centric analysis of TGF-α-regulated sites from Francavilla et al. (2016) dataset.



B

Data	Analysis	No. terms
EGF	ORA	134
TGF	ORA	122
EGF	RWHN	36
TGF	RWHN	29

Figure 2-7. Cluster containing sites annotated to "cell growth, inhibited" in PhosphoSitePlus, spiked into Francavilla et al. networks.
(A) Terms in common between ORA and RWHN results with both the EGF and TGF-α networks. (B) Total number of terms in the output of both types of analyses on both networks.

2.4.5 Experimental Case Study 2: Phosphorylation-Dependent Response to Lapatinib Treatment and Resistance in Breast Cancer

To verify the applicability of the algorithm with non-temporal shotgun phosphoproteomics data, we applied it to the experimental dataset from Ruprecht *et al.* (Ruprecht *et al.*, 2017), describing changes in the phosphoproteome upon treatment with the breast cancer drug lapatinib in sensitive ("parental" BT-474) or lapatinib-resistant (BT-474-J4) cells. The paper uncovered and experimentally validated the role of several metabolic enzymes and signaling pathways that were driving lapatinib resistance in breast cancer. In particular, proteins that form the spliceosome, those involved in glycolysis and glycogen catabolism, and PI3K/AKT/mTOR pathway members were differentially phosphorylated in lapatinib-resistant cells compared to those in parental lapatinib-treated cells.

We first constructed a multilayer heterogeneous network, as described in the Experimental Procedures, which included 1603 phosphorylated sites that showed significant changes in response to lapatinib treatment in either the parental or resistant cell lines (Table 2-4). Data were grouped into five clusters (referred to as clusters 1–5) using the k-means method. The number of clusters to be used was determined using the elbow plot method (Zambelli, 2016) (Appendix 1 Figure 4A), confirming that five clusters were sufficient to capture all features found in the data. Indeed, sites sharing similar regulation profiles between the resistant and parental cells were clustered together (Appendix 1 Figure 4B). As the original publication

used KEGG pathways rather than GO terms, we incorporated these in the functional annotation layer of our network to investigate the flexibility of our approach.

Table 2-4. Multilayer heterogeneous network, constructed from Ruprecht et al. lapatinib-regulated phosphorylated sites in parental and lapatinib-resistant cell lines. KEGG pathways were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	75688	1599
Protein-protein	12752	3129
Function-function	15	152
Site-protein	1599	1599 sites, 930 proteins
Protein-function	692	103 proteins, 152 pathways

The results of RWHN applied to the multilayer heterogeneous network (Figure 2-8A) show that when the seed nodes are set to sites belonging to cluster 4, representing sites that are less phosphorylated in lapatinib-resistant cells, the spliceosome pathway was highly ranked. This cluster interestingly contained signaling sites such as ERBB2_Y1233, ERBB2_T1227, ERBB2_Y1233, ERBB3_S627, and MAPK1_Y187, alongside sites on known spliceosome factors such as SRRM2 (S454 and S2449), SRSF6 (S301), and CWC25 (S170). Figure 2-8A shows that there was no enrichment for any pathways in this cluster using standard ORA. However, ORA did show enrichment for metabolic terms in several clusters, unlike RWHN, despite this being one of the clear conclusions of the original paper. Metabolic pathways were enriched in clusters 3 ("glycolysis/gluconeogenesis") and 5 ("central carbon metabolism in cancer") using ORA. Since much of the investigation and experimental validation by Ruprecht *et al.*(2017) was done on the sites found to be regulated only in lapatinib-treated resistant cells, we investigated whether our approach could uncover more nuanced roles of sites within these datasets. We constructed a second network using only those sites that were regulated in lapatinib-treated resistant cells (Table 2-5). These were grouped into four clusters (referred to as clusters 1–4) by k-means clustering as mentioned before (Appendix 1 Figure 4A,C), and KEGG pathways were used again in the functional annotation layer.

Table 2-5. Multilayer heterogeneous network, constructed from Ruprecht et al. lapatinib-regulated phosphorylated sites in lapatinib-resistant cell line. KEGG pathways were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	668491	1910
Protein-protein	14770	3880
Function-function	15	187
Site-protein	1910	1910 sites, 1087 proteins
Protein-function	874	96 proteins, 187 pathways

As described in Figure 2-8B, two sets of seed nodes (clusters 2 and 3, representing sites substantially upregulated and downregulated, respectively, Appendix 1 Figure 4C) tended to rank metabolic pathways more highly ("starch and sucrose metabolism", "galactose metabolism", "pentose phosphate pathway", "purine metabolism", and "amino sugar and nucleotide sugar metabolism"). Cluster 3, along with cluster 1 (representing moderately downregulated sites), ranked "spliceosome" highly. These initial interpretations indicate that using a specific subset of the regulated sites in the data yielded more specific and functionally relevant terms being ranked highly by the RWHN algorithm. Moreover, only cluster 2 showed enrichment for any metabolic pathways using ORA (Figure 2-8B). As this is a large cluster containing all the sites shown to be significantly upregulated in the resistant cell line, there is a considerable amount of "noise" in the ORA results. The enrichment of two metabolic pathways ("glycolysis/gluconeogenesis" and "central carbon metabolism in cancer") potentially corroborates the role of these sites as metabolism regulators, as indicated by RWHN. This cluster contained several of the sites classified as metabolic drivers of the resistance phenotype, such as ENO1_Y44, PKM_Y175, and PDHA_S293/S300. Intriguingly, S17 on the diverse signaling regulator SRC is also found in this cluster; this site is annotated to induction of enzymatic activity in PSP.

A closer look at cluster 3 uncovered sites on spliceosome factors (RBMX_S58, CDK13_S525, and CWC25_S17) and known signaling proteins (ERBB2_Y1233 and FGFR4_S505). A site on the adaptor protein IRS2 (S1176) was also found in this cluster, as a known player in glucose and lipid metabolism (Neukamm *et al.*, 2012). This points toward the link between rewired signaling and metabolism investigated by Ruprecht *et al.*(2017) It is also annotated to "metastatic potential" in PSP, indicating its therapeutic potential. ORA also showed enrichment for the "spliceosome" in cluster 1. Indeed, this cluster contained several of the sites highlighted in the paper. For example, the known lapatinib-responsive sites SF3B2_S309 and SF3B1_T261 could be found in this cluster, alongside 17 of the 30 SRRM2 sites that were significantly regulated by lapatinib.

Figure 2-8. Site-centric and gene-centric analysis of Ruprecht et al. datasets.
A. *RWHN and ORA results related to sites regulated by lapatinib in lapatinib-resistant and parental cell lines.* ***B****. RWHN and ORA results related to sites regulated by lapatinib in lapatinib-resistant cell lines only.*

From the comparison of ORA and RWHN methods, we have extracted meaningful associations between specific phosphorylated sites and their functions in driving lapatinib resistance. Our algorithm also highlighted potential further sites, such as IRS2_S1176, compared to the ORA method, which may serve as hubs or intersections of crosstalk between key signaling pathways and the rewired metabolic processes. It can be noted that the RWHN output was substantially more refined than the ORA output, with the conclusions of the original paper being captured in 14 terms in our analysis of the sites regulated in resistance, as opposed to 51 using ORA.

2.4.6 Experimental Case Study 3: Spatially Resolved Activity of HRAS

The final dataset we analyzed described immunoprecipitation (IP) samples, in order for us to see if our algorithm would be useful for analyzing local and global phosphoproteomes. The subcellular localization of signaling proteins has been established as crucial in regulating cascades and cellular processes. There are now numerous methods that allow experimental biologists to detect spatial effects of proteins and phosphorylation, for example, through proximity labeling and targeted proteomics (Gingras, Abe and Raught, 2019; Orre *et al.*, 2019). Recent work by Santra *et al.*(2019) took a similar approach to investigate the differential roles of a mutant, constitutively active form of HRAS, HRASV12, at different subcellular locations. HRASV12 constructs tagged with different signal peptides (targeting them to different subcellular locations) were stably transfected into HeLa cells; this allowed the authors to collect multi-omics samples with endogenous HRAS, unlocalized HRASV12 and HRASV12 localized to either the plasma membrane in disordered membrane regions (DM) or lipid rafts (LR) and the Golgi apparatus (GA) and the endoplasmic reticulum (ER). Using this approach, it was confirmed that the majority of HRASV12 activity is mediated from the plasma membrane, with ER- and GA-localized HRASV12 involved in regulating some of HRASV12's role in cell survival.

The phosphoproteomics data from this study were clustered using k-means clustering, with an optimal cluster number of six determined using an elbow plot (Appendix 1 Figure 5A,B). Table 2-6 describes the properties of the multilayer network constructed with these clusters. The first three clusters represented dynamics discussed in the original paper: cluster 1 represented phosphorylated events of mutant HRAS regardless of localization (location-independent); cluster 4 represented HRAS in the secretory pathway, at the ER and GA (ER/GA); and cluster 6 represented HRAS at the plasma membrane, in the DM or LR (DM/LR). As these experimental conditions were best characterized in the original paper, downstream analysis focused on these three clusters.

Table 2-6. Multilayer heterogeneous network constructed using regulated sites from the Santra et al. dataset. GOBP terms were included as functional annotations.

Subnetwork	Edges	Nodes
Site-site	766	621
Protein-protein	7262	2163
Function-function	370	452
Site-protein	621	621 sites, 474 proteins
Protein-function	900	122 proteins, 442 functions

Using RWHN (Figure 2-9, in purple), we found that the ER/GA and DM/LR clusters both highly ranked key signaling pathways such as “ERK1 and ERK2 cascades” and “TOR signaling”, alongside signaling regulation such as “receptor-mediated endocytosis”. This is consistent with the finding from the original paper that the majority of HRAS-mediated effects derive from the localized population. Specifically, the DM/LR cluster ranks mitogenic processes highly, such as “regulation of epithelial cell proliferation” and “positive regulation of metaphase/anaphase transition of the cell cycle”. The original paper also concluded that the majority of HRAS’ mitogenic effects were mediated from the plasma membrane. The ER/GA cluster independently ranks the “positive regulation of growth” and “Notch signaling pathway”. This potentially implicates that ER/GA localized HRAS, in the known developmental roles of HRAS in development; this was touched upon but not investigated in the original paper.

Conversely, when a term is ranked highly in all three clusters, this could indicate that this is part of the localization-independent regulation of HRAS. A total of 20 of the highly regulated terms fit this definition; of these, several are related to immunological processes, such as T-help cell lineage commitment” and “cellular response to interleukin-25”. The location-independent cluster also independently ranks the “interleukin-27-mediated signaling pathway” and “STAT cascade”. This suggests that HRAS’ role in regulating cytokine and immune response is not location-dependent.

An obvious outcome of the results using RWHN and ORA on these data is the difference in the number of terms associated to each cluster. There were 132 terms describing these clusters with ORA (Figure 2-9, in green), compared to 86 with RWHN, with many of the enriched terms related to the over-represented categories depicted in Figure 2-4, such as “gene expression”. Several comparisons can be drawn between the results of the two analyses to add credence to the accuracy of results from RWHN. For example, cluster 1 shows enrichment for the term “interleukin-6-mediated signaling pathway”, and cluster 3 is enriched for the term “regulation of the mitotic cell cycle”. These results from ORA are consistent with the results from RWHN and the original paper, however, with a significant amount of noise and redundancy in the ORA output relative to that of RWHN.

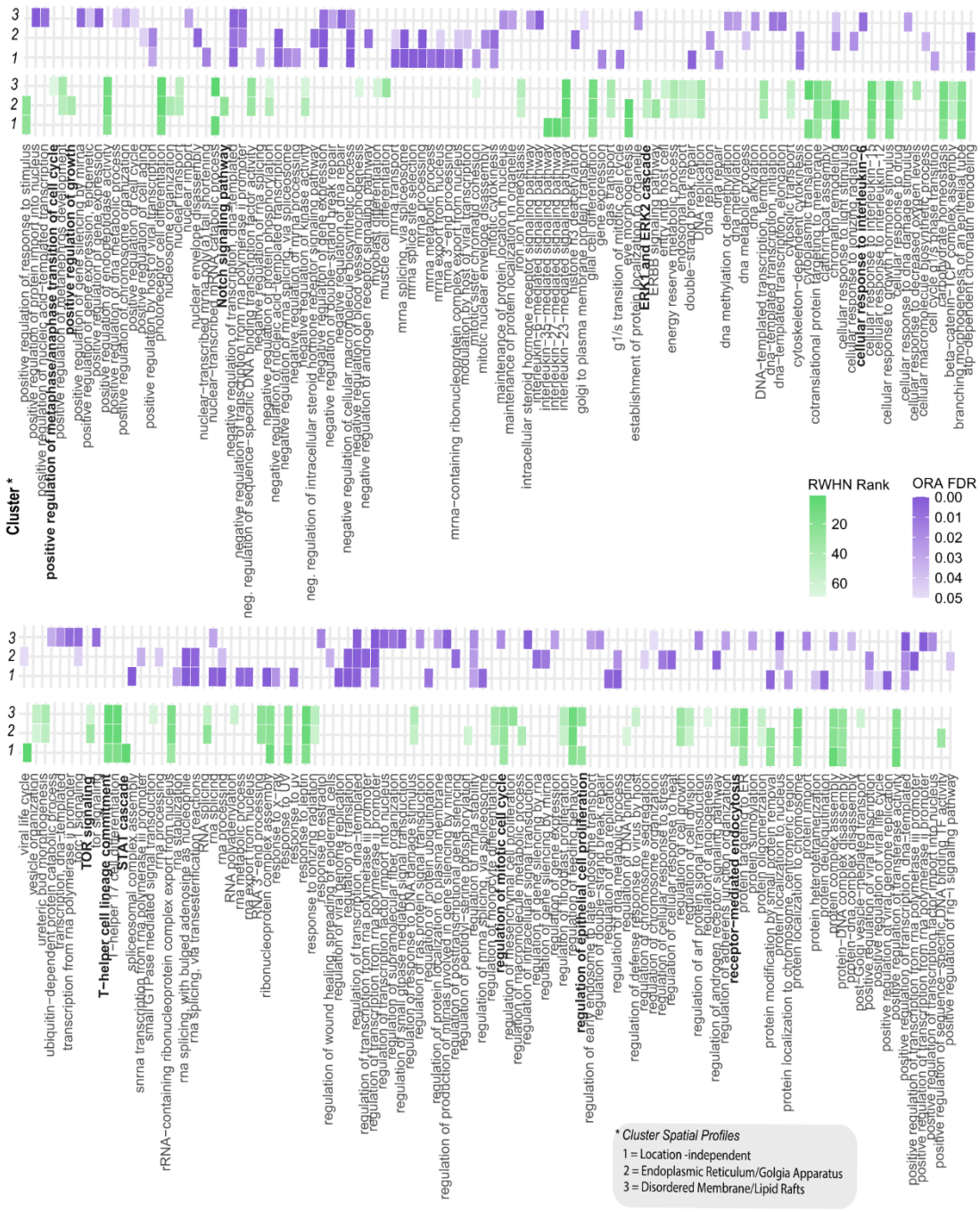


Figure 2-9. Site-centric and gene-centric analysis of Santra et al. datasets.
Results from RWHN and ORA are in green and purple, respectively. Clusters shown in the figure refer to phosphorylated sites that are location-independent or found at the endoplasmic reticulum (ER)/Golgi apparatus (GA) or disordered membrane (DM)/lipid rafts (LR) (marked in the figure as clusters 1, 2, and 3, respectively).

2.5 Discussion

Phosphorylation has an important impact on protein functions and thus the cellular behavior. Within a network of kinase–substrate interactions, phosphorylation modulates the flow of information and regulation of disparate processes throughout the cell.

Understanding the impact of phosphorylation on protein functions traditionally required the experimental investigation and manipulation of individual sites; however, with the advent of high-throughput phosphoproteomics, thousands of novel phosphorylation sites of unknown functions have been discovered. As there is limited experimentally validated information available on how regulation at individual phosphorylated sites impacts the downstream cellular output, current methods for analyzing high-throughput phosphoproteomics data rely on general descriptors of phosphorylated protein functions such as GO terms or involvement in known pathways. Their use in proteomics and transcriptomics analysis, although widespread, has come under recent criticism. Maertens *et al.* (2020) demonstrated that annotations are lacking for a substantial number of genes known to be important in cancer, while Haynes *et al.* (2018) demonstrate a "rich-getting-richer" effect, with well-studied genes having higher annotation rates despite having less omics or molecular evidence for the association with processes or diseases. These problems are exacerbated when used in analyzing phosphoproteomics data, as enrichment analyses also disregard site-specific or multiple-site regulation, masking the roles of proteins in a particular modification state (Krug *et al.*, 2019; Needham *et al.*, 2019).

Here, we developed and tested a method that associates phosphorylated sites to potential functions using RWHN. By incorporating the pattern of phosphorylation upon perturbation, we consider more of the information available in phosphoproteomics datasets and establish the phosphorylated sites as the key drivers in the functional analysis. The algorithm is capable of recapturing experimentally proven functions of phosphorylated sites in a non-gene-centric manner, refining analysis and extracting more information from phosphoproteomics data.

To prove the utility of our algorithm, we applied it to a validation network, which represented a simple model of phosphorylation dynamics in the MAPK/ERK pathway, and to two previously published phosphoproteomics dataset describing temporal cellular signaling events and the impact of resistance to drug treatment in breast cancer. Our algorithm successfully distinguished between differentially regulated phosphorylated sites from the same protein and associated them to both known and previously uninvestigated functions. For example, when applying this approach to the data from Francavilla *et al.* (2016) we can recapture the roles of EGFR sites with functions described in PSP. Both EGFR_Y1110 and EGFR_Y1197 are annotated to receptor internalization and cell motility in PSP; in our analysis, we found that these sites are both found in clusters that associate with "regulation

of clathrin-dependent endocytosis"; however, only the TGF-α cluster also ranked "regulation of cell migration" highly, reproducing the experimental evidence from Francavilla *et al.* While the authors of the original manuscript associated these sites with their functions using a combination of gene-centric approaches and experimental validation, here, we use a single site-centric algorithm to extract these associations directly from the phosphoproteomics dataset, demonstrating the power of our approach in quickly and more specifically narrowing down candidates for further functional studies. By associating ligand-dependent regulation of phosphorylated sites, it is possible to disentangle their multifaceted role in regulating cellular signaling networks and downstream cellular behaviors.

From the Ruprecht *et al.* data, we can use the example of SRRM2 to illustrate the validity of our approach. This component of the spliceosome is known to be highly phosphorylated, with 675 phosphorylation sites recorded in PSP. At the time of writing, none of these sites were ascribed to a function in the database, despite the importance of phosphorylation and dephosphorylation events in orchestrating splicing events (Naro and Sette, 2013). It is also regularly mutated (>5% of cases in TGCA) in lung, stomach, bladder, endometrial, and colorectal cancers (Hornbeck *et al.*, 2015). Ruprecht *et al.* uncovered a new role of the spliceosome in modulating lapatinib resistance, highlighting several sites on spliceosome proteins (including SRRM2_S1132, S1987, and S970) that were differentially modified in resistant and parental cell populations. Previous work has described how SRRM2 depletion in HER2-positive breast and ovarian cancer cells reduced the rate of migration (Mukherji *et al.*, 2006); the spliceosome plays myriad roles in the breast cancer environment, but site-specific analysis is lacking. Using our algorithm, we can begin to uncover the nuanced modification-specific roles of proteins such as SRRM2 in metabolic rewiring. We suggest that 11 of the SRRM2 sites found in cluster 2 of the lapatinib-resistant sites (Figure 2-8B) could be at the intersection between the spliceosome and metabolic processes; the conjunction between these processes was central to the work of Ruprecht *et al.* The sites in this cluster could be investigated as points of crosstalk between these different cellular processes. This example highlights the value of our approach in generating hypotheses of modified protein functions and signaling pathway crosstalk.

Although the associations predicted here between sites and functions cannot be interpreted as causative, phosphorylation sites can act as indirect and context-dependent regulators in cellular processes. This is demonstrated by the clustering of SRC_S17, a protein that is typically associated with signaling downstream of cell–surface receptors, with a cluster that

ranked metabolic pathways higher than cellular processing or signaling pathways. Critical to this is a robust method of defining clusters, to capture meaningful regulatory patterns at different phosphorylated sites. Here, we define clusters using established techniques (*e.g.*, elbow method for k-means) or visually identifying them with differing time-resolved behaviors. Using clusters to define edges between phosphorylated sites in the multilayer heterogeneous network, the pattern of experimentally perturbed modules in the cellular system can be investigated. Although having fixed edges between sites, as with the PPI layer of the multilayer network, would likely allow for a closer representation of the whole cellular system, this would not be computationally viable or in the interest of experimental biologists. This algorithm is targeted toward datasets that describe the effect of specific treatments or perturbations, allowing for a focused view on the impacted processes to identify candidates to investigate further. Quantitatively defining the edges between phosphorylated sites based on the dataset collected also prevents the issues plaguing standard enrichment analyses as discussed before; fixed edges, information provided in databases or previously published data, would almost certainly skew the analysis toward well-studied phosphorylated sites and biological processes.

Multilayer heterogeneous networks are increasingly being used to integrate omics data; here, their use allows phosphoproteomics data to be incorporated with PPI networks and functional annotations, overcoming the issue of considering phosphoproteomics data primarily on the protein level. A potential drawback of our approach is the reliance on large semi-curated resources such GO or STRING. For instance, different clusters may have many terms or pathways in common given the involvement of many proteins in the same biological functions and the high proportion of frequently used GO terms (Gaudet and Dessimoz, 2017). We theorize that this may be rectified if the data were clustered into more groups, in order to capture more nuanced phosphorylation patterns. However, enrichment of non-specific terms or false positives remains an issue when analyzing high-throughput omics data by other commonly used methods too. By only including edges in the protein layer of the network with a STRING experimental confidence score of greater than 0.4 and only including non-redundant terms in the function layer, we have reduced the noise in the resulting list of associated terms. This could be taken further by filtering for only high-confidence functional annotations or incorporating annotations from multiple sources into the network. Moreover, RWHN does not result in a different set of terms in the output compared to the accepted standard method ORA but rather provides a refinement that will allow for biological insights to be reached more clearly.

Our method is flexible enough to be used with any discovery phosphoproteomics data that describe a change between conditions. This is an improvement in the previously published use of RWHN using multiple sources of phosphoproteomics data to uncover disease-dependent regulation (Xu and Wang, 2016). Moreover, it could easily be generalized to any post-translational modification proteomics dataset, as it incorporates readily available PPI and functional annotation data, as demonstrated here. The fundamental aspect would be maintained with any of these expansions: specific patterns of regulation at modified sites dictate movement through the multilayer heterogeneous network.

2.6 Conclusions

We have proposed a site-centric approach to analyze phosphoproteomics data that provides a robust alternative to gene-centric methods of analysis. We integrated clustered quantitative phosphoproteomics data, a context-specific PPI network, and functional annotations into a multilayer heterogeneous network and used the RWHN method to predict the functions of phosphorylation sites with similar regulatory patterns. Using our algorithm, we extracted experimentally validated associations between phosphorylated sites and their role in cellular processes, which could not be captured using the typical gene-centric methods. Moreover, our algorithm has the potential to be used by researchers in predicting novel site–function associations and generating hypotheses to be experimentally validated.

3 Spatially Resolved Phosphoproteomics Reveals FGFR2b Recycling-driven Regulation of Autophagy and Cell Survival

3.1 Abstract

The precise spatiotemporal regulation of Receptor Tyrosine Kinase (RTK) signalling during endocytosis, specifically during recycling to the plasma membrane, remains poorly elucidated. RTK recycling and signalling drive cell proliferation and motility during development, in homeostasis, and are dysregulated in disease, such as cancer. Focusing on Fibroblast Growth Factor Receptor 2b (FGFR2b) recycling, we generated a signalling signature in proximity of the recycling endosomes (REs) by developing a sensitive and Spatially Resolved Phosphoproteomics (SRP) approach based on ascorbate peroxidase APEX2-driven biotinylation. Combining this approach with traditional phosphoproteomics, bioinformatics, and targeted assays, we uncovered novel FGFR2b signalling partners at the REs. FGF10-stimulated FGFR2b activated mTOR-dependent signalling when localized at the REs during recycling. By regulating mTOR and ULK1 from the REs, FGFR2b supressed autophagy and increased cell survival in response to FGF10, but not to FGF7. Therefore, SRP and our systems biology approach identifies REs-specific signalling partners that adds to the growing function of RTK trafficking in orchestrating cellular fate and suggests a therapeutically targetable vulnerability in ligand-responsive cancer cells

3.2 Introduction

Endocytosis is the process by which surface molecules, including Receptor Tyrosine Kinases (RTKs), undergo internalization from the plasma membrane (PM). RTKs are internalized into the early endosome (EEs) within seconds of ligand binding, followed by sorting either to the lysosome via the late endosome (LEs) for degradation, or to the recycling endosome (REs) for recycling back to the PM (Miaczynska, 2013; Stasyk and Huber, 2016; Schmid, 2017; Cullen and Steinberg, 2018; Naslavsky and Caplan, 2018; O'Sullivan and Lindsay, 2020; Sigismund *et al.*, 2021). Recycling is not merely a process happening by default after receptor internalization to control receptor availability at the cell surface (Hsu, Bai and Li, 2012). Rather, recycling actively regulates signalling duration and signalling outputs (Miaczynska, 2013; Stasyk and Huber, 2016; Schmid, 2017; Cullen and Steinberg, 2018; Naslavsky and Caplan, 2018; O'Sullivan and Lindsay, 2020; Sigismund *et al.*, 2021). For instance, we and

others have established a link between the recycling of RTKs, such as Fibroblast and Epidermal Growth Factor Receptors (FGFR, EGFR) or integrin and the sustained signalling activation that regulates cell motility (Caswell *et al.*, 2008; Francavilla *et al.*, 2013, 2016). Thus, recycling and its known regulators (e.g. RAB11, TTP, RCP) (Francavilla *et al.*, 2013, 2016) maintain homeostasis in health and their dysregulation leads to several diseases, including cancer, diabetes, and neurodegeneration (Lanzetti and Di Fiore, 2017; O'Sullivan and Lindsay, 2020; Yarwood *et al.*, 2020). However, we still do not know which RTK signalling partners are recruited to the REs during recycling and thereby how they modulate downstream cellular responses.

The view of endocytosis as a way to attenuate signalling by receptor down-modulation and by controlling receptor availability at the cell surface has been challenged by data linking endocytosis to the propagation of RTK signalling from endosomes (Miaczynska, 2013; Schmid, 2017; Naslavsky and Caplan, 2018). For instance, EGFR signalling from EEs leads to AKT phosphorylation and cell survival (Wang *et al.*, 2002), suggesting that EGFR internalization is required for the full spectrum of signalling activation. The recruitment of specific scaffold proteins to EEs regulates a certain branch of signalling, exemplified by the p14/MP1 complex engaging the kinase ERK (Teis, Wunderlich and Huber, 2002; Brüggemann *et al.*, 2021). Another example of the crucial role of endosomes as signalling regulators is the recruitment of the LAMTOR complex to the lysosomes, which regulates the kinase mTOR in response to nutrients and growth factors with consequences for signalling, cell growth, metabolism, and autophagy (Saxton and Sabatini, 2017). Much less is however known about the role of REs as signalling regulators compared to EEs or lysosomes (Savini, Zhao and Wang, 2019; Yuan and Song, 2020). Knowing this would allow modulating signalling responses in a specific manner. For instance, depleting the recycling adaptor RCP in cancer cells which switches EGFR trafficking from recycling to degradation decreases cell proliferation and migration (Francavilla *et al.*, 2016). Recently, we found a reciprocal regulation between FGFR2b and EGFR signalling outputs from the REs, which lead to FGFR2b-dependent phosphorylation of EGFR on threonine 696 (T693) and of the cell cycle regulator CDK1 on T161 leading to cell cycle progression (Smith *et al.*, 2021). This data suggests that the REs can integrate and propagate signals, prompting us to further investigate here FGFR2b signals at the REs.

The FGFR family is a useful system for studying the contribution of trafficking to signalling outputs (Szybowska *et al.*, 2021). There are four FGFRs, with FGFR1-3 having splice-variants

denoted as b and c isoforms, and 21 FGF ligands, with each FGFR/FGF pair regulating signalling specificity in a context-dependent manner during development, in adult homeostasis, and in several diseases such as cancer (Ornitz and Itoh, 2015). One stark example of such functional selectivity is given by FGFR2b which is expressed on epithelial cells (Ornitz and Itoh, 2015). Stimulation of FGFR2b with its exclusive ligand FGF7 induced receptor degradation in contrast to stimulation with FGF10 which resulted in recycling of FGFR2b via the REs (Belleudi *et al.*, 2007; Francavilla *et al.*, 2013; Smith *et al.*, 2021). These two different trafficking routes of FGFR2b were associated with different phosphorylation dynamics within the signalling cascade and an increase in cell proliferation and proliferation/migration, respectively (Francavilla *et al.*, 2013; Smith *et al.*, 2021). Therefore, the duration and location of FGFR signalling must be strictly regulated to modulate the right cellular outputs (Ferguson, Smith and Francavilla, 2021; Szybowska *et al.*, 2021).

Here, to investigate FGFR2b signalling at the REs, we developed a Mass Spectrometry (MS)-based phosphoproteomics approach which allowed us to distinguish sites globally phosphorylated upon FGF10 binding to FGFR2b from sites phosphorylated in the proximity of the REs during receptor recycling. This Spatially Resolved Phosphoproteomics (SRP) approach is based on proximity-dependent biotinylation, which has been recently developed to profile the interactome of internalized receptors such as G-protein coupled receptors (GPCRs) (Lobingier *et al.*, 2017; Paek *et al.*, 2017). Proximity-dependent biotinylation occurs when a bait protein, tagged with a biotin ligase such as BioID or a peroxidase such as APEX2, encounters other proteins within the labelling radius of 10 nm or 20 nm, respectively (Han *et al.*, 2017, 2019). Combined with biotin enrichment using streptavidin beads and MS analysis, interactomes of bait proteins can be identified even for all subcellular compartments (Gingras, Abe and Raught, 2019; Go *et al.*, 2021). Using peroxidases is preferable for investigating short-acting, dynamic processes, due to their short labelling time of 1 minute. APEX2-based approach was successfully used to define the interactors of GPCRs and selective autophagy receptor-dependent cargo (Lobingier *et al.*, 2017; Zellner, Schifferer and Behrends, 2021) and was therefore also our choice. However, we expanded the method by adding a phosphorylated peptide enrichment step after biotin enrichment using streptavidin beads, which allowed us to uncover FGFR2b signalling partners localized at the REs and to study their impact on FGFR2b responses. To dissect the spatially restricted signalling modules regulated by FGF10/FGFR2b during recycling, we combined the SRP approach with traditional quantitative phosphoproteomics of epithelial cells expressing non-functional variants of the PM-fission protein dynamin-2 (DNM2) or of the recycling regulator

RAB11 (RAB11), which maintained FGFR2b at the PM or in the EEs, respectively (Smith *et al.*, 2021). We found that localized FGFR2b signalling at the REs during recycling regulates mTOR signalling with functional consequences on autophagy and cell survival.

3.3 Results

3.3.1 Inhibiting FGFR2b trafficking alters the FGF10 phosphoproteome in epithelial cells

To investigate FGFR2b recycling-dependent signalling in response to the recycling ligand FGF10, we transiently expressed Dynamin_K44A-eGFP (dominant negative Dynamin, GFP-DnDNM2) or eGFP-RAB11_S25N (dominant negative RAB11, GFP-DnRAB11), which maintained FGFR2b at the PM or in the EEs, respectively (Smith et al., 2021), in the HeLa epithelial cell line stably expressing FGFR2b (HeLa_FGFR2bST) (Francavilla *et al.*, 2009, 2013). The expression of DnDNM2 and DnRAB11 inhibited FGFR2b internalization and recycling to the PM in cells treated with FGF10 for 40 min, a time point where FGFR2b localized in the REs in cells expressing wild-type e-GFP-RAB11a (GFP-RAB11) (Figure 3-1)(Francavilla *et al.*, 2013; Smith *et al.*, 2021). This finding confirms what we previously observed for endogenous FGFR2b in the T47D epithelial breast cancer cell line (Smith *et al.*, 2021) and for FGFR1 in HeLa cells (Francavilla *et al.*, 2009). Immunoblot analysis of HeLa, T47D, and BT20 (another breast cancer cell line expressing endogenous FGFR2b (Smith *et al.*, 2021)) in the same experimental conditions showed that impeding FGFR2b trafficking did not alter FGFR2b activation or the phosphorylation of ERK1/2 downstream of FGF10 (Figure 3-2, Appendix 2 Figure 1A) (Francavilla *et al.*, 2013; Ornitz and Itoh, 2015). Therefore, we used a traditional Mass Spectrometry (MS)-based quantitative phosphoproteomics approach to comprehensively investigate changes in FGFR2b signalling when FGF10-dependent FGFR2b trafficking was impaired. We analysed HeLa cells transiently expressing FGFR2b and either eGFP (as control, HeLa-FGFR2b GFP), DnRAB11 (HeLa-FGFR2b DnRAB11) or Dn-DNM2 (HeLa-FGFR2b DnDNM2) and stimulated with FGF10 for 40 min (Figure 3-3A). Firstly, we checked that the transient expression of dominant negative proteins did not alter the HeLa proteome using Pearson correlation, which was indeed high among all the experimental conditions (Appendix 2 Figure 1B-C). The quality of the 6000 identified phosphorylated sites was consistent with previous publications (Smith *et al.*, 2021) (Appendix 2 Figure 1D-H). Principal component analysis (PCA) separated the MS runs based on the experimental conditions (Figure 3-3B), suggesting that the location of FGFR2b during FGF10-dependent FGFR2b trafficking affects signalling activation. To characterize this, we utilised Fuzzy c-means

clustering of phosphorylated sites significantly dysregulated across the four conditions (ANOVA, *p*-value < 0.001) and identified 11 clusters (Figure 3-3C, Figure 3-4D).

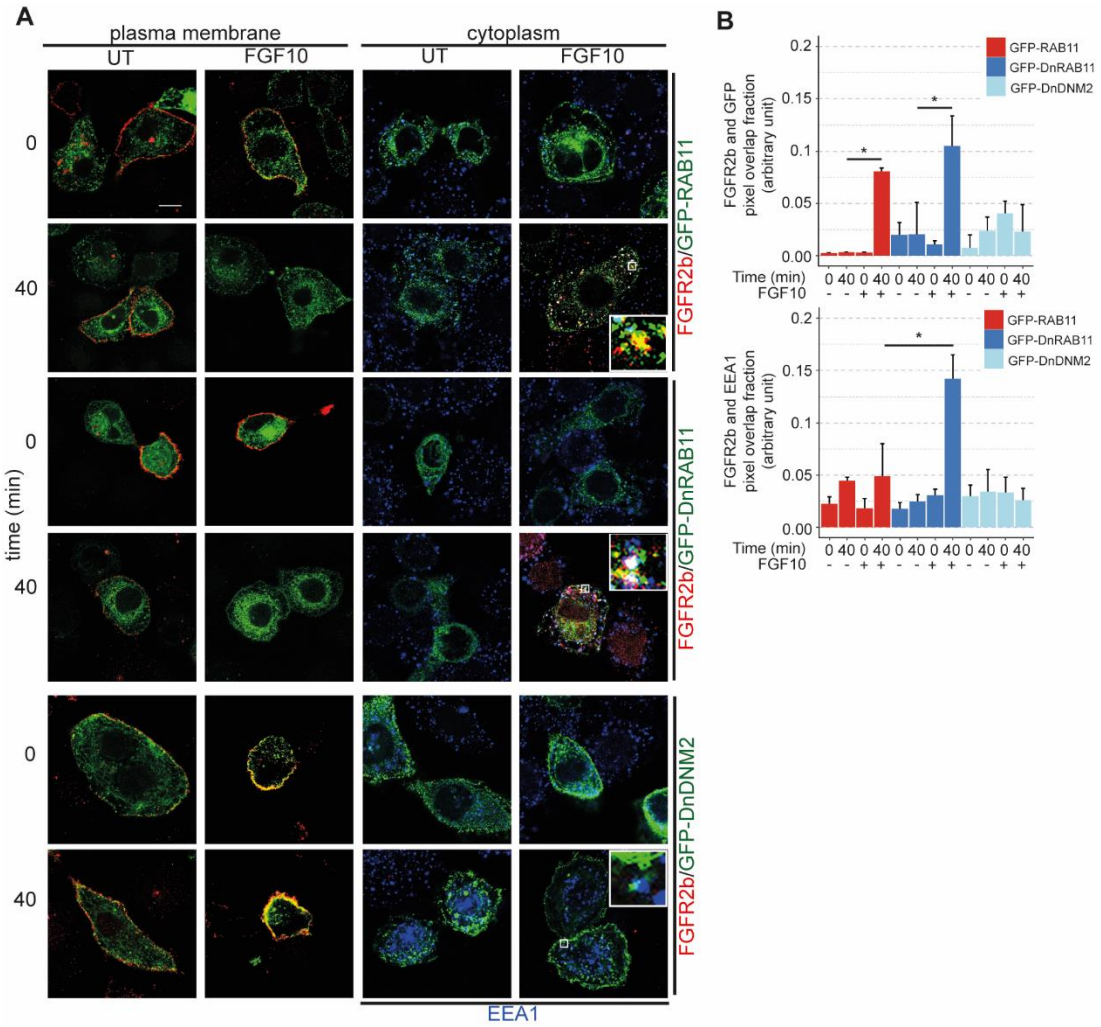


Figure 3-1. Analysis of FGFR2b trafficking reveals FGFR2b activation is not affected by receptor sub-cellular localization.
A. *Assessment of FGFR2b (red) internalization (cytoplasm) and FGFR2b recycling (plasma membrane) in HeLa cells stably transfected with FGFR2b-HA (HeLa_FGFR2bST), expressing eGFP-RAB11a (GFP-RAB11), dominant negative eGFP-RAB11a_S25N (GFP-DnRAB11), or dominant negative dynamin_K44A-eGFP (GFP-DnDNM2) (green, and left either untreated (UT) or treated with FGF10 for 0 and 40 min. Early endosome antigen 1 (EEA1) (blue) is a marker for EEs (Smith* et al.*, 2021). Scale bar, 5μm.* ***B.*** *Co-localization of FGFR2b (red pixels) with GFP-tagged proteins (green pixels) (top panel) indicated by red-green pixel overlap. Co-localization of FGFR2b (red pixels) with EEA1 (blue pixels) indicated by red-blue pixel overlap (bottom panel). Representative images are shown in 1A. Values represent median ± SD from N=3; * p value < 0.005 (students t-test) (Smith* et al.*, 2021).*

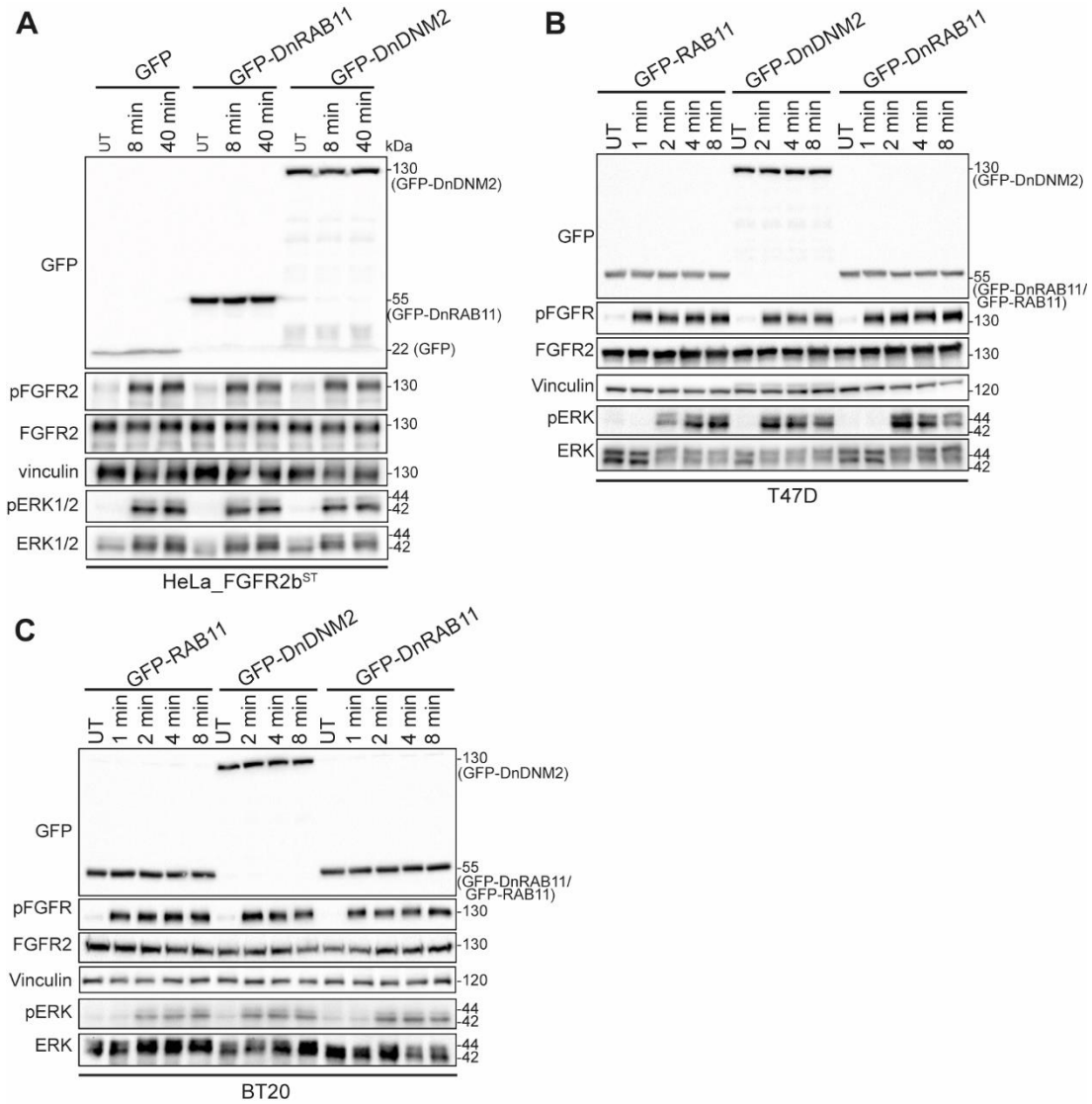


Figure 3-2. Immunoblot analysis reveals FGFR2b activation is not affected by receptor sub-cellular localization.
*Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST cells expressing GFP, GFP-DnRAB11 or GFP-DnDNM2 and left untreated (UT) or treated with FGF10 for 8 and 40 minutes (**A**), T47D (**B**) or BT20 (**C**) breast cancer cells expressing endogenous FGFR2b and transfected with GFP-RAB11, GFP-DnRAB11 or GFP-DnDNM2, and either left untreated or treated with FGF10 for 1, 2, 4 and 8 minutes*

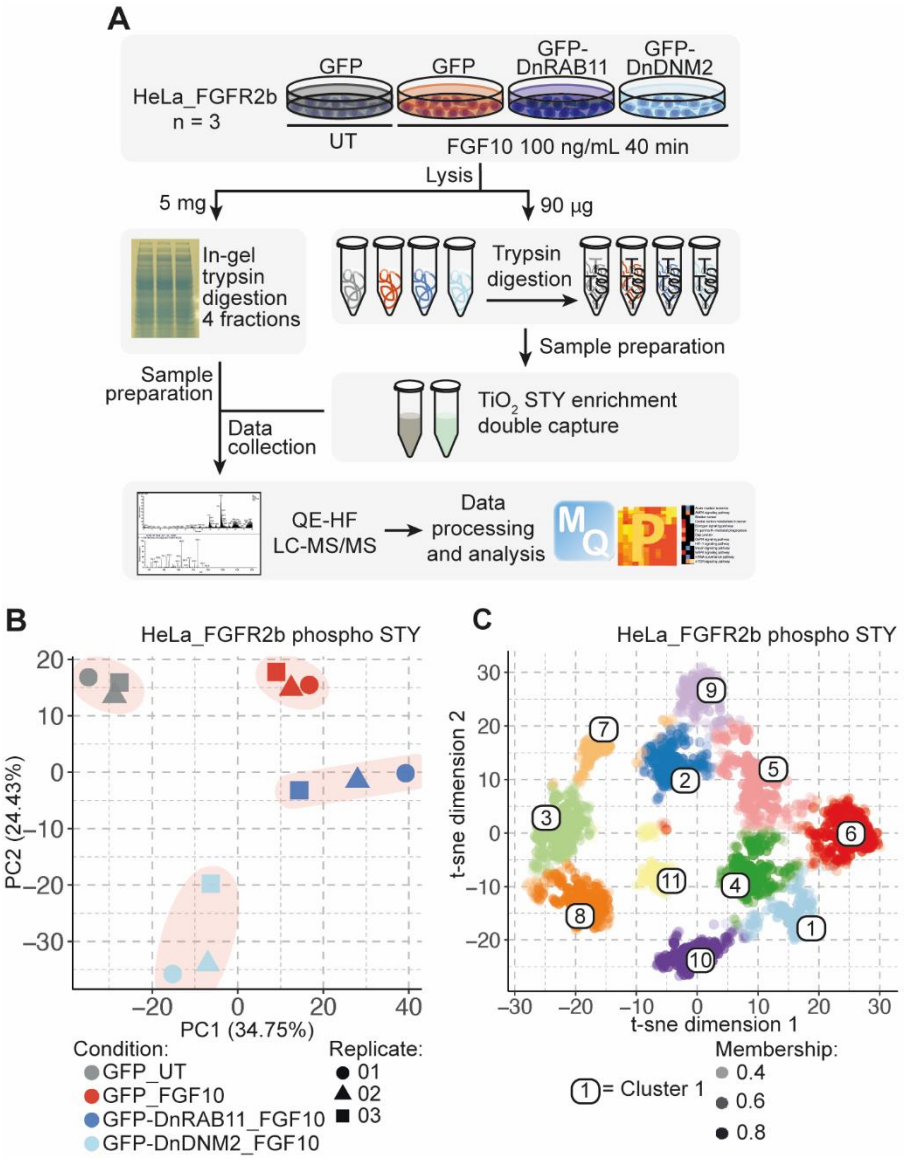


Figure 3-3. Phosphoproteomics analysis identifies FGFR2b internalization- and recycling-dependent signalling pathways.
A. *Workflow of the phosphoproteomics experiment in HeLa cells transiently expressing FGFR2b (HeLa-FGFR2b) and either GFP, GFP-DnRAB11 or GFP-DnDNM2 and left untreated (UT) or treated with FGF10 for 40 min.* ***B.*** *Principal Component Analysis (PCA) of the HeLa phosphoproteome from 2A showed small variation between technical replicates and separated samples based on experimental conditions.* ***C.*** *t-distributed stochastic neighbour embedding (t-sne) analysis identified 11 clusters corresponding to phosphorylated peptides differentially regulated among four conditions. Colour corresponds to the cluster with highest membership score, determined using fuzzy c-means clustering based on the median z-score of the four conditions. Each cluster is identified by a unique colour and corresponding number. Opacity corresponds to the membership score assigned to each phosphorylated site within its most likely cluster.*

We focused on phosphorylated sites either regulated by FGF10 but not affected by the expression of DnRAB11 or DnDNM2 (clusters 3 and 4), regulated by FGF10 but dysregulated by the expression of DnDNM2 (clusters 5 and 8), or regulated by FGF10 but dysregulated by the expression of both DnRAB11 and DnDNM2 (clusters 9 and 11), hereby defined as the membrane, the internalization, and the recycling response clusters, respectively (Figure 3-4A). We then checked the relevance of these responses in T47D, which express endogenous FGFR2b (Smith *et al.*, 2021) We analysed the proteome and the phosphoproteome of FGF10-stimulated T47D transiently expressing RAB11, DnRAB11, or DnDNM2 (Appendix 2 Figure 2A,B). The quality controls for the T47D datasets were in line with previous findings (Smith *et al.*, 2021) (Appendix 2 Figure 2C-I), and the analysis of the phosphoproteome showed a T47D-specific pattern. However, when we used the 11 clusters identified in the HeLa-FGFR2b phosphoproteome corresponding to the membrane, internalization, and recycling responses to extract phosphorylated sites with corresponding patterns of regulation in the T47D phosphoproteome, we found similar responses in the two cell lines (Figure 3-4A-B). Over-representation analysis (ORA) of KEGG pathways identified mTOR signalling as a pathway enriched for proteins dysregulated in the recycling response common to both HeLa-FGFR2b and T47D cells (Figure 3-4C). However, we did not find any signalling pathways specifically enriched upon inhibition of FGFR2b internalization (cells expressing DnDNM2 and stimulated with FGF10) (Figure 3-4C). This suggests that the FGFR2b recycling route and not merely the FGFR2b presence in the cytoplasm regulates FGFR2b signalling. We analysed the subcellular localization of proteins in the mTOR signalling pathway belonging to the recycling response cluster and surprisingly found an enrichment not only for vesicles and the REs, but also for the late endosome in both cell lines (Figure 3-4D). Altogether, this data confirms that recycling is crucial for FGFR2b signalling regulation (Smith *et al.*, 2021) and identifies the mTOR pathway as specifically enriched downstream of FGFR2b recycling. However, this approach did not reveal FGFR2b signalling partner was recruited to and specifically phosphorylated at the REs.

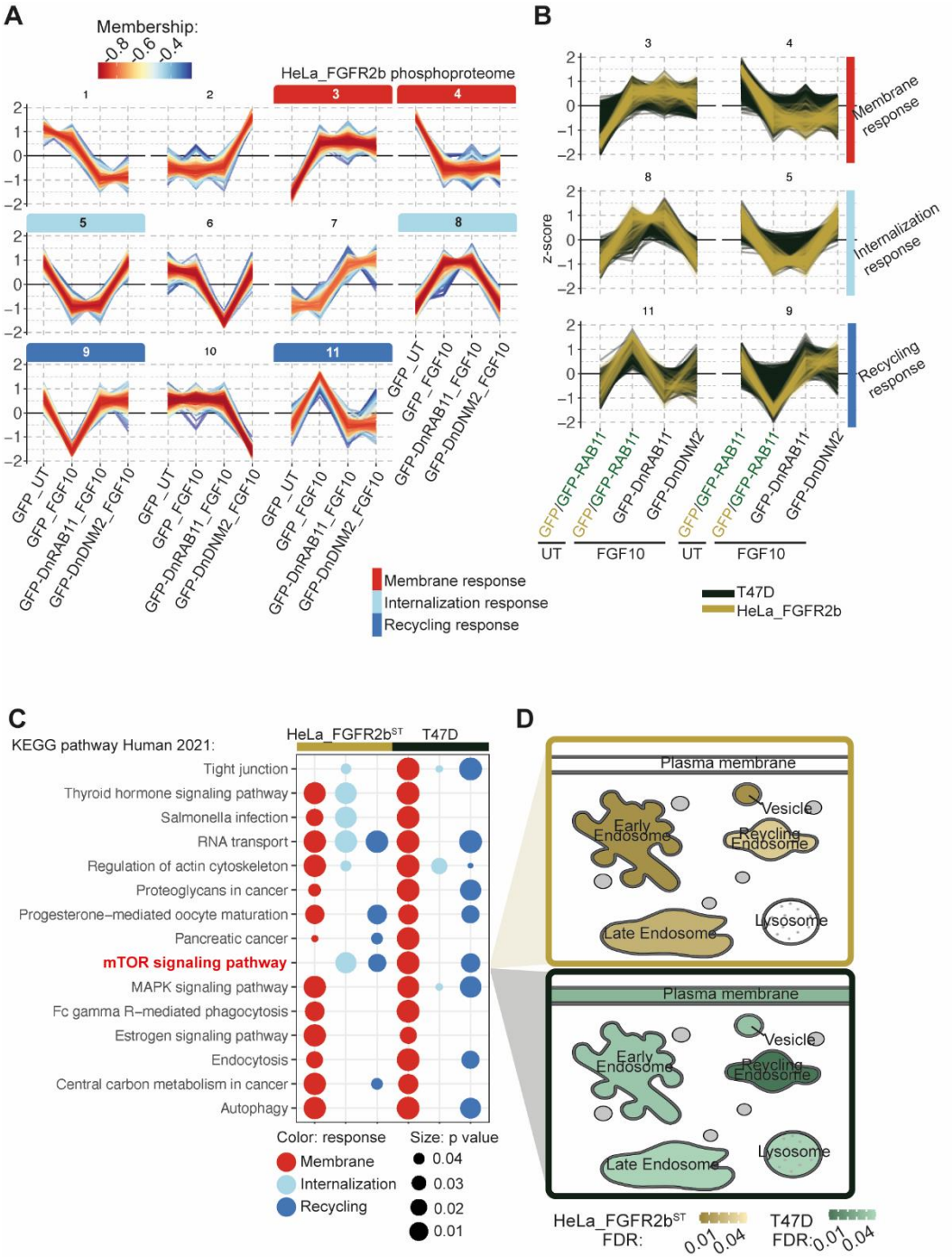


Figure 3-4. Characterisation of FGFR2b internalization- and recycling- signatures in HeLa_FGFR2b and T47D.

A. Plots of the median z-scored intensities of phosphorylated sites based on the 11 clusters from Figure 3-3C identified membrane response (red; clusters 3 and 4), internalization response (light blue; clusters 5 and 8), and recycling response (dark blue; clusters 9 and 11). Colour key indicates membership value assigned by Fuzzy c-means clustering. B. Plots of the median z-scored intensities of phosphorylated sites based on the three main clusters identified in A. HeLa_FGFR2bST (yellow) and T47D (dark green) were treated as in Figure 3-3A and Appendix 2 Figure 1J, respectively. C. KEGG pathway over representation analysis (ORA) between HeLa_FGFR2bST and T47D phosphorylated proteins within the membrane (red), internalization (light blue) and recycling (dark blue) responses (A) identified mTOR signalling as associated with FGFR2b recycling. The size of dot indicates statistical significance based

on p-value. D. Visualisation of the sub-cellular localization of the phosphorylated proteins belonging to the "mTOR signalling pathway" KEGG term in HeLa_FGFR2bST (yellow) and T47D (dark green) using SubCellularVis.

3.3.2 APEX2-based proximity biotinylation assay enriches for phosphorylated proteins at the RE

To investigate recycling-dependent FGFR2b signalling in a spatially resolved (at the REs) and temporally sensitive (40 min simulation with FGF10) manner, we used the APEX2-based proximity labelling system (Gingras, Abe and Raught, 2019) (Figure 3-5A). The APEX2 system involves the addition of a 27 kDa enzyme to a bait protein (FGFR2b-HA, eGFP-RAB11, and eGFP in this study) to biotin-label proteins within 20 nm of the bait protein in less than one minute following addition of biotin-phenol (BP) and hydrogen peroxidase (H_2O_2) as an oxidant (Han *et al.*, 2019). Biotinylated proteins can then be pulled-down using streptavidin beads and analysed using MS-based proteomics (Gingras, Abe and Raught, 2019), as recently shown for different sub-cellular compartments using known localized bait proteins (Smith *et al.*, 2021). Therefore, we designed an APEX2-based experiment to investigate events downstream of FGF10-dependent FGFR2b recycling localized at the RAB11a-positive REs. We stably transfected HeLa or T47D cells with FGFR2b-HA-APEX2 (HeLa_FGFR2b-APEX2ST, T47D_FGFR2^{KO}_FGFR2b-APEX2) and verified that FGFR2b signalling and trafficking were not altered by the presence of APEX2 upon FGF10 stimulation over time (Figure 3-5B, Figure 3-6, Appendix 2 Figure 3A). As controls, we expressed either eGFP-RAB11-APEX2 or eGFP-APEX2 in HeLa-FGFR2bST (HeLa-FGFR2bST_RAB11-APEX2; HeLa-FGFR2bST_GFP-APEX2) to exclude events occurring at the REs and in the cytoplasm, respectively, independent of FGFR2b activation (Figure 3-5A, Appendix 2 Figure 3B). For all three APEX2-tagged bait proteins, biotin-phenol treatment for 40 minutes and H2O2 incubation for 1 minute followed by streptavidin-bead pulldown (hereby pulldown) enriched for the bait, biotinylated proteins and phosphorylated/biotinylated proteins (Figure 3-7A, Appendix 2 Figure 3C-E). Known interactors of FGFR2b, such as PLCγ and FRS2, following 1 and 8 min FGF10 treatment, but not histone (H3), were identified after the pulldown without changing their basal-level activation (Figure 3-7A, Appendix 2 Figure 3B). Finally, to confirm that RAB11-APEX2 successfully enriched biotinylated proteins in the proximity of the REs during FGFR2b recycling, we immunoblotted the RAB11-APEX2 pulldowns from cells stimulated with FGF10 for 40 min for RAB11 (using the GFP antibody), known markers of the REs (e.g. RAB25), and for FGFR2b (using the anti-HA antibody) (Figure 3-7B,C). RAB11-APEX2 also associated with other known markers of recycling, including RCP (Francavilla *et al.*, 2016) (Appendix 2 Figure 3F), confirming the ability of our approach to detect proteins in the proximity of the REs.

Taken together, this data supports the use of APEX2 as a system to reveal signalling partners downstream of FGF10-stimulated FGFR2b at the REs.

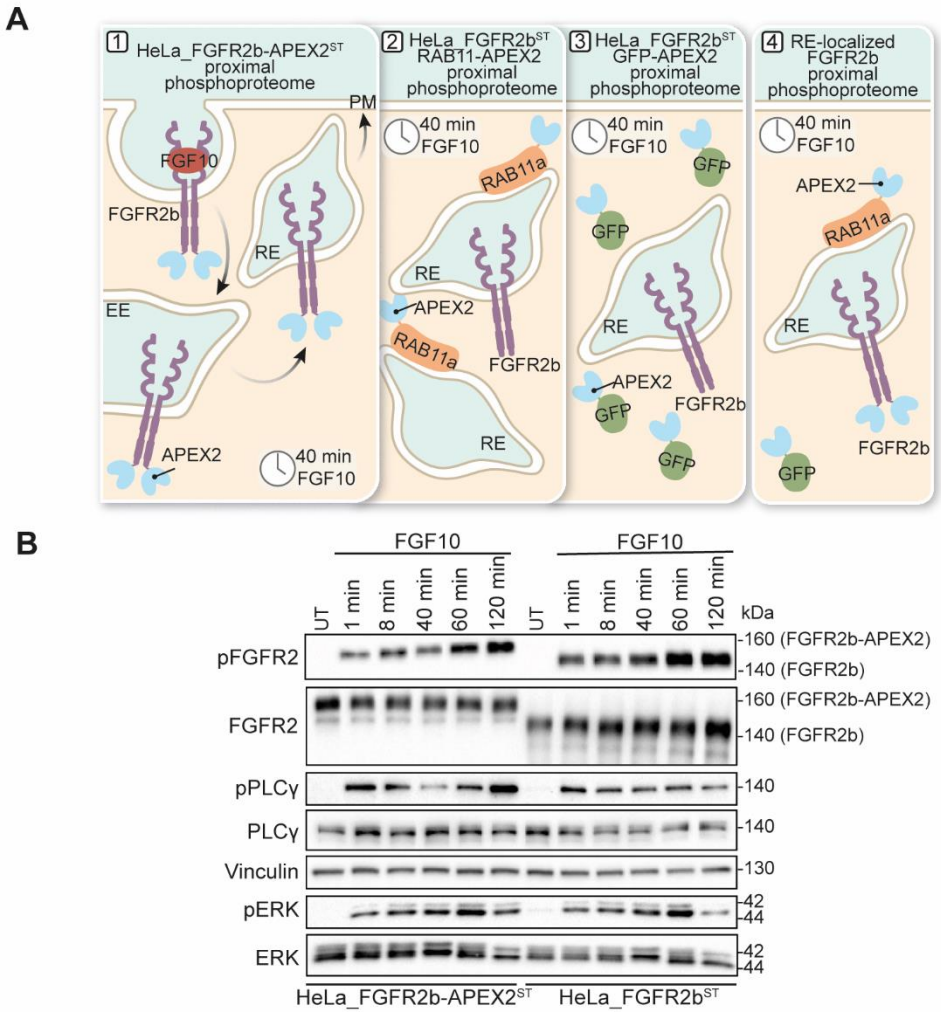


Figure 3-5. Design of spatially resolved phosphoproteome (SRP) experiment to elucidate FGFR2b and RAB11a proximal interactors.
A. *Schematic underlying the Spatially Resolved Phosphoproteomics (SRP) approach. Panel 1 represents the trafficking of FGFR2b-APEX2 stimulated with FGF10 in HeLa_FGFR2b-APEX2ST and subsequent FGFR2b-APEX2 proximal phosphoproteome; panels 2 and 3 represent the localization of FGFR2b and of the APEX2-tagged proteins in cells expressing either FGFR2b and Rab11-APEX2 (HeLa_FGFR2bST RAB11-APEX2) or FGFR2b and GFP-APEX2 (HeLa_FGFR2bST GFP-APEX2) stimulated for 40 min with FGF10, and the proximal phosphoproteomes to the bait. Panel 4 represents the phosphorylated events occurring at the RAB11- and FGFR2b-positive REs upon 40 min FGF10 stimulation after subtracting cytosolic events using HeLa_FGFR2bSTGFP-APEX2 phosphoproteome.* ***B.*** *Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST (right) or HeLa_FGFR2b-APEX2ST (left) stimulated with FGF10 for 1, 8, 40, 60, or 120 min.*

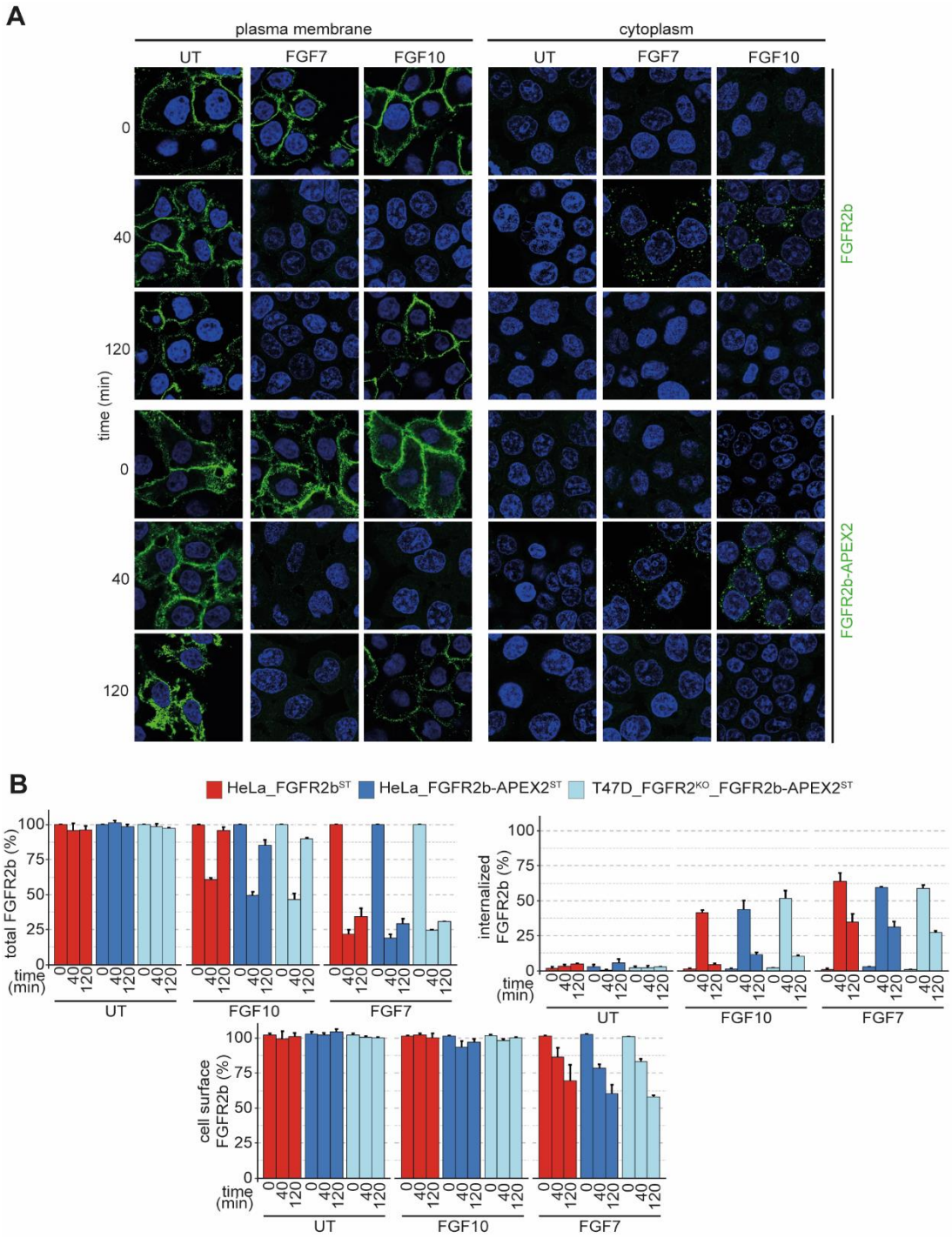


Figure 3-6. APEX2 tagged-FGFR2b does not have an altered trafficking response downstream of FGF7 or FGF10 stimulation.

A. *Assessment of FGFR2b (top panels) and FGFR2b-APEX2 (bottom panels) (green) internalization (cytoplasm) and recycling (plasma membrane) in HeLa cells untreated (UT) or in response to FGF7 or FGF10 treatment for 0, 40 or 120 min. Nuclei are stained with DAPI (blue). Scale bar, 5 µm.* ***B****. Quantification of FGFR2b trafficking in HeLa_FGFR2bST (red), HeLa_FGFR2b-APEX2ST (dark blue) and T47D_FGFR2KO_FGFR2b-APEXST (light blue) cells, showing the presence (total), internalized (internalized) and recycled (cell-surface) FGFR2b upon stimulation. Values represent mean ± standard error of mean (SEM) of at least three independent experiments. Representative images are shown in B.*

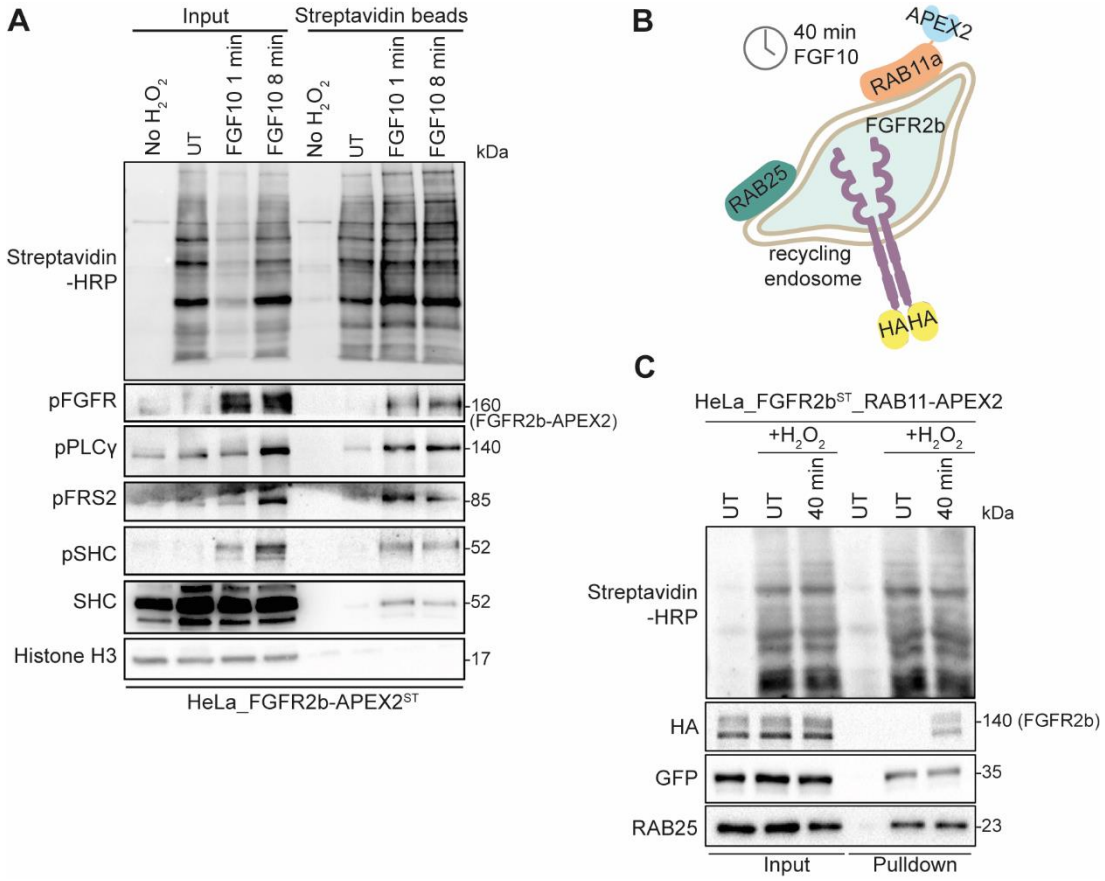


Figure 3-7. APEX2 tagged-FGFR2b and RAB11a identifies compartment-specific signalling partners upon FGF10 stimulation.
A. *Immunoblot analysis with the indicated antibodies of input or biotinylated proteins enriched with Streptavidin beads from HeLa_FGFR2b-APEXST treated either with H_2O_2 or with FGF10 for 1 and 8 minutes.* ***B.*** *Schematic of RE-localised FGFR2b, following 40 minutes of FGF10 treatment. Both RAB11-APEX2 and RAB25 localize at the REs.* ***C.*** *Immunoblot analysis with the indicated antibodies of biotinylated proteins enriched with Streptavidin beads from HeLa_FGFR2bST RAB11-APEX2 stimulated with either H_2O_2 or with FGF10 for 40 minutes.*

3.3.3 **Spatially Resolved Phosphoproteomics (SPR) reveals FGFR2b signalling enriched at the REs**

To uncover FGFR2b signalling partners at the REs, we designed a phosphoproteomics approach based on the detection of phosphorylated proteins in the pulldowns from cells expressing FGFR2b-APEX2 or RAB11-APEX2 (Figure 3-7), hereby referred to as the Spatially Resolved Phosphoproteomics (SRP) approach (Figure 3-8). HeLa_FGFR2b-APEX2ST expressing RAB11, HeLa-FGFR2bST RAB11-APEX2 and HeLa-FGFR2bST GFP-APEX2 were treated with FGF10 for 40 minutes alongside HeLa_FGFR2b-APEX2ST expressing RAB11 and HeLa-FGFR2bST GFP-APEX2 left untreated, as controls. We collected both the global proteome and phosphoproteome (obtained after TiO2-based chromatography enrichment of phosphorylated peptides) and the proximal proteome and phosphoproteome obtained after enrichment of biotinylated proteins proximal to APEX2-tagged protein baits with streptavidin

beads (proteome) followed by TiO_2-based chromatography enrichment of phosphorylated peptides (phosphoproteome) (Figure 3-8). We first analysed the global and the proximal proteomes, which showed strong correlation between replicates and a clear distinction between the global and the proximal samples as assessed by Pearson correlation and PCA plots (Figure 3-9A, Appendix 2 Figure 4A-B). We noticed that the stimulated HeLa-FGFR2bST RAB11-APEX2 and the HeLa-FGFR2bST GFP-APEX2 proteomes showed partial overlap in the PCA plots (Figure 3-9A). To assess whether these were accurate spatial profiles of the REs and cytoplasmic background respectively, we checked which proteins were differentially abundant in each condition and performed an enrichment analysis for subcellular compartments on the resulting proteins lists (Figure 3-9B-D). Proteins that were more abundant in the stimulated HeLa-FGFR2bST RAB11-APEX2 samples, including transferrin receptor (TFRC), VPS51 and VPS52(Francavilla *et al.*, 2009; Schindler *et al.*, 2015), were enriched for the REs, whilst the proteins found in the stimulated HeLa-FGFR2bST GFP-APEX2 samples localized to the whole endosomal system (Figure 3-9B-D). Therefore, our approach captured the spatial profile of REs at the proteome level.

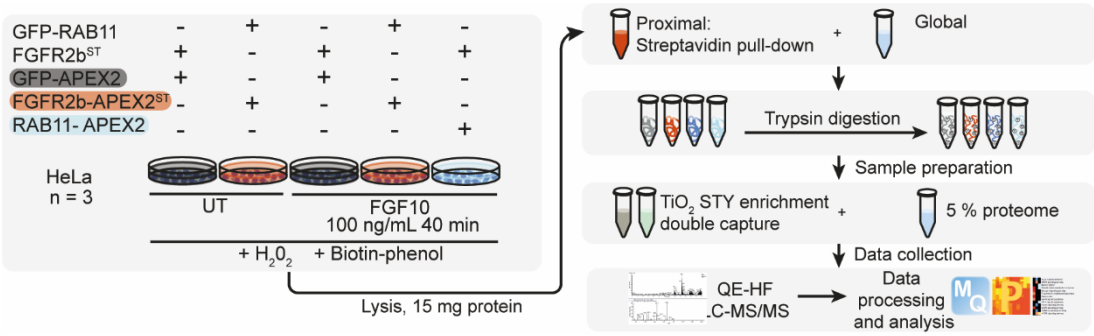


Figure 3-8. Workflow of the spatially resolved proteomics and phosphoproteomics experiments.

Next, we analysed the phosphoproteomics data. Pearson correlation and PCA analysis revealed again strong correlation between replicates and a striking distinction between the global and the proximal samples (Figure 3-9E, Appendix 2 Figure 4C-D). Interestingly, the number of phosphorylated sites quantified in the proximal samples (1565 - 2951) was substantially lower than the number quantified in the global samples (8312 - 8862) (Appendix 2 Figure 4E-F). This reduction was not seen in the number of identified protein groups, indicating that the enrichment for phosphorylated peptides that follows the enrichment for biotinylated proteins had a substantial effect on the identifications (Appendix 2 Figure 4E-F). To further assess how the double enrichment for biotinylated proteins and for phosphorylated peptides affected the raw data analysis, we checked the confidence of

identification and the distribution of intensities (Appendix 2 Figure 4G-H). We found that there was a left-shift in the distribution of the intensities of the double-enriched proximal samples compared to the global samples, and therefore we normalized the global and the proximal phosphoproteome samples separately (Appendix 2 Figure 4H). However, this analysis did not affect the overall quality of the proximal phosphoproteome data, as we found that >89% (6224 and 11726 for global and proximal, respectively) of the phosphorylated sites identified were Class I (≥ 0.75 localisation probability (Olsen *et al*, 2006)) and had the expected proportions of single or multiple phosphorylated sites or serine, threonine, or tyrosine residues (Appendix 2 Figure 4I) (Smith *et al.*, 2021).

To reveal the phosphorylated interactome of FGFR2b localized at the REs we normalized the quantified phosphorylated sites from the control and FGF10-treated FGFR2b-APEX2 and RAB11-AEPX2 samples against the corresponding time points of the GFP-APEX2 samples. Hierarchical clustering of the normalized data revealed a cluster of phosphorylated sites enriched in both the FGFR2b-APEX2 and the RAB11-APEX2 samples treated with FGF10, hereby the FGFR2b Recycling Proximal Signalling Cluster (Figure 3-9F). This cluster was enriched for proteins localized to the EEs and the REs (representing the trafficking route taken by FGFR2b (Smith *et al.*, 2021)) (Figure 3-9G) and included known regulators of trafficking like RCP, also known as RAB11FIP1 (Francavilla *et al.*, 2016). Therefore, the stimulated proximal phosphoproteome captures phosphorylated proteins in the proximity of the REs. We reasoned that the phosphorylated sites found on proteins identified in the proximal proteome were the highest confidence proximal phosphorylated interactors of FGFR2b. Therefore, looking at the phosphorylated proteins would act as a quality-control step for our double enrichment approach. We found 273 proteins from the -proximal proteome with corresponding phosphorylated sites in the proximal phosphoproteome dataset (Figure 3-9H) and 71.3% of these sites (273) were found also in the FGFR2b Recycling Proximal Signalling Cluster (Figure 3-9I). Unexpectedly, both the phosphorylated proteins belonging to the FGFR2b Recycling Proximal Signalling Cluster and the 273 proteins found in the proximal proteome- were enriched for autophagy and mTOR signalling pathways (Figure 3-9J). These findings confirm the robustness and specificity of the SRP approach to identify spatially resolved signalling pathways downstream of FGFR2b recycling (Ornitz and Itoh, 2015; Smith *et al.*, 2021; Szybowska *et al.*, 2021).

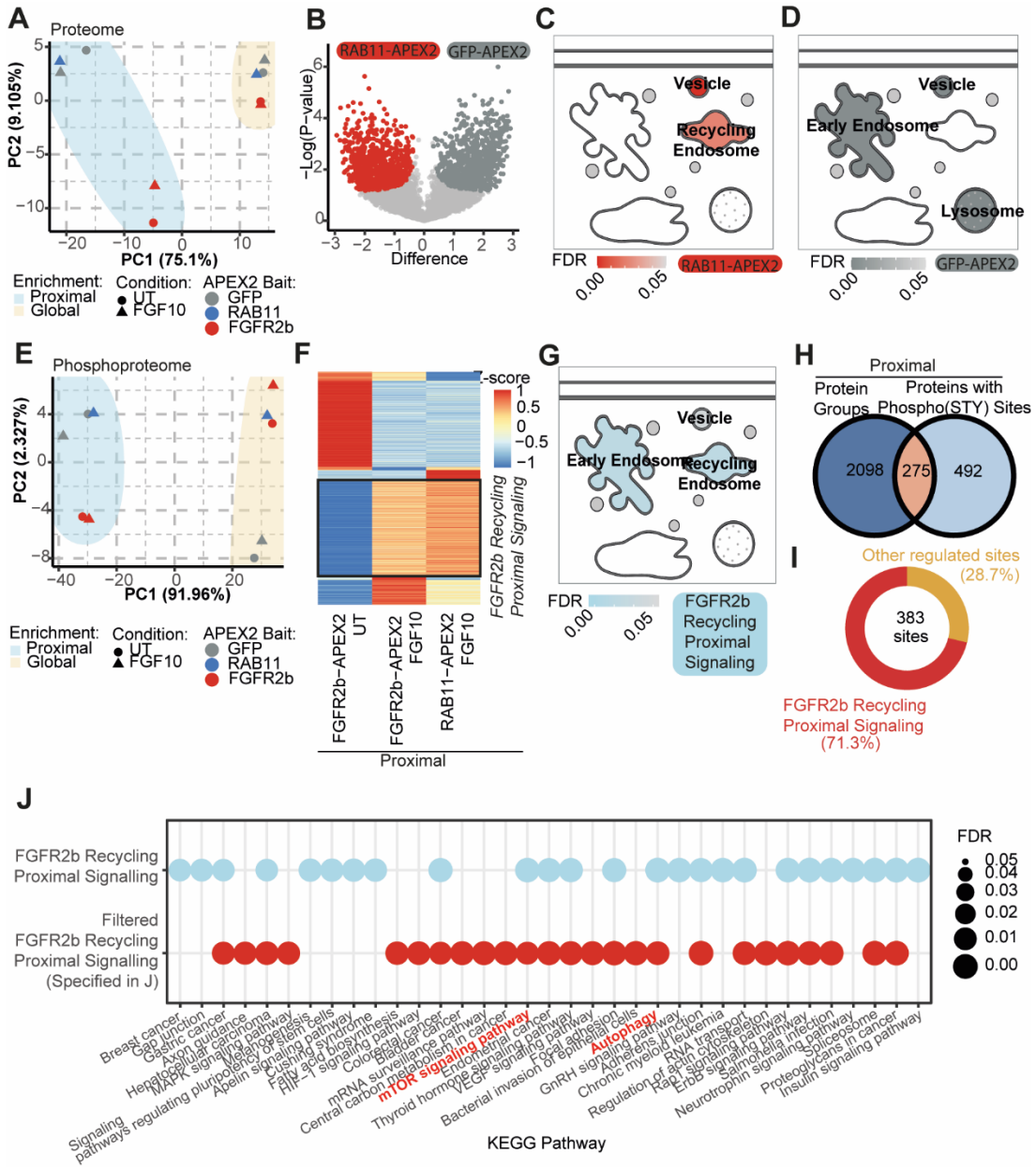


Figure 3-9. Spatially resolved proteomics and phosphoproteomics reveal FGFR2b-dependent regulation of mTOR signalling and autophagy.

A. *PCA plot of the proteome data based on median values.* ***B.*** *Volcano plot of the proximal proteome comparing HeLa_FGFR2bST RAB11-APEX2 and HeLa_FGFR2bST GFP-APEX2 stimulated with FGF10 for 40 min.* ***C, D****. Visualisation of the sub-cellular localization of the differentially abundant proteins from C using SubCellularVis.* ***E.*** *PCA plot of the phosphoproteome data based on median values.* ***F.*** *Cluster analysis of the proximal phosphoproteome from the indicated conditions normalized to the proximal phosphoproteome of HeLa_FGFR2bST GFP-APEX2 for each timepoint. Phosphorylated sites upregulated at 40 min stimulation with FGF10 in both HeLa_FGFR2bST RAB11-APEX2 and HeLa_FGFR2b-APEX2ST RAB11 are marked as the FGFR2b Recycling Proximal Signalling Cluster.* ***G.*** *Visualisation of the sub-cellular localization of phosphorylated proteins found in the FGFR2b Recycling Proximal Signalling cluster using SubcellulaRVis.* ***H.*** *Overlap of proteins and phosphorylated proteins detected in the proximal proteome and phosphoproteome samples, respectively.* ***I.*** *Distribution of the 383 phosphorylated sites from 275*

phosphorylated proteins, 71.3% of which were found in the FGFR2b Recycling Proximal Signalling Cluster. ***J.*** *KEGG pathway over representation analysis (ORA) of the phosphorylated sites found in the FGFR2b Recycling Proximal Signalling Cluster (blue light) and among the phosphorylated sites on proteins quantified at the proteome level from (I).*

To verify whether the link between FGFR2b recycling, mTOR signalling and autophagy identified by the SRP approach was also regulated in the global phosphoproteome, we performed Volcano plot analysis of untreated and FGF10-stimulated global phosphoproteomes. We found differentially regulated sites in the two conditions as expected (Francavilla *et al.*, 2013; Smith *et al.*, 2021) (Figure 3-10A, Appendix 2 Figure 4J), including FGFR2 phosphorylated sites (S587, Y586, Y656, Y657). However, there was only a small overlap (69 phosphorylated sites) between the FGF10-regulated phosphorylated sites from the global phosphoproteome and the FGFR2b Recycling Proximal Signalling Cluster, indicating that the SRP approach capably distinguished the proximal and the global phosphoproteome profiles (Figure 3-10B). Indeed, whereas enrichment analysis of the global upregulated phosphorylated sites revealed proteins involved in canonical signalling pathways, such as PI3K-AKT signalling, FoxO signalling and Ras signalling, terms like autophagy or mTOR signalling were enriched specifically in the FGFR2b Recycling Proximal Signalling Cluster, which included densely connected phosphorylated proteins (443 compared to 224 found in the global phosphoproteome) (Figure 3-10C-D). FGFR2 and EGFR were found phosphorylated in both the proximal and the global phosphoproteome (Figure 3-10E). One of the catalytic sites of FGFR2 (Y656) (Ornitz and Itoh, 2015) was also identified as part of the internalization response (Figure 3-4), corroborating the role of this site for FGFR2 trafficking (Francavilla *et al.*, 2013). Interestingly, T693 on EGFR was found phosphorylated only in the proximal phosphoproteome (Figure 3-10E), consistent with its role in regulating FGFR2b recycling at the REs (Smith *et al.*, 2021). Therefore, the SRP approach selectively enriches for phosphorylation proteins localized at the REs which could not be detected in the global phosphoproteome.

3.3.4 FGFR2b recycling suppresses autophagy in a mTOR- and ULK1-dependent manner

Among the phosphorylated proteins annotated to the mTOR signalling pathway/autophagy in the enrichment analysis of the FGFR2b Recycling Proximal Signalling Cluster (Figure 3-10C) we found several canonical regulators upstream of mTOR signalling, including RAF1, RRAS2, MAP2K2, RPS6 and five proteins known to regulate autophagy either directly or downstream of mTOR signalling: the AMPK subunit PRKAA1, the mTOR subunit RPTOR, TSC1, TSC2 and

the kinase ULK1 phosphorylated on S638, the latter known to prevent autophagy (Saxton and Sabatini, 2017; Shin *et al.*, 2018) (Figure 3-10F).

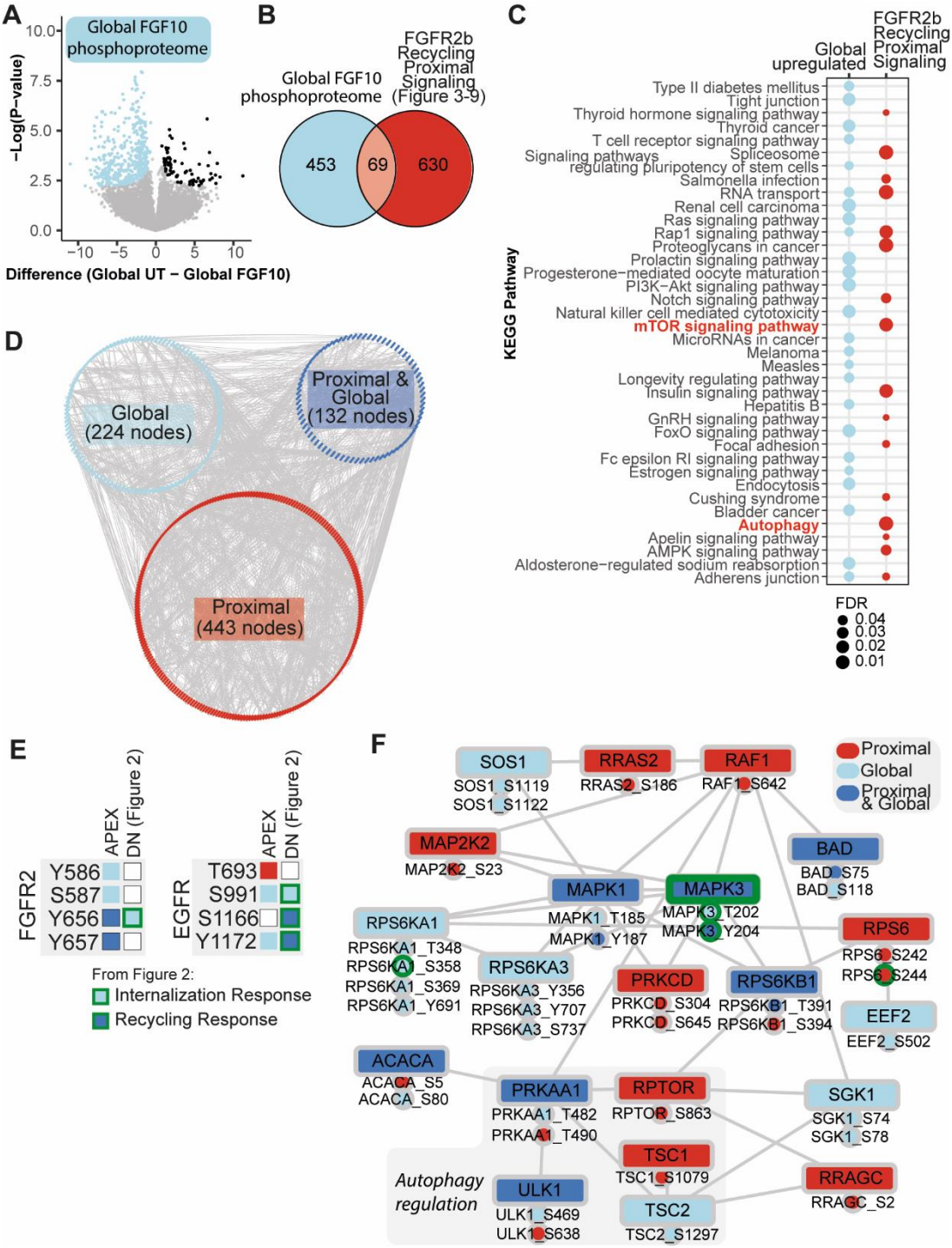


Figure 3-10. FGFR2b regulates mTOR signalling and autophagy from the REs.
A. *Volcano plot of the phosphorylated sites from the global phosphoproteome in unstimulated and FGF10 40 min stimulated cells.* ***B.*** *Overlap between the phosphorylated sites upregulated in the global phosphoproteome upon FGF10 stimulation based on (A) and the phosphorylated sites upregulated in the FGFR2b Recycling Proximal Signalling cluster from the proximal phosphoproteome (Figure 3-9F).* ***C****. KEGG pathway over representation*

analysis (ORA) of proteins with phosphorylated sites from (B). ***D.*** *Protein-protein interaction network based on the STRING database and visualized with Cytoscape of the phosphorylated proteins shown in (B). The number of phosphorylated protein nodes is indicated in parentheses. Unconnected nodes were removed.* ***E****. Phosphorylated sites identified on FGFR2 and EGFR in the global or in the proximal phosphoproteome, and in the phosphoproteome from HeLa_FGFR2bST cells expressing GFP, GFP-DnRAB11 or GFP-DnDNM2 (Figure2A). Light blue with no border indicates upregulated in global SRP phosphoproteome only. Dark blue with no border indicates updated in both global and proximal SRP phosphoproteome. Red indicates upregulated in proximal SRP phosphoproteome only. Light blue with green border indicates found in Internalization response cluster (as in Figure 3-4). Dark blue with green border indicates found in recycling response cluster (as in Figure 3-4).* ***F****. Subnetwork from D including proteins annotated to mTOR pathway or autophagy based on KEGG (C). Node colouring indicates whether the phosphorylated protein or the phosphorylated sites were found in global, proximal phosphoproteome or both. Proteins involved in autophagy regulation are highlighted in grey. Sites and proteins also quantified in Figure 3-4 have a green border.*

To test whether FGF10-mediated FGFR2b recycling regulates autophagy, we assessed autophagy in HeLa-FGFR2bST, T47D and BT20 treated with FGF10 and with FGF7 (as a negative control for FGFR2b recycling (Francavilla *et al.*, 2013)) for 2 hr. FGF10 treatment significantly decreased autophagy compared to control in all cell lines, whereas FGF7 did not (Figure 3-11A). As we starved cells before stimulation with FGFs and starvation is known to increase autophagy (Ichimiya *et al*, 2020), we checked the levels of autophagy markers in starved cells followed or not by stimulation with either serum (as control), FGF7 and FGF10. The autophagosome-formation associated microtubule-associated proteins 1A/1B light chain 3B (LC3B) (Schaaf *et al.*, 2016) was supressed to levels seen in serum-treated cells by FGF10 treatment alone (Figure 3-11B). This FGF10-, but not FGF7-dependent decrease in LC3B levels was seen in HeLa-FGFR2bST and BT20 cells as well, alongside a decrease in active BECLIN1 (Figure 3-11C), another mediator of autophagosome formation and maturation (Ashkenazi *et al.*, 2017). These results suggest that autophagy regulation is FGFR2b-recycling dependent. To confirm the importance of FGFR2b in autophagy regulation, we compared autophagy in parental T47D, T47D depleted of *Fgfr2*, and T47D depleted of *Fgfr2* and overexpressing FGFR2b and found upregulation of autophagy in the absence of FGFR2b and less autophagy in T47D expressing high levels of FGFR2b compared to parental T47D (Appendix 2 Figure 5A-B).

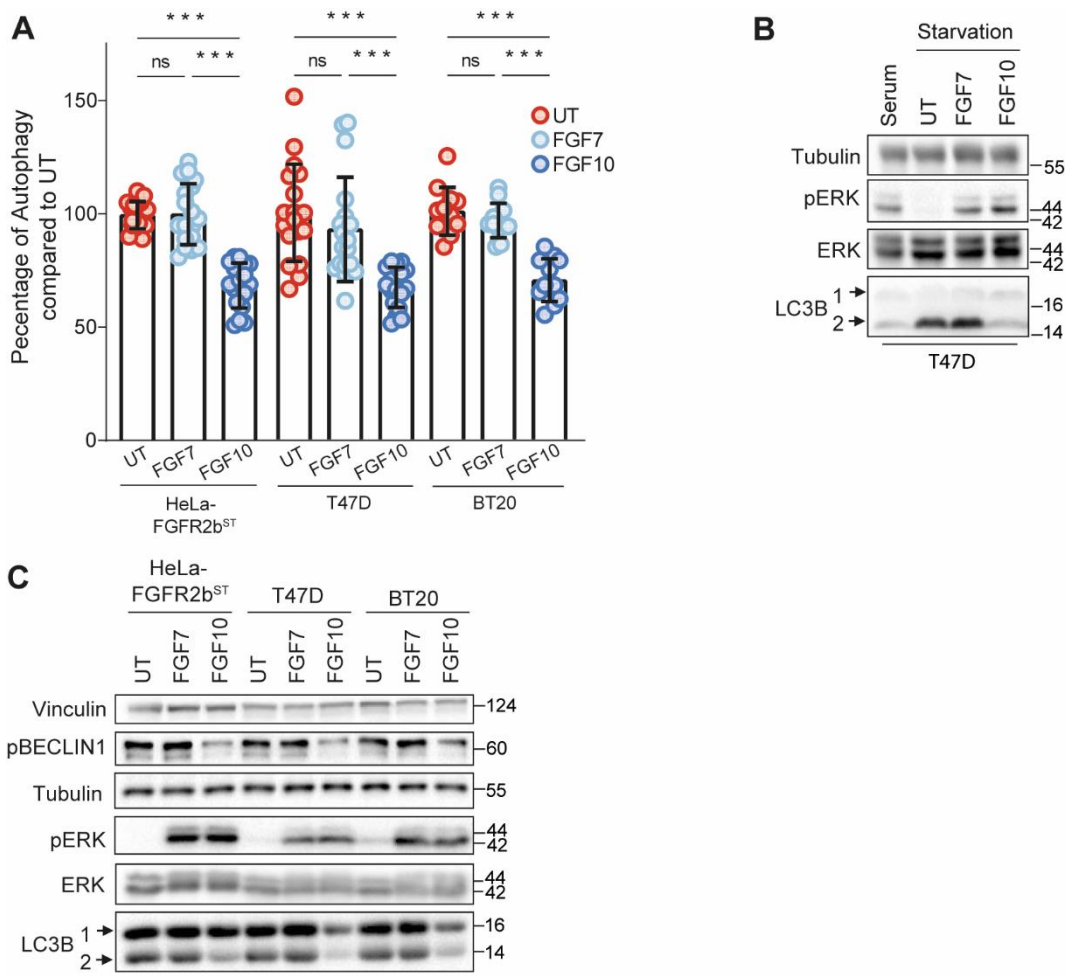


Figure 3-11. FGFR2b regulates autophagy from the REs
A. *Autophagy measured by acridine orange of HeLa_FGFR2bST, T47D, and BT20 treated with FGF7 or FGF10. N = 12, P =< 0.001*** (one-way ANOVA with Tukey test)* ***B.*** *Immunoblot analysis with the indicated antibodies of the effect of serum starvation and FGF treatment on autophagic markers in T47D.* ***C.*** *Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST, T47D, and BT20 treated with FGF7 or FGF10 for 40 minutes*

Using the same cell model, we also investigated whether autophagy downstream of FGFR2b required mTOR signalling. To test this, we compared autophagy in cells subjected to either starvation, glucose removal and glucose-6-phosphate treatment (which are known to induce mTOR-dependent autophagy) or sodium valproate and fluspiriline treatment (which are known to induce mTOR-independent autophagy). We found that mTOR-dependent but not mTOR-independent autophagy was affected by FGFR2b levels, as cells depleted of *Fgfr2* showed the highest levels of mTOR-dependent autophagy, whilst high levels of FGF2b expression (T47D FGFR2b^{KO}-FGFR2b) induced the lowest levels of mTOR-dependent autophagy (Appendix 2 Figure 5A-B). As high expression of RTKs is generally associated with higher levels of internalisation in basal conditions and potential difference in signalling

regulation (Resat *et al*, 2003) (Appendix 2 Figure 1I, Figure 3-6), we concluded that internalized FGFR2b was required for mTOR-dependent autophagy. Moreover, mTOR-dependent, but not mTOR-independent, autophagy and the autophagy markers LC3B and Beclin were regulated by FGF10 treatment in parental T47D (Appendix 2 Figure 5C-D). Finally, both mTOR signalling and the known autophagy regulator ULK1 were required for FGF10-dependent regulation of autophagy in T47D and HeLa cells (Appendix 2 Figure 5E-F). These findings indicate that suppression of autophagy is FGFR2b recycling-, mTOR- and ULK1-dependent in FGF10-stimulated epithelial cells.

As we identified ULK1 phosphorylation on S638 in the local phosphoproteome (Figure 3-10F) and this phosphorylation event is known to regulate autophagy (Zachari and Ganley, 2017), we investigated whether the phosphorylation of ULK1 on S638 occurred at the REs during FGFR2b recycling. In both HeLa-FGFR2bST and T47D cells ULK1 was phosphorylated on S638 in proximity of both FGFR2b and RAB11 (Figure 3-12, Appendix 2 Figure 6A-B). Furthermore, confocal analysis of T47D_FGFR2^{KO}_FGFR2b-APEX cells expressing RAB11 and stimulated with FGF10 for 40 min showed a significant co-localization between phosphorylated ULK1 and FGFR2b at the REs (Figure 3-13). This findings confirm that ULK1 is associated to REs (Longatti *et al.*, 2012) and suggests that the presence of FGFR2b at the REs is necessary for the recruitment of phosphorylated ULK1. Indeed, we did not visualize any ULK1 phosphorylation on S638 when FGFR2b recycling was impaired by expressing DnRAB11 (Figure 3-13).

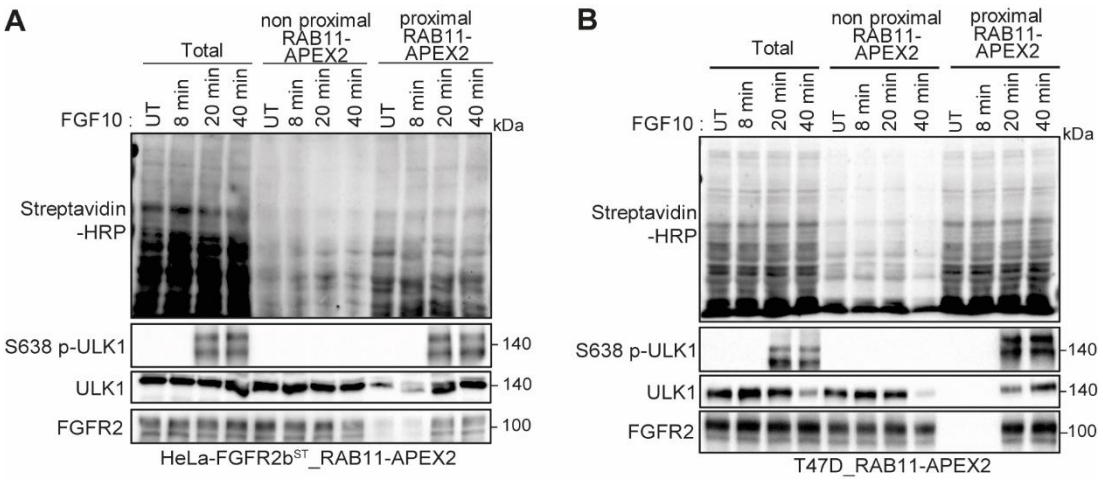


Figure 3-12. Phosphorylated ULK1 is found at RAB11-positive REs.
*Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST RAB11-APEX2 (**A**) or T47D transfected with RAB11-APEX2 (T47D RAB11-APEX2) (**B**) stimulated with FGF10 for the indicated timepoints. Non proximal and proximal samples represent the supernatant and the pulldown following enrichment of biotinylated samples with streptavidin beads, respectively, and run against total lysates (total).*

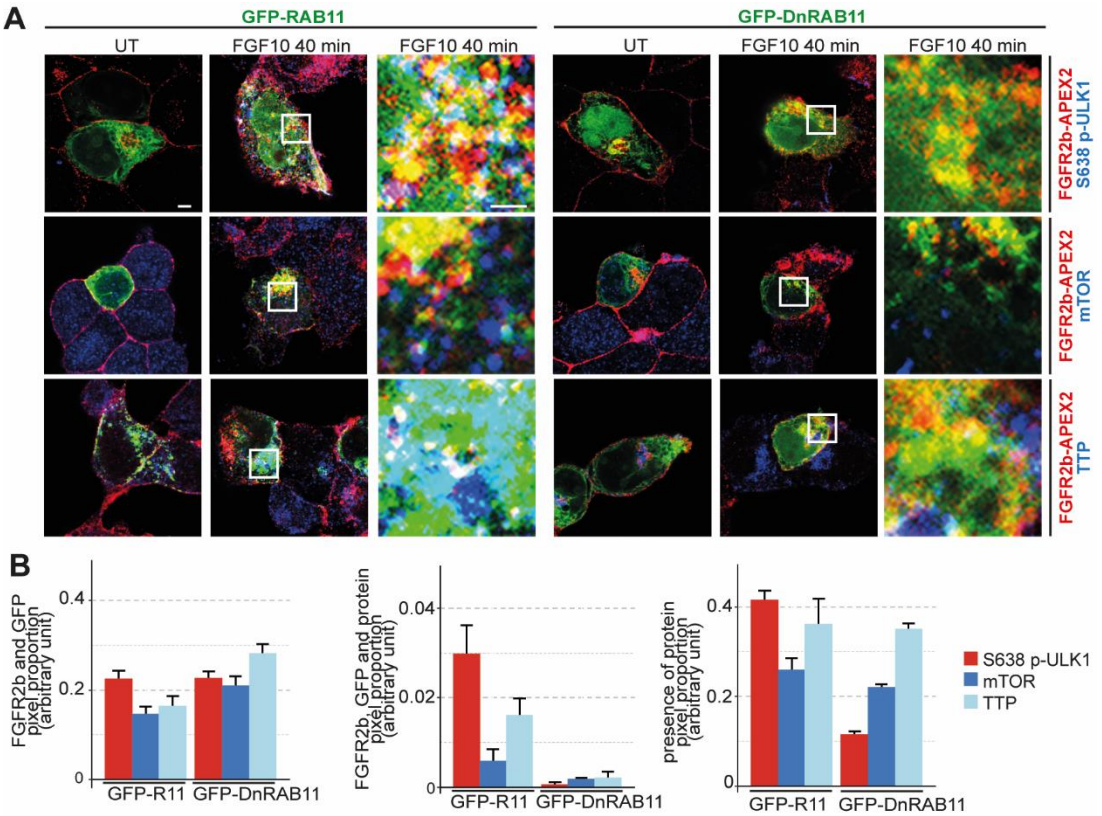


Figure 3-13. Phosphorylated ULK1 colocalises with FGFR2b at the RE.
A. *Co-localization of FGFR2b-APEX2 (red) with phosphorylated ULK1 on S638, mTOR or TTP (blue) in T47D_FGFR2KO_FGFR2b-APEXST transfected with RAB11 or GFP-DnRAB11 (green) and stimulated or not with FGF10 for 40 min as indicated. Scale bar, 5 µm. The image on the right is a magnification of the FGF10-stimulted samples, scale bar 50 µm.* ***B.*** *Quantification of pixels overlap of the conditions shown as representative images in C (see colour code) and of the proportion of each protein present in each condition based on pixels (Smith et al., 2021). Values represent the median ± SD of at least three independent experiments*

The phosphorylation of ULK1 downstream of FGFR2b recycling is a specific event, as other FGFR2b downstream pathways were only marginally affected in cells expressing either DnRAB11 or DnDNM2, treated with the primaquine and dynasore compounds, or stimulated with FGF7, all conditions that impaired FGFR2b trafficking (van Weert *et al.*, 2000; Macia *et al.*, 2006; Francavilla *et al.*, 2013; Smith *et al.*, 2021) (Figure 3-14, Appendix 2 Figure 6C-D). Intriguingly, inhibiting FGFR2b localization at the REs also misplaced the FGFR2b recycling regulator TTP (Francavilla *et al.*, 2013) from REs to LAMP1-positive lysosomes (Figure 3-13, Figure 3-15), where it has previously been shown to negatively regulate mTOR signalling (Kim *et al.*, 2012). Therefore, we checked mTOR localization and activation in our experimental conditions. mTOR was localized on lysosomes in both RAB11- and DnRAB11-expressing cells (Figure 3-13, Figure 3-15), but its activation decreased in cells with impaired FGFR2b trafficking (Appendix 2 Figure 6E), confirming the link between FGFR2b recycling and mTOR signalling. Therefore, we checked how inhibiting FGFR2b recycling by expressing DnRAB11

affected mTOR signalling partners, including RPTOR and AMPK (Saxton and Sabatini, 2017). Inhibiting FGFR2b recycling prevented RPTOR and AMPK dephosphorylation on S863 and T172, respectively, and induced the phosphorylation of S638 on ULK1 up to 2 hours after FGF10 stimulation (Figure 3-14C, Appendix 2 Figure 6C).

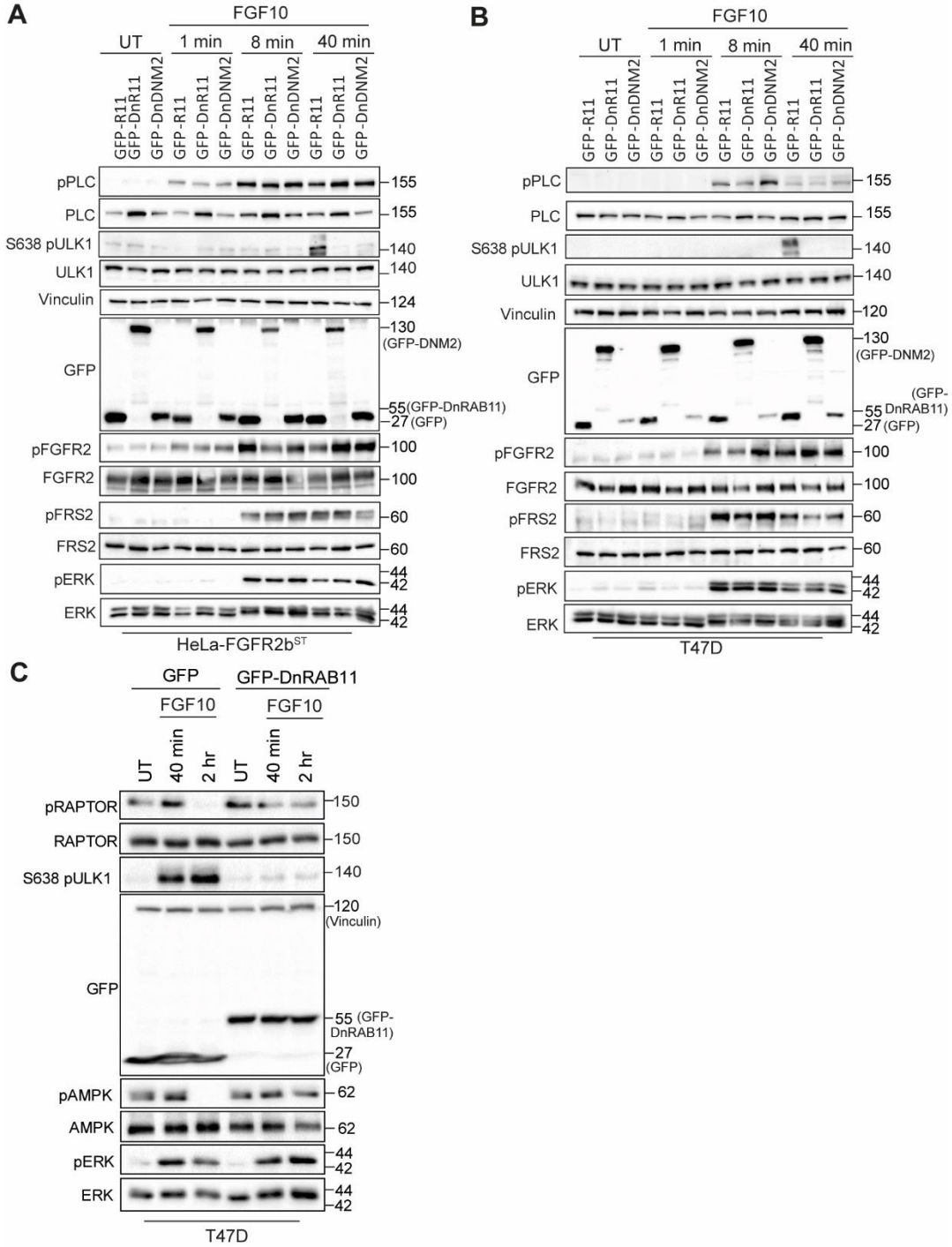


*Figure 3-14. Phosphorylated ULK1 recruitment at the REs depends on FGFR2b recycling. Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST (**A**) or T47D (**B,C**) transfected either with RAB11, GFP-DnRAB11, or GFP-DnDNM2 (**A,B**) or with GFP and GFP-DnRAB11 (**C**) and left either untreated (UT) or treated with FGF10 for the indicated time points.*

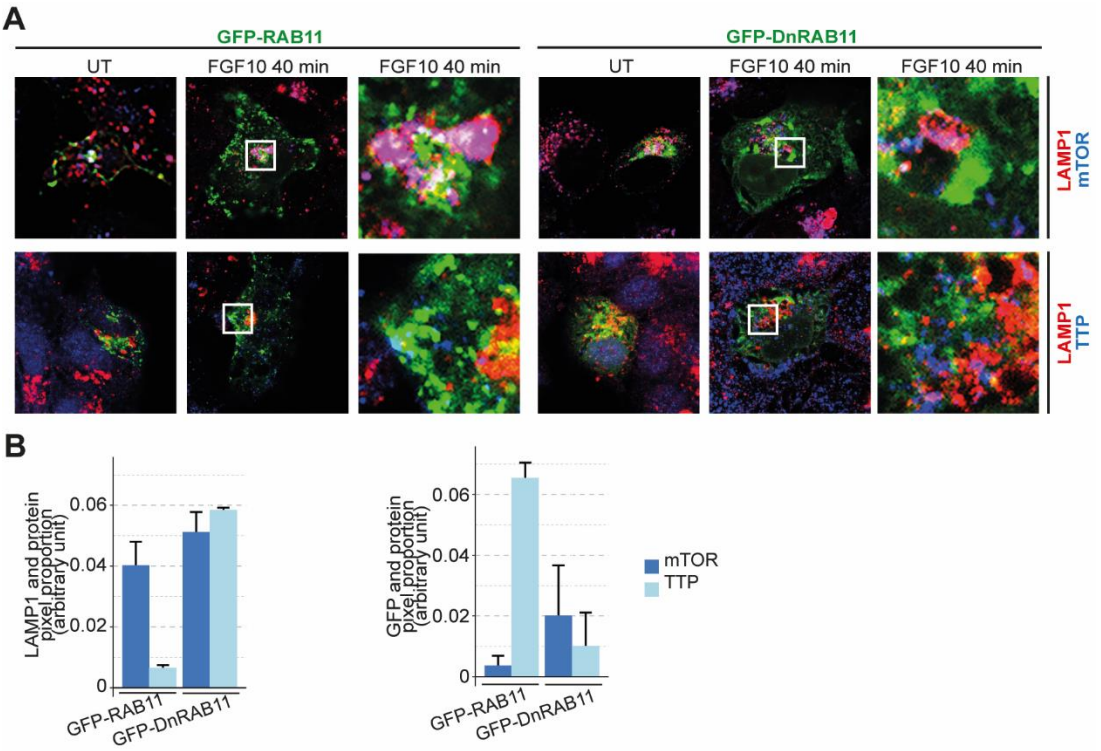


Figure 3-15. mTOR activation is dependent on FGFR2b recycling.
***A**. Co-localization of LAMP1 (red) with mTOR or TTP (blue) in T47D_FGFR2KO_FGFR2b-APEXST transfected with RAB11 or GFP-DnRAB11 (green) and stimulated or not with FGF10 for 40 min as indicated. Scale bar, 5 µm. The image on the right is a magnification of the FGF10-stimulted samples, scale bar 50 µm. **B**. Quantification of pixel overlap of conditions shown as representative images in E, Values represent the median ± SD of at least three independent experiments*

In conclusion, FGFR2b recycling regulates mTOR activation and ULK1 phosphorylation, and these signalling events are crucial for autophagy suppression downstream of FGF10.

3.3.5 **FGFR2b recycling regulates mTOR- and ULK1-dependent cell survival**

To investigate how FGFR2b signalling partners at the REs affected long-term FGFR2b responses during recycling, we tested the impact of impaired FGFR2b trafficking on FGF10-induced autophagy, cell proliferation and apoptosis. Firstly, we found that autophagy did not change or was slightly increased in FGF10-stimulated cells expressing DN-RAB11 or treated with the trafficking inhibitors monensin (Francavilla *et al.*, 2009) and dynasore (Macia *et al.*, 2006) and in cells stimulated with FGF7 (Figure 3-16A, Appendix 2 Figure 7A). Immunoblot analysis showed reduced levels of LC3B and RPTOR phosphorylation and increased phosphorylation of ULK1 on S638 2 hours following stimulation with FGF10 in wild type cells but not in cells expressing DN-RAB11 or DN-DNM2 (Figure 3-16B, Appendix 2 Figure 7B). Consistent with previous data, FGF7 did not affect any of these autophagy markers (Figure

3-16B, Figure 3-11G-I, Figure 3-14E-G, Appendix 2 Figure 6C-D, Appendix 2 Figure 7B). Therefore, autophagy is regulated downstream of FGFR2b recycling in response to FGF10. Next, we tested cell proliferation by EdU incorporation and apoptosis by caspase3 staining and found that both increased in cells with impaired FGFR2b trafficking compared to wild type cells upon stimulation with FGF10 (Figure 3-16C). As the link between increased progression into S-phase, increased apoptosis and increased autophagy is established (Mathiassen, De Zio and Cecconi, 2017; Yamasaki, Jin and Ohsumi, 2020), as we observed in FGF10-stimulated epithelial cells with impaired FGFR2b recycling (Figure 3-16A-C, Appendix 2 Figure 7A-B), we asked whether the FGFR2b signalling partners at the REs regulated these responses. Therefore, we treated cells with inhibitors of ULK1 (ULK101, SBI0206965), FGFR (PD173074) or mTOR (rapamycin) as controls, and investigated autophagy, apoptosis and proliferation following 24 hours stimulation with FGF10 (Figure 3-16D). ULK1 inhibition reduced autophagy compared to untreated cells but the level of autophagy was comparable to that of FGF10-treated cells, consistent with our data after 2 hours stimulation (Appendix 2 Figure 5E-F). As expected, both mTOR and FGFR inhibitors increased autophagy and its markers in response to FGF10 (Figure 3-16D, Appendix 2 Figure 5E-F, Appendix 2 Figure 7C). In addition, ULK and mTOR, but not FGFR, inhibition induced higher levels of apoptosis and of the apoptotic marker caspase 3 than in FGF10-treated cells (Figure 3-16D, Appendix 2 Figure 7C). All the inhibitors however induced a decrease in overall cell proliferation (Figure 3-16D). Altogether, this data suggests a functional link between FGFR2b recycling, the activation of mTOR and ULK1 in proximity of the REs and long-term responses.

In conclusion, we showed the importance of REs as signalling platforms to co-ordinate cellular fate by revealing that the inability of active FGFR2b to reach the REs disengages the link between mTOR/ULK1 signalling, autophagy and overall cell survival.

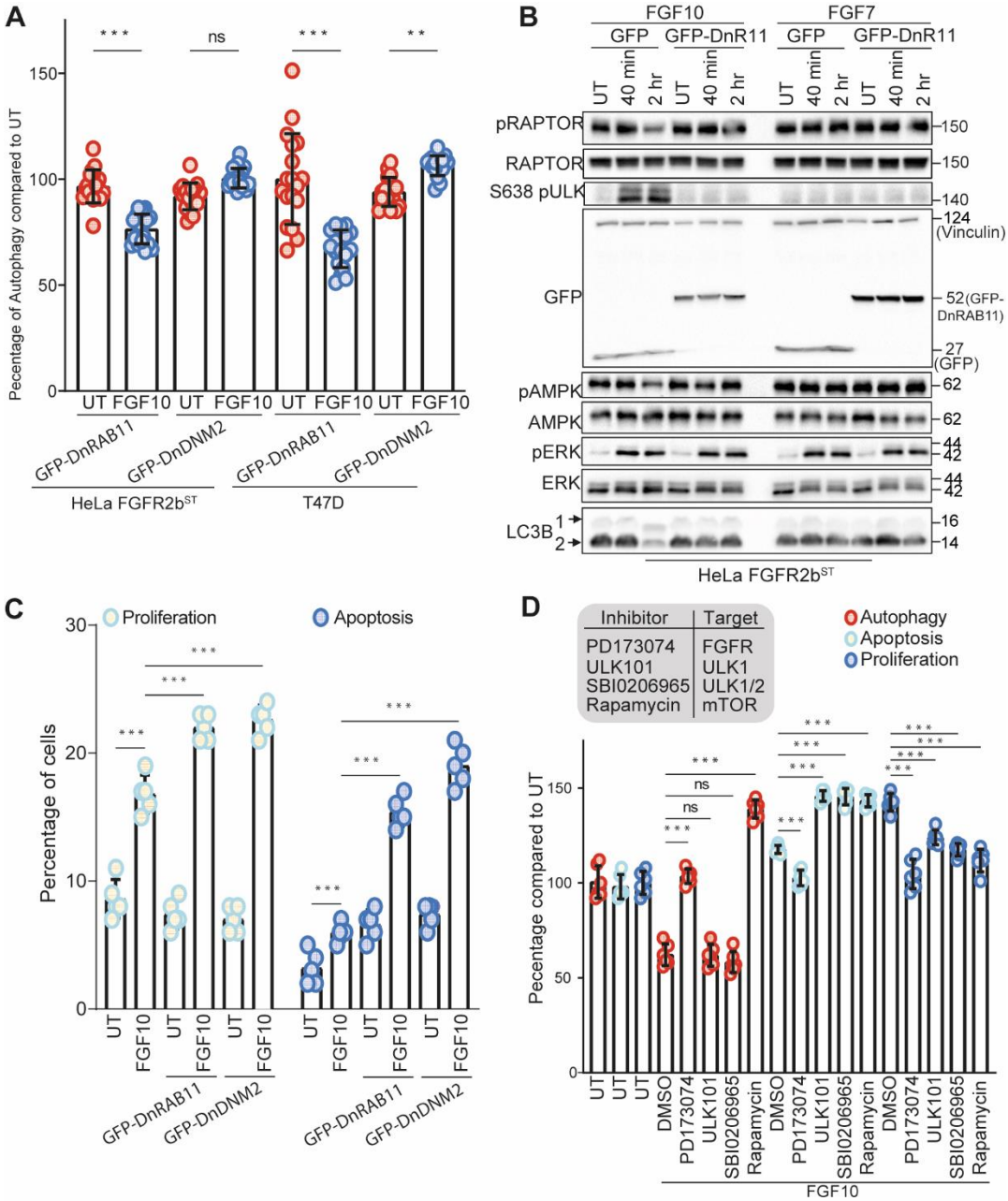


Figure 3-16. Inhibiting the recycling of FGFR2b causes dysregulated autophagy and alters the balance of proliferation and cell death.

A. *Autophagy measured by acridine orange of HeLa_FGFR2bST (left) or T47D (right) transfected either with RAB11, GFP-DnRAB11, or GFP-DnDNM2 and incubated or not with FGF10 for 2h. N = 18, P =< 0.001*** (one-way ANOVA with Tukey test).* ***B****. Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST transfected either with GFP or GFP-DnRAB11 and left either untreated (UT) or treated as indicated.* ***C****. Measurement of T47D cell proliferation by EdU incorporation and cell apoptosis by cleaved caspase 3 activated dye and presented as a percentage of cells following transfection either with RAB11, GFP-DnRAB11, or GFP-DnDNM2, starvation, and stimulation with FGF10. N = 5, P =< 0.001*** (one-way ANOVA with Tukey test).* ***D.*** *Measurement cell proliferation by EdU incorporation, cell apoptosis by cleaved caspase 3 activated dye, and autophagy by acridine orange staining of T47D treated with the with PD173074, ULK101, SBI0206965, or Rapamycin (targets of inhibition specified in inset table) and stimulated or not with FGF10 for 2hindicated inhibitors and stimulated or not with FGF10 for 2h. Data are presented as percentage compared to untreated cells. N = 6, P =< 0.001*** (one-way ANOVA with Tukey test).*

3.4 Discussion

The importance of endocytosis in regulating selected RTK signalling cascades to drive cell fate in different contexts, including development or cancer, is now recognised (Stasyk and Huber, 2016; Lanzetti and Di Fiore, 2017; Budayeva and Kirkpatrick, 2020; O'Sullivan and Lindsay, 2020; Wang, Bian and Li, 2021). Here, we developed and utilised Spatially Resolved Phosphoproteomics to uncover FGFR2b signalling partners localized at the REs during receptor recycling and we showed that such partners suppressed mTOR-dependent autophagy and controlled cell survival (Figure 3-17). We found that the autophagy regulator ULK1 phosphorylated on S638 was recruited to FGFR2b- and RAB11-positive REs to prevent autophagy in FGF10-stimulated cells. This recruitment of phosphorylated ULK1 was prevented in the absence of FGFR2b recycling, resulting in impaired autophagy. Chemical inhibition of ULK1 and of mTOR, one of ULK1 known upstream regulator (Zachari and Ganley, 2017) as well as of endosomal trafficking not only released FGF10/FGFR2b-dependet suppression of autophagy but also perturbed the longer-term effects on cell behaviour downstream of FGFR2b activation, including the balance between apoptosis and proliferation (Figure 3-17).

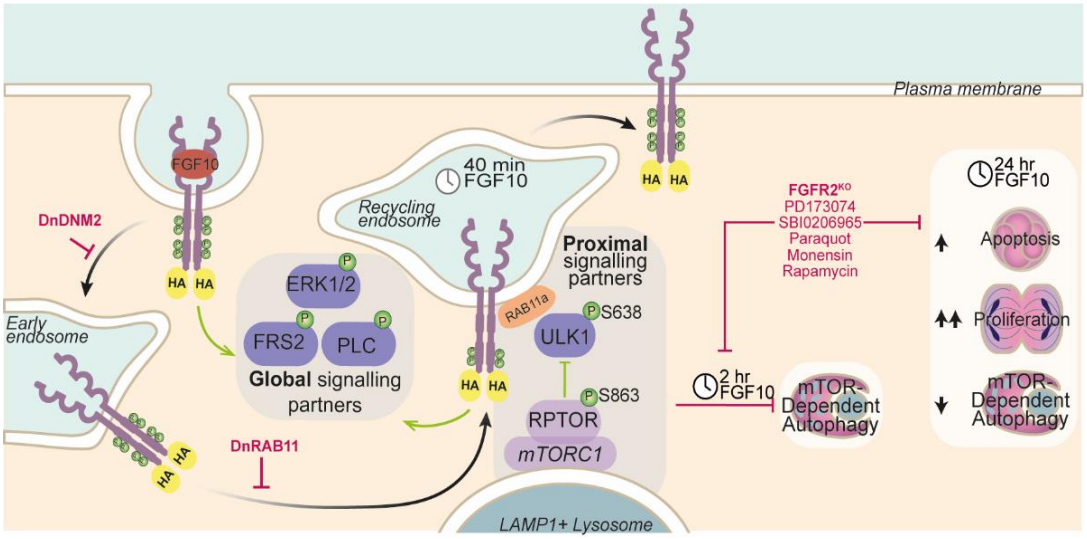


Figure 3-17. Model of FGFR2b global and proximal signalling partners during recycling to the plasma membrane.
Long-term responses are also indicated based on the data of this study. The black arrow indicates FGFR2b trafficking.

Inhibiting FGFR2b trafficking by genetic means alters the global phosphorylation programme in response to FGF10 (Figure 3-3), confirming the crucial role of receptor internalization and recycling in driving long-term responses (Francavilla *et al.*, 2013; Smith *et al.*, 2021). Indeed, 23% and 12.8% of FGF10-dependent regulated phosphorylated sites were part of the internalization and recycling response of epithelial cells, respectively. This data high lights

that certain signalling cascades are activated only when the receptor is "at the right place at the right time". The FGFR2b internalization-dependent phosphorylated sites discovered here and shared with the community could inform us on how dynamin regulates FGFR2b and, more broadly, RTK functions, for instance in cancer (Chernikova *et al.*, 2018). However, our genetic approach did not reveal signalling partners specifically recruited to and phosphorylated at the REs during FGFR2b recycling. To reveal the contribution of internalised FGFR2b localized at the REs to signalling, we developed a biotinylation-driven approach that we named Spatially Resolved Phosphoproteomics (SRP). This approach enabled us to generate a snapshot of spatially and temporally resolved signalling events downstream of FGFR2b (Figure 3-9, Figure 3-10), expanding the analysis from partners in the proximity of a protein bait as shown for GPCRs (Lobingier *et al.*, 2017) to phosphorylated partners in the proximity of FGFR2b at the REs. We first enriched for biotinylated proteins, and then for phosphorylated peptides (following protein digestion), a method already proven efficient with a BioID-based protocol biotinylation (Liu *et al.*, 2021), but that differ s from most of the published work in which enrichment for phosphorylated and APEX2-biotinylated proteins was performed prior to protein digestion (Qin *et al.*, 2021). Our method is an improvement on the latter due to phosphorylated site, rather than phosphorylated protein, resolution in the resulting dataset. We envision that our method (double biotinylated protein-modified peptide enrichment) could be easily adapted to study the localized dynamics of other modifications, thus enlarging the recently published BioID organelle interactome libraries (Go *et al.*, 2021). The advantage of using APEX2-driven biotinylation over BioID or TurboID for defining signalling events in subcellular compartments is the tighter time frame that can be defined (e.g. seconds, minutes), which is essential for defining discrete signalling events. In contrast to other spatial phosphoproteomics methods (Martinez-Val, Dorte B. Bekker-Jensen, *et al.*, 2021) that enrich for organelles and then for global phosphorylation events, the SRP approach allows the enrichment of phosphorylated events at one (or more) organelle under acute stimulation, thus revealing unique phosphoproteome signatures in a spatio-temporal defined manner. Finally, our SRP method provides the exciting opportunity to investigate endosome-proximal phosphorylation events in a high-throughput manner as opposed to signalling partners identified at EEs using biochemical and low-throughput methods. For example, populations of phosphatidylinositol 3-kinase (PI3K) have been shown to be activated at the EEs downstream of epidermal growth factor (EGF) stimulation (Thapa *et al.*, 2020). Signalling from the LEs has also been described, via MAPK and JNK (Jiang, Pan and Low, 2017). However, signalling partners localized at RAB11-positive REs are less known.

Here, SRP allows us to identify ULK1 phosphorylated at S638 as localized to FGFR2b- and RAB1-positive REs (Figure 3-10F, Figure 3-12).

Phosphorylated ULK1 at S638 is primarily seen at the REs when active FGFR2b is present. ULK1 has been previously shown to localize to RAB11-positive REs, but its phosphorylation state and the intersect with RTK trafficking has not been previously investigated (Longatti *et al.*, 2012; Zachari and Ganley, 2017). We also showed that upstream regulators of ULK1, such as mTOR or AMPK (Zachari and Ganley, 2017), are in the proximity of the REs, suggesting that the REs may be a site of mTOR/ULK1 regulation downstream of FGFR2b. The link between mTOR and endocytic trafficking processes and lysosomal transport has been previously documented (Takahashi *et al.*, 2012; Dauner *et al.*, 2017; Kvainickas *et al.*, 2019). A potential dynamic interface between the REs and lysosomes, which are localized in close proximity one to each other (Miller *et al.*, 1986), would favour the interaction of mTOR complexes and ULK1 within the REs. This intriguing possibility is worthy of further investigation and would have important implications for the understanding of RTK trafficking and the behaviour of RTKs within the endosomal system.

RTK signalling and endocytosis have previously been linked to regulation of autophagy (Fraser *et al.*, 2017) and EGFR recycling has been shown to decrease in cells lacking autophagy regulators (Fraser *et al.*, 2019). Signals from growth factors are known to converge on inhibitors of mTORC1 complexes at the lysosomal membrane to inhibit autophagy and catabolic processes (Saxton and Sabatini, 2017). Focusing on the FGFR family, the FGFR2b selective ligand FGF7 has been shown to induce autophagy in keratinocytes after 24 hours stimulation (Nanni, Ranieri, Rosato, *et al.*, 2018). However, within the 2 hours timeframe used in our experiments, FGF7 fails to alter ULK1/mTOR signalling or the downstream autophagy response, in contrast to the responses achieved in FGF10-stimulated cells (Figure 3-11C, Appendix 2 Figure 5E, F). The stark difference between FGF7 and FGF10 high lights the role of FGFR2b recycling as the instigator of the exclusive FGF10/ULK1/autophagy regulation. How this is orchestrated from the REs remains however unclear. One possibility is the involvement of EGFR signalling, as we have recently shown that EGFR is phosphorylated downstream of FGF10/FGFR2b recycling at the REs (Smith *et al.*, 2021). Thus, a picture of REs as a point of convergence for several signalling pathways and for coordination of long-term responses is clearly emerging.

Recycling is known to control cellular responses, including proliferation, migration, invasion and, as shown in this study, the rate of autophagy (Caswell *et al.*, 2008; Francavilla *et al.*,

2013, 2016). It is therefore not surprising that impeding the recycling of FGFR2b leads to dysregulated cellular proliferation and enhanced cell death. How the inhibition of active trafficking leads to enhanced apoptosis is not well characterised and deserve further investigations. However, we have started dissecting how cell proliferation is tightly regulated by multiple converging mechanisms downstream of FGFR2b, including receptor recycling and its duration (Francavilla *et al.*, 2013; Smith *et al.*, 2021), EGFR and ULK1 phosphorylation occurring at the REs (Smith *et al.*, 2021) (Figure 3-10), and suppression of autophagy (Figure 3-16). The G1/S checkpoint is known to be controlled by the homeostatic balances of nutrients, such as amino acids and sugars, all downstream of mTOR signalling (Mathiassen, De Zio and Cecconi, 2017). Thus, FGF10 may specifically alter the balance at this checkpoint by supressing the negative regulation of autophagy on cell cycle progression via mTOR/ULK1 during receptor recycling (Mathiassen, De Zio and Cecconi, 2017; Nowosad *et al.*, 2020; Yamasaki, Jin and Ohsumi, 2020). The combination of pharmacological inhibition of signalling and autophagy inhibitors has shown greater cytotoxic effects in several diseases (Aveic *et al.*, 2018). It is therefore time to speculate that such a combination may prove efficient in FGFR2b-driven genetic diseases or cancer, including breast cancer (Porębska *et al.*, 2018).

In conclusion, we have discovered a role for internalized FGFR2b in regulating autophagy from the REs. The approach described here and the datasets collected provide a resource to the cell signalling research community that can be used to further study the role of internalized, activated RTKs in modulating signalling cascades.

3.5 Methods

3.5.1 Reagents and Tools Table

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Experimental Models		
HeLa C3	ATCC	B6.129P2Gpr37tm1Dgen/J
BT20	LGC ATCC®	HTB-19
T47D	LGC ATCC®	HTB-133
NEB® 10-beta Competent E. coli (High Efficiency)	New England Biolabs	Cat. No: C3019H
Recombinant DNA		
eGFP-Rab11	Addgene	Plasmid #12674
eGFP-Rab11_S52N	Mutagenesis of eGFP-Rab11	Plasmid #12674
eGFP	Addgene	Plasmid #34680

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Dynamin_K44a-eGFP	Mutagenesis of Dynamin-eGFP, Addgene	Plasmid #34680
HA-FGFR2b	Francavilla et al. (2013)	PI: Prof Olsen
HA-FGFR2b-PCDH1	Generated for this study	
APEX2	Addgene	Plasmid #49386
HA-FGFR2b-APEX2-PCDH1	Generated for this study	
HA-FGFR2b-APEX2	Generated for this study	
eGFP-RAB11-APEX2	Generated for this study	
eGFP-APEX2	Generated for this study	
Antibodies		
p44/42 MAPK (Erk1/2) (137F5) Rabbit mAb	Cell Signaling Technology	4695
FGF Receptor 2 (D4L2V) Rabbit mAb	Cell Signaling Technology	23328
GFP (D5.1) Rabbit mAb	Cell Signaling Technology	2956
Anti-ERK 1/2 Antibody (MK1) (mouse monoclonal)	Santa Cruz Biotechnology	sc-135900
Monoclonal Anti-γ-Tubulin antibody produced in mouse	Sigma-Aldrich	T5326
Anti-Vinculin antibody, Mouse monoclonal	Sigma-Aldrich	V9264
GFP (D5.1) Rabbit mAb	Cell Signaling Technology	2956
Purified Mouse Anti-EEA1 (monoclonal)	BD Bioscience	610457
Phospho-FGF Receptor (Tyr653/654) (55H2) Mouse mAb	Cell Signaling Technology	3476

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Phospho-PLCγ1 (Tyr783) Antibody (rabbit polyclonal)	Cell Signaling Technology	2821
PLCγ1 Antibody (rabbit polyclonal)	Cell Signaling Technology	2822
Phospho-Shc (Tyr239/240) Antibody (rabbit polyclonal)	Cell Signaling Technology	2434S
Shc Antibody (rabbit polyclonal)	Cell Signaling Technology	2432
Phospho-FRS2-α (Tyr196) Antibody (rabbit polyclonal)	Cell Signaling Technology	3864
Anti-FRS2 Antibody (A-5) (mouse monoclonal)	Santa Cruz Biotechnolo gy	sc-17841
Anti-Histone H3 antibody - Nuclear Marker and ChIP Grade (rabbit polyclonal)	Abcam	ab-1791
Anti-HA (12CA5) (mouse monoclonal)	Sigma-Aldrich	11583816001
RAB25	Cell Signaling Technology	4314S
LC3B Antibody (rabbit polyclonal)	Cell Signaling Technology	2775
Phospho-Beclin-1 (Ser93) (D9A5G) Rabbit mAb	Cell Signaling Technology	14717
Beclin-1 (D40C5) Rabbit mAb	Cell Signaling Technology	3495
Phospho-ULK1 (Ser638) (D8K9O) Rabbit mAb	Cell Signaling Technology	14205
ULK1	Cell Signaling Technology	6439S
ANTI-PHOSPHO-RAPTOR(S863) (rabbit polyclonal)	Sigma-Aldrich	SAB1305088
Raptor (24C12) Rabbit mAb	Cell Signaling Technology	2280
pAMPK T172	Cell Signaling Technology	50081S

Reagent/Resource	Reference or Source	Identifier or Catalog Number
AMPK	Cell Signaling Technology	5831S
mTOR (7C10) Rabbit mAb	Cell Signaling Technology	
SH3BP4 Polyclonal antibody (rabbit)	Proteintech	17691-1-AP
Mouse LAMP-1/CD107a Lumenal Domain Antibody (polyclonal)	R&D systems	AF4320
Anti-LAMP1 antibody - Lysosome Marker (rabbit polyclonal)	Abcam	ab24170
CHMP1b	sc-514013	Santa Cruz Biotech
Rab7 Antibody (rabbit polyclonal)	Cell Signaling Technology	2094S
FIP1/RCP Antibody (rabbit polyclonal)	Novus Biologicals	NBP2-20033
Cleaved Caspase-3 (Asp175) Antibody (rabbit polyclonal)	Cell Signaling Technology	9661S
Peroxidase-AffiniPure F(ab')2 Fragment Goat Anti-Mouse IgG (H + L) (min X Hu, Bov, Hrs Sr Prot)	Stratech	115-036-045-JIR-0.5ml
Peroxidase-AffiniPure F(ab')2 Fragment Goat Anti-Rabbit IgG (H + L) (min X Hu Sr Prot)	Stratech	115-036-062-JIR-0.5ml
Goat anti-Rabbit IgG (H + L) Secondary Antibody, Alexa Fluor® 488 conjugate	Invitrogen	A11034
Goat anti-Mouse IgG (H + L) Secondary Antibody, Alexa Fluor® 488 conjugate	Invitrogen	A11001
Goat anti-Rabbit IgG (H + L) Secondary Antibody, Alexa Fluor® 568 conjugate	Invitrogen	A11011
Donkey Anti-Mouse IgG (H + L) Secondary Antibody, Alexa Fluor® 647 conjugate	Invitrogen	A31571
Donkey Anti-Rabbit IgG (H + L) Secondary Antibody, Alexa Fluor® 647 conjugate	Invitrogen	A31573
Oligonucleotides and sequence-based reagents		
ACADM F- CGTTTTCATTGGAGATCACAGC R-CCAAGACCTCCACAGTTCTCT	In house	n/a
ACC1	QuantiTect	QT00033838
ACOX1 F- CCCATAAGCCTTTGCCAGGA R-GGCTTCACCTGGGCATACTT	In house	n/a
ACOX2 F - GCACCCCGACATAGAGAGC R-CTGCGGAGTGCAGTGTTCT	In house	n/a
ACTB F- GCAAGCAGGAGTATGACGAG R-CAAATAAAGCCATGCCAATC	In house	n/a
APEX2 F-ATATATATATGTCGACGACTACAAGGATGACGACG ATAAGGG R-	In house	n/a

Reagent/Resource	Reference or Source	Identifier or Catalog Number
ATATATATATGTCGACTTAGTCCAGGGTCAGGCGCT CCAGGGG		
BIRC5 F- GAGGCTGGCTTCATCCACTG R- CTTTTTGCTTGTTGTTGGTCTCC	In house	n/a
CCND1 F- CATCTACACCGACAACTCCATC R- TCTGGCATTTTGGAGAGGAAG	In house	n/a
CD36 F- GACCTGCTTATCCAGAAGAC R- TTGCTGCTGTTCATCATAC	In house	n/a
FASN F- CTTCCGAGATTCCATCCTACGC R: TGGCAGTCAGGCTCACAAACG	In house	n/a
GLUT1 F- CCAGCTGCCATTGCCGTT R- GACGTAGGGACCACACAGTTGC	In house	n/a
LPL F- GACACAGCTGAGGACACTTG R- TGGAGTCTGGTTCTCTCTTG	In house	n/a
PPARA	QuantiTect	QT01006243
SCD F - TCCAGAGGAGGTACTACAAACCT R - GCACCACAGCATATCGCAAG	In house	n/a
TWIST F-GTTTTGCAGGCCAGTTTGAT R- TGCATGCATTCTCAAGAGGT	In house	n/a
VEGF	QuantiTect	QT01682072
Chemicals, enzymes and other reagents		
Trypsin porcine pancreas (proteomics grade)	Sigma-Aldrich	T6567
Lysyl Endopeptidase	FUJIFILM Wako Chemicals	2541
TiO beads "Titanspheres"	GL Sciences	5020-75000
Pre-cast gradient gel: Nu-PAGE 4-12% Bis-Tris Gel 1.0mm 10 well	Invitrogen	NP0321BOX
Sep-Pak Classic C18 cartridges	Waters	WAT051910
Solid Phase Extraction Disk “Empore” C18 (Octadecyl) 3 M	Agilent Technologies	2215
Solid Phase Extraction Disk “Empore” C8 (Octyl) 3 M	Agilent Technologies	2214
2,5-Dihydroxybenzoic acid	Sigma-Aldrich	85707
Acetonitrile	VWR Internation Ltd	1.00030.2500
Glacial Acetic Acid (HPLC Grade)	Fisher Scientific UK	10060000
Formic Acid (HPLC Grade)	Sigma-Aldrich	5438040250
Trifluoracetic Acid (Spectroscopy Grade)	Sigma-Aldrich	1082621000
Biotin phenol	Iris Biotech	

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Hydrogen peroxide	Sigma-Aldrich	
Trolox	Sigma-Aldrich	
Sodium ascorbate	VWR	
Streptavidin Sepharose High Performance 5 ml	Cytvia	17511301
TRIzol™ Reagent	ThermoFisher Scientific	Cat #15596026
Q5 High Fidelity 2x mastermix	NEB	M0429S
GoScript Reverse Transcription Mix, Random Primers	Promega	A2801
qPCRBIO SyGreen Mix Separate-ROX	pcr biosystems	PB20.14-50
exoSAP-IT	Life Technologies	78250.40.ul
Xtra-Clear Flat 8-Strip Caps	Star labs	I1400-0900-C
96-Well PCR Plate Non-Skirted Low Profile Natura	Star labs	E1403-0200-C
PURELINK QUICK MINI	NEB	K210010
T4 DNA Ligase 20,000 u	NEB	M0202S
Escort IV	SLS	L3287-1ML
Lipofectamine Transfection Readent	Life Technologies	18324020
RPMI 1640 Medium Glutamax Supplement (500ml)	Gibco	61870010
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma-Aldrich	D6429-500ML
Fetal Bovine Serum, South American origin	Life Technologies Ltd	10270106
Penicillin-Streptomycin (10,000 U/mL)	Life Technologies Ltd	15140122
Sodium Pyruvate soltution 100mM (100 mL)	Sigma-Aldrich	V5265-100ML
opti-MEM glutamax reduced serum media	Thermo Fisher Scientific	
Human FGF10	Francavilla et_al 2013	PI: Prof Olsen
Human FGF7	Francavilla et_al 2013	PI: Prof Olsen
Transferrin From Human Serum, Alexa Fluor™ 647 Conjugate	Invitrogen	T23366
Transferrin From Human Serum, Tetramethylrhodamine Conjugate	Invitrogen	T2872

Reagent/Resource	Reference or Source	Identifier or Catalog Number
DMSO	Sigma-Aldrich	276855-250ml
PD173074	Selleckchem	S1264
ULK-101	Selleckchem	S8793
SBI0206965	Selleckchem	S7885
Rapamycin	Selleckchem	S1039
Dynasore	Selleckchem	S8047
Monensin	Selleckchem	S2324
Glucose-6-phosphate	Selleckchem	S5518
Sodium Valporate	Spectrum Chemical	S2479
Fluspirilene	Merck	F100
Pierce Protease Inhibitor Tablets-20 tables	Life Technologies	A32963
Prestained Protein Standard Broad Range	Sigma-Aldrich	SDS7B2
Color Prestained Protein Standard Broad Range	NEB	P7712S
BSA	Melford Biolaboratories Ltd	A30075
Bradford Reagent	Bio-Rad	5000205
Clarity ECL	Bio-Rad	1705061
Albumin, Bovine Fraction V (BSA), 100 Grams Cat No: A30075-100.0	Melford Biolaboratories	A30075-100.0
MEMBRANE PROTRAN 0,45uM NC 300MMX4 M	VWR	10600002
Prolong Diamond Antifade Mountant-2 mL	Life Technologies	965
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Life Technologies	
Other		
Venor®GeM Classic Mycoplasma PCR Detection Kit(100 tests)	Cambridge Bioscience	11-1100
ProtoScript; II First Strand cDNA Synthesis Kit	New England Biolabs	E6560L
Isolate II PCR and Gel kit	Bioline	BIO-52059
Isolate II plasmid mini kit	Bioline	BIO-52056
Rnase-Free Dnase Set	Qiagen	
Confocal Microscope Leica Sp8 Inverted	Leica	
Mx3000P qPCR machine	Agilent	
Universal Hood II Gel Molecular Imaging System ChemiDoc XRS+	Bio-Rad	Look up
Synergy H1 microplate reader	BioTek	Look up
Utimate® 3000 Rapid Separation LC	Dionex	

Reagent/Resource	Reference or Source	Identifier or Catalog Number
QE-HF LC-MS/MS	Thermo Fisher Scientific	

3.5.2 Experimental Models

3.5.2.1 Cell Culture

Human epithelial cell lines were purchased from ATCC, authenticated through short tandem repeat (STA) analysis of 21 markers by Eurofins Genomics, checked monthly for mycoplasma via a PCR-based detection assay (Venor®GeM – Cambio), and grown in the indicated media supplemented with 2 mM L-glutamine and 100 U/ml penicillin, 100 µg/ml streptomycin and 10% fetal bovine serum. HeLa and BT20 were grown in DMEM. T47D were grown in RPMI.

3.5.2.2 Transfection protocol

All transfections were carried out in Gibco opti-MEM glutamax reduced serum media (ThermoFisher Scientific). HeLa cells were transfected using Lipofectamine 2000 (ThermoFisher Scientific) according to the manufacturer's instructions, 24 h after RNA interference transfection where indicated. T47D and BT20 cells were transfected using Escort IV according to manufacturer instructions, same as above. Assays were performed 36 h after transfection, as previously described (Smith *et al.*, 2021).

HeLa cells stably expressing FGFR2b or FGFR2b-APEX2 are denoted as follows: HeLa_FGFR2bST, HeLa_FGFR2b-APEX2ST. All cells transfected with the following constructs were transiently transfected: GFP, GFP-DnDNM2, GFP-DnRAB11, GFP-RAB11, RAB11-APEX2, GFP-APEX2.

3.5.3 Quantitative Proteomics

3.5.3.1 Sample Preparation

For all experiments, each treatment was analysed in biological triplicates.

<u>HeLa samples for phosphoproteomics:</u> Cells were washed with PBS and lysed in ice-cold 1% triton lysis buffer supplemented with Pierce protease inhibitor tablet (Life Technologies) and phosphatase inhibitors: 5 nM Na3VO4, 5 mM NaF and 5 mM b-glycerophosphate. 5 mg of protein was obtained for each experimental condition. Proteins were precipitated overnight at -20°C in four-fold excess of ice-cold acetone. The acetone-precipitated proteins were solubilized in denaturation buffer (6 M urea, 2 M thiourea in 10 mM HEPES pH 8). Cysteines were reduced with 1 mM dithiothreitol (DTT) and alkylated with 5.5 mM chloroacetamide (CAA). Proteins were digested with endoproteinase Lys-C (Wako, Osaka, Japan) and

sequencing grade modified trypsin (modified sequencing grade, Sigma) followed by quenching with 1% trifluoroacetic acid (TFA). Peptides were purified using reversed-phase Sep-Pak C18 cartridges (Waters, Milford, MA) and eluted with 50% ACN and enriched for phosphoserine-, phosphothreonine- and phosphotyrosine-containing peptides, with Titansphere chromatography. Six mL of 12% TFA in ACN was added to the eluted peptides and subsequently enriched with TiO2 beads (5 µm, GL Sciences Inc., Tokyo, Japan). The beads were suspended in 20 mg/mL 2,5-dihydroxybenzoic acid (DHB), 80% ACN, and 6% TFA and the samples were incubated in a sample to bead ratio of 1:2 (w/w) in batch mode for 15 min with rotation. After 5 min centrifugation the beads were washed with 10% ACN, 6% TFA followed by 40% ACN, 6% TFA and collected on C8 STAGE-tips and finally washed by 80% ACN, 6% TFA. Elution of phosphorylated peptides was done with 20ul 5% NH3 followed by 20 µL 10% NH3 in 25% ACN, which were evaporated to a final volume of 5 µL in a sped vacuum. The concentrated phosphorylated peptides were acidified with addition of 20 µL 0.1% TFA, 5% ACN and loaded on C18 STAGE-tips. Peptides were eluted from STAGE-tips with 20 µL of 40% ACN followed by 10 µL 60% ACN and ACN and reduced to 5 µL by SpeedVac and 5 µL 0.1% FA, 5% ACN was added (Smith *et al.*, 2021).

<u>HeLa samples for proteomics:</u> 30 ug of protein was collected as a proteome sample for in-gel digestion (Francavilla *et al.*, 2016). Samples were prepared in lysis buffer as above containing 10 mM DTT, alkylated with 5.5 mM CAA and run on 1.00 mm Invitrogen NuPAGE 4-12% Bis-Tris Gel (Invitrogen) in NuPAGE MOPs running buffer (Invitrogen). Gels were washed in ddH_2O, and proteins subsequently fixed and stained with Colloidal Coomassie Stain (Invitrogen). Each sample was equally separated into four fractions, and de-stained using 50% 20 mM ammonium bicarconate (NH_4HCO_3) + 50% ethanol (EtOH). Peptides were digested by trypsin in 50 mM NH_4HCO_3, neutralised and extracted using increasing concentrations of ACN, starting with 50% ACN, 3% TFA and ending with 100% ACN. Digested peptides were evaporated to a final volume of 100 µL in a speed vacuum and loaded on C18 STAGE-tips. Peptides were eluted from STAGE-tips with 20 µL of 40% ACN followed by 10 µL 60% ACN and ACN and reduced to 5 µL by SpeedVac and 5 µL 0.1% FA, 5% ACN added.

<u>T47D samples for proteomics and phosphoproteomics:</u> Cells were washed with PBS and prepared as described above up to elution off Sep-Pak C18 cartridges (Waters, Milford, MA) with 50% ACN. Prior to enrichment for phosphoserine-, phosphothreonine- and phosphotyrosine-containing peptides, with Titansphere chromatography, a small amount of the eluted peptides (1%) was taken for proteome analysis: after evaporation in a speed vacuum, peptides were resuspended in 40 µl of 0.1% TFA, 5% ACN and loaded on C18 STAGE-

tips. Six mL of 12% TFA in ACN was added to the eluted peptides and subsequently enriched with TiO2 beads (5 µm, GL Sciences Inc., Tokyo, Japan). The beads were suspended in 20 mg/mL 2,5-dihydroxybenzoic acid (DHB), 80% ACN, and 6% TFA and the samples were incubated in a sample to bead ratio of 1:2 (w/w) in batch mode for 15 min with rotation. After 5 min centrifugation the supernatant were collected and incubated a second time with a two-fold dilution of the previous bead suspension. Sample preparation continued as described above.

HeLa sample for global and proximal proteomics and phosphoproteomics: Cells were pre-incubated for 40 minutes with Biotin Phenol (Iris Biotech) and either left untreated or treated with FGF10 (100 ng/mL) for 40 minutes. Hydrogen peroxide (Sigma Aldrich) was added for 1 minute before quenching with Trolox (Sigma Aldrich) and sodium ascorbate (VWR) during ice cold lysis. Cells were lysed using APEX-RIPA buffer (50mM Tris-HCl pH 7.5. 150mM NaCl. 0.1% SDS. 1% Triton. 0.5% sodium deoxycholate. 1x protease inhibitor cocktail. 1mM PMSF. 10mM sodium ascorbate. 10mM sodium azide. 5mM Trolox) and protein extracted as described above. For each sample, 15 mg of protein was collected. 120 µg was taken and run on a gradient gel to acquire the global proteome analysis. 5 mg was separated from the lysate and precipitated in acetone and processed as above to obtain the global phosphoproteome. The rest of the lysate was enriched for biotinylation using a 2-hour RT pull-down with streptavidin beads, to generate the proximal proteome. A fifth of the bead slurry was stripped using boiling 4x sample buffer enrich with biotin, the supernatant was run on a gradient gel to acquire the proximal proteome. The remaining streptavidin bead slurry was stripped using boiling 8M guanidine ph1.5 supplemented in 5mM TCEP and 10mM CAA. Reduced samples where then processed with LysC for 60min RT, diluted to 1M guanidine using Tris 25mM pH8.5, before digestion with trypsin and enrichment for phosphorylated peptides as outlined above.

3.5.3.2 Mass Spectrometry

Purified peptides were analysed by LC-MS/MS using an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to a QE-HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometer (Smith *et al.*, 2021). Mobile phase A was 0.1% FA in water and mobile phase B was 0.1% FA in ACN and the column was a 75 mm x 250 µm inner diameter 1.7 mM CSH C18, analytical column (Waters). A 1ml aliquot of the sample (for proteome analysis) or a 3ml aliquot was transferred to a 5ml loop and loaded on to the column at a flow of 300nl/min at 5% B for 5 and 13 min, respectively. The loop was then taken out of line and the flow was reduced from 300nl/min to 200nl/min in 1 min., and to

7% B. Peptides were separated using a gradient that went from 7% to 18% B in 64 min., then from 18% to 27% B in 8 min. and finally from 27% B to 60% B in 1 min. The column was washed at 60% B for 3 min. and then REs-equilibrated for a further 6.5 minutes. At 85 minutes the flow was increased to 300nl/min until the end of the run at 90min. Mass spectrometry data was acquired in a data dependent manner for 90 min. in positive mode. Peptides were selected for fragmentation automatically by data dependent analysis on a basis of the top 8 (phosphoproteome analysis) or top 12 (proteome analysis) with m/z between 300 to 1750Th and a charge state of 2, 3 or 4 with a dynamic exclusion set at 15 sec. The MS Resolution was set at 120,000 with an AGC target of 3e6 and a maximum fill time set at 20ms. The MS2 Resolution was set to 60,000, with an AGC target of 2e5, and a maximum fill time of 110 ms for Top12 methods, and 30,000, with an AGC target of 2e5, and a maximum fill time of 45 ms for Top8 analysis. The isolation window was of 1.3Th and the collision energy was of 28.

3.5.3.3 Raw Files Analysis

Raw data were analysed by the MaxQuant software suite (https://www.maxquant.org; version 1.6.2.6 and 1.5.6.5) using the integrated Andromeda search engine (Cox *et al.*, 2011). Proteins were identified by searching the HCD-MS/MS peak lists against a target/decoy version of the human Uniprot Knowledgebase database that consisted of the complete proteome sets and isoforms (v.2019; https.//uniprot.org/proteomes/UP000005640_9606) supplemented with commonly observed contaminants such as porcine trypsin and bovine serum proteins. Tandem mass spectra were initially matched with a mass tolerance of 7 ppm on precursor masses and 0.02 Da or 20 ppm for fragment ions. Cysteine carbamidomethylation was searched as a fixed modification. Protein N-acetylation, N-pyro-glutamine, oxidized methionine, and phosphorylation of serine, threonine, and tyrosine were searched as variable modifications for the phosphoproteomes. Protein N-acetylation, oxidized methionine and deamidation of asparagine and glutamine were searched as variable modifications for the proteome experiments. Biotinylation by BP (Y; (C18H23N3O3S)) was included as a variable and fixed modification for global and proximal raw files. Label-free parameters were used for all the analysis as described (Cox *et al.*, 2014). False discovery rate was set to 0.01 for peptides, proteins, and modification sites. Minimal peptide length was six amino acids. Site localization probabilities were calculated by MaxQuant using the PTM scoring algorithm (Olsen *et al.*, 2006).The dataset were filtered by posterior error probability to achieve a false discovery rate below 1% for peptides, proteins and modification sites. Only peptides with Andromeda score >40 were included.

3.5.3.4 Data and Statistical Analysis

All statistical and bioinformatics analyses were done using the freely available software Perseus, version 1.6.5.0 or 1.6.2.1. (Tyanova *et al.*, 2016), R framework and Bioconductor (Huber *et al*, 2015), Python framework (available at http://www.python.org), SubcellulaRVis (available at http://phenome.manchester.ac.uk/subcellular/), STRING v11.5 (Szklarczyk *et al.*, 2019, p. 11), Cytoscape (version 3.7.2) (Shannon *et al*, 2003). Over-representation analysis (ORA) of KEGG terms was performed using Enrichr and the EnrichR R inferface (Kuleshov *et al.*, 2016).

All measured peptide intensities were normalized using the "normalizeQuantiles" function from the Bioconductor R-package *limma*. Potential contaminant proteins or phosphorylated-peptides and peptides or phosphorylated-peptides matching the reverse sequence database were removed. For all datasets, phosphorylated peptides with a localization score greater than 0.75 were included in the downstream bioinformatics analysis. Pearson correlation was calculated in R.

For both the HeLa and the T47D phosphoproteomics datasets, samples were normalized separately and were grouped based on treatment and only phosphorylated peptides with values in all three replicates of at least one treatment were included in further analysis. Missing values were subsequently imputed from a normal distribution using Perseus default settings. The HeLa-FGFR2b dataset was separated into eleven clusters by fuzzy c-means clustering using the "fanny" function from the R package "cluster" performed after multi-sample ANOVA test with FDR < 0.001 in Perseus. Python (version: 3.7.4) were used for clustering of phosphorylated sites with the same pattern of regulation in T47D as seen in HeLa-FGFR2bST, Kernelised Parzen window classifier was chosen as a distance-based non-parametric algorithm which given it followed Bayesian probability theorem enabled close comparison the fuzzy c-means method of clustering. Over-representation analysis (ORA) of KEGG terms was performed using Enrichr and the EnrichR R inferface (Kuleshov *et al.*, 2016) and significantly over-represented terms within the data were represented in dot plots. The SubCellulaRVis tool to visualize the sub-cellular localization of proteins.

For the SRP datasets, phosphorylated site intensities acquired from global and proximal MS runs were normalized separately. Replicates were summarised by calculating the median of normalized intensities. Missing values were imputed using random draws from a truncated distribution with the *impute.QRLIC()* function from the R CRAN package *imputeLCMD.* PCA, Students' t-test and one-way ANOVA were calculated using the prcomp(), t.test() and aov()

functions in R, respectively. Hierarchical clustering was performed using the *hclust()* function in R. ORA and use of SubCellulaRVis tool were performed as described above. Networks were constructed using STRING; only interactions with an experimental confidence > 0.4 were included. KEGG pathways were extracted using the *EnrichmentBrowser* package from Bioconductor.

3.5.4 Biochemical and Functional Assays

3.5.4.1 Cell Lysis and Western Blotting

Cells were serum starved overnight in serum-free medium and stimulated for the indicated time points with 100 ng/mL of FGF7 or FGF10. Where indicated, cells were pre-incubated for 2 hours with 100 nM PD173074, 0.5 µM Rapamycin, 2 µM ULK101 and 10 µM SBI-0206965. Control cells were pre-incubated with DMSO alone. After stimulation, cell extraction and immunoblotting were performed as previously described (Smith *et al.*, 2021). Proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Protran, Biosciences). Proteins of interest were visualized using specific antibodies, followed by peroxidase-conjugated secondary antibodies and by an enhanced chemiluminescence kit (Amersham Biosciences). Blots were visualised using the Universal Hood II Gel Molecular Imaging System (Bio-Rad). Each experiment was repeated at least three times and produced similar results.

3.5.4.2 Biotinylation assays

Biotinylation pull downs experiments were performed as described previously (Smith *et al.*, 2021). Briefly, cells were pre-incubated (40 min.) with Biotin Phenol (Iris Biotech) after stimulation with ligands. Hydrogen peroxide (Sigma Aldrich) was added for 1 min. before quenching with Trolox (Sigma Aldrich) and Sodium ascorbate (VWR) during ice cold lysis. A 2 hour RT pull-down with streptavidin beads was then performed and the supernatant was run against the bound proteins and the total lysates.

3.5.4.3 RNA Isolation and real-time PCR analysis

RNA from cell lines was isolated with TRIZOL® (Invitrogen). After chloroform extraction and centrifugation, 5 µg RNA was DNase treated using RNase-Free DNase Set (Qiagen) and 1 µg of DNase treated RNA was then taken for cDNA synthesis using the Protoscript I first strand cDNA synthesis kit (New England Biolabs). Selected genes were amplified by quantitative real time PCR (RT-qPCR) using Sygreen (PCR Biosystems). Relative expression was calculated using the delta-delta CT methodology and beta-actin was used as reference housekeeping gene. qPCR machine used was Applied Biosystems MX300P.

3.5.4.4 Functional Assays

EdU Incorporation. Indicated cells were labelled with 20 µM 5-ethynyl-2'-deoxyuridine (EdU) for 4 h and processed following the manufacturer's protocol (Click-iT® EdU Alexa Fluor® 488 Imaging Kit, Thermo Fisher). Prior to imaging cells were then stained with 5ng/ml Hoecsht 3342 for 15 minutes. Stained cells were analysed using a using a Leica microscope system. Statistical analysis was performed at the endpoint across repeats, as indicated in the figure legends.

Cleaved caspase assay. Apoptosis was measured in cells receiving either 24-hour treatment with FGF10Appropriately treated cells were incubated with 20 mM CellEvent™ Caspase-3/7 Green Detection Reagent (Invitrogen) made to 100X in PBS for 4h in darkness then washed thoroughly in 1X PBS. Fluorescence was measured at 502 nm excitation and 530 nm emission. Statistical analysis was performed at the endpoint across repeats, as indicated in the figure legends.

Autophagy. Cells were assayed for autophagy using 5 mM Acridine Orange (Sigma) for 30 minutes after which excess was removed by thorough washing with 1X PBS. This fluorophore appears green when diffuse but is shifted to the red end of the spectrum when accumulated in acidic vesicles (SenthilKumar *et al*, 2019). As such, excitation/emission wavelengths of 500/526 nm were used to measure intensity of diffuse acridine orange (non-specific) and 460/650 nm to assess autophagic staining. The ratio of these values represents stained autophagosomes. Statistical analysis was performed at the endpoint across repeats, as indicated in the figure legends.

3.5.5 Immunofluorescence and Quantification

Immunofluorescence staining was performed as previously described (Smith *et al.*, 2021). To detect FGFR2 cells were incubated with anti-FGFR2 antibody (Cell Signalling) for 45 minutes with gentle agitation. The binding of the antibody did not activate receptor signalling in untreated cells nor induced receptor internalization as previously reported (Smith *et al.*, 2021). After stimulation cells were incubated at 37°C for different time points. At each time point, non-permeabilized cells were either fixed to visualize the receptor on the cell surface (plasma membrane) or acid-washed in ice-cold buffer (50 mM glycine, pH 2.5) to remove surface-bound antibody. Acid-washed cells were then fixed and permeabilized to visualize the internalized receptor (cytoplasm). Finally, to detect FGFR2b cells were stained with AlexaFluor488-conjugated donkey anti-mouse or anti-rabbit (Jackson ImmunoResearch Laboratories). Nuclei were stained with DAPI. Coverslips were then mounted in mounting medium (Vectashield; Vector Laboratories).

For co-localization experiments, cells were acid washed, fixed, permeabilized with 0.02% saponin (Sigma), treated with the indicated primary antibody for 60 min at 37 °C, and stained with AlexaFluor488 (or 568 or 647)-conjugated donkey anti-mouse or anti-rabbit. Samples expressing GFP-tagged proteins were kept in the dark. Nuclei were stained with DAPI. Coverslips were then mounted in mounting medium (Vectashield; Vector Laboratories).

All the images were acquired at room temperature on a Leica TCS SP8 AOBS inverted confocal using a 100x oil immersion objective and 2.5x or 3x confocal zoom. The confocal settings were as follows: pinhole, 1 airy unit, format, 1024 x 1024. Images were collected using the following detection mirror settings: FITC 494-530nm; Texas red 602-665nm; Cy5 640-690nm. The images were collected sequentially. Raw images were exported as .lsm files, and adjustments in image contrast and brightness were applied identical for all images in a given experiment using the freely available software Image J v. 1.52p (Schindelin *et al*, 2012).

Quantification of rFGFR2b ecycling, Co-localization, and Expression Fraction was performed as recently described in detail (Smith *et al.*, 2021). The scripts for the quantification of co-localization were written in the Python language and the code for Costes-adjusted MCC was taken verbatim from the CellProfiler code base. Statistical analysis was performed across repeats, as indicated in the figure legends.

3.5.6 Data Availability

The mass spectrometry proteomics data in Thermo Scientific's *.raw format have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al*, 2019) partner repository with the following dataset identifiers: Submission details: Project Name: Signalling and recycling endosome_dataset 01. Project accession: PXD028370. At the time of thesis submission data was only available to manuscript reviewers and will be made available upon publication. All other data supporting the findings of this study are available within the article and its supplemental information files or from the Lead Contact upon reasonable request.

4 SubcellulaRVis: an app for the visualization of groups of proteins in subcellular compartments

4.1 Abstract

High-throughput 'omics methods result in lists of differentially regulated or expressed genes or proteins, whose function is generally studied through statistical methods such as enrichment analyses. One aspect of protein regulation is subcellular localization, which is crucial for their correct processing and function and can change in response to various cellular stimuli. Enrichment of proteins for subcellular compartments is often based on Gene Ontology Cellular Compartment annotations. Results of enrichment are typically visualized using bar-charts, however enrichment analyses can result in a long list of significant annotations which are highly specific, preventing researchers from gaining a broad understanding of the subcellular compartments their proteins of interest may be located in. Schematic visualization of known subcellular locations has become increasingly available for single proteins via the UniProt and COMPARTMENTS platforms. However, it is not currently available for a list of proteins (e.g. from the same experiment) or for visualizing the results of enrichment analyses.

To generate an easy-to-interpret visualization of protein subcellular localization after enrichment we developed the SubcellulaRVis web app, which visualizes the enrichment of subcellular locations of gene lists in an easy and impactful manner. SubcellulaRVis projects the results of enrichment analysis on a graphical representation of a eukaryotic cell. Implemented as a web app and an R package, this tool is user-friendly, provides exportable results in different formats, and can be used for gene lists derived from multiple organisms.

Here, we show the power of SubcellulaRVis to assign proteins to the correct subcellular compartment using gene list enriched in previously published spatial proteomics datasets. We envision SubcellulaRVis will be useful for cell biologists with limited bioinformatics expertise wanting to perform precise and quick enrichment analysis and immediate visualization of gene lists.

4.2 Author Summary

Cells contain intracellular compartments, such as membrane-bound organelles and the nucleus, and are surrounded by a plasma membrane. Proteins can be found in different cellular compartments; depending on the subcellular compartment they are localized to, they can be differentially regulated or perform location-specific functions. High-throughput 'omics experiments result in a list of proteins or genes of interest; one way in which their functional role can be understood is through their subcellular localization, as deduced through statistical enrichment for Gene Ontology Cellular Component (GOCC) annotations or similar. We have designed a bioinformatic tool, named SubcellulaRVis, that compellingly visualizes the results of protein localization after enrichment and simplifies the results for a quick interpretation. We demonstrate that SubcellulaRVis precisely describes the subcellular localization of gene lists whose locations have been previously ascertained in publications. SubcellulaRVis can be accessed via the web or as a stand-alone app. SubcellulaRVis will be useful for experimental biologists with limited bioinformatics expertise analyzing data related to cellular signalling, protein (re)localization and regulation, and location-specific functional modules within the cell.

4.3 Introduction

The localization of proteins within cells is critical for their processing, function, and regulation. Cellular homeostasis and response to environmental signals are also dependent on sequestering or dynamic re-localization of proteins in specific compartments (Hung and Link, 2011). Regardless of whether a protein is relocated in the cell or remains in a single compartment for the duration of its life, spatial regulation is critical for protein function. For example, upon translation, proteins bound for secretion will be trafficked through multiple compartments that compose the endomembrane system; contained within these compartments are "quality control" proteins that mediate correct folding or degradation for malformed proteins (Phillips, Gomez-Navarro and Miller, 2020). Cellular processes also occur in discrete, non-membrane bound locations, such as degradation within the proteasome, or transcriptional regulation in stress granules (Protter and Parker, 2016; Rousseau and Bertolotti, 2018). It is also increasingly appreciated that proteins which are dynamically recruited to different subcellular locations form functionally active modules that may be location specific, such as those that form and disperse at the lysosomal surface in response to growth factor or nutrient sensing (Inpanathan and Botelho, 2019).

Co-regulated proteins or genes are commonly identified using high-throughput 'omics experiments, such as those based on transcriptomics, proteomics and phosphoproteomics.

The regulated proteins or genes identified in these experiments are often assessed based on enrichment for annotations to particular biological characteristics or participation in biological pathways. Annotations based on subcellular localization are stored in the Gene Ontology aspect Cellular Component (GOCC) or the Jensen COMPARTMENTS database (Binder *et al.*, 2014; The Gene Ontology Consortium, 2018). Enrichment for these annotations can provide initial indications of location-specific roles for a protein or gene list. Moreover, when analyzing spatially resolved data, annotations describing subcellular localization also provide an important data quality control. This is exemplified in the analysis of proteomics data that utilize proximity dependent biotinylation, in which bait proteins (known, characterized proteins tagged with a biotin ligase or peroxidase) covalently modify neighboring proteins (termed "prey") through the addition of biotin. Isolating the biotinylated proteins generates a spatial interactome of the bait protein. Recently, BioID-based proximity-dependent biotinylation was used to generate cell-wide maps of proteins localized to particular organelles (Go *et al.*, 2021). As demonstrated in this work, calculating enrichment for GOCC terms is a useful method for understanding whether the bait has been correctly tagged, in which case one would expect enrichment for the subcellular compartment the bait is commonly found.

Several web-based tools, such as EnrichR (Kuleshov *et al.*, 2016), can be used to perform enrichment analyses on protein or gene lists based on the GOCC or compartments databases. These analyses are accessible to biologists with minimal bioinformatics skills and are useful for understanding the subcellular compartments overrepresented in a list of proteins or genes of interest. However, a significant barrier in the interpretation of the results is that annotations are often highly specific, meaning it can be difficult to extract general trends from the potentially long list of returned annotations. For example, the enrichment for GOCC terms (calculated using EnrichR) on a list of GPI-anchored proteins (extracted from UniProt), results in a list of terms that are difficult to immediately generalize (Table 4-1). In contrast, certain cellular compartments are poorly characterized, for example the individual compartments of the endosomal system (Naslavsky and Caplan, 2018), and therefore analyses of proteins localized in such compartments may be noisy and incomplete.

Table 4-1. Results of GOCC enrichment from EnrichR on a list of 139 GPI-anchored proteins

GOCC Term	Overlap	P-value	Adjusted P-value
anchored component of plasma membrane (GO:0046658)	33/46	3.16E-63	2.31E-61
intrinsic component of external side of plasma membrane (GO:0031233)	17/24	1.46E-32	5.35E-31
anchored component of external side of plasma membrane (GO:0031362)	16/20	3.50E-32	8.52E-31
membrane raft (GO:0045121)	15/163	3.20E-13	5.84E-12
secretory granule membrane (GO:0030667)	16/274	5.34E-11	7.79E-10
Golgi lumen (GO:0005796)	8/100	3.86E-07	4.70E-06
lysosomal lumen (GO:0043202)	7/86	1.88E-06	1.96E-05
specific granule membrane (GO:0035579)	6/91	3.52E-05	3.22E-04
vacuolar lumen (GO:0005775)	7/161	1.12E-04	9.11E-04
plasma membrane raft (GO:0044853)	5/82	2.34E-04	0.001708

We have implemented a Shiny web app, named SubcellulaRVis, that provides graphical visualization of subcellular compartment enrichment from a list of multiple genes or proteins. By performing the enrichment test using condensed annotations, so that there is a single term for each major intracellular compartment in the results, the analyses can be concisely visualized and interpreted. SubcellulaRVis represents an alternative to the visualization provided on the web interfaces of UniProt and COMPARTMENTS which only visualize the localization of a single protein (Binder *et al.*, 2014; The UniProt Consortium, 2021). A major novelty of our approach is the interactive visualization on a schematic of the eukaryotic cell of GOCC enrichment (Fig. 1), allowing for user-friendly interpretation, exploration, and presentation of the data. The visualization can be exported as a static image or as a table of the data behind the visualization. SubcellulaRVis provides a solution for non-bioinformaticians to investigate the subcellular localization of the proteins within a dataset of interest and standardizes the visualization of these results. The app is available at: http://phenome.manchester.ac.uk/subcellular/.

4.4 Design and Implementation

Our tool aims to simplify the interpretation of GOCC enrichment analyses through visualization of a standardized schematic of the cell. We selected subcellular compartments that are non-transient and common to eukaryotes. Though it is a non-exhaustive list, we anticipate it will meet the needs of the majority of cell biology researchers.

We first defined each cellular compartment visualized on the schematic (as shown in Figure 4-2) in respect to the GOCC terms that could describe it. We did this by utilizing the hierarchical structure of GO. The GO terms are hierarchically organized, with GO terms towards the "top" of the hierarchy being more general descriptors (e.g. GO:0005886, plasma membrane.) whilst terms towards the "bottom" of the hierarchy are more specific (e.g. GO:0005901, caveola) (Ashburner *et al.*, 2000). Each term in the GO hierarchy has parent terms (those that are higher in the hierarchy) and may also have child terms (those that are lower in the hierarchy). We identified the highest level GOCC term that would best describe each of the compartments visualized on the schematic (Table 4-2). Then, using the GO.db package in Bioconductor (version 3.13) (Huber *et al.*, 2015), the child terms of the high-level GOCC terms were extracted and associated together. We call these grouped, compartment-specific terms the SubcellulaRVis compartment and use the high-level parent term to describe them, as shown in Table 1, hereafter. We extracted all the genes annotated to the terms in each SubcellulaRVis compartment using the species related AnnotationData libraries in Bioconductor (version 3.13), to create gene sets against which to calculate enrichment (Table 4-2). The annotation lists were generated for multiple species: *H. sapiens, M. musculus, D. melanogaster, S. cerevisiae, R. rattus,* and X. laevis.

Table 4-2. Number of child terms for each general parent term selected.

Compartment	GOID	Number of child terms	Number of genes annotated					
			H. sapiens	M. musculus	D. melanogaster	S. cerevisiae	R. rattus	X. laevis
Cytoplasm	GO:0005737	1230	11847	11262	4700	4083	10436	4884
Cytoskeleton	GO:0005856	273	2302	2213	523	255	2111	889
Endoplasmic reticulum	GO:0005783	110	1961	1777	499	635	1700	685
Endosome	GO:0005768	77	978	906	240	189	846	296
Extracellular region	GO:0005576	94	4570	2697	1036	110	2072	1102
Golgi apparatus	GO:0005794	47	1608	1468	379	285	1398	622
Intracellular vesicle	GO:0097708	294	2405	1965	423	297	1848	555
Lysosome	GO:0005764	38	697	472	106	6	456	145
Mitochondrion	GO:0005739	96	1627	1852	863	1129	1570	650
Nucleus	GO:0005634	492	8072	7127	2827	2302	6585	3673
Peroxisome	GO:0005777	19	136	146	98	88	149	66
Plasma membrane	GO:0005886	471	5726	5343	1383	507	5779	1878
Ribosome	GO:0005840	20	243	231	285	267	295	239

Enrichment is calculated on the user supplied list of HUGO Gene Nomenclature Committee (HGNC) symbols. The genes are annotated to the GOCC terms based on the annotation lists generated for each species, and the annotations are then grouped based on the SubcellulaRVis compartment. A standard enrichment test is performed using the hypergeometric probability function, *phyper(),* in R, to calculate the enrichment for each SubcellulaRVis compartment in the user-supplied gene list and corrected using the false discovery rate (FDR). Though it has limitations (Naeem *et al.*, 2012; Cao and Zhang, 2014), the hypergeometric test was selected as values associated to the genes or proteins (e.g. expression or abundance values) do not need to be supplied in order to perform the test. To improve the accuracy of this calculation, the user can also input a background population of expressed genes or proteins in their sample, if known (as discussed in (Grossmann *et al.*, 2007)). In the absence of the user-supplied background population list, the calculation is performed based on all the genes in the reference genomes (from the Bioconductor AnnotationData databases, as described above).

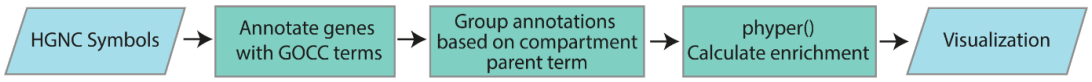


Figure 4-1. SubcellulaRVis workflow

The SubcellulaRVis tool has been implemented as a Shiny web app (http://phenome.manchester.ac.uk/subcellular/) and a standalone app with R package (https://github.com/JoWatson2011/subcellularvis). The R package will be submitted to Bioconductor upon acceptance of the article. The visualization within the app has been implemented using ggplot2 and Plotly, allowing interactive exploration of the results when viewed on a web browser. All dependencies are described in the DESCRIPTION file of the R package (https://github.com/JoWatson2011/subcellularvis) and will be installed along with the package, as described in the README file.

4.5 Results

4.5.1 Features of SubcellulaRVis

The key feature of the SubcellulaRVis app is the visualization of GOCC enrichment results on a graphical representation of a eukaryotic cell (as seen in Figure 4-2), allowing for rapid and simple interpretation of the predominant localization of proteins of interest. By providing

visualizing for a group of proteins, we provide an alternative to similar tools that visualize the localization of single proteins (Binder *et al.*, 2014; The UniProt Consortium, 2021).

The app provides an interactive view of protein localization, and a static image of the plot can be exported; the characteristics of both the image and the plot (such as text size or color scale) can be chosen by the user. This is demonstrated by the test data, available in the app, which is a list of 139 proteins which are annotated as GPI-anchored in UniProt (The UniProt Consortium, 2021). Figure 4-2 visualizes the enrichment of these proteins for the membrane bound organelles of the secretory pathway, plasma membrane and for the extracellular space (Table 4-3). Cellular compartments in white are not enriched for the gene list.

Using the app or the associated R package, the user can view the results of the enrichment analysis graphically, as discussed above, or in tabular form. A standard enrichment including all the GOCC terms (rather than the summarized terms) can also be calculated and can be viewed in a separate tab on the app; also included is the SubcellulaRVis compartment each of these terms is assigned to, allowing the user full understanding of the way their data has been analyzed and visualized.

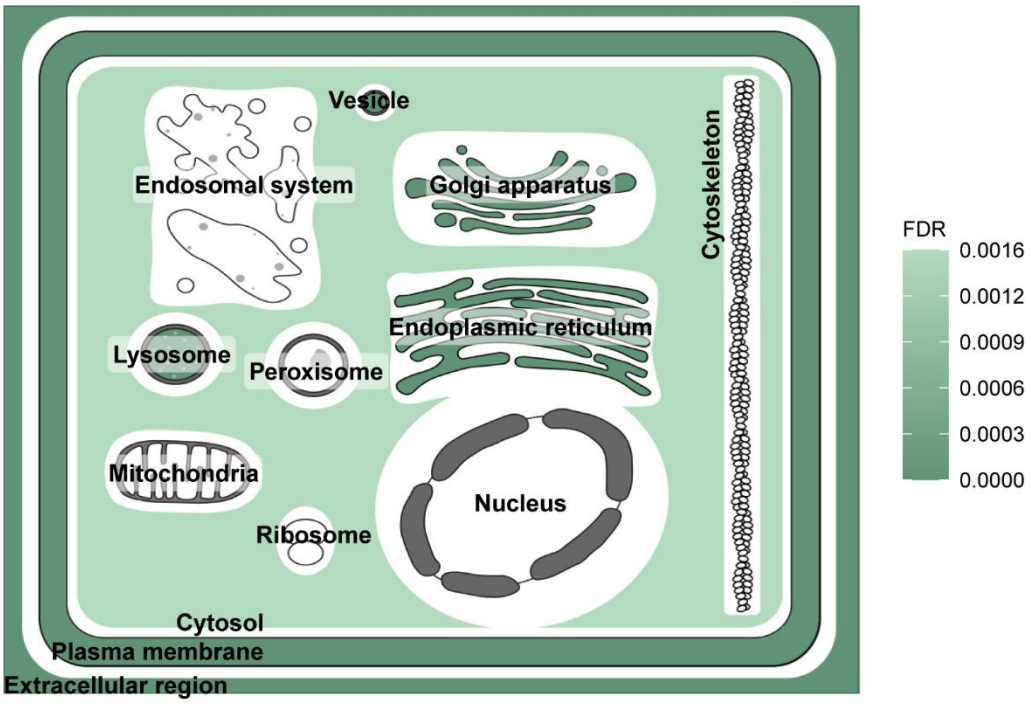


Figure 4-2. Visualization of enrichment results using SubcellulaRVis on a gene list of GPI anchored proteins (The UniProt Consortium, 2021).

Table 4-3. Enrichment results using SubcellulaRVis for gene list of 139 GPI-anchored proteins. Break in table indicates non-significant enrichment.

Compartment	p-value	FDR	Number of genes in user list annotated to compartment
Plasma membrane	4.70E-120	6.58E-119	137
Extracellular region	7.01E-83	9.11E-82	111
Golgi apparatus	1.15E-08	1.38E-07	21
Intracellular vesicle	1.34E-08	1.47E-07	26
Endoplasmic reticulum	3.61E-07	3.61E-06	21
Lysosome	4.84E-07	4.36E-06	12
Vacuole	1.97E-06	1.57E-05	12
Cytoplasm	0.000223	0.001559	57
Endosome	0.040296	0.241778	6
Nucleus	0.999908	1	10
Cytoskeleton	0.995838	1	1
Mitochondrion	0.995369	1	0
Ribosome	0.546016	1	0
Peroxisome	0.356874	1	0

4.5.2 Validation of GO Cellular Compartment Summarization

To assess the precision of the SubcellulaRVis compartment assignments, we validated the SubcellulaRVis tool using two previously published data sets describing subcellular localization of proteins. Both datasets utilized different spatial proteomics technologies in order to separate or label proteins found at different subcellular compartments.

The first dataset analyzed was from Go *et al.* (2021) (from Supplementary table 8 of original publication). The authors used proximity dependent biotinylation to label proteins in the vicinity of different organellar markers which were tagged with the biotin ligase enzyme BioID. In total, 4424 proteins were identified and associated to different subcellular compartments. We calculated the enrichment using SubcellulaRVis of the proteins associated to the different organelles in the datasets, treating them as organelle-specific protein lists. The most significantly enriched SubcellulaRVis compartment for each of the twenty experimentally determined, organelle-associated protein lists were directly or closely matched (Table 4-4). For example, proteins experimentally associated to the cell junction, nucleolus and microtubules were enriched for the plasma membrane, nucleus, and cytoskeleton compartments, respectively by the SubcellulaRVis tool. The exception was the "Mitochondrial Membrane, Peroxisome", as the proteins in this list were significantly enriched for the Cytoplasm. However, 21 / 172 proteins in this subset were annotated to both the mitochondria and the cytoplasm SubcellulaRVis compartments, and 6 / 172 were annotated to the peroxisome and the cytoplasm SubcellulaRVis compartments; none of the

proteins in this subset were solely annotated to the mitochondria or the peroxisome, pointing to few location-specific markers for these compartments in the protein list. Moreover the 'Mitochondrial Matrix' and 'Mitochondrial Inner Membrane/ Intermembrane Space' showed significant enrichment for the mitochondria, indicating that our tool can associate known organelle markers to the correct compartment. This case study also provides a useful example of how analysis based on standardized, generic cellular compartments names, such as the SubcellulaRVis compartments, could be useful for comparison between different spatial studies.

Table 4-4. SubcellulaRVis enrichment results for each of the protein lists associated to cellular compartments by Go et al. (2021). The final column represents the overlap between the user-supplied gene list and genes annotated to the highest enriched SubcellulaRVis compartment.

Experimentally determined Cellular Compartment	**Highest enriched SubcellulaRVis compartment**	**FDR**	**No. User genes annotated to comp. / No. user genes**
Cell junction	Plasma membrane	1.88E-74	152/235
Chromatin	Nucleus	1.08E-216	352/383
ER membrane	Endoplasmic reticulum	7.11E-83	85/127
Mitochondrial matrix	Mitochondrion	8.78E-132	280/299
Actin, cytosol	Cytoplasm	6.19E-104	261/300
ER lumen	Endoplasmic reticulum	2.17E-127	122/165
Endosome, lysosome	Endosome	1.82E-97	88/173
Nucleolus	Nucleus	1.46E-85	146/164
Nucleoplasm	Nucleus	1.43E-131	222/247
Nuclear body	Nucleus	2.91E-179	278/295
Plasma membrane	Plasma membrane	1.04E-120	245/376
Centrosome	Cytoskeleton	7.48E-93	129/258
Mitochondrial Membrane, Peroxisome	Cytoplasm	3.78E-76	172/189
Golgi	Golgi apparatus	1.14E-100	96/146
Nuclear outer membrane-ER membrane network	Endoplasmic reticulum	5.58E-144	159/262
Cytoplasmic RNP granule	Cytoplasm	1.00E-43	119/143
Microtubules	Cytoskeleton	1.42E-86	87/115
Mitochondrial inner membrane, Mitochondrial inter membrane space	Mitochondrion	1.23E-134	106/122
Misc.	Cytoplasm	3.35E-86	234/279
Early endosome, Recycling endosome	Intracellular vesicle	1.03E-60	80/146

To assess the precision of the SubcellulaRVis tool for the analysis of a non-human data set we extracted the data from Nightingale et al. (2019) (from Supplementary Table 3 of original

publication) which used the hyperplexed Localisation of Organelle Proteins by Isotope Tagging (hyperLOPIT) method (Mulvey *et al.*, 2017) to reconstruct spatial profiles of organelle proteomes in *Saccharomyces cerevisiae*. As with the previous dataset, SubcellulaRVis found the highest enrichment for the compartment most closely describing the experimentally associated cellular compartment, with the exception of the mitochondrion (Table 4-5). For example, the proteins lists associated to the endoplasmic reticulum and Golgi apparatus were associated with the respective SubcellulaRVis compartment. We concluded that SubcellulaRVis accurately replicates the protein localization determined in *S. cerevisiae*. Moreover, the SubcellulaRVis visualization (as demonstrated in Figure 1) would be an impactful way to display the enrichment results for each of the organelle-specific lists.

Table 4-5. SubcellulaRVis enrichment results for each of the protein lists associated to cellular compartments by Nightingale et al. (2019). The final column represents the overlap between the user-supplied gene list and genes annotated to the highest enriched SubcellulaRVis compartment.

Experimentally determined Cellular Compartment	**Highest enriched SubcellulaRVis compartment**	**FDR**	**No. User genes annotated to compartment / No. user genes**
Cytosol	Cytoplasm	9.47E-243	743/894
ER	Endoplasmic reticulum	4.55E-168	148/170
Golgi	Golgi apparatus	1.60E-111	80/109
Mitochondrion	*No significant enrichment*	-	12/186
Nucleus	Nucleus	1.55E-169	535/865
Plasma membrane	Plasma membrane	2.56E-118	160/401

4.6 Availability and Future Directions

We have introduced an app named SubcellulaRVis, which provides easy visualization of the enriched subcellular locations from a user-supplied list of HGNC symbols. The visualization is more biologically insightful than a bar chart and will aid in efficient interpretation of the characteristics of gene lists and standardization of subcellular localization analyses. We anticipate our software will be particularly useful for experimental biologists with limited bioinformatics expertise analyzing data related to cellular signalling, protein (re)localization and regulation, and spatially resolved functional modules within the cell. For instance, our tool will provide impactful visualization of how protein localization varies upon two different extracellular perturbations, or for providing confirmation of correct tagging of proteins for spatial proteomics experiments.

SubcellulaRVis also provides an alternative view for the visualization of the endosomal system (Figure 4-3) which will be validated once appropriate data, such as spatial proteomics profiling of individual endosomal compartments (e.g. the early or recycling endosomes), becomes available. The endosomal system is notoriously hard to characterize (Naslavsky and Caplan, 2018), therefore our tool will be a valuable resource for distinguishing gene lists based on enrichment for different endosomal compartments. One potential use will be for researchers studying dynamic protein recycling to or from the plasma membrane.

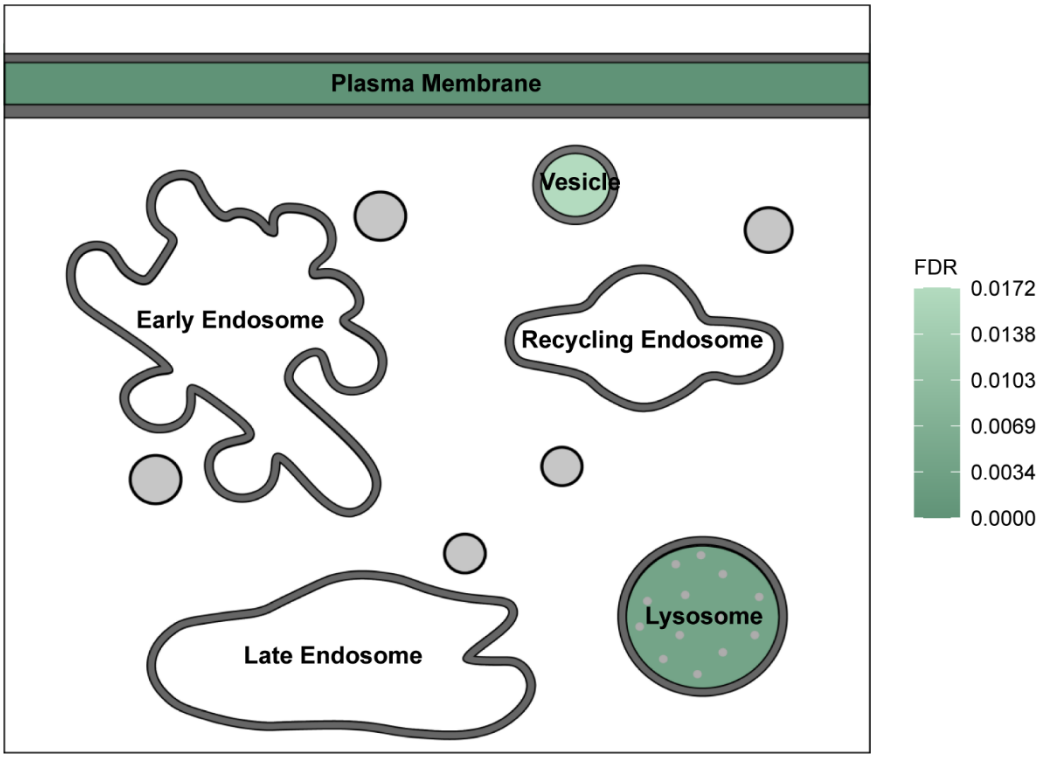


Figure 4-3. Visualization of enrichment results using SubcellulaRVis, using the trafficking-specific endosomal view, on a gene list of GPI anchored proteins (The UniProt Consortium, 2021).

The software can be installed at https://github.com/JoWatson2011/subcellularvis and the Shiny app is hosted at http://phenome.manchester.ac.uk/subcellular/. The R package will be submitted to Bioconductor upon acceptance of the article.

5 Ligand identity and concentration determine RTK signalling architecture

5.1 Introduction

Growth factor binding to one of the known 58 receptor tyrosine kinases (RTKs), which are divided into 20 classes, induces RTK dimerization and the activation of signalling cascades through a set of core pathways (Wintheiser and Silberstein, 2021). The activation of the receptors and signalling cascade are driven by the sequential phosphorylation of associated proteins. Ultimately, these phosphorylation-mediated cascades impact cellular fate such as the regulation of proliferation, differentiation, or cell survival (Lemmon and Schlessinger, 2011). RTKs are differentially regulated depending on the identity and the availability of their ligands in the extracellular environment. Functional selectivity describes the specificity of receptors for multiple ligands, each of which will have a differential effect on downstream signalling and cellular outcomes (Karl *et al.*, 2020). For instance, activation of the epidermal growth factor (EGF) receptor (EGFR) with transforming growth factor (TGF)-α has been shown to induce proliferation and migration to a greater extent than EGF in stimulated epithelial cells (Francavilla *et al.*, 2016). Ligand abundance or concentration in the extracellular environment also determines cell behaviour in multicellular organisms. Indeed, growth factors show different distributions in tissue; for example, EGF has been reported to have concentrations of 1 – 20 ng/mL in blood plasma and serum, but up to 500 ng/mL in tissues and organs (Sigismund *et al.*, 2005). Growth factor or morphogen gradients are considered critical drivers of pattern formation in development, such as FGF gradients in embryogenesis (Dorey and Amaya, 2010; Balasubramanian and Zhang, 2016). Finally, the surplus or surfeit of ligand that comes from over- or repressed-expression can be a key stage in disease progression, such as the overexpression of TGF-β inducing renal fibrosis (Meng, Nikolic-Paterson and Lan, 2016).

EGFR has traditionally been considered the prototypical RTK (Chen *et al.*, 2016), however the diversity of mechanisms through which RTKs can be activated and regulated is being increasingly appreciated. Despite this, the effect of ligand availability on the regulation of RTKs signalling outputs has been most notably investigated using the EGFR ligand, EGF. Low concentrations of EGF (<10 ng/mL) were shown to induce comparatively little AKT activation compared to higher concentrations of EGF (> 10 ng/mL, < 26 ng/mL). An investigation using a range of low EGF concentrations (0.01 ng/mL, 0.05 ng/mL, 0.2 ng/mL, and 2 ng/mL)

demonstrated the potency of this ligand, with 0.2 ng/mL EGF inducing high proportions of activated ERK, alongside activation of the transcription factors Fra1 and c-Myc, and the cell cycle regulator retinoblastoma protein (Albeck, Mills and Brugge, 2013). Moreover, increasing EGF concentrations have been shown to induce stronger phenotypic changes in cells. Though phosphorylation of EGFR plateaued with concentrations of EGF higher than 1 ng/mL, increasing EGF concentration (40 ng/mL, 80 ng/mL and 400 ng/mL) induced higher migration in connective-tissue derived fibroblast cells (Kong, Majeska and Vazquez, 2011). However, speed and directionality of migrating cells was shown to be non-linear, with treatment of 50 ng/mL EGF inducing the most rapid and direct migration of MDA-MB-231 breast cancer cells compared to 25 ng/mL and 100 ng/mL (Wang *et al.*, 2004). The stability of active EGFR has also been shown to be dose-dependent, as a high concentration of EGF leads to EGFR degradation after internalization, whilst low concentrations reduce degradation by approximately two thirds (Sigismund *et al.*, 2005, 2008).

As EGFR is not a prototypical RTK but has been the primary focus of ligand abundance studies, a comparative study with other RTKs is warranted. An obvious candidate is FGFR, given evidence of crosstalk between the receptors (Zakrzewska *et al.*, 2013; Smith *et al.*, 2021). The influence of ligand abundance has only been briefly investigated for the FGFR family, in the context of FGF2, which was shown to have a biphasic effect on cell proliferation; cell proliferation was shown to be highest at the lower and higher concentrations of a 0.1 – 100 ng/mL range (Garcia-Maya *et al.*, 2006). Mid-range FGF2 concentrations were also associated with higher FRS2 and ERK activation (Garcia-Maya *et al.*, 2006). However, the effect of ligand concentration on, for example, FGFR receptor trafficking and downstream signalling hasn't been investigated. FGF1, which activates all 7 FGFR isoforms, is a good candidate for comparison to EGF:EGFR concentration-dependent behaviours, as FGF1:FGFR receptor and signalling dynamics have been well characterised (Haugsten *et al.*, 2005, 2014).

We performed a phosphoproteomics experiment to study short-term phosphoproteomics dynamics and long-term proteomics changes downstream of high (100 ng/mL) and low (5 ng/mL) concentrations of EGF and FGF1 in the epithelial HeLa cells. Phosphoproteomics is a powerful tool for investigating global phosphorylation events downstream of extracellular stimuli, and has been previously used to dissect ligand-based regulation through functional selectivity (Francavilla *et al.*, 2013, 2016), however a systems-level overview to compare the mechanisms by which ligand abundance impacts upon phosphorylation-mediated signalling cascades has yet to be constructed. The high and low concentrations were selected based

on several lines of reasoning: 100 ng/ml being used widely in the literature as a generic, saturating concentration of ligand; the work of Sigismund et al. (2005, 2008), describing the divergence in EGFR stability upon different ligand concentration; and existing knowledge on the dynamic range of both EGF and FGF ligands, as described above. Combining several bioinformatics tools, the dose- and ligand-specific changes that occur through phosphorylation up to 120 minutes following stimulation were resolved, as well as changes in protein abundance 36 hours after stimulation. Using a conceptually original analysis method, we linked short-term phosphorylation dynamics to long-term changes in protein abundance, downstream of different ligands and ligand concentrations, via modified transcription factors. This provided the foundation for hypotheses to be experimentally validated in the future, such as the ligand- and concentration-dependent regulation of cell growth downstream of FGF1 and cell matrix adhesion downstream of EGF. Our results provide insight into the multiple cellular processes regulated by RTK ligand abundance and identity.

5.2 Methods

5.2.1 Reagents & Consumables

Table 5-1. List of reagents and consumables.

Reagent or Resource	Source	Product/Catalogue ID
Cell Culture		
RPMI 1640 Medium Glutamax Supplement (500ml)	Gibco	61870010
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma-Aldrich	D6429-500ML
SILAC DMEM	Thermo Scientific	88394
Sodium Pyruvate solution 100mM (100ml)	Sigma-Aldrich	S8636-100ML
Penicillin-Streptomycin (10,000 U/mL)	Life Technologies Ltd	15140122
Fetal Bovine Serum, South American origin	Life Technologies	10270106
Dialysed Fetal Bovine Serum	Merck	F0392
L-ARGININE:HCL	Cambridge Isotope Laboratories	CLM-2265-H-0.25
L-ARGININE:HCL	Cambridge Isotope Laboratories	CNLM-539-H-0.5
L-ARGININE:HCL	Sigma-Aldrich	A6969
L-LYSINE:2HCL	Cambridge Isotope Laboratories	DLM-2640-0.5

L-LYSINE:2HCL	Cambridge Isotope Laboratories	CNLM-291-H-0.5
L-LYSINE:2HCL	Sigma-Aldrich	L8662
Antibodies		
Rabbit anti Phospho-EGF Receptor (Tyr1068)	Cell Signalling Technology	2234S
Rabbit polyclonal anti EGFR	Merck	06-847
Mouse monoclonal phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (E10)	Cell Signalling Technology	9106L
Rabbit monoclonal p44/42 MAPK (Erk1/2) (137F5)	Cell Signalling Technology	4695S
Mouse monoclonal Vinculin	Sigma Aldrich	V 9264-200UL
Peroxidase-AffiniPure F(ab')2 Fragment Goat Anti-Mouse IgG (H + L) (min X Hu, Bov, Hrs Sr Prot)	Stratech	115-036-062-JIR-0.5ml
Peroxidase-AffiniPure F(ab')2 Fragment Goat Anti-Rabbit IgG (H + L) (min X Hu Sr Prot)	Stratech	111-036-045-JIR-0.5m
Peptides		
Human EGF	Sigma-Aldrich	E9644-.2MG
Human FGF1	Pepro Tech Limited	100-17A-50
Mass-spectrometry sample preparation		
Trypsin porcine pancreas (proteomics grade)	Sigma-Aldrich	T6567
Lysyl Endopeptidase	FUJIFILM Wako Chemicals	2541
TiO beads “Titanspheres”	GL Sciences	5020-75000
Pre-cast gradient gel: Nu-PAGE 4-12% Bis-Tris Gel 1.0mm 10 well	Invitrogen	NP0321BOX
Sep-Pak Classic C18 cartridges	Waters	WAT051910
Solid Phase Extraction Disk “Empore” C18 (Octadecyl) 3 M	Agilent Technologies	2215
Solid Phase Extraction Disk “Empore” C8 (Octyl) 3 M	Agilent Technologies	2214
Pierce Protease Inhibitor Tablets-20 tablets	Life Technologies	A32963

5.2.2 Cell Culture and SILAC labelling

HeLa and T47D cells were purchased from ATCC. Cell lines authenticated by Eurofins Genomic and confirmed not to have mycoplasma infection through monthly tests using a PCR-based detection assay (Venor®GeM—Cambio). HeLa cells were grown in DMEM and T47D cells were grown in RPMI, both supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate and 10% foetal bovine serum, and incubated in a humidified cabinet at 37°C.

HeLa Cells were labelled using SILAC DMEM supplemented with dialysed foetal bovine serum, as previously described (Francavilla *et al.*, 2013). Three SILAC labels were used: light, using L-arginine and L-lysine; medium, using L-arginine-U-$^{13}C_6{}^{14}N_4$ and L-lysine-$^{2}H_4$; and heavy, using L-arginine-U-$^{13}C_6$-$^{15}N_4$ and L-lysine-U-$^{13}C_6$-$^{15}N_2$.

5.2.3 Western blotting

Cells were serum starved overnight in serum-free medium and stimulated for 8, 40, or 120 minutes with either 5 ng/ml or 100 ng/ml of EGF or FGF1. Stimulation, cell extraction and immunoblotting were performed as previously described (Francavilla *et al.*, 2016). Proteins were resolved by SDS–PAGE and transferred to nitrocellulose membranes (Protran, Biosciences). Proteins of interest were visualized using specific antibodies, followed by peroxidase-conjugated secondary antibodies. Blots were visualized using the Universal Hood II Gel Molecular Imaging System (Bio-Rad). Each experiment was repeated at least three times and produced similar results.

5.2.4 Sample Preparation & Mass Spectrometry

Each mass spectrometry experiment was performed with technical duplicates.

SILAC-labelled HeLa cells for phosphoproteome analysis were prepared as described in Smith et al., 2021. Briefly, cells were washed with PBS and lysed at 4°C in ice-cold 1% triton lysis buffer supplemented with Pierce protease inhibitor tablet (Life Technologies) and phosphatase inhibitors: 5 nM Na3VO4, 5 mM NaF and 5 mM β-glycerophosphate. 600 µg of proteins were obtained from each sample. Proteins were precipitated overnight at −20°C in fourfold excess of ice-cold acetone. SILAC triplicates were mixed in a 1:1:1 ratio. Precipitated proteins were dissolved in denaturation buffer (6 M urea, 2 M thiourea in 10 mM HEPES pH 8). Cysteines were reduced with 1 mM dithiothreitol (DTT) and alkylated with 5.5 mM chloroacetamide (CAA). Proteins were digested with endoproteinase Lys-C (Wako, Osaka, Japan) and sequencing grade modified trypsin (modified sequencing grade, Sigma) followed by quenching with 1% trifluoroacetic acid (TFA). Peptides were purified using reversed-phase Sep-Pak C18 cartridges (Waters, Milford, MA) and eluted with 50% acetonitrile (ACN). After removing ACN by vacuum concentrator at 60°C, peptides were suspended in phosphopeptide immunoprecipitation buffer (50 mM MOPS pH 7.2, 10 mM sodium phosphate, 50 mM NaCl) and dissolved overnight. Clarified peptides were purified using reversed-phase Sep-Pak C18 cartridges (Waters, Milford, MA) and eluted with 50% ACN and enriched for phosphorylated peptides with Titansphere chromatography. 12% TFA in ACN was added to the eluted peptides (1:1 dilution) and subsequently enriched with TiO2 beads (5 µm, GL Sciences Inc., Tokyo, Japan). The beads were suspended in 20 mg/mL 2,5-

dihydroxybenzoic acid (DHB), 80% ACN, and 6% TFA and the samples were incubated in a sample to bead ratio of 1:2 (w/w) in batch mode for 15 min with rotation. After 5-min centrifugation, the supernatant was collected and incubated a second time with a twofold dilution of the previous bead suspension. The beads were washed with 10% ACN, 6% TFA followed by 40% ACN, 6% TFA and collected on C8 STAGE-tips and finally washed by 80% ACN, 6% TFA. Elution of phosphorylated peptides was done with 20ul 5% NH3 followed by 20 µL 10% NH3 in 25% ACN, which were evaporated to a final volume of 5 µL in a sped vacuum. The concentrated phosphorylated peptides were acidified with addition of 20 µL 0.1% TFA, 5% ACN and loaded on C18 STAGE-tips. Peptides were eluted from STAGE-tips with 20 µL of 40% ACN followed by 10 µL 60% ACN and ACN and reduced to 5 µL by SpeedVac and 5 µL 0.1% FA, 5% ACN added. A small amount of the eluted peptides (1%) was taken for proteome analysis before enrichment of phosphorylated peptides: after evaporation in a speed vacuum, 40 µl of 0.1% TFA, 5% ACN were added followed by MS analysis.

For analysis of the long-term proteome changes, SILAC-labelled HeLa cells were lysed, and 50 µg protein from each SILAC condition were mixed before fractionation with SDS–PAGE and in-gel digestion with trypsin as described in Francavilla (2016).

Mass spectrometry analysis of prepared samples was performed by the BioMS facility at the University of Manchester. Purified peptides were analysed by LC-MS/MS using an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to a QE-HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. Mobile phase A was 0.1% FA in water, and mobile phase B was 0.1% FA in ACN and the column was a 75 mm x 250 µm inner diameter 1.7 µM CSH C18, analytical column (Waters). A 1 µl aliquot of the sample (for proteome analysis) or a 3 µl aliquot was transferred to a 5 µl loop and loaded on to the column at a flow of 300 nl/min at 5% B for 5 and 13 min, respectively. The loop was then taken out of line and the flow was reduced from 300 nl/min to 200nl/min in 1 min., and to 7% B. Peptides were separated using a gradient that went from 7% to 18% B in 64 min., then from 18% to 27% B in 8 min. and finally from 27% B to 60% B in 1 min. The column was washed at 60% B for 3 min. and then re-equilibrated for a further 6.5 min. At 85 min, the flow was increased to 300nl/min until the end of the run at 90min. Mass spectrometry data were acquired in a data dependent manner for 90 min in positive mode. Peptides were selected for fragmentation automatically by data-dependent analysis on a basis of the top 8 (phosphoproteome analysis) or top 12 (proteome analysis) with m/z between 300 and 1750Th and a charge state of 2, 3 or 4 with a dynamic exclusion set at 15 s. The MS resolution

was set at 120,000 with an AGC target of 3e6 and a maximum fill time set at 20ms. The MS2 resolution was set to 60,000, with an AGC target of 2e5, and a maximum fill time of 110 ms for Top12 methods, and 30,000, with an AGC target of 2e5, and a maximum fill time of 45 ms for Top8 analysis. The isolation window was of 1.3Th (2.6 Th for SILAC-labelled samples), and the collision energy was of 28.

5.2.5 Data & Statistical Analysis

Raw data were analysed by the MaxQuant software suite (Cox and Mann, 2008) (https://www.maxquant.org; version 1.6.2.6) using the integrated Andromeda search engine (Cox *et al.*, 2011). Proteins were identified by searching the HCD-MS/MS peak lists against a target/decoy version of the human UniProt Knowledgebase database that consisted of the complete proteome sets and isoforms (v.2016; https://uniprot.org/proteomes/UP000005640_9606) supplemented with commonly observed contaminants such as porcine trypsin and bovine serum proteins. Tandem mass spectra were initially matched with a mass tolerance of 7 ppm on precursor masses and 0.02 Da or 20 ppm for fragment ions. Cysteine carbamidomethylation was searched as a fixed modification. Protein N-acetylation, N-pyro-glutamine, oxidized methionine and phosphorylation of serine, threonine and tyrosine were searched as variable modifications. Protein N-acetylation, oxidized methionine and deamidation of asparagine and glutamine were searched as variable modifications for the proteome experiments. For the quantification of SILAC-labelled samples, labelled lysine and arginine were specified as fixed or variable modification, depending on prior knowledge about the parent ion (MaxQuant SILAC triplet identification). False discovery rate was set to 0.01 for peptides, proteins and modification sites. Minimal peptide length was seven amino acids. Site localization probabilities were calculated by MaxQuant using the PTM scoring algorithm (Olsen *et al.*, 2006). The datasets were filtered by posterior error probability to achieve a false discovery rate below 1% for peptides, proteins, and modification sites. Only peptides with Andromeda score > 40 were included.

All statistical and bioinformatics analyses were performed using freely available packages in the *R* programming language. For phosphoproteome analysis, only peptides with localization probabilities higher than 0.75 (class I, Olsen et al (2006)) were included in the downstream bioinformatics analysis. Normalized SILAC ratios from MaxQuant output .txt files were used. Median SILAC ratios were calculated and used for subsequent analysis. Regulated phosphorylated sites and differentially abundant proteins were determine by applying a 0.05 and 0.95 quantile threshold (for at least one point in the time-course, in the case of

phosphorylated sites). Phosphorylated sites with incomplete time courses (I.e. ≥1 missing value) were excluded from analysis.

The clustering methods used in Figure 5-4 are described in Table 5-2. The Rand index was calculated using the *fossil* package. Enrichment analyses were performed using the *EnrichR R* interface (Kuleshov *et al.*, 2016). Protein-protein interactions were extracted from STRING v11 (Szklarczyk *et al.*, 2019); interactions were filtered to have an experimental confidence score > 0.4. Kinase-substrate interactions were extracted from PhosphositePlus v6.6.0 (Hornbeck *et al.*, 2015). Network visualisation was performed with Cytoscape (v3.8.2).

Table 5-2. R packages used for clustering.

Method	**R package**
K-Means	*stats* (Base R)
Fuzzy C-means	*MFuzz* (Bioconductor), *e1071* (CRAN)
Hierachical	*stats* (Base R)
CLUE	*ClueR* (CRAN)
Silhouette analysis	*cluster* (CRAN)
Fuzzy Silhouette analysis	*fclust* (CRAN)
Within sum of squares	*stats* (CRAN)

5.3 Results

5.3.1 Summary of Experimental Design & Quality Control Assessments

To reveal the concentration dependent effects on signalling downstream of EGF and FGF1 in epithelial cells, a SILAC-based phosphoproteomics experiment was designed (Figure 5-1A). These ligands were selected as their concentration dependent effects have been briefly investigated in the literature. Moreover, HeLa cells are known to endogenously express both EGFR and FGFR1/4 (Francavilla *et al.*, 2009; Smith *et al.*, 2021). We compared untreated cells with cells treated with either of the ligands at the high or low concentration. Three timepoints were included: 8 minutes, 40 minutes, and 120 minutes. These timepoints were selected as they reflect important points in RTK signalling and internalisation: by 8 minutes signalling cascades have been initiated from the plasma membrane and RTKs have been internalised (Francavilla *et al.*, 2013, 2016), by 40 minutes RTKs bound for degradation are not detectable by western blot (Sigismund *et al.*, 2008; Francavilla *et al.*, 2016) and by 120 minutes regulatory loops and changes in protein abundance would be expected (Bahrami and Drabløs, 2016).

Quality assessment of the data were consistent with previous publications (Smith *et al.*, 2021). A high number (> 3500) of phosphorylated sites were quantified across the different

experimental conditions (Figure 51-B), and the Andromeda search scores were high (median of 113), indicating high confidence in peptide identification (Figure 5-1C). Of the identified phosphorylated peptides, 69.7% had a Class I localisation of the phosphorylated peptide (>0.75 localisation probability) (Figure 5-1D). The distribution of modifications found on serine (S), threonine (T) and tyrosine (Y) residues or were a single, double or more than double phosphorylation at the same site were consistent with previous publications (Figure 5-1E, F) (Sharma *et al.*, 2014). The Pearson correlation (R) between experimental conditions and principal component analysis (PCA) revealed that replicates were generally well correlated (R > 0.7) (Figure 5-1G,H).

To study how the short-term dynamic phosphorylation events could be mediating long-term systematic changes, a proteome experiment was performed in which SILAC labelled HeLa cells were stimulated with 100 ng/mL or 5 ng/mL EGF or FGF1 for 36 hours (Figure 5-1I). Replicates had a correlation of >0.6 (Figure 5-1J) and over 3000 proteins were quantified in each sample (Figure 5-1K).

A relatively low correlation or high variance was observed between replicate quantitative values of some of the experimental conditions: 40 minute 100 ng/mL EGF for phosphoproteome (Figure 5-1G,H) and 5 ng/mL EGF for 36 hour proteome (Figure 5-1J,M). Given the generally high confidence in identification (Figure 5-1C,K), potential technical reasons underlying this were investigated based on MaxQuant output files using the PTXQC tool in R (Bielow, Mastrobuoni and Kempa, 2016). The results indicated that there were a high number of peptides with missing cleavages in certain MS runs but no other issues with the data quality were indicated by the tool. Separate investigation of the missed cleavages found that indeed the level of missed cleavages was high across all the samples, with greater than 50% of peptides having at least one missed cleavage site. Peptides from replicate 2 (R2) of 40-minute 100 ng/mL EGF had a higher number of peptides with missed cleavages than the other phosphoproteome (STY) conditions (Figure 5-1L), consistent with the previous quality control measures (Figure 5-1G,H). However, the 36-hour proteome samples had relatively consistent levels of missed cleavages in peptides. This indicated that the missed cleavages may be causing poor correlation of the 40-minute 100 ng/mL EGF phosphoproteome replicates, but this was the unlikely the cause for the 5 ng/mL EGF 36 hour proteome replicates. Regardless, the number of missed cleavages was cause for concern, and influenced the stringency of the downstream analysis, which has specified below at the relevant points.

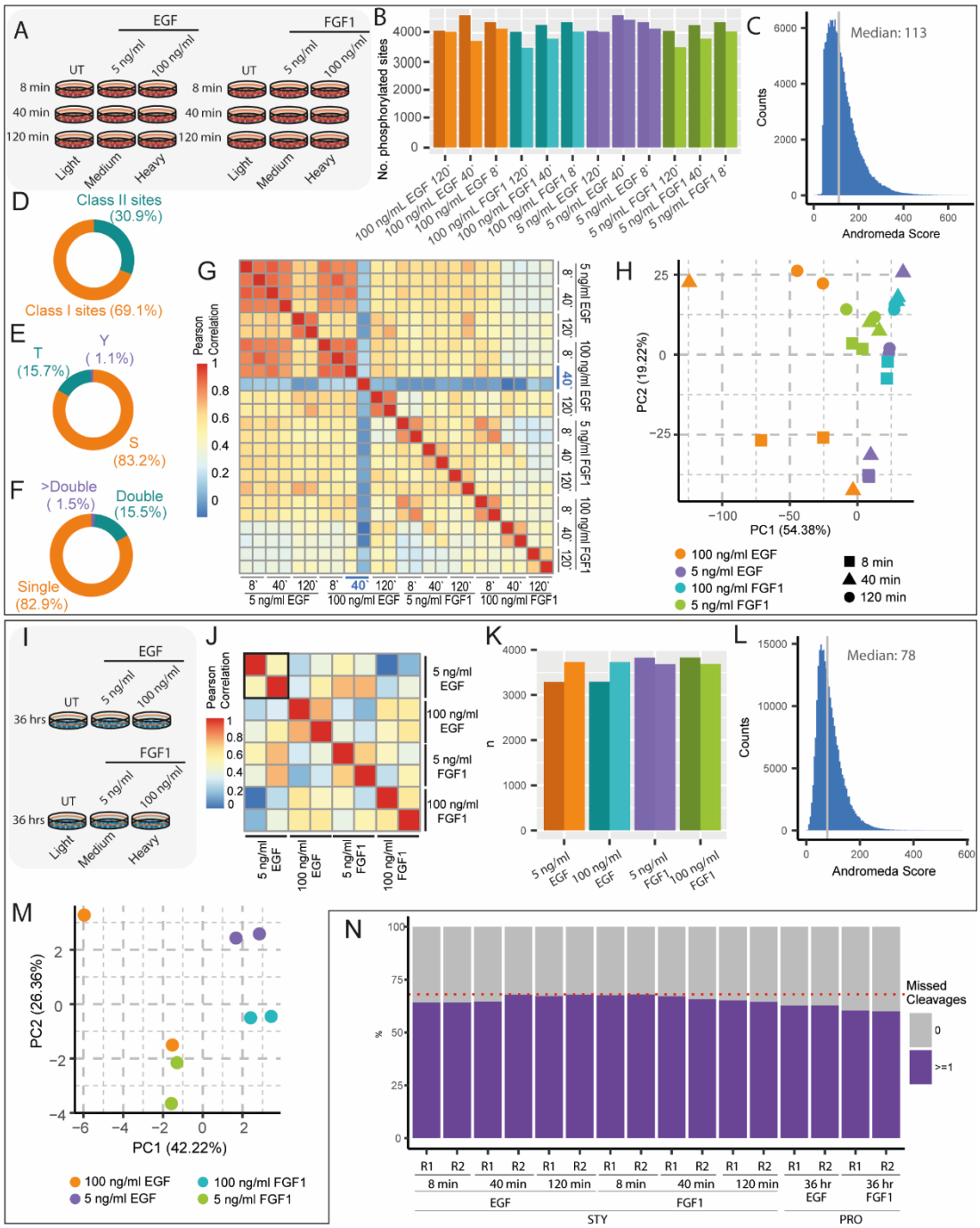


Figure 5-1. Set up and quality assessment of phosphoproteomics investigation of concentration-dependent regulation of intracellular EGF and FGF1 signalling.
A. *Experimental design. Light, medium, and heavy refer to the SILAC labels used.* ***B.*** *Number of phosphorylated sites quantified in each phosphoproteomic experimental condition.* ***C.*** *Andromeda score of identified phosphorylated peptides in mass-spectrometry runs.* ***D.*** *Proportion of phosphorylated sites which are singly, doubly, or more than doubly phosphorylated.* ***E.*** *Proportion of phosphorylated sites which are class I (with a localisation probability > 0.7) or class II.* ***F.*** *Proportion of phosphorylated S, T, or Y phosphorylated sites.* ***G.*** *Pearson correlation between phosphoproteome mass spectrometry samples. Highlighted with a black box and blue text are experimental conditions with poor correlation between replicates.* ***H.*** *PCA of phosphoproteome mass spectrometry samples.* ***I.*** *Experimental design for proteomics investigation of long-term impacts of concentration dependent EGF and FGF1*

response, 36 hours after signalling. ***J.*** *Pearson correlation between 36-hour proteome experimental conditions. Highlighted with a black box and blue text are experimental conditions with poor correlation between replicates.* ***K.*** *Number of peptides quantified in each experimental condition.* ***L.*** *Andromeda score of identified peptides in 36-hour proteome mass-spectrometry samples.* ***M.*** *PCA of 36-hour proteome mass spectrometry samples.* ***N.*** *Number of mis-cleaved peptides, and the number of missed cleavages, in each mass spectrometry runs. The highest number of peptides with >1 missed cleavages are marked by the shaded areas.*

As a further check of the quality of phosphoproteomics data western blot was used an alternative, low-throughput method. Consistency between phosphorylation of the well-studied signalling protein MAPK was assessed in two epithelial cell lines, HeLa and T47D (Figure 5-2A,B), and compared to the MS quantitation of single-phosphorylated peptides associated to MAPK1-T181 and MAPK3-T202 (Figure 5-2C). T181 and T202 are the activation sites on MAPK (also known as ERK) (Courcelles *et al.*, 2013; Lavoie, Gagnon and Therrien, 2020). Higher levels of phosphorylation can be seen downstream of all stimuli and concentrations in HeLa cells compared to T47D, however there is consistency in the phosphorylation dynamics between the two cell lines. Both high and low concentrations of FGF1 initially induced high levels of MAPK phosphorylation which declined by 40 minutes; interestingly, the response was stronger with the lower concentration of FGF1. High concentrations of EGF induced a steadily decreasing level of MAPK phosphorylation over the 120 minutes, which was consistent with 100 ng/mL EGF stimulation performed by Francavilla et al. (2016). By focusing on MAPK phosphorylation, the poor correlation between replicates in the mass spectrometry data at 40 minutes following 100n g/ml EGF (Figure 5-1G) was reproduced, with only one of the replicates following the phosphorylation pattern seen by western blot of a slow decline in MAPK1/3 phosphorylation up to 120 minutes. Based on these two pieces of evidence, the decision was made to remove the 40-minute timepoint with 100 ng/mL EGF from the analysis.

Given the known relationship between ligand concentration and EGFR stability, total levels of EGFR were also probed for. This relationship was confirmed in both HeLa and T47D: 100 ng/mL of EGF induced receptor degradation by 120 minutes whilst stimulation with 5 ng/mL EGF or FGF1 did not alter receptor abundance. Despite this, EGFR phosphorylation was stronger and more sustained downstream of 100 ng/mL EGF stimulation. A comparison with FGFR could not be made due to issues with antibody specificity for total and phosphorylated protein.

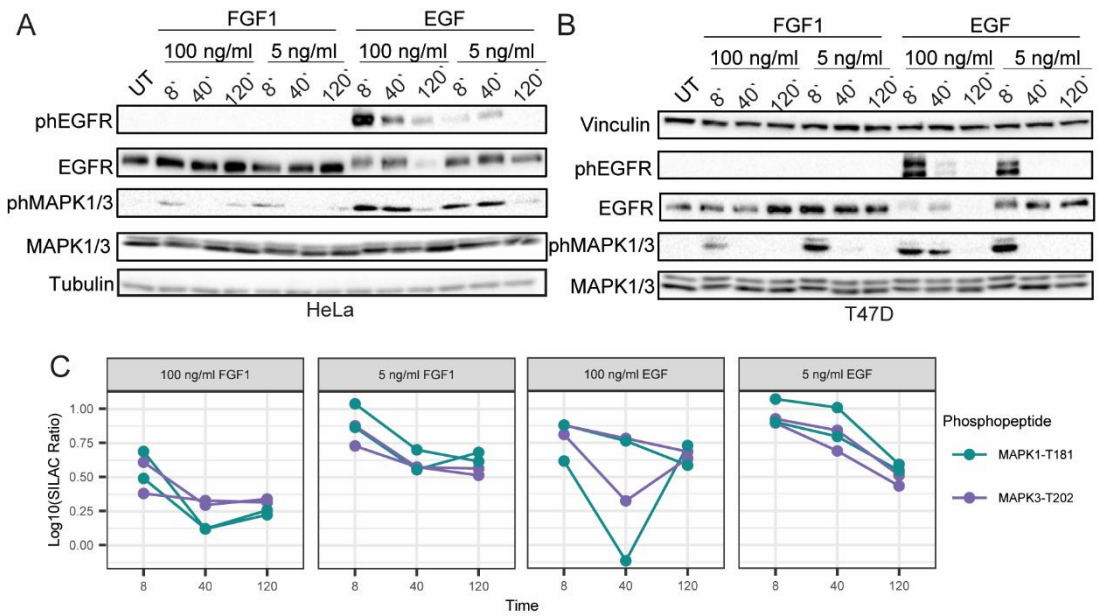


Figure 5-2. Validation of phosphoproteomics approach.
*Western blot analysis of stimulation with high (100 ng/ml) and low (5 ng/ml) concentrations of EGF and FGF1 for 8 minutes, 40 minutes and 120 minutes in HeLa (**A**) and T47D (**B**) cells. **C.** SILAC MS quantification profile of single-phosphorylated peptides for MAPK1 and MAPK3; replicates shown as separate lines.*

Next, regulated sites and proteins were determined by calculating a threshold based on 0.05 and 0.95 quantiles as an alternative to the use an absolute upper and lower bound, as has been used previously in the analysis of SILAC data (Francavilla *et al.*, 2016). Thresholds were used rather than a statistical test due to the underlying issues with the data quality. Moreover, given the differences in phosphorylation amplitude dependent on stimulant condition, as demonstrated in Figure 5-2, setting the threshold using upper and lower quantiles in each experimental condition allowed for condition-specific determination of regulated phosphorylated sites. Phosphorylated sites were filtered based on whether one of the quantitative values lay outside of the upper or lower quantiles for at least one time point (Table 5-3). The 100 ng/mL EGF and 100 ng/mL FGF1 stimulant conditions had the highest number of regulated sites (434 and 478, respectively), however all stimulant conditions had > 380 regulated sites.

Table 5-3. Number of regulated sites (following quantile filtering) in each stimulant condition.

Stimulant condition	Number of regulated sites
100 ng/ml EGF	434
5 ng/ml EGF	382
100 ng/ml FGF1	478
5 ng/ml FGF1	412

A well-accepted outcome of RTK-induced signalling cascades is the modification and regulation of transcription factors, ultimately affecting the gene expression programme of the cell (Weidemüller *et al.*, 2021). Therefore, we designed a bioinformatics approach focussed on these final players in the signalling cascades, and how their phosphorylation may long-term changes in protein expression (Figure 5-3). First, the phosphoproteomics data were analysed in a site-centric manner using the RWHN method (Watson, Schwartz and Francavilla, 2021). Next, modified transcription factors were identified amongst the parent proteins of the regulated phosphorylated sites (Table 5-3) and their target genes identified amongst the differentially abundant proteins in the 36 hr proteome data. Over-representation analysis was performed on these proteins. Finally, the overlap between functional terms from the RWHN and over-representation analysis was found and subsequent analysis performed on the phosphorylated proteins associated to these terms.

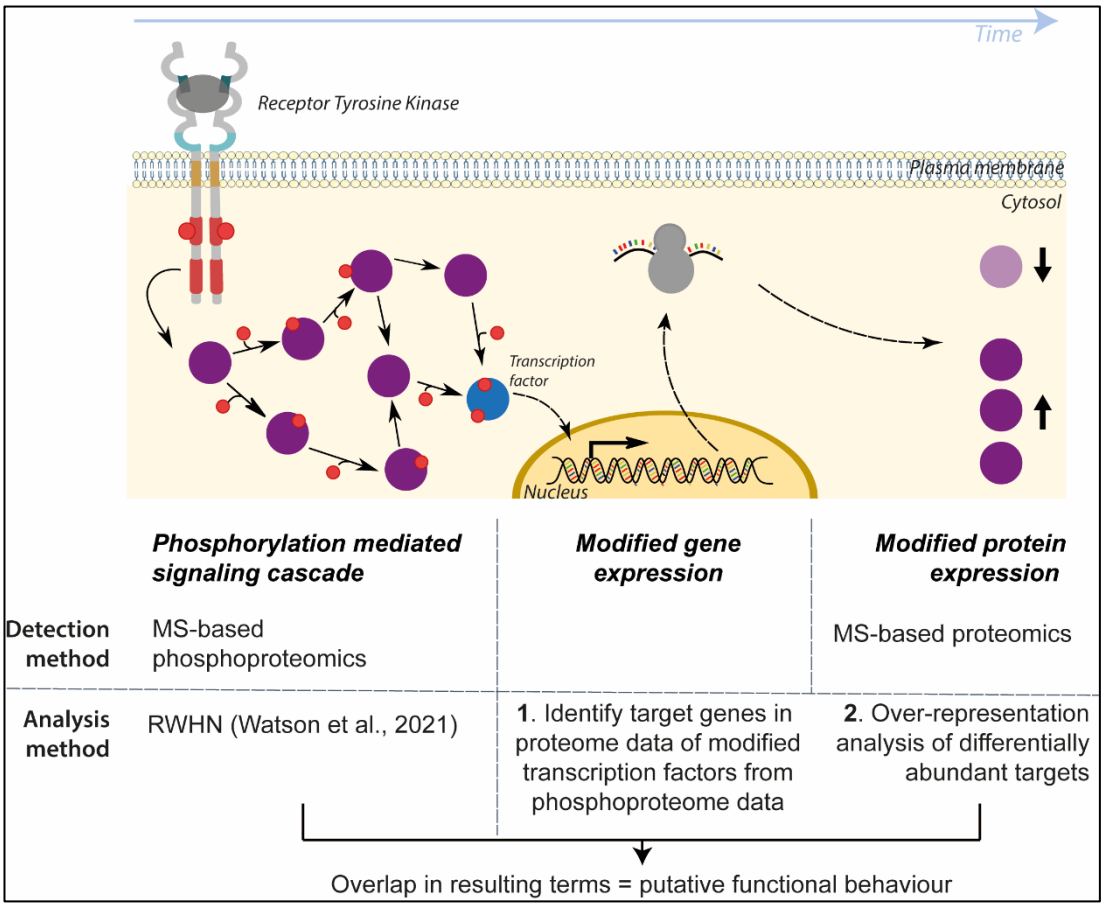


Figure 5-3. Analysis pipeline of phosphoproteome and 36-hour proteome data.

5.3.2 Functional impact of early phosphoproteome changes predicted with clustering and RWHN method

To reveal a functional connection between growth factor-specific signalling dynamics and the processes downstream of the regulated transcription factors, the short term

phosphoproteome was analysed in a site centric manner, using the Random Walk on Heterogeneous Networks (RWHN) for phosphoproteomics method (Watson, Schwartz and Francavilla, 2021). As RWHN relies on clustering of the data, a clustering method has to be selected based on factors such as overlap of the data or nonuniform density (Everitt *et al.*, 2011). A small-scale assessment was performed of the consistency between several clustering methods and associated metrics to predict the optimal number of clusters for my phosphoproteomics datasets. First, the overlap between cluster partitions by three clustering methods was examined. Hierarchical, K-means and fuzzy C-means clustering methods were selected as they all have been applied to phosphoproteomics datasets in the literature. Data were clustered with 3 to 20 partitions and the Rand Index, a common metric for assessing cluster overlap, was used to calculate the overlap (Figure 5-4A). K-means and Fuzzy C-means clustering had the highest Rand Index, indicating a stronger overlap in cluster partitions. An exception was seen when calculated for 100 ng/mL EGF datapoints, as hierarchical clustering and K-means had a slightly higher Rand Index.

Methods to determine the number of clusters to partition the data into were then considered. Commonly, the elbow-plot method is used to select an optimal cluster number for K-mean clustering, where the within-cluster sum of squares (a read-out for variance) is plotted as a function of the number of cluster partitions, and the "elbow" in the resulting curve identified (Thorndike, 1953). Silhouette analysis is a measure of how similar each cluster is to the others; a higher value indicates more unique clusters (Rousseeuw, 1987). A development of the method was proposed specifically for fuzzy C-means clustering, named fuzzy silhouette analysis (Campello and Hruschka, 2006). The final method considered, knowledge-based CLUster Evaluation (CLUE), is phosphoproteomics specific and uses enrichment of kinase substrates to infer the optimal partitioning of the data (P. Yang *et al.*, 2015). In general, each method predicted an ideal number of partitions within 1-2 of each other for each of the stimulant conditions (Figure 5-4B). The silhouette width calculated for the K-means clustering of 5 ng/mL FGF1 was the only outlier, suggesting an ideal cluster number of 19 compared to the 6 or 8 suggested by the other prediction methods. The final number of partitions was selected based on the median of the values from all the prediction methods (Table 5-4). The clusters were used as input for the RWHN method, to associate clustered sites to biological functions (Watson, Schwartz and Francavilla, 2021). Several clusters returned greater than 100 GOBP terms in their results (Figure 5-5). We therefore decided to filter the terms that came from the RWHN analysis using the results from the proteome analysis, described in the next section.

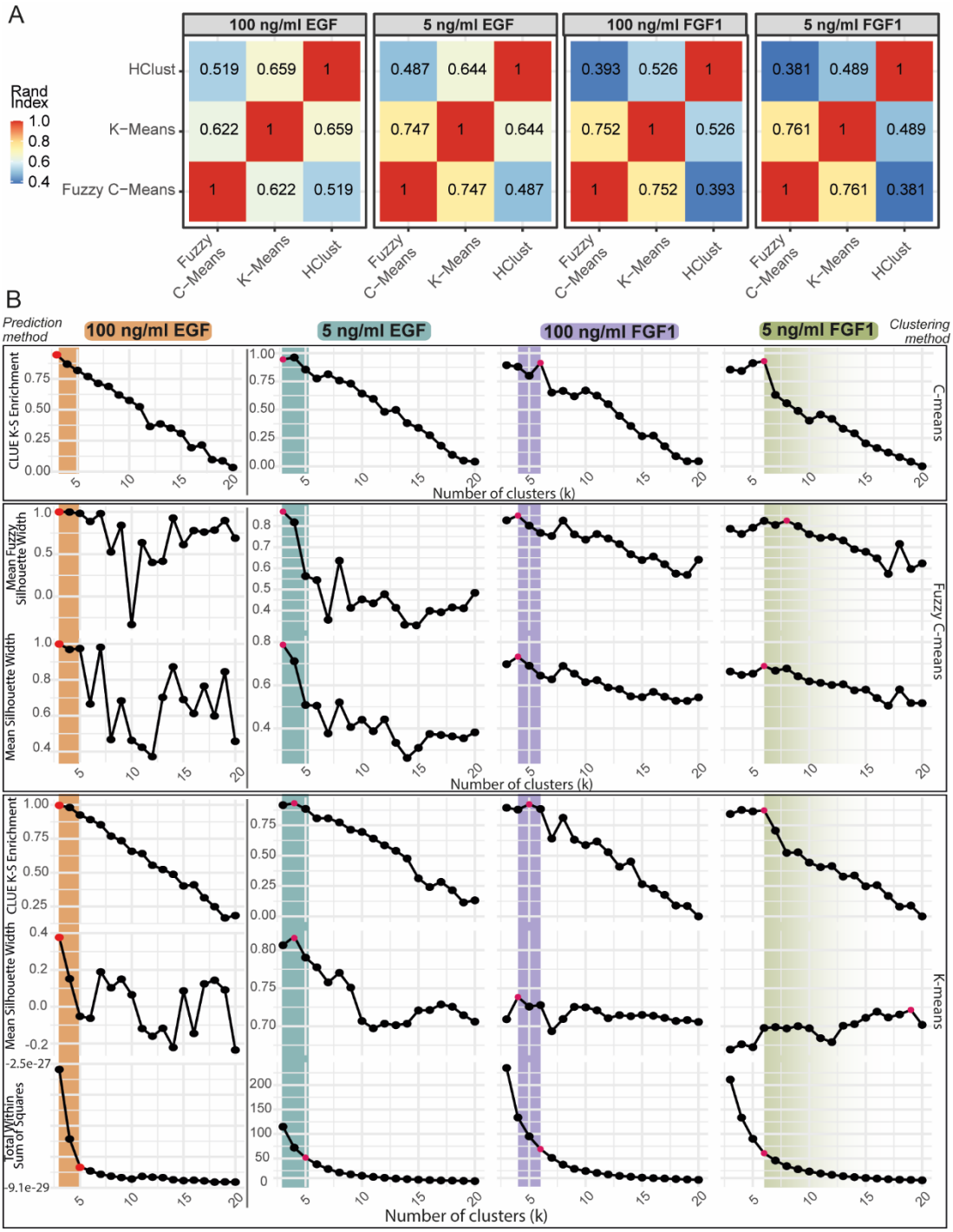


Figure 5-4. Similarity between different clustering methods and measures to determine optimal number of clusters.
A. *Average intersect between hierarchical, K-means and Fuzzy C-means clustering methods, calculated using the Rand index, over a range of 3 to 20 clusters.* ***B.*** *Comparison of methods to determine optimal cluster numbers for three different cluster algorithms (C-means, Fuzzy C means and K-means). Optimal cluster number determined by each method is indicated with a red point.*

Table 5-4. Number of clusters used for each stimulant condition.

Experimental condition	Cluster partitions
100 ng/ml EGF	3
5 ng/ml EGF	3
100 ng/ml FGF1	4
5 ng/ml FGF1	6

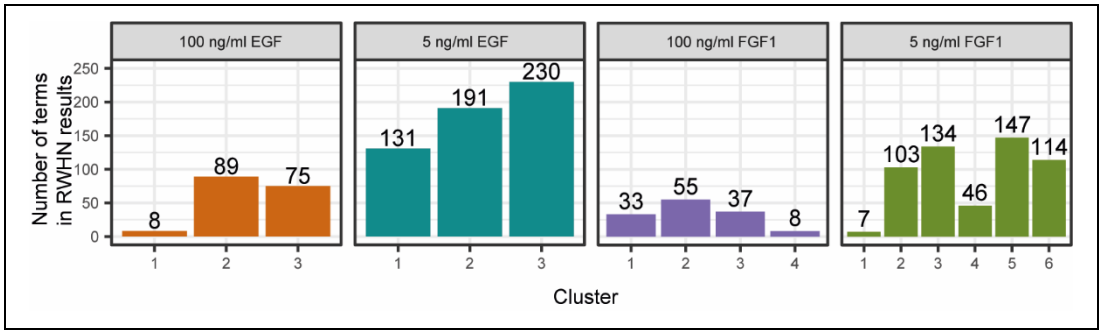


Figure 5-5. Number of GOBP terms resulting from RWHN applied to short-term phosphoproteome.

5.3.3 Associating long-term proteomic changes with the phosphorylation of transcription factors.

To connect the short-term phosphoproteome to long-term proteome changes, we identified modified transcription factors in the phosphoproteome data and their respective target genes in the proteome data. There are several databases containing known transcription factor target genes (Weidemüller *et al.*, 2021). The ChIP Enrichment Analysis (CHEA) and Transcriptional Regulatory Relationships Unravelled by Sentence-based Text mining (TRRUST) databases use data from CHIP-seq and text-mining followed by manual curation, respectively (Lachmann *et al.*, 2010; Han *et al.*, 2018). These two databases were chosen based on their inclusion in the popular enrichment analysis suite EnrichR, indicating confidence in the resources, and that they had been experimentally or manually curated (Kuleshov *et al.*, 2016). Firstly, regulated sites from each stimulant condition were filtered to find phosphorylated transcription factors which were also present in the CHEA or TRRUST databases (Figure 5-6A). The 15 phosphorylated sites on 11 transcription factors were then cross-referenced with PhosphoSitePlus to identify phosphorylation events with known regulatory effects (Figure 5-6A). Most of the transcription factor phosphorylated sites identified in our dataset did not have a known functional role, however my analysis identified phosphorylated sites on the well-studied transcription factors JUNB, ERF and SMAD2; the latter two are known targets of MAPK signalling (Lavoie, Gagnon and Therrien, 2020). Sites with a negative regulatory effect on JUNB (e.g. JUNB-S259, JUNB-T255) were upregulated in

all the experimental conditions except for 5 ng/mL FGF1. The 5 ng/mL FGF1 condition had the lowest number of transcription factor modifications identified, which was unexpected given the stronger signalling through MAPK1/3 observed for the lower concentration of FGF1 compared to the higher (Figure 5-6A) This suggests that FGF1 at low concentrations mainly regulates events upstream of transcription factors.

Target genes of the modified transcription factors were then identified in the list of differentially abundant proteins from the 36-hour proteome data (Figure 5-6B). Targets of JUN were found to be differentially abundant in all four stimulant conditions, whilst others were only found in some. Targets of SUZ12, for example, were only found upon stimulation with EGF of either concentration. SMAD2 targets were only differentially abundant downstream of 5 ng/mL EGF. To get a simple overview of processes that may be regulated by the differentially abundant proteins, enrichment for gene ontology biological process (GOBP) terms was performed.

The proteins were separated based on whether their expression was increased or decreased at 36 hours (Figure 5-6C). Proteins with decreased abundance following 36 hours of stimulation with 5 ng/mL FGF1 were generally involved in metabolic processes, whilst those with 100 ng/mL FGF1 had an association with general cellular processes such as RNA processing, cell motility and cell growth. Only proteins with increased abundance following EGF stimulation were enriched for GOBP terms. The 5 ng/mL EGF condition mostly showed enrichment for general transcription or DNA/RNA maintenance processes, but also was enriched for more general cellular behaviours such as ‘cell matrix adhesion’ and ‘mesodermal cell differentiation’. Proteins with increased abundance downstream of 5ng/mL FGF1 were enriched for ‘negative regulation of cellular proliferation’. Therefore, long term changes in the proteome caused by ligand- and concentration-dependent regulation may impact cellular outcomes.

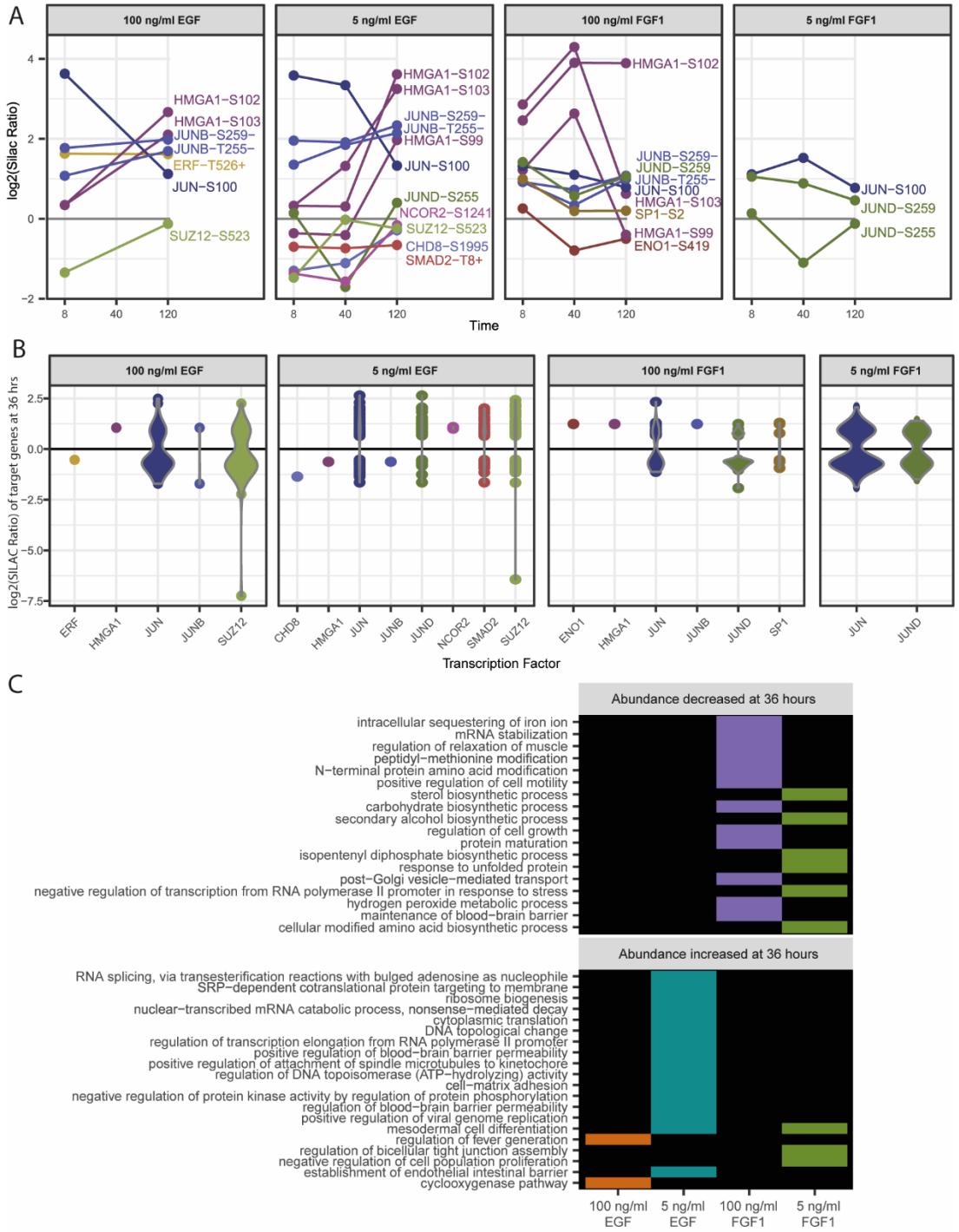


Figure 5-6. Gene targets of differentially phosphorylated transcription factors have altered protein abundance 36 hours following stimulation with high or low concentrations of EGF and FGF1.

A. *Phosphorylation profiles of transcription factors found in the CHEA or TRRUST databases. Phosphorylated sites with a known effect on induction or inhibition of transcription factor activity are marked with a + or a -, respectively, at the end of the phosphorylated site identifier.* ***B.*** *Protein abundance of target genes of differentially phosphorylated transcription factors 36 hours after stimulation.* ***C.*** *Enrichment for GOBP terms of transcription factor target genes whose abundance is up or down at 36 hours in each experimental condition.*

5.3.4 Connecting dynamic phosphorylation with long-term changes

To connect the results from the short-term phosphoproteome and 36-hour proteome analysis, the overlap between the GOBP terms from both sets of results was found (Figure 5-7A). Only terms in the 5ng/mL EGF and 100 ng/mL FGF1 stimulant conditions had GOBP term(s) in common; therefore, only the clusters associated to these terms were further analysed (Figure 5-7B). Of particular interest were the broader cellular behaviours that were highly ranked. 'Cell-matrix adhesion' was ranked highly by clusters 2 and 3 of the 5 ng/mL EGF condition. In the 36-hour data, this term was enriched within the targets that had increased abundance (Figure 5-6C). In contrast, 'Regulation of cell growth', ranked highly by cluster 1 sites regulated by the 100 ng/mL FGF1 condition, was enriched among modified transcription factor targets with decreased abundance at 36 hours (Figure 5-6C).

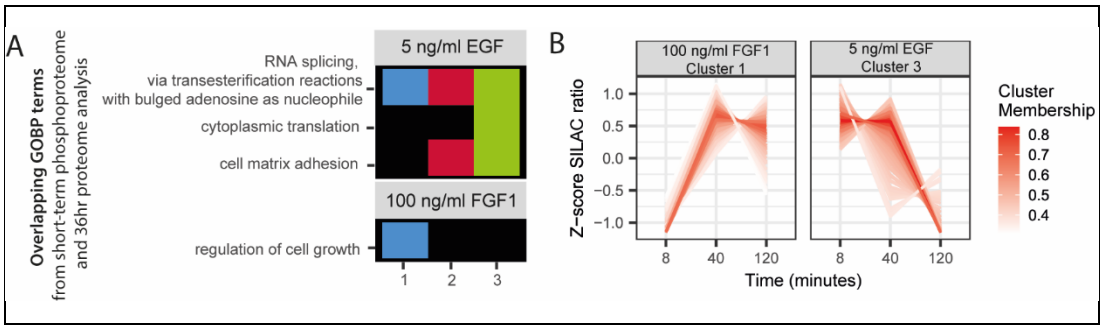


Figure 5-7. Overlap of terms found in phosphoproteome and proteome analysis.
A. *Subset of RWHN terms (Figure 5-5) also found in ORA of regulated gene targets at 36 hours (Figure 5-6).* ***b.*** *Phosphorylation dynamics of sites part of clusters 1 and 3 of 100 ng/ml FGF1 and 5 ng/ml EGF, respectively. Cluster membership refers to fuzzy C-means calculation of fit to cluster trend.*

To reveal how the phosphorylated sites in those clusters influenced long term behavioural trends, networks were constructed based on kinase-substrate and protein-protein interactions (PPI). Sites regulated by 5 ng/mL EGF found in cluster 3 (Figure 5-8A) and 100 ng/mL FGF1 found in cluster 1 (Figure 5-8B) were used as the basis of the network. Thus, the proteins on which the sites were found were connected based on whether they had a regulated phosphorylated site that was the substrate of another protein in the network (as recorded in PhosphoSitePlus) or have another known interaction (as recorded in STRING) (Hornbeck *et al.*, 2015; Szklarczyk *et al.*, 2019).

The network constructed based on 5 ng/mL EGF cluster 3 had a large, connected component containing well-known signalling kinases, adaptors, and transcription factors, including MAPK1/3, PIK3C2A, SHC1, SOS1 and JUN. The latter was shown to have targets with increased and decreased abundance at 36 hours (Figure 5-6B). The dual response could be

explained by the presence of ERF, a transcriptional repressor, in this cluster (Hou *et al.*, 2020). A notable but small component of the network contains RAB7A, CCZ1 and HNRNPK. RAB7A is a marker for the late endosome and responsible for diverting endosomal material from the early to late endosome (Naslavsky and Caplan, 2018). It is in its active form upon the exchange of guanine diphosphate to guanine triphosphate by guanine exchange factors such as CCZ1 (Vaites *et al.*, 2017). It is known that though the majority of EGFR is recycled to the membrane when stimulated with low concentrations of EGF, approximately a third is still degraded (Sigismund *et al.*, 2005, 2008). The RAB7A site S72 is found in both the high and low EGF stimulant conditions and follows a matching regulation trend at a lower amplitude, however phosphorylation on CCZ1 was not detected upon 100 ng/mL EGF stimulation (Figure 5-9A). The role of CCZ1-266 phosphorylation is unknown; depending on the effect it has on CCZ1 activity, it may be key to modulating RAB7A in response to lower EGF concentration, with consequences on EGFR fate.

The network constructed from 100 ng/mL FGF1 Cluster 1 is far less connected (Figure 5-8B), however some indications on the molecular connection between short-term phosphorylation and long-term trends may still be drawn. MAPK3 and PI3KC2A are part of small components of the network, indicating that these are the predominant signalling pathways through which long-term trends are coordinated, as with the 5ng/mL EGF cluster 3 network. Interestingly, ERBB2 is found in this cluster, indicating a potential point of cross-talk, such as that described by Smith et al. (2021) between FGFR2b-EGFR. The largest cluster has several proteins involved in translation regulation downstream of mTOR. RPS6 is a known effector of mTOR and regulator of cell growth and size (Ruvinsky and Meyuhas, 2006). The elongation initiation factors (EIF), EIF3A, EIF4G1, EIF4G2 are not known to be direct targets of mTOR signalling but work in concert with other EIFs that are (Harris *et al.*, 2006; LeFebvre *et al.*, 2006; Xie and Proud, 2014). Given the work described in chapter 3, this cluster, and the presence of RPTOR (phosphorylated at S877), is interesting. RPTOR-S877 is listed in PhosphositePlus as inhibiting enzymatic activity, with several putative kinases: TBK1, GSK3B and ULK1. Inhibition of RPTOR would also limit mTORC1 activity and be associated to the promotion of autophagy and the suppression of the translation initiation proteins. This may be through a subcellular mechanism similar to that described downstream of FGF10 in chapter 2.

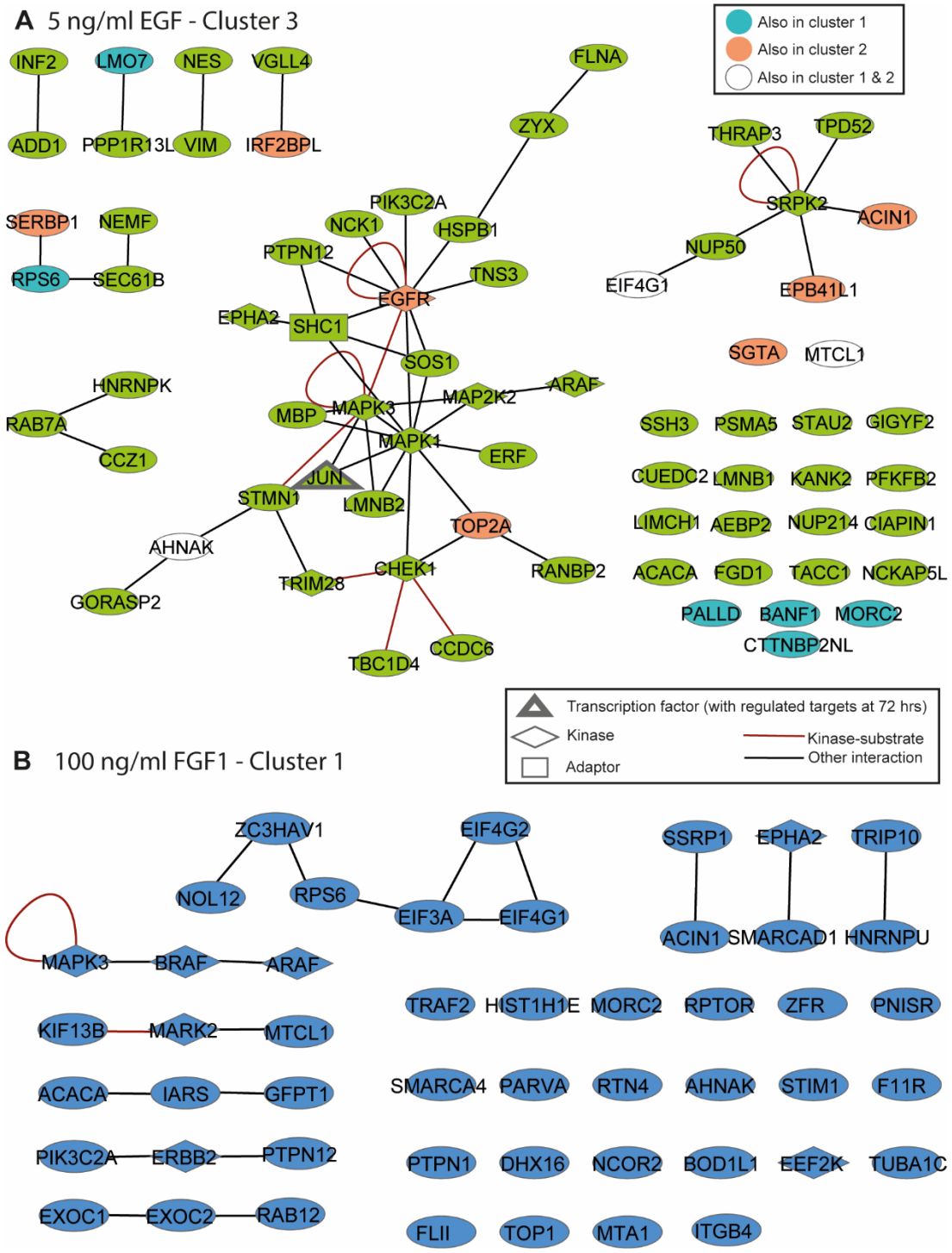


Figure 5-8. Networks based on kinase-substrate relationships (from PhosphoSitePlus) and other protein interactions (from STRING database).
*Constructed from proteins with phosphorylated sites found in 5 ng/ml EGF cluster 3 (**A**) and 100 ng/ml FGF1 cluster 1 (**B**).*

Another of the small network components contains RAB12, EXOC1 and EXOC2. RAB12 has been associated to a novel trafficking route from the recycling endosome to the lysosome, and may also be involved in the suppression of autophagy via mTORC1 in amino acid-rich conditions (Matsui and Fukuda, 2011, 2013; Matsui, Itoh and Fukuda, 2011). There is a moderate effect on the phosphorylation of the RAB12 site S21 in the high and low FGF1 stimulant conditions (Figure 5-9B). Whether this is a trafficking route utilised by FGFR1 has not been previously investigated.

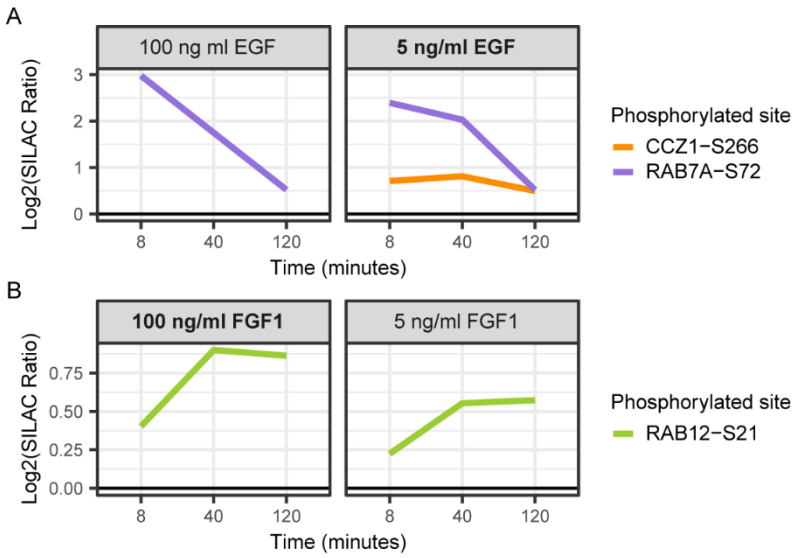


*Figure 5-9. Phosphorylation profiles of RAB7A-S72 and CCZ1-S266 downstream of EGF (**A**) and RAB12-S21 downstream of FGF1 (**B**).*

5.4 Discussion

The proteomics experiment described in this chapter brings together several fundamental concepts in RTK signalling: the comparison of two ligands, EGF and FGF1, at two different concentrations; the reconstruction of early phosphorylation events downstream of each growth factor; and the long-term changes in protein abundance induced by signalling cascades.

Given potential technical problems in the acquisition of the dataset, the work presented in this chapter may be best considered as a framework for how to analyse similar datasets. The novelty of the analysis method described is the association of short term phosphoproteomics response to longer term proteome changes via differentially modified transcription factors (Figure 5-3). We utilised GOBP term enrichment to connect differentially abundant proteins to putative short-term instigators of those processes. In this way, a link between stimulus and cellular output can be drawn entirely with the combination of phosphoproteomics and proteomics. A thorough search of the literature has not revealed a similar approach being taken by other studies, despite the concept of short-term phosphoproteomics changes

yielding longer-term changes in the proteome and protein abundance being well-appreciated in cellular signalling (Pope and Medzhitov, 2018). However, we note that the "missing link" in this approach is the incorporation of transcriptomics data, to validate whether changes in protein abundance are indeed due to phosphorylation-modulated transcription factor activity. It could be argued that assessing the 36-hour proteome is more beneficial than transcriptomics for investigating the effects of signalling cascades, as transcript and protein abundances are not directly correlated (Gygi *et al.*, 1999; Cox and Mann, 2012; Lundberg and Borner, 2019). Another 'omic layer that could be integrated is transcription factor-specific genomics through ChIP-seq. This would overcome overrepresentation in the databases utilised here of well-studied transcription factors such as JUN and JUN-D, which may be skewing the results (Mellis *et al.*, 2021). Changes in protein abundance may also be caused by other factors. Homeostatic processes that alter gene expression would be accounted for by SILAC normalisation. However, aside from transcription and translation, other cellular processes that alter protein abundance *in vitro* and may be regulated by the presence of EGF or FGF1 include stress response, proliferation, or migration.

A second methodological concept explored in this chapter is clustering of phosphoproteomics data. The inherent noise and density of phosphoproteomics data mean that selecting a clustering method, and moreover the number of clusters, is non-trivial. A comprehensive assessment on clustering of temporal phosphoproteomics data has yet to be performed, though some attention has been paid to the clustering of proteomics data (Albaum *et al.*, 2011). How to measure the quality or appropriateness of a particular clustering method is difficult without a suitable means of validating the partitioning of the data or a "gold-standard" dataset with pre-validated phosphorylated sites to benchmark against. In biological studies, the former may be achieved through experimental approaches, which is not feasible when investigating several differentially regulated phosphorylated sites. However, this is often a critical step in the analysis of phosphoproteomics data that can lead to distinctly different outcomes in the analysis. Considering that small changes in the stoichiometry of phosphorylated sites can alter protein function and fate (Olsen *et al.*, 2010), a crucial phosphorylated site being included in one cluster or another may highlight it as a candidate or be missed. Without the ability to validate the partitions nor a benchmarking dataset, two aspects of clustering methods were focussed on. The first was the overlap between different clustering methods. Generally, the two stochastic methods tested, k-means and Fuzzy C-means, had the highest overlap, but this was altered considerably based

on the number of points associated to each phosphorylated site (two time points, 8 and 120 mins, for 100 ng/mL EGF regulated phosphorylated sites, three time points for all other stimulant conditions), indicating that the dimensionality of the data has an impact on the clustering. The second assessment made was on the overlap between methods to estimate optimal number of cluster partitions. Several were assessed, including a phosphoproteomics specific method, named CLUE, that bases the estimation on enrichment for kinase-substrate interactions (P. Yang *et al.*, 2015). The estimates were generally within a small range for each stimulant condition; this level of conformity between the different methods is perhaps expected give the moderate overlap in cluster partitioning with the different methods.

The clusters selected were used as input for the RWHN algorithm. A large number of GOBP terms were highly ranked by all stimulant conditions, but particular by 5 ng/mL EGF and 5 ng/mL FGF1. This was unexpected, given that one of the interesting aspects of the algorithm highlighted in the original paper was the more refined number of GOBP terms returned. This could be because the data presented in this chapter was too noisy or were poorly clustered, despite the extensive evaluation. The resulting terms from the RWHN analysis were filtered for those that were enriched in the 36-hour transcription factor targets for each stimulant condition, to screen for processes relevant to this analysis. However, this would not be possible when analysing a stand-alone phosphoproteomics dataset. Therefore, the cause of so many terms being highly ranked requires further investigation.

Several biological insights have been gleaned from this analysis and could be further validated by designing appropriate experiments. It is clear that ligand abundance has an important impact on cell fate. For example, that proteins with increased abundance downstream of 5ng/mL FGF1 were enriched for 'negative regulation of cellular proliferation' would indicate a ligand and concentration dependent effect on cell cycle progression. This would also indicate a similar role for FGF1 in regulating proliferation as FGF2, which inhibits proliferation at low and high concentrations (Garcia-Maya *et al.*, 2006); our results do not indicate a role for high FGF1 concentrations in inhibiting proliferation but this could be validated experimentally (e.g. using a proliferation assay based on EdU incorporation). Previous unstudied behaviours associated to differences in ligand abundance were cell matrix adhesion upon treatment with 5 ng/mL EGF and regulation of cell growth upon treatment with 100 ng/mL FGF1. Whether the dynamic regulation of the signalling proteins identified in the clusters with GOBP terms in common with 36 hr transcription factor targets is important for mediating the long-term processes could also be investigated. For example,

MAPK3 and PI3KC2A are both phosphorylated components of the 5 ng/mL EGF and 100 ng/mL FGF1 clusters investigated in Figure 5-8, and potential targets of mTOR are highlighted downstream of 100 ng/mL FGF1 Figure 5-8B. Assays measuring cellular growth or adhesion downstream of high or low EGF or FGF1 upon perturbation of the sites and pathways highlighted in our analysis would provide stronger evidence for dynamic phosphorylation being a key regulator of these cellular behaviours. Perturbation could include pharmaceutical kinase inhibitors of the the MAPK, PI3K and mTOR pathways, or mutagenesis of the sites on transcription factors (e.g. JUN, JUNB), core pathway members (E.g. RPTOR, RPS6) or non-essential kinases (e.g. EGFR, ERBB2).

The effect on receptor trafficking also warrants further biochemical and functional investigation. The link between EGF concentration and EGFR has long been known: high concentration of EGF predominantly promotes internalisation by clathrin-independent endocytosis and receptor degradation, low concentration of EGF predominantly promotes internalisation by clathrin-mediated endocytosis and receptor recycling to the cell surface. Our results indicate that this may be mediated by RAB7A and its activator CCZ1. The association of RAB12 to 'regulation of cell growth' downstream of 5 ng/mL FGF1 is also intriguing, given that RAB12 primarily coordinates a trafficking pathway between the recycling endosome and the lysosome. Cell growth and autophagy via AMPK and mTOR are generally considered to be regulated from the lysosome, however there is evidence (including that described in chapter 3) indicating other endosomal compartments may also be involved (Fraser *et al.*, 2019; Lamming and Bar-Peled, 2019). However, a trafficking route for FGFR family members between the lysosome and recycling endosome has yet to be described; whether this is biologically relevant is worthy of further investigation. Co-localisation of EGFR with RAB7A and FGFR with RAB12 downstream of high or low EGF and FGF1 using a microscopy-based immunofluorescence assay would be a necessary first step for validating these conclusions. The specific dynamics could then be interrogated. To investigate the role of CCZ1, the protein could be knocked-out or knocked-down using siRNA or the phosphorylated site (S266) mutated and the immunofluorescence assay repeated, to observe whether the trafficking pathway or level of degradation of EGFR is altered. A useful experiment in all case may be to observe the role of intracellular signalling downstream of high or low EGF or FGF1, by retaining the receptors at the plasma membrane using a mutation in the membrane fission dynamin (dominant-negative dynamin, as used in chapter 3) and immunoblotting for the changes in phosphorylation of the signalling proteins highlighted in our analysis (MAPK3, PI3KCA, mTOR).

6 Discussion

6.1 Summary of findings

Phosphorylation drives signalling cascades through the successive modification of proteins downstream of extracellular stimuli. The advent, increasing accessibility and continuing improvement of phosphoproteomics methodology has created a wealth of data describing phosphorylation events in signalling cascades. However, a few core questions remain in the bioinformatics analysis of phosphoproteomics data: how to determine whether a phosphorylated site is contributing to the functional role of a protein or towards a biological process; and, since every potential candidate cannot be experimentally validated, how functionally relevant phosphorylated sites can be identified or predicted *in-silico*.

These questions are particularly crucial in the context of phosphoproteomics experiments describing RTK signalling. Small changes in phosphorylation site stoichiometry may have a large impact or be non-functional unless in tandem with other modifications. Moreover, many questions remain in understanding how phosphorylation regulates RTK signalling on the systems level, downstream of ligand-based regulation such as functional selectivity or ligand abundance. RTK signalling can be modelled well with networks, however it is difficult to integrate the phosphorylation state of proteins (as discovered in phosphoproteomics) into typical PPI network representations.

The research presented in this thesis describes the application of original and existing network-based methods to understand the functional output of phosphorylation in RTK signalling cascades. A key contribution of the thesis was the development of a method that aims to infer the functions of phosphorylated sites based on shotgun phosphoproteomics datasets, as described in chapter 2. This method first constructs a multilayer network based on shotgun phosphoproteomics data; this allows for multiple aspects of the data to be integrated into the analysis. Our approach describes site centric means to analyse phosphoproteomics data and is an improvement on widely used enrichment analyses.

In Chapter 3, work from a multi-phosphoproteomics study investigating the subcellular signalling modules of FGFR2b is described. My contribution to this work was the analysis of a phosphoproteomics data set based on APEX2 proximity-dependent biotinylation. At the time of submission, a phosphoproteomics dataset collected through proximity dependent biotinylation had not been described. The analysis of the data revealed that proximal phosphoproteomes of internalised RTKs could be uncovered using APEX2 proximity

dependent biotinylation. This approach revealed localised suppression of autophagy via the mTOR pathway by FGFR2b in the recycling endosome. The work in Chapter 3 revealed a need to better visualise and understand how protein or gene lists derived from high throughput 'omic experiments relate to subcellular localization. Therefore, in Chapter 4, I described the development of a web app to address this need. SubcellularVis is an efficient tool to visualise subcellular compartment enrichment.

Finally, in chapter 5 I address another means of RTK regulation: ligand abundance. I collected and analysed a high dimension (phospho)proteomics dataset describing short-term phosphorylation and long-term proteomics changes downstream of high and low concentrations of EGF and FGF1. This analysis combined the site-centric method described in chapter 2 and a novel approach that assessed the impact of transcription factor modification using changes in protein abundance.

Taken together, these key findings address the objectives and aims outlined in section 1.6.

6.2 Interpreting, predicting, and validating phosphorylation-mediated function

A definitive view of biological function of proteins has long been debated. The selective-effect definition of function is "why a molecule does what it does", whilst the causal-effect definition is "what the molecule does" (Keeling *et al.*, 2019). Projects such as the Gene Ontology have created a widely accepted definition of function among molecular biologists to encompass both of these, with the biological process aspect addressing the selective-effect and the molecular function aspect addressing the causal-effect of gene products (Thomas, 2017). However, GO annotations are not the only way molecular biologists may consider function. High-throughput methods, particularly those describing the role of modified protein dynamics, force us to consider whether the selective- or causal-effect of biological function is relevant to the experimental question.

Selective-effect functions can be driven by modifications on multiple proteins. This is the central assumption of the method described in chapter 2, where the input data is phosphorylated sites that have been clustered based on similar patterns of regulation. The clusters are then described based on the output of the RWHN algorithm. The RWHN method is very different from standard enrichment analyses such as ORA, but the output is essentially interpreted in the same way. As with enrichment analyses, the next step is the functional validation of phosphorylated sites identified in a cluster of interest. The gold-standard remains wet-lab experimental validation that the biological function is indeed being

differentially regulated by the sites (Handler *et al.*, 2018), however not every candidate site within a cluster associated with functionally interesting terms can be validated. For instance, previously unstudied phosphorylated sites may not have suitable reagents targeting them, or the cluster may be too large to allow feasible experimental validation for all of them. For bioinformatics validation of sites within a cluster of interest, cross-referencing with phosphorylated site-specific databases, such as PhosphoSitePlus, can help to identify "true" functional associations. However, focussing only on those annotated sites feeds into the rich-getting-richer effect (briefly discussed in chapter 2). One of the benefits of the RWHN method is in inference of the functions of phosphorylated sites which can drive hypothesis generation, however a bottleneck remains in validating these hypotheses.

Causal-effect function in the context of phosphoproteomics can be thought of as phosphorylation induced interactions, modulation of protein activity and induced protein relocalisation. I use this meaning of biological function to define the subcellular localization of FGFR2b in Chapter 3. The background of the project is based in the causal-effect of phosphorylation of Y734 on FGFR2b, which switches the trafficking route of the receptor towards recycling (Francavilla *et al.*, 2013). However, it was not known whether the modification-induced relocalisation of FGFR2b contributed to the system-level changes in function (I.e., the shift towards a more proliferative cellular state). By defining the interactors of FGFR2b at the recycling endosome I could differentiate the role of global FGFR2b signalling versus internalised signalling. We uncovered a module of autophagy inhibition regulated by FGF10-activated FGFR2b from the recycling endosome. Moreover, the phosphorylation state of the receptor itself was different in the global profile vs. the recycling endosome profile, as two sites on FGFR2b (Y586 and S587) were dephosphorylated in the proximal phosphoproteome. This indicates that at 40 minutes post-stimulation with FGF10, there may be two distinct populations of FGFR2b defined by their phosphorylation state and proximity to certain phosphatases or kinases.

There are occasions where these causal- and selective-effects are not so distinct, such as when an individual protein modification can semi-independently induce a systems level change. This is demonstrated by the phosphorylation of transcription factors. I investigate this concept in chapter 5 by identifying phosphorylated transcription factors in each of the stimulant conditions and then extracting their target genes from the 36-hour proteome data. Enrichment for GOBP terms was calculated for those proteins that had increased or decreased abundance in each stimulant condition. This analysis was inspired by the work of

Mellis *et al.* (2021), who used transcriptomic profiling to associate transcription factors that strongly responded to drug perturbations as crucial for maintaining cell identity; siRNA knock down of those transcription factors increased the efficiency of reprogramming the cells and differentiating them to another cell type. The responsiveness to perturbation was assessed by calculated enrichment for GOBP terms in up or downregulated gene targets. It can be noted that transcriptomic changes are often dampened on the proteome level due to translational regulation, which is why I wanted to explore the perturbation using the same approach with proteomics data. The main difficulties were in assessing which transcription factors were having the biggest impact and how to relate this back to the phosphorylation dynamics within the signalling cascade. I propose one way of doing the latter by filtering the results from the site centric RWHN method for those also found in the enrichment results of the regulated proteins. The results would have to be experimentally validated in order to confirm how successful this analysis method is.

6.3 Scaling proteomics analysis methods to phosphoproteomics

Phosphoproteomics is a field still undergoing maturation, relative to proteomics. Reproducibility, missingness, identification difficulties and efficient sample preparation are all aspects that continue to affect phosphoproteomics and have implications for data analysis. It is not uncommon for methods developed in more mature omics fields to be applied to less established ones. However, each 'omics methodology is often highly specific to the properties of the molecules being addressed. In the case of applying protein-level analysis methods to phosphoproteomics, or indeed any modified proteome, the compromise of ignoring site-specific information or the modification state of the protein is inherent.

The method described in chapter 2 is an important development to complement existing site-centric methods for the analysis of shotgun phosphoproteomics data. Methods such as ssGSEA and PHOTON are useful for the functional analysis of phosphoproteomic datasets where well studied proteins and sites are being investigated or are the focus of the analysis. However, it is limited by study bias found in PTMsigDB (on which ssGSEA is based) and PPI networks (on which PHOTON is based). Our approach aims to reduce some of this bias by indirectly associating phosphorylated sites to functions, via a PPI network. The site-specificity of our method could be further improved by integrating site-specific data, for example kinase-substrate interactions in place of PPIs in the protein sublayer and functional annotations from databases such as PhosphositePlus or PTMsigDB. In the initial version of the algorithm, I chose not to integrate these data types since they have low coverage of the

phosphoproteome, but as phosphorylated site functions and kinase-substrate relationships are uncovered these resources will become more easily integrated into data-dependent analysis methods. Another advantage of existing methods over RWHN is speed and computational intensity. In its current form, the method takes several hours to run on a relatively complex data set.

The use of protein-level or proteomics specific methods was an issue encountered when analysing the APEX dataset acquired in chapter 3. In previous proximity dependent biotinylation studies, the SAINT algorithm has been a commonly used tool for identifying high-confidence interactors of the bait protein by calculating the probability of "true" interactions based on intensity or spectral counts. Lobingier et al. (2017) used SAINT to identify the interactors of GPCRs at different subcellular locations. However, SAINT relies on protein-level identifiers. Therefore, in order to use the SAINT algorithm, the quantitative phosphoproteomics values would have had to be summarised to the level of the protein. This is problematic; if two sites on the same protein have diverging phosphorylation patterns, summarising these will nullify their effect. The use of SAINT in this context would also have to be validated using a benchmark phosphoproteomics dataset, given that it was designed for proteomics data. Our study proves that proximity dependent biotinylation methods can be used in conjunction with phosphoproteomics to gain biological insights, however statistical methods are lacking and need to be developed or scaled from proximity dependent biotinylation-proteomics.

I chose to use an alternative protein centric method to analyse the APEX2 dataset. By using PPI networks and protein-level enrichment analysis to understand the proximally regulated phosphoproteome, I avoided the issue of having to summarise site-specific data. In the case of this dataset, the quantification was less relevant than the identification, and therefore was not required for the downstream analysis, avoiding the issue of summarising quantitative values on phosphorylated sites. I represented the PPI network with the phosphorylated sites depicted at the protein nodes, highlighting the role of modification in modulating these interactions (Figure 3-10F). It would have been useful, however, to encode in the network the functional roles of the sites on the proteins. For example, ULK1 is identified in the global and proximal phosphoproteome analysis; in the global phosphoproteome site S469 is identified and in the proximal phosphoproteome S638 is identified. Both are annotated on PhosphoSitePlus as inhibitory sites, adding detail to the regulation of the protein within the network. Moreover, only the latter is a putative mTOR/AMPK substrate. It would have been

interesting to integrate kinase-substrate analyses with the network representation to uncover the differential patterns of ULK1 phosphorylation in the global and proximal phosphoproteome, which was later confirmed by wet-lab experimental analysis.

The analysis of the data in Chapter 3 inspired the development of the tool described in Chapter 4. The use of enrichment analysis for phosphoproteomics data, as previously stated, requires disregard of the identity and quantitation of individual phosphorylated sites. Therefore, this tool is more appropriate for proteomics and transcriptomics analysis. Given that a detailed map of modified proteins and their subcellular locations is not currently available, though may not be far off given technological advances that have presented broad proteome level maps of this kind (Go *et al.*, 2021; Huttlin *et al.*, 2021), a generalised picture of the subcellular location of the protein which is modified will add some description alongside other data analysis.

In chapter 3, the difference in Andromeda identification scoring and intensity distribution between global and proximal phosphoproteomes indicated that the double enrichment step (for biotinylated and phosphorylated peptides) during sample preparation was somewhat skewing the data. This was not seen on the proteome level and has not been reported in previous proteomics studies that used APEX proximity dependent biotinylation. There are several potential reasons for this. First, it may be that the biotin-phenol group on the peptide skews the identification and quantification of the peptide. Evidence for this hypothesis is that MaxQuant, with Andromeda database search, did not find a significant number of biotinylated peptides in the data. A second potential reason for the divergence in intensities and identification confidence is that double enrichment of peptides samples is highly stringent and causes many more peptides to be lost during sample preparation. A final reason could be the use of LFQ. It has been suggested that LFQ is not ideal for APEX2 based MS studies (Hung *et al.*, 2016), though has been used to define proteomes with success (Lobingier *et al.*, 2017; Del Olmo *et al.*, 2019). It is hard to say whether an isobaric or metabolic labelling strategy would have prevented the difference in identification confidence and intensity distribution for the phosphoproteomics samples. This is something to be investigated to ensure correct use of the methodology.

6.4 Capturing dynamic signalling processes using phosphoproteomics

Phosphorylation mediated signalling cascades are a continuous process that has to be discretized in order to be studied with ‘omics technologies. Given the expense of resources

and time required to run phosphoproteomics experiments, the appropriate selection of time points to study dynamic signalling is particularly relevant. This highlights the importance of preliminary wet lab work for informing the experimental design of MS-based phosphoproteomics experiments. For example the time points selected for the experiments described in chapter 3 and 5 were selected based on previous work (Francavilla *et al.*, 2013, 2016). The concentrations of EGF and FGF1 were chosen based on previous work, and points in the dynamic range of both ligands that would be comparable, (I.e. what could be considered "high" and "low" for each) (Sigismund *et al.*, 2005, 2008; Reddy *et al.*, 2016).

PPI network representations of signalling networks are static portrayals of a dynamically evolving system. As discussed in the introduction, network-based modelling is one way of projecting dynamic relationships onto the network. This is the basis of work from the Saez-Rodriguez group in the development of the CellNOptR and PHONEMeS software suites (Terfve *et al.*, 2012, 2015; Gjerga *et al.*, 2021). I represent dynamics differently using a multilayer network approach. Phosphorylation dynamics are encoded by clustering in pre-processing of the data; phosphorylated sites that are clustered together are connected in the multilayer heterogeneous network. Though our network representation is not useful for understanding the flow of phosphorylation through the network nor the influences on network topology, it is more useful for understanding the functional relevance of the different phosphorylation dynamics than traditional network modelling approaches.

Spatial regulation is an aspect of signalling regulation of RTKs that has been discussed at length in this thesis. Given that relocalisation of proteins is not necessarily coordinated by phosphorylation it cannot be fully captured by standard shotgun phosphoproteomics experiments alone. Therefore, as the signalling community takes a holistic view of the regulation of signalling cascades, integration of phosphoproteomics datasets with other proteomics methodologies, such as the APEX2 spatial profiling described in chapter 3, will become a vital tool. This would allow easier reconciliation between phosphorylation-based regulation and spatial regulation that is only available through low throughput methods such as immunofluorescent assays and western blots. In its most useful and broadest form, the approach described in chapter 3 could be expanded to include a panel of subcellular markers that can be used to define dynamic spatial profiles of any molecule that relocalises upon perturbation.

The quality of the data in chapter 5 also highlights how making improvements in phosphoproteomics technologies accessible to the wider community will allow dynamics of

signalling cascades to be better studied. Currently, state-of-the-art instrumentation is limited to specialist laboratories. This limits the power of phosphoproteomics datasets collected in conventional laboratory set-ups. Given a time course of three time points, as in the dataset from Chapter 5, missing values at any point in the time course cannot be interpreted fully: does the null value indicate that the phosphorylation was at such low stoichiometry at that timepoint that it was below the limit of detection, or was it missing at random? Since the time course is an ordered vector of quantitative values, imputation can drastically change the interpretation of the dynamic regulation at the site and skew analysis. However, removing all sites with a missing value at one time point is a drastic filtering measure and could also be biasing the analysis (Bartlett, Harel and Carpenter, 2015). With more time and resources, this experiment would have been repeated and more replicates would have been included to improve the correlation. This would have allowed for more confidence in the results.

An aspect of dynamic signalling processes not investigated in this work is the contribution of signalling crosstalk, either from the activation of other signalling cascades or from PTMs. *In vitro*, the complexity of cross-activation of signalling cascades or pathways can be controlled for, by only activating one receptor or receptor family at one time (e.g. stimulation with EGF in serum-starved cells will induce EGFR signalling cascade). However, this is not realistic of the *in vivo* environment, in which cells may receive and respond to many extracellular signals. Indeed, work from our group has shown that EGFR is reciprocally primed by FGFR2b in the recycling endosome downstream of FGF10 stimulation (Smith *et al.*, 2021). Within the signalling cascades themselves, crosstalk can be introduced through other transient PTMS, such as ubiquitination. One mode is through the creation of recognition motifs, for example phosphodegrons, which are associated with E3 ubiquitin ligase recruitment and subsequent protein degradation (Cardozo and Pagano, 2004). Conversely, phosphorylation can regulate the activity of deubiquitinases such as USP44 (Stegmeier *et al.*, 2007). Phosphorylation dependent SUMOylation, acetylation, or methylation and protease activation have also been described (Roux and Thibault, 2013). Phosphoproteomics methodologies do not capture these correlated PTMs nor their functional effects. Multi-omics techniques that integrate multiple PTMs will be necessary to integrate the signalling cascade crosstalk, as has been previously utilised (Francavilla *et al.*, 2016; Rivera *et al.*, 2021). The RWHN method described in chapter 2 could be expanded to include the other modifications from such experiments, adding and more biologically descriptive detail to the multilayer heterogeneous network.

6.5 Biological insights on the regulation of FGFR and EGFR

The studies presented in Chapters 3 and 5 of this thesis describe different aspects of ligand and receptor-based regulation of downstream RTK signalling.

Chapter 3 provides the most comprehensive overview to date of the spatial regulation of FGFR2b. Through the multi-phosphoproteomics approach, the consequences of maintaining FGFR2b at the plasma membrane and the recycling endosome describe the consequences of receptor regulation on the FGFR2b signalling cascade. In combination with other published work, our study also describes the diverging roles of FGF7 and FGF10 in regulating autophagy. Previous work has described the functional selectivity of FGFR2b to FGF7 and FGF10, which induce migration and proliferation respectively (Francavilla *et al.*, 2013). FGF7 has previously been described as a promotor of autophagy (Nanni, Ranieri, Raffa, *et al.*, 2018; Nanni, Ranieri, Rosato, *et al.*, 2018). Our study therefore demonstrates another contrasting behaviour between FGF7 and FGF10 with consequences for cell fate. It remains to be seen whether FGF7 mediates its effects on autophagy from the plasma membrane or intracellularly, for example in the late endosome or lysosome. This could be tested using the FGFR2b-APEX2 system described, with an APEX-tagged marker for late endosome/lysosome (e.g. RAB7, LAMP1) used in place of the RAB11-APEX2 recycling endosome marker we describe.

The work in chapter 5 is still in progress. However, initial analysis predicts which experiments could be performed, which are described in the discussion of the chapter. Firstly, functional assays to test changes in autophagy/cell growth, proliferation, and adhesion downstream of high or low concentrations of EGF or FGF1 should be performed, in order to validate the hypotheses and the approach of assessing long term trends through modified transcription factors and subsequent translational programmes. Introducing perturbations of key signalling pathways and proteins, through kinase inhibitors or site-specific mutagenesis, would confirm some of our molecular predictions, for example the contribution of downstream mTOR effectors upon high concentration FGF1 stimulation. Secondly, the role of RAB7A and RAB12 in mediated dose-dependent responses and unique trafficking pathways, downstream of EGF and FGF1 respectively, could be investigated using immunofluorescence assays and biochemical perturbation of expected trafficking routes. Finally, the role of specific transcription factors could be assessed by evaluating changing expression of their targets with either qPCR or RNAseq. This would add credibility to the analysis approach described.

There is considerable value in comparing the signalling cascades downstream of FGFR and EGFR as described in Chapter 5. EGFR continues to be considered the model RTK, however this has been questioned based on structural studies that highlight the differences in how EGFR is primed prior to ligand activation. Moreover, the focus on EGFR in studying trafficking of RTKs has been shown by our study to not represent all modes of RTK trafficking. Unpublished observations from the APEX2 dataset from chapter 3 also indicate that FGFR2b may be present in distinct populations in the endosomal system downstream of FGF10 signalling, with a proportion present at the Golgi apparatus or endoplasmic reticulum where they may interact with retromer or retriever complexes. EGFR, on the other hand, was shown to interact with retromer/retriever in unstimulated conditions. A comparative study of internalised EGFR and FGFR2b would likely reveal further points of distinction between the regulations of these important RTKs.

6.6 Future perspectives

The RWHN method described in Chapter 2 provides an alternative to protein-centric analysis methods for the phosphoproteomics community. Multilayer methods have been increasing in popularity in biological network modelling and have been shown to be useful in associating molecular actions to phenotypes (Lee *et al.*, 2020). However, they had not been applied to shotgun phosphoproteomics data sets before; I proved here that multilayer heterogeneous networks are a useful framework for representing shotgun phosphoproteomics data sets in order to uncover the functional outputs of differentially phosphorylated sites. For routine integration of the method into the analysis of phosphoproteomics data several improvements could be made. The integration of site-specific annotations and kinase-substrate interactions need to be tested within the network. Using such data would be reliant on their increased coverage of the phosphoproteome. Testing the integration of other annotations and interaction types would ideally be done using a gold standard benchmarking dataset, which was lacking at the time the original work was published. An appropriate dataset to benchmark against would undoubtedly help to optimise and improve the method. Another improvement would be the accessibility of the method to researchers without a programming background, perhaps into a pipeline of other computational tools for the pre-processing of the data. The method is currently accessible as a package downloadable from GitHub but would ideally be stored on a package repository such as CRAN or Bioconductor in order to be well-utilised and trusted by bioinformaticians. Finally, implementation in another programming framework could improve the computational efficiency of the method and therefore be worth investigating.

Multi-omics approaches – that is, the integration of datasets derived from different 'omic technologies – are continuing to increase in popularity (Dugourd and Saez-Rodriguez, 2019). The multilayer network representation is a particularly useful framework for multi-omics, as demonstrated by Santra *et al.* (2019). Our method could be further expanded to incorporate several layers of 'omic information to gain insights on functional regulation. An interesting application could be in studying PTM crosstalk, with a separate layer of the network for each type of PTM that converge on the protein layer (Leutert, Entwisle and Villén, 2021). Though running RWHN on a network containing more layer's worth of nodes would require significant computational resources and time, it would be worth investigating whether modules of functional regulation across molecular layers could be uncovered with our approach.

The APEX2 dataset described in Chapter 3 is a unique and invaluable resource for understanding the internalised, phosphorylated interactome of FGFR2b. Though one biological insight is discussed in detail in the paper -the proximal suppression of autophagia bye internalised FGFR2b - there are more discoveries to be found from this data set. One example is identification of potential populations of FGFR2b internalised in separate subcellular compartments. This could be further investigated through comparison with recently published protein atlases that describe high confidence markers of subcellular compartments and experimental work to assess weather internalised FGFR2b co-localises with markers of the endoplasmic reticulum, Golgi apparatus or members of retromer/retriever complexes. We also believe there is a link between actin cytoskeletal remodelling (through the actin binding protein cortactin, CTTN) and internalised FGFR2b. More time will be needed to investigate these potential hypotheses fully. It also remains to be seen whether other RTKs may follow a similar internalisation response or coordinate similar internalised signalling modules.

As previously alluded, the investigation of spatially restricted, proximal phosphoproteomes through APEX2-based proximity dependent biotinylation provides researchers a powerful methodology through which to analyse intracellular signalling modules. This method achieves a more holistic view of RTK signalling, by integrating the ligand, receptor, and protein network levels of regulation through experimental design, spatial profiling, and phosphoproteomics profiling, respectively. The role of signalling from the receptor at subcellular compartments is being increasingly appreciated, and there is tremendous scope for the APEX2-phosphoproteomics to be applied to many questions in this field. However,

the limitations need to be addressed, such as the questions regarding statistical analysis and correct quantification method, discussed above.

As with the RWHN method, I envision the SubcellularVis tool described in Chapter 4 to be useful to cell biology researchers aiming to get a general understanding of protein subcellular localisation from protein or gene lists. Not only can the tool be used for hypothesis generation and data exploration, but the tool also provides a visualisation of enrichment of cellular compartments that can be used to standardise the communication of such information in publications. This is demonstrated in Chapter 3 where SubcellulaRVis visualisations are well-utilised. Further improvements to the tool include the ability for users to upload a background gene list, that would improve the accuracy of the enrichment calculations. In future I hope to implement the graphical representation in the JavaScript package D3, which would provide more a rapid, flexible, and professional final visualisation.

Finally, the work in chapter 5 can be further expanded in several ways. Firstly, the issues with the quality of the dataset could be addressed by repeating the experiment. However, given the time and expense involved in mass-spectrometry based phosphoproteomics, the preliminary conclusions I have drawn could be further investigated with the experiments suggested in section 6.5. It would also be interesting to explore a multi-omics approach and include transcriptomics data in the analysis.

6.7 Conclusions

This thesis presents an original contribution of multiple bioinformatics analyses and tools to better understand the spatial and phosphorylation-based coordination of RTK signalling, and achieves the aims outlined in section 1.6. First, a phosphorylated site-centric analysis method was developed based on the RWHN algorithm, that successfully associates phosphorylated sites to their known functions. Secondly, complex phosphoproteomics datasets were analysed with network-based approaches to uncover regulation of cellular processes downstream of RTKs. My analysis uncovered an internalised FGFR2b signalling module at the recycling endosome that is responsible for the suppression of autophagy in an mTOR dependent manner and led to the development of a tool for visualising enrichment of subcellular localisation within protein lists. I also highlighted the role of ligand abundance in regulating RTK signalling cascades and proposed further experiments to investigate this. The work in this thesis demonstrates the valuable role of integrating network-based methods in the analysis phosphoproteomics datasets in order to better understand RTK signalling cascades.

References

Adam, K. and Hunter, T. (2018) 'Histidine kinases and the missing phosphoproteome from prokaryotes to eukaryotes', *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 98(2), pp. 233–247. doi:10.1038/labinvest.2017.118.

Adlung, L. *et al.* (2017) 'Protein abundance of AKT and ERK pathway components governs cell type-specific regulation of proliferation', *Molecular Systems Biology*, 13(1), p. 904. doi:10.15252/msb.20167258.

Äijö, T., Granberg, K. and Lähdesmäki, H. (2013) 'Sorad: a systems biology approach to predict and modulate dynamic signaling pathway response from phosphoproteome time-course measurements', *Bioinformatics*, 29(10), pp. 1283–1291. doi:10.1093/bioinformatics/btt130.

Akhmanova, A. and Hammer, J.A. (2010) 'Linking molecular motors to membrane cargo', *Current Opinion in Cell Biology*, 22(4), pp. 479–487. doi:10.1016/j.ceb.2010.04.008.

Albanese, C. *et al.* (1995) 'Transforming p21ras Mutants and c-Ets-2 Activate the Cyclin D1 Promoter through Distinguishable Regions (∗)', *Journal of Biological Chemistry*, 270(40), pp. 23589–23597. doi:10.1074/jbc.270.40.23589.

Albaum, S.P. *et al.* (2011) 'A guide through the computational analysis of isotope-labeled mass spectrometry-based quantitative proteomics data: an application study', *Proteome Science*, 9(1), p. 30. doi:10.1186/1477-5956-9-30.

Albeck, J.G., Mills, G.B. and Brugge, J.S. (2013) 'Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals', *Molecular Cell*, 49(2), pp. 249–261. doi:10.1016/J.MOLCEL.2012.11.002.

Alter, O., Brown, P.O. and Botstein, D. (2000) 'Singular value decomposition for genome-wide expression data processing and modeling', *Proceedings of the National Academy of Sciences of the United States of America*, 97(18), pp. 10101–10106. doi:10.1073/pnas.97.18.10101.

Ancot, F. *et al.* (2009) 'Proteolytic cleavages give receptor tyrosine kinases the gift of ubiquity', *Oncogene*, 28(22), pp. 2185–2195. doi:10.1038/onc.2009.88.

Aoidi, R., Maltais, A. and Charron, J. (2016) 'Functional redundancy of the kinases MEK1 and MEK2: Rescue of the Mek1 mutant phenotype by Mek2 knock-in reveals a protein threshold effect', *Science Signaling*, 9(412), pp. ra9–ra9. doi:10.1126/scisignal.aad5658.

Aoki, K. *et al.* (2011) 'Processive phosphorylation of ERK MAP kinase in mammalian cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(31), pp. 12675–80. doi:10.1073/pnas.1104030108.

Arkun, Y. and Yasemi, M. (2018) 'Dynamics and control of the ERK signaling pathway: Sensitivity, bistability, and oscillations', *PLOS ONE*. Edited by M. Thattai, 13(4), p. e0195513. doi:10.1371/journal.pone.0195513.

Ashall, L. *et al.* (2009) 'Pulsatile Stimulation Determines Timing and Specificity of NF-κB-Dependent Transcription', *Science*, 324(5924), pp. 242–246. doi:10.1126/science.1164860.

Ashburner, M. *et al.* (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25–29. doi:10.1038/75556.

Ashkenazi, A. *et al.* (2017) 'Polyglutamine tracts regulate beclin 1-dependent autophagy', *Nature*, 545(7652), pp. 108–111. doi:10.1038/nature22078.

Auciello, G. *et al.* (2013) 'Regulation of fibroblast growth factor receptor signalling and trafficking by Src and Eps8', *Journal of Cell Science*, 126(2), pp. 613–624. doi:10.1242/jcs.116228.

Aveic, S. *et al.* (2018) 'Autophagy inhibition improves the cytotoxic effects of receptor tyrosine kinase inhibitors', *Cancer Cell International*, 18(1), pp. 1–12. doi:10.1186/s12935-018-0557-4.

Avet-Loiseau, H. *et al.* (1998) 'High Incidence of Translocations t(11;14)(q13;q32) and t(4;14)(p16;q32) in Patients with Plasma Cell Malignancies', *Cancer Research*, 58(24), pp. 5640–5645. Available at: https://cancerres.aacrjournals.org/content/58/24/5640 (Accessed: 30 July 2021).

Babina, I.S. and Turner, N.C. (2017) 'Advances and challenges in targeting FGFR signalling in cancer', *Nature Reviews. Cancer*, 17(5), pp. 318–332. doi:10.1038/nrc.2017.8.

Backer, J.M. (2011) 'The Regulation of Class IA PI 3-Kinases by Inter-Subunit Interactions', in Rommel, C., Vanhaesebroeck, B., and Vogt, P.K. (eds) *Phosphoinositide 3-kinase in Health and Disease: Volume 1*. Berlin, Heidelberg: Springer (Current Topics in Microbiology and Immunology), pp. 87–114. doi:10.1007/82_2010_52.

Baffi, T.R. *et al.* (2021) 'mTORC2 controls the activity of PKC and Akt by phosphorylating a conserved TOR interaction motif', *Science Signaling*, 14(678). doi:10.1126/scisignal.abe4509.

Bahleda, R. *et al.* (2020) 'Phase I, first-in-human study of futibatinib, a highly selective, irreversible FGFR1–4 inhibitor in patients with advanced solid tumors', *Annals of Oncology*, 31(10), pp. 1405–1412. doi:10.1016/j.annonc.2020.06.018.

Bahrami, S. and Drabløs, F. (2016) 'Gene regulation in the immediate-early response process', *Advances in Biological Regulation*, 62, pp. 37–49. doi:10.1016/j.jbior.2016.05.001.

Bak, R.O., Gomez-Ospina, N. and Porteus, M.H. (2018) 'Gene Editing on Center Stage', *Trends in Genetics*, 34(8), pp. 600–611. doi:10.1016/j.tig.2018.05.004.

Balasubramanian, R. and Zhang, X. (2016) 'Mechanisms of FGF gradient formation during embryogenesis', *Seminars in cell & developmental biology*, 53, pp. 94–100. doi:10.1016/j.semcdb.2015.10.004.

Balázsi, G. and Oltvai, Z.N. (2005) 'Sensing Your Surroundings: How Transcription-Regulatory Networks of the Cell Discern Environmental Signals', *Science's STKE*, 2005(282), pp. pe20–pe20. doi:10.1126/stke.2822005pe20.

Ballif, B.A. *et al.* (2005) 'Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors', *Proceedings of the National Academy of Sciences*, 102(3), pp. 667–672. doi:10.1073/pnas.0409143102.

Barabási, A.L. and Albert, R. (1999) 'Emergence of scaling in random networks', *Science*, 286(5439), pp. 509–512. doi:10.1126/science.286.5439.509.

Barkovits, K. *et al.* (2020) 'Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition', *Molecular & Cellular Proteomics : MCP*, 19(1), pp. 181–197. doi:10.1074/mcp.RA119.001714.

Bartlett, J.W., Harel, O. and Carpenter, J.R. (2015) 'Asymptotically Unbiased Estimation of Exposure Odds Ratios in Complete Records Logistic Regression', *American Journal of Epidemiology*, 182(8), pp. 730–736. doi:10.1093/aje/kwv114.

Batrouni, A.G. and Baskin, J.M. (2020) 'A MAP for PI3K activation on endosomes', *Nature Cell Biology*, 22(11), pp. 1292–1294. doi:10.1038/s41556-020-00597-3.

Batth, T.S. *et al.* (2018) 'Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling', *Cell Reports*, 22(10), pp. 2784–2796. doi:10.1016/j.celrep.2018.02.038.

Battiston, F. *et al.* (2017) 'Multilayer motif analysis of brain networks', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4), p. 047404. doi:10.1063/1.4979282.

Bauer, M. *et al.* (2014) 'Evaluation of Data-Dependent and -Independent Mass Spectrometric Workflows for Sensitive Quantification of Proteins and Phosphorylation Sites', *Journal of Proteome Research*, 13(12), pp. 5973–5988. doi:10.1021/pr500860c.

Baumdick, M. *et al.* (2015) 'EGF-dependent re-routing of vesicular recycling switches spontaneous phosphorylation suppression to EGFR signaling', *eLife*. Edited by S.R. Pfeffer, 4, p. e12223. doi:10.7554/eLife.12223.

Beck, S. *et al.* (2015) 'The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics', *Molecular & cellular proteomics: MCP*, 14(7), pp. 2014–2029. doi:10.1074/mcp.M114.047407.

Behar, M. and Hoffmann, A. (2010) 'Understanding the temporal codes of intra-cellular signals', *Current Opinion in Genetics & Development*, 20(6), pp. 684–693. doi:10.1016/J.GDE.2010.09.007.

Bekker-Jensen, D.B. *et al.* (2020) 'Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries', *Nature Communications*, 11(1), p. 787. doi:10.1038/s41467-020-14609-1.

Belle, L. *et al.* (2015) 'The tyrosine phosphatase PTPN14 (Pez) inhibits metastasis by altering protein trafficking', *Science Signaling*, 8(364), p. ra18. doi:10.1126/scisignal.2005547.

Belleudi, F. *et al.* (2007) 'Keratinocyte Growth Factor Receptor Ligands Target the Receptor to Different Intracellular Pathways', *Traffic*, 8(12), pp. 1854–1872. doi:https://doi.org/10.1111/j.1600-0854.2007.00651.x.

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)*, 57(1), pp. 289–300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Ben-Sahra, I. *et al.* (2016) 'mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle', *Science (New York, N.Y.)*, 351(6274), pp. 728–733. doi:10.1126/science.aad0489.

Bento, C.F. *et al.* (2016) 'Mammalian Autophagy: How Does It Work?', *Annual Review of Biochemistry*, 85(1), pp. 685–713. doi:10.1146/annurev-biochem-060815-014556.

Bessman, N.J. *et al.* (2014) 'Complex relationship between ligand binding and dimerization in the epidermal growth factor receptor', *Cell reports*, 9(4), pp. 1306–1317. doi:10.1016/j.celrep.2014.10.010.

Beyer, E.M. and MacBeath, G. (2012) 'Cross-talk between Receptor Tyrosine Kinase and Tumor Necrosis Factor-α Signaling Networks Regulates Apoptosis but not Proliferation *', *Molecular & Cellular Proteomics*, 11(6). doi:10.1074/mcp.M111.013292.

Bezdek, J.C. (1973) 'Cluster Validity with Fuzzy Sets', *Journal of Cybernetics*, 3(3), pp. 58–73. doi:10.1080/01969727308546047.

Bielow, C., Mastrobuoni, G. and Kempa, S. (2016) 'Proteomics Quality Control: Quality Control Software for MaxQuant Results', *Journal of Proteome Research*, 15(3), pp. 777–787. doi:10.1021/acs.jproteome.5b00780.

Bill, C.A. and Vines, C.M. (2020) 'PHOSPHOLIPASE C', *Advances in experimental medicine and biology*, 1131, pp. 215–242. doi:10.1007/978-3-030-12457-1_9.

Binder, J.X. *et al.* (2014) 'COMPARTMENTS: unification and visualization of protein subcellular localization evidence', *Database*, 2014(0), pp. bau012–bau012. doi:10.1093/database/bau012.

Birrer, M.J. *et al.* (2007) 'Whole Genome Oligonucleotide-Based Array Comparative Genomic Hybridization Analysis Identified Fibroblast Growth Factor 1 As a Prognostic Marker for Advanced-Stage Serous Ovarian Adenocarcinomas', *Journal of Clinical Oncology*, 25(16), pp. 2281–2287. doi:10.1200/JCO.2006.09.0795.

Bjørbaek, C., Zhao, Y. and Moller, D.E. (1995) 'Divergent functional roles for p90rsk kinase domains', *The Journal of Biological Chemistry*, 270(32), pp. 18848–18852. doi:10.1074/jbc.270.32.18848.

Black, A.R. and Black, J.D. (2013) 'Protein kinase C signaling and cell cycle regulation', *Frontiers in Immunology*, 3. doi:10.3389/fimmu.2012.00423.

Blagoev, B. *et al.* (2003) 'A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling', *Nature Biotechnology*, 21(3), pp. 315–318. doi:10.1038/nbt790.

Blatti, C. and Sinha, S. (2016) 'Characterizing gene sets using discriminative random walks with restart on heterogeneous biological networks', *Bioinformatics*, 32(14), pp. 2167–2175. doi:10.1093/bioinformatics/btw151.

Blondel, V.D. *et al.* (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), p. P10008. doi:10.1088/1742-5468/2008/10/P10008.

Boissan, M. *et al.* (2014) 'Nucleoside diphosphate kinases fuel dynamin superfamily proteins with GTP for membrane remodeling', *Science (New York, N.Y.)*, 344(6191), pp. 1510–1515. doi:10.1126/science.1253768.

Bollaert, E., Serra, A. de R. and Demoulin, J.-B. (2019) 'The HMG box transcription factor HBP1: a cell cycle inhibitor at the crossroads of cancer signaling pathways', *Cellular and Molecular Life Sciences*, 76(8), pp. 1529–1539. doi:10.1007/s00018-019-03012-9.

Bork, P. *et al.* (1996) 'Structure and distribution of modules in extracellular proteins', *Quarterly Reviews of Biophysics*, 29(2), pp. 119–167. doi:10.1017/S0033583500005783.

Bork, P., Holm, L. and Sander, C. (1994) 'The immunoglobulin fold. Structural classification, sequence patterns and common core', *Journal of Molecular Biology*, 242(4), pp. 309–320. doi:10.1006/jmbi.1994.1582.

Borle, A.B. (1969) 'Kinetic analyses of calcium movements in HeLa cell cultures. I. Calcium influx', *The Journal of General Physiology*, 53(1), pp. 43–56. doi:10.1085/jgp.53.1.43.

Boyle, E.I. *et al.* (2004) 'GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes', *Bioinformatics*, 20(18), pp. 3710–3715. doi:10.1093/bioinformatics/bth456.

Bramer, L.M. *et al.* (2021) 'A Review of Imputation Strategies for Isobaric Labeling-Based Shotgun Proteomics', *Journal of Proteome Research*, 20(1), pp. 1–13. doi:10.1021/acs.jproteome.0c00123.

Brancho, D. *et al.* (2003) 'Mechanism of p38 MAP kinase activation in vivo', *Genes & Development*, 17(16), pp. 1969–1978. doi:10.1101/gad.1107303.

Brankatschk, B. *et al.* (2012) 'Regulation of the EGF Transcriptional Response by Endocytic Sorting', *Science Signaling*, 5(215), pp. ra21–ra21. doi:10.1126/scisignal.2002351.

Branon, T.C. *et al.* (2018) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–887. doi:10.1038/nbt.4201.

Braun, S. *et al.* (2004) 'Fibroblast growth factors in epithelial repair and cytoprotection', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1445), pp. 753–757. doi:10.1098/rstb.2004.1464.

Brenes, A. *et al.* (2019) 'Multibatch TMT Reveals False Positives, Batch Effects and Missing Values.', *Molecular & cellular proteomics : MCP*, 18(10), pp. 1967–1980. doi:10.1074/mcp.RA119.001472.

Brent, R. (2009) 'Cell signaling: What is the signal and what information does it carry?', *FEBS Letters*, 583(24), pp. 4019–4024. doi:10.1016/J.FEBSLET.2009.11.029.

Brewer, M.R. *et al.* (2013) 'Mechanism for activation of mutated epidermal growth factor receptors in lung cancer', *Proceedings of the National Academy of Sciences*, 110(38), pp. E3595–E3604. doi:10.1073/pnas.1220050110.

Browaeys-Poly, E. *et al.* (2010) 'Grb14 inhibits FGF receptor signaling through the regulation of PLCγ recruitment and activation', *FEBS Letters*, 584(21), pp. 4383–4388. doi:https://doi.org/10.1016/j.febslet.2010.09.048.

Brown, G.C. and Kholodenko, B.N. (1999) 'Spatial gradients of cellular phospho-proteins', *FEBS Letters*, 457(3), pp. 452–454. doi:10.1016/S0014-5793(99)01058-3.

Brüggemann, Y. *et al.* (2021) 'Growth factor-dependent ErbB vesicular dynamics couple receptor signaling to spatially and functionally distinct Erk pools', *Science Signaling*, 14(683), p. eabd9943. doi:10.1126/scisignal.abd9943.

Buchler, N.E., Gerland, U. and Hwa, T. (2003) 'On schemes of combinatorial transcription logic', *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), pp. 5136–5141. doi:10.1073/pnas.0930314100.

Buchtova, M. *et al.* (2015) 'Instability restricts signaling of multiple fibroblast growth factors', *Cellular and Molecular Life Sciences*, 72(12), pp. 2445–2459. doi:10.1007/s00018-015-1856-8.

Budayeva, H.G. and Kirkpatrick, D.S. (2020) 'Monitoring protein communities and their responses to therapeutics', *Nature Reviews Drug Discovery*, 19(6), pp. 414–426. doi:10.1038/s41573-020-0063-y.

Buffard, M. *et al.* (2019) 'Network Reconstruction and Significant Pathway Extraction Using Phosphoproteomic Data from Cancer Cells', *PROTEOMICS*, 19(21–22), p. 1800450. doi:10.1002/pmic.201800450.

Burack, W.R. and Sturgill, T.W. (1997) 'The Activating Dual Phosphorylation of MAPK by MEK Is Nonprocessive', *Biochemistry*, 36(20), pp. 5929–5933. doi:10.1021/bi970535d.

Burd, C. and Cullen, P.J. (2014) 'Retromer: A Master Conductor of Endosome Sorting', *Cold Spring Harbor Perspectives in Biology*, 6(2), p. a016774. doi:10.1101/cshperspect.a016774.

Burke, J.E. and Williams, R.L. (2015) 'Synergy in activating class I PI3Ks', *Trends in Biochemical Sciences*, 40(2), pp. 88–100. doi:10.1016/j.tibs.2014.12.003.

Campa, C.C. *et al.* (2018) 'Rab11 activity and PtdIns(3) P turnover removes recycling cargo from endosomes', *Nature Chemical Biology*, 14(8), pp. 801–810. doi:10.1038/s41589-018-0086-4.

Campello, R.J.G.B. and Hruschka, E.R. (2006) 'A fuzzy extension of the silhouette width criterion for cluster analysis', *Fuzzy Sets and Systems*, 157(21), pp. 2858–2875. doi:10.1016/J.FSS.2006.07.006.

Cánovas, B. *et al.* (2018) 'Targeting p38α Increases DNA Damage, Chromosome Instability, and the Anti-tumoral Response to Taxanes in Breast Cancer Cells', *Cancer Cell*, 33(6), pp. 1094-1110.e8. doi:10.1016/j.ccell.2018.04.010.

Cao, J. and Zhang, S. (2014) 'A Bayesian extension of the hypergeometric test for functional enrichment analysis', *Biometrics*, 70(1), pp. 84–94. doi:10.1111/biom.12122.

Capuani, F. *et al.* (2015) 'Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells', *Nature Communications*, 6(1), p. 7999. doi:10.1038/ncomms8999.

Cardozo, T. and Pagano, M. (2004) 'The SCF ubiquitin ligase: insights into a molecular machine', *Nature Reviews. Molecular Cell Biology*, 5(9), pp. 739–751. doi:10.1038/nrm1471.

Cargnello, M. and Roux, P.P. (2011) 'Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases', *Microbiology and Molecular Biology Reviews*, 75(1), pp. 50–83. doi:10.1128/MMBR.00031-10.

Carlson, M. (2019) *GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.8.2.*, *Bioconductor 3.10*.

Casado, P. *et al.* (2013) 'Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells.', *Science signaling*, 6(268), p. rs6. doi:10.1126/scisignal.2003573.

Casar, B. *et al.* (2009) 'Ras subcellular localization defines extracellular signal-regulated kinase 1 and 2 substrate specificity through distinct utilization of scaffold proteins', *Molecular and Cellular Biology*, 29(5), pp. 1338–1353. doi:10.1128/MCB.01359-08.

Casar, B., Pinto, A. and Crespo, P. (2008) 'Essential Role of ERK Dimers in the Activation of Cytoplasmic but Not Nuclear Substrates by ERK-Scaffold Complexes', *Molecular Cell*, 31(5), pp. 708–721. doi:10.1016/j.molcel.2008.07.024.

Caswell, P.T. *et al.* (2008) 'Rab-coupling protein coordinates recycling of alpha5beta1 integrin and EGFR1 to promote cell migration in 3D microenvironments', *The Journal of Cell Biology*, 183(1), pp. 143–155. doi:10.1083/jcb.200804140.

Chatr-aryamontri, A. *et al.* (2017) 'The BioGRID interaction database: 2017 update', *Nucleic Acids Research*, 45(D1), pp. D369–D379. doi:10.1093/nar/gkw1102.

Chellaiah, A.T. *et al.* (1994) 'Fibroblast growth factor receptor (FGFR) 3. Alternative splicing in immunoglobulin-like domain III creates a receptor highly specific for acidic FGF/FGF-1', *The Journal of Biological Chemistry*, 269(15), pp. 11620–11627.

Chen, H. *et al.* (2007) 'A Molecular Brake in the Kinase Hinge Region Regulates the Activity of Receptor Tyrosine Kinases', *Molecular Cell*, 27(5), pp. 717–730. doi:10.1016/j.molcel.2007.06.028.

Chen, J. *et al.* (2016) 'Expression and Function of the Epidermal Growth Factor Receptor in Physiology and Disease', *Physiological Reviews*, 96(3), pp. 1025–1069. doi:10.1152/physrev.00030.2015.

Chen, K.-E., Healy, M.D. and Collins, B.M. (2019) 'Towards a molecular understanding of endosomal trafficking by Retromer and Retriever', *Traffic*, 20(7), pp. 465–478. doi:https://doi.org/10.1111/tra.12649.

Chen, R.H., Sarnecki, C. and Blenis, J. (1992) 'Nuclear localization and regulation of erk- and rsk-encoded protein kinases.', *Molecular and Cellular Biology*, 12(3), pp. 915–927. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC369523/ (Accessed: 20 February 2021).

Chernikova, S.B. *et al.* (2018) 'Dynamin impacts homology-directed repair and breast cancer response to chemotherapy', *The Journal of Clinical Investigation*, 128(12), pp. 5307–5321. doi:10.1172/JCI87191.

Chesi, M. *et al.* (1997) 'Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3', *Nature Genetics*, 16(3), pp. 260–264. doi:10.1038/ng0797-260.

Chioni, A.-M. and Grose, R. (2012) 'FGFR1 cleavage and nuclear translocation regulates breast cancer cell behavior', *The Journal of Cell Biology*, 197(6), pp. 801–817. doi:10.1083/jcb.201108077.

Chiu, V.K. *et al.* (2002) 'Ras signalling on the endoplasmic reticulum and the Golgi', *Nature Cell Biology*, 4(5), pp. 343–350. doi:10.1038/ncb783.

Cho, H. *et al.* (2007) 'Statistical identification of differentially labeled peptides from liquid chromatography tandem mass spectrometry', *PROTEOMICS*, 7(20), pp. 3681–3692. doi:10.1002/pmic.200601034.

Choi, H. *et al.* (2011) 'SAINT: Probabilistic scoring of affinity purificationg-mass spectrometry data', *Nature Methods*, 8(1), pp. 70–73. doi:10.1038/nmeth.1541.

Chou, M.F. and Schwartz, D. (2011) 'Biological Sequence Motif Discovery Using motif-x', *Current Protocols in Bioinformatics*, 35(1), p. 13.15.1-13.15.24. doi:10.1002/0471250953.bi1315s35.

Choudhary, C. and Mann, M. (2010) 'Decoding signalling networks by mass spectrometry-based proteomics', *Nature Reviews Molecular Cell Biology*, 11(6), pp. 427–439. doi:10.1038/nrm2900.

Christ, L. *et al.* (2017) 'Cellular Functions and Molecular Mechanisms of the ESCRT Membrane-Scission Machinery', *Trends in Biochemical Sciences*, 42(1), pp. 42–56. doi:10.1016/j.tibs.2016.08.016.

Chung, I. *et al.* (2010) 'Spatial control of EGF receptor activation by reversible dimerization on living cells', *Nature*, 464(7289), pp. 783–787. doi:10.1038/nature08827.

Cohen, S. (1962) 'Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal', *Journal of Biological Chemistry*, 237(5), pp. 1555–1562. doi:10.1016/S0021-9258(19)83739-0.

Colella, B. *et al.* (2020) 'mTOR Inhibition Leads to Src-Mediated EGFR Internalisation and Degradation in Glioma Cells', *Cancers*, 12(8). doi:10.3390/cancers12082266.

Collins, M.O. *et al.* (2008) 'Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder.', *Molecular & cellular proteomics : MCP*, 7(7), pp. 1331–48. doi:10.1074/mcp.M700564-MCP200.

Collins, T.N. *et al.* (2018) 'Crk proteins transduce FGF signaling to promote lens fiber cell elongation', *eLife*, 7, p. e32586. doi:10.7554/eLife.32586.

Colvin, J.S. *et al.* (2001) 'Lung hypoplasia and neonatal death in Fgf9-null mice identify this gene as an essential regulator of lung mesenchyme', *Development (Cambridge, England)*, 128(11), pp. 2095–2106.

Compeer, E.B. *et al.* (2018) 'A mobile endocytic network connects clathrin-independent receptor endocytosis to recycling and promotes T cell activation', *Nature Communications*, 9(1), p. 1597. doi:10.1038/s41467-018-04088-w.

Corbit, K.C. *et al.* (2000) 'Different protein kinase C isoforms determine growth factor specificity in neuronal cells', *Molecular and Cellular Biology*, 20(15), pp. 5392–5403. doi:10.1128/mcb.20.15.5392-5403.2000.

Corbit, K.C. *et al.* (2003) 'Activation of Raf-1 Signaling by Protein Kinase C through a Mechanism Involving Raf Kinase Inhibitory Protein *', *Journal of Biological Chemistry*, 278(15), pp. 13061–13068. doi:10.1074/jbc.M210015200.

Corvera, S., D'Arrigo, A. and Stenmark, H. (1999) 'Phosphoinositides in membrane traffic', *Current Opinion in Cell Biology*, 11(4), pp. 460–465. doi:10.1016/S0955-0674(99)80066-0.

Courcelles, M. *et al.* (2013) 'Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions.', *Molecular systems biology*, 9, p. 669. doi:10.1038/msb.2013.25.

Cox, J. *et al.* (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. *et al.* (2014) 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ', *Molecular and Cellular Proteomics*, 13(9), pp. 2513–2526. doi:10.1074/mcp.M113.031591.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372. doi:10.1038/nbt.1511.

Cox, J. and Mann, M. (2012) '1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data.', 13(Suppl 16), p. S12. doi:10.1186/1471-2105-13-S16-S12.

Cross, M.J. *et al.* (2000) 'Tyrosine 766 in the fibroblast growth factor receptor-1 is required for FGF-stimulation of phospholipase C, phospholipase D, phospholipase A(2), phosphoinositide 3-kinase and cytoskeletal reorganisation in porcine aortic endothelial cells', *Journal of Cell Science*, 113 (Pt 4), pp. 643–651.

Crudden, C. *et al.* (2019) 'Below the Surface: IGF-1R Therapeutic Targeting and Its Endocytic Journey', *Cells*, 8(10). doi:10.3390/cells8101223.

Csanaky, K., Hess, M.W. and Klimaschewski, L. (2019) 'Membrane-Associated, Not Cytoplasmic or Nuclear, FGFR1 Induces Neuronal Differentiation', *Cells*, 8(3). doi:10.3390/cells8030243.

Csermely, P. *et al.* (2012) 'Structure and dynamics of molecular networks: A novel paradigm of drug discovery. A comprehensive review', *Pharmacology and Therapeutics*, 138(3), pp. 333–408. doi:10.1016/j.pharmthera.2013.01.016.

Cullen, P.J. and Steinberg, F. (2018) 'To degrade or not to degrade: mechanisms and significance of endocytic recycling', *Nature Reviews Molecular Cell Biology*, 19(11), pp. 679–696. doi:10.1038/s41580-018-0053-7.

Damian, D. and Gorfine, M. (2004) 'Statistical concerns about the GSEA procedure', *Nature Genetics*, 36(7), pp. 663–663. doi:10.1038/ng0704-663a.

D'Angelo, G. *et al.* (2017) 'Statistical Models for the Analysis of Isobaric Tags Multiplexed Quantitative Proteomics', *Journal of Proteome Research*, 16(9), pp. 3124–3136. doi:10.1021/acs.jproteome.6b01050.

Dannhauser, P.N. and Ungewickell, E.J. (2012) 'Reconstitution of clathrin-coated bud and vesicle formation with minimal components', *Nature Cell Biology*, 14(6), pp. 634–639. doi:10.1038/ncb2478.

Dauner, K. *et al.* (2017) 'mTOR complex 1 activity is required to maintain the canonical endocytic recycling pathway against lysosomal delivery', *Journal of Biological Chemistry*, 292(14), pp. 5737–5747. doi:10.1074/jbc.M116.771451.

David, C.C. and Jacobs, D.J. (2014) 'Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins', *Methods in molecular biology (Clifton, N.J.)*, 1084, pp. 193–226. doi:10.1007/978-1-62703-658-0_11.

Day, E.K., Sosale, N.G. and Lazzara, M.J. (2016) 'Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process', *Current Opinion in Biotechnology*, 40, pp. 185–192. doi:10.1016/j.copbio.2016.06.005.

De Martino, A. and De Martino, D. (2018) 'An introduction to the maximum entropy approach and its application to inference problems in biology', *Heliyon*, 4(4), p. e00596. doi:10.1016/j.heliyon.2018.e00596.

Deak, M. *et al.* (1998) 'Mitogen- and stress-activated protein kinase-1 (MSK1) is directly activated by MAPK and SAPK2/p38, and may mediate activation of CREB', *The EMBO Journal*, 17(15), pp. 4426–4441. doi:10.1093/emboj/17.15.4426.

Decker, S.J. (1990) 'Epidermal growth factor and transforming growth factor-alpha induce differential processing of the epidermal growth factor receptor', *Biochemical and Biophysical Research Communications*, 166(2), pp. 615–621. doi:10.1016/0006-291x(90)90853-f.

Del Olmo, T. *et al.* (2019) 'APEX2-mediated RAB proximity labeling identifies a role for RAB21 in clathrin-independent cargo sorting', *EMBO reports*, 20(2). doi:10.15252/embr.201847192.

Demetriades, C., Plescher, M. and Teleman, A.A. (2016) 'Lysosomal recruitment of TSC2 is a universal response to cellular stress', *Nature Communications*, 7(1), p. 10662. doi:10.1038/ncomms10662.

Dempsey, P.J., Garton, K. and Raines, E.W. (2002) 'Emerging roles of TACE as a key protease in ErbB ligand shedding', *Molecular Interventions*, 2(3), pp. 136–141. doi:10.1124/mi.2.3.136.

Deng, H. *et al.* (2011) 'Phosphorylation of Bcl-associated death protein (Bad) by erythropoietin-activated c-Jun N-terminal protein kinase 1 contributes to survival of erythropoietin-dependent cells', *The International Journal of Biochemistry & Cell Biology*, 43(3), pp. 409–415. doi:10.1016/j.biocel.2010.11.011.

Dérijard, B. *et al.* (1995) 'Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms', *Science (New York, N.Y.)*, 267(5198), pp. 682–685. doi:10.1126/science.7839144.

Dibble, C.C. *et al.* (2012) 'TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1', *Molecular Cell*, 47(4), pp. 535–546. doi:10.1016/j.molcel.2012.06.009.

Dickhut, C. *et al.* (2014) 'Impact of Digestion Conditions on Phosphoproteomics', *Journal of Proteome Research*, 13(6), pp. 2761–2770. doi:10.1021/pr401181y.

Dickinson, R.J. *et al.* (2002) 'Expression of the ERK-specific MAP kinase phosphatase PYST1/MKP3 in mouse embryos during morphogenesis and early organogenesis', *Mechanisms of Development*, 113(2), pp. 193–196. doi:10.1016/S0925-4773(02)00024-2.

Diwanji, D., Thaker, T. and Jura, N. (2019) 'More than the Sum of the Parts: Towards Full-Length Receptor Tyrosine Kinase Structures', *IUBMB life*, 71(6), pp. 706–720. doi:10.1002/iub.2060.

Doll, S., Gnad, F. and Mann, M. (2019) 'The Case for Proteomics and Phospho-Proteomics in Personalized Cancer Medicine', *PROTEOMICS – Clinical Applications*, 13(2), p. 1800113. doi:10.1002/prca.201800113.

Domanova, W. *et al.* (2016) 'Unraveling Kinase Activation Dynamics Using Kinase-Substrate Relationships from Temporal Large-Scale Phosphoproteomics Studies', *PLoS ONE*, 11(6). doi:10.1371/journal.pone.0157763.

Donatis, A.D. *et al.* (2008) 'Proliferation Versus Migration in Platelet-derived Growth Factor Signaling THE KEY ROLE OF ENDOCYTOSIS', *Journal of Biological Chemistry*, 283(29), pp. 19948–19956. doi:10.1074/jbc.M709428200.

Dong, J.-M. *et al.* (2016) 'Proximity biotinylation provides insight into the molecular composition of focal adhesions at the nanometer scale', *Science Signaling*, 9(432), p. rs4. doi:10.1126/scisignal.aaf3572.

Dorey, K. and Amaya, E. (2010) 'FGF signalling: diverse roles during early vertebrate embryogenesis.', *Development (Cambridge, England)*, 137(22), pp. 3731–42. doi:10.1242/dev.037689.

Dow, L.E. *et al.* (2008) 'Loss of human Scribble cooperates with H-Ras to promote cell invasion through deregulation of MAPK signalling', *Oncogene*, 27(46), pp. 5988–6001. doi:10.1038/onc.2008.219.

Dreisewerd, K. (2014) 'Recent methodological advances in MALDI mass spectrometry', *Analytical and Bioanalytical Chemistry*, 406(9–10), pp. 2261–2278. doi:10.1007/s00216-014-7646-6.

Dufour, C. *et al.* (2008) 'FGFR2-Cbl interaction in lipid rafts triggers attenuation of PI3K/Akt signaling and osteoblast survival', *Bone*, 42(6), pp. 1032–1039. doi:10.1016/j.bone.2008.02.009.

Dugourd, A. and Saez-Rodriguez, J. (2019) 'Footprint-based functional analysis of multiomic data', *Current Opinion in Systems Biology*, 15, pp. 82–90. doi:10.1016/j.coisb.2019.04.002.

Dunham, W.H., Mullin, M. and Gingras, A.-C. (2012) 'Affinity-purification coupled to mass spectrometry: Basic principles and strategies', *PROTEOMICS*, 12(10), pp. 1576–1590. doi:10.1002/pmic.201100523.

Dunn, J.D., Reid, G.E. and Bruening, M.L. (2009) 'Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry', *Mass Spectrometry Reviews*, 29(1), p. n/a-n/a. doi:10.1002/mas.20219.

Düvel, K. *et al.* (2010) 'Activation of a metabolic gene regulatory network downstream of mTOR complex 1', *Molecular Cell*, 39(2), pp. 171–183. doi:10.1016/j.molcel.2010.06.022.

Earley, L. *et al.* (2013) 'Front-End Electron Transfer Dissociation: A New Ionization Source', *Analytical Chemistry*, 85(17), pp. 8385–8390. doi:10.1021/ac401783f.

Eblaghie, M.C. *et al.* (2003) 'Negative Feedback Regulation of FGF Signaling Levels by Pyst1/MKP3 in Chick Embryos', *Current Biology*, 13(12), pp. 1009–1018. doi:10.1016/S0960-9822(03)00381-6.

Ebner, R. and Derynck, R. (1991) 'Epidermal growth factor and transforming growth factor-alpha: differential intracellular routing and processing of ligand-receptor complexes', *Cell Regulation*, 2(8), pp. 599–612. doi:10.1091/mbc.2.8.599.

Eden, E.R. *et al.* (2010) 'Membrane contacts between endosomes and ER provide sites for PTP1B–epidermal growth factor receptor interaction', *Nature Cell Biology*, 12(3), pp. 267–272. doi:10.1038/ncb2026.

Ekerot, M. *et al.* (2008) 'Negative-feedback regulation of FGF signalling by DUSP6/MKP-3 is driven by ERK1/2 and mediated by Ets factor binding to a conserved site within the DUSP6/MKP-3 gene promoter', *Biochemical Journal*, 412(2), pp. 287–298. doi:10.1042/BJ20071512.

Eldar, A. and Elowitz, M.B. (2010) 'Functional roles for noise in genetic circuits', *Nature*, 467(7312), pp. 167–173. doi:10.1038/nature09326.

Elias, J.E. and Gygi, S.P. (2007) 'Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry', *Nature Methods*, 4(3), pp. 207–214. doi:10.1038/nmeth1019.

Ellison, S.L.R., Barwick, V.J. and Farrant, T.J.D. (2009) *Practical Statistics for the Analytical Scientist: A Bench Guide*. Royal Society of Chemistry.

Eng, J.K. *et al.* (2011) 'A Face in the Crowd: Recognizing Peptides Through Database Search', *Molecular & Cellular Proteomics : MCP*, 10(11). doi:10.1074/mcp.R111.009522.

Enslen, H., Raingeaud, J. and Davis, R.J. (1998) 'Selective Activation of p38 Mitogen-activated Protein (MAP) Kinase Isoforms by the MAP Kinase Kinases MKK3 and MKK6 *', *Journal of Biological Chemistry*, 273(3), pp. 1741–1748. doi:10.1074/jbc.273.3.1741.

Entwisle, S.W. *et al.* (2020) 'Proteome and Phosphoproteome Analysis of Brown Adipocytes Reveals That RICTOR Loss Dampens Global Insulin/AKT Signaling *', *Molecular & Cellular Proteomics*, 19(7), pp. 1104–1119. doi:10.1074/mcp.RA120.001946.

Esteban-Villarrubia, J. *et al.* (2020) 'Tyrosine Kinase Receptors in Oncology', *International Journal of Molecular Sciences*, 21(22), p. 8529. doi:10.3390/ijms21228529.

Ettenberg, S.A. *et al.* (1999) 'cbl-b inhibits epidermal growth factor receptor signaling', *Oncogene*, 18(10), pp. 1855–1866. doi:10.1038/sj.onc.1202499.

Everitt, B.S. *et al.* (2011) *Cluster Analysis*. Chichester, UK: John Wiley & Sons, Ltd (Wiley Series in Probability and Statistics). doi:10.1002/9780470977811.

Fabregat, A. *et al.* (2018) 'The Reactome Pathway Knowledgebase', *Nucleic Acids Research*, 46(D1), pp. D649–D655. doi:10.1093/nar/gkx1132.

Faust, D. *et al.* (2012) 'Differential p38-dependent signalling in response to cellular stress and mitogenic stimulation in fibroblasts', *Cell Communication and Signaling*, 10(1), pp. 1–13. doi:10.1186/1478-811X-10-6.

Favelyukis, S. *et al.* (2001) 'Structure and autoregulation of the insulin-like growth factor 1 receptor kinase', *Nature Structural Biology*, 8(12), pp. 1058–1063. doi:10.1038/nsb721.

Ferguson, H.R., Smith, M.P. and Francavilla, C. (2021) 'Fibroblast Growth Factor Receptors (FGFRs) and Noncanonical Partners in Cancer Signaling', *Cells*, 10(5), p. 1201. doi:10.3390/cells10051201.

Fernández-Costa, C. *et al.* (2020) 'Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results', *Journal of Proteome Research*, 19(8), pp. 3153–3161. doi:10.1021/acs.jproteome.0c00153.

Fernández-Nogueira, P. *et al.* (2020) 'Tumor-Associated Fibroblasts Promote HER2-Targeted Therapy Resistance through FGFR2 Activation', *Clinical Cancer Research*, 26(6), pp. 1432–1448. doi:10.1158/1078-0432.CCR-19-0353.

Fey, D. *et al.* (2012) 'Crosstalk and Signaling Switches in Mitogen-Activated Protein Kinase Cascades', *Frontiers in Physiology*, 3. doi:10.3389/fphys.2012.00355.

Fischer, C. *et al.* (2019) 'Signals trigger state-specific transcriptional programs to support diversity and homeostasis in immune cells', *Science Signaling*, 12(581), p. eaao5820. doi:10.1126/scisignal.aao5820.

Fisher, T.L. and Blenis, J. (1996) 'Evidence for two catalytically active kinase domains in pp90rsk', *Molecular and Cellular Biology*, 16(3), pp. 1212–1219. doi:10.1128/mcb.16.3.1212.

Fong, C.W. *et al.* (2003) 'Tyrosine Phosphorylation of Sprouty2 Enhances Its Interaction with c-Cbl and Is Crucial for Its Function *', *Journal of Biological Chemistry*, 278(35), pp. 33456–33464. doi:10.1074/jbc.M301317200.

Formstecher, E. *et al.* (2001) 'PEA-15 Mediates Cytoplasmic Sequestration of ERK MAP Kinase', *Developmental Cell*, 1(2), pp. 239–250. doi:10.1016/S1534-5807(01)00035-1.

Fortian, A. and Sorkin, A. (2014) 'Live-cell fluorescence imaging reveals high stoichiometry of Grb2 binding to the EGF receptor sustained during endocytosis', *Journal of Cell Science*, 127(2), pp. 432–444. doi:10.1242/jcs.137786.

Francavilla, C. *et al.* (2009) 'The binding of NCAM to FGFR1 induces a specific cellular response mediated by receptor trafficking.', *The Journal of cell biology*, 187(7), pp. 1101–16. doi:10.1083/jcb.200903030.

Francavilla, C. *et al.* (2013) 'Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs.', *Molecular cell*, 51(6), pp. 707–22. doi:10.1016/j.molcel.2013.08.002.

Francavilla, C. *et al.* (2016) 'Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking', *Nature Structural & Molecular Biology*, 23(6), pp. 608–618. doi:10.1038/nsmb.3218.

Franco, I. *et al.* (2014) 'PI3K Class II α Controls Spatially Restricted Endosomal PtdIns3P and Rab11 Activation to Promote Primary Cilium Function', *Developmental Cell*, 28(6), pp. 647–658. doi:10.1016/j.devcel.2014.01.022.

Fraser, J. *et al.* (2017) 'Interplay of autophagy, receptor tyrosine kinase signalling and endocytic trafficking', *Essays in Biochemistry*, 61(6), pp. 597–607. doi:10.1042/EBC20170091.

Fraser, J. *et al.* (2019) 'Targeting of early endosomes by autophagy facilitates EGFR recycling and signalling', *EMBO reports*, 20(10), p. e47734. doi:10.15252/embr.201947734.

Freed, D.M. *et al.* (2017) 'EGFR ligands differentially stabilize receptor dimers to specify signaling kinetics', *Cell*, 171(3), pp. 683-695.e18. doi:10.1016/j.cell.2017.09.017.

Freed, D.M., Alvarado, D. and Lemmon, M.A. (2015) 'Ligand regulation of a constitutively dimeric EGF receptor', *Nature Communications*, 6. doi:10.1038/ncomms8380.

Frémin, C. *et al.* (2015) 'Functional Redundancy of ERK1 and ERK2 MAP Kinases during Development', *Cell Reports*, 12(6), pp. 913–921. doi:10.1016/j.celrep.2015.07.011.

French, A.R. *et al.* (1995) 'Intracellular trafficking of epidermal growth factor family ligands is directly influenced by the pH sensitivity of the receptor/ligand interaction', *The Journal of Biological Chemistry*, 270(9), pp. 4334–4340. doi:10.1074/jbc.270.9.4334.

Frieden, B.R. and Gatenby, R.A. (2011) 'Information dynamics in living systems: prokaryotes, eukaryotes, and cancer', *PloS One*, 6(7), p. e22085. doi:10.1371/journal.pone.0022085.

Fruman, D.A. *et al.* (2017) 'The PI3K Pathway in Human Disease', *Cell*, 170(4), pp. 605–635. doi:10.1016/j.cell.2017.07.029.

Fu, Y. and Qian, X. (2014) 'Transferred Subgroup False Discovery Rate for Rare Post-translational Modifications Detected by Mass Spectrometry', *Molecular & Cellular Proteomics : MCP*, 13(5), pp. 1359–1368. doi:10.1074/mcp.O113.030189.

Fukazawa, T. *et al.* (1996) 'Tyrosine Phosphorylation of Cbl upon Epidermal Growth Factor (EGF) Stimulation and Its Association with EGF Receptor and Downstream Signaling Proteins *', *Journal of Biological Chemistry*, 271(24), pp. 14554–14559. doi:10.1074/jbc.271.24.14554.

Furcht, C.M., Buonato, J.M. and Lazzara, M.J. (2015) 'EGFR-activated Src family kinases maintain GAB1-SHP2 complexes distal from EGFR', *Science Signaling*, 8(376), pp. ra46–ra46. doi:10.1126/scisignal.2005697.

Furdui, C.M. *et al.* (2006) 'Autophosphorylation of FGFR1 Kinase Is Mediated by a Sequential and Precisely Ordered Reaction', *Molecular Cell*, 21(5), pp. 711–717. doi:10.1016/j.molcel.2006.01.022.

Fürthauer, M. *et al.* (2002) 'Sef is a feedback-induced antagonist of Ras/MAPK-mediated FGF signalling', *Nature Cell Biology*, 4(2), pp. 170–174. doi:10.1038/ncb750.

Futschik, M.E. and Carlisle, B. (2005) 'Noise-robust soft clustering of gene expression time-course data', *Journal of Bioinformatics and Computational Biology*, 3(4), pp. 965–988. doi:10.1142/S0219720005001375.

le Gallic, L. *et al.* (1999) 'Transcriptional Repressor ERF Is a Ras/Mitogen-Activated Protein Kinase Target That Regulates Cellular Proliferation', *Molecular and Cellular Biology*, 19(6), pp. 4121–4133. doi:10.1128/MCB.19.6.4121.

Gan, H.K., Cvrljevic, A.N. and Johns, T.G. (2013) 'The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered', *The FEBS journal*, 280(21), pp. 5350–5370. doi:10.1111/febs.12393.

Gan, X. *et al.* (2012) 'PRR5L degradation promotes mTORC2-mediated PKC-δ phosphorylation and cell migration downstream of Gα12', *Nature Cell Biology*, 14(7), pp. 686–696. doi:10.1038/ncb2507.

Garcia-Closas, M. *et al.* (2008) 'Heterogeneity of Breast Cancer Associations with Five Susceptibility Loci by Clinical and Pathological Characteristics', *PLoS Genetics*, 4(4). doi:10.1371/journal.pgen.1000054.

Garcia-Maya, M. *et al.* (2006) 'Ligand concentration is a driver of divergent signaling and pleiotropic cellular responses to FGF', *Journal of Cellular Physiology*, 206(2), pp. 386–393. doi:10.1002/jcp.20483.

Garrido, A., Brandt, M. and Djouder, N. (2016) 'Transport to Rhebpress activity', *Small GTPases*, 7(1), pp. 12–15. doi:10.1080/21541248.2015.1120846.

Gatenby, R.A. (2019) 'The Role of Cell Membrane Information Reception, Processing, and Communication in the Structure and Function of Multicellular Tissue', *International Journal of Molecular Sciences*, 20(15), p. 3609. doi:10.3390/ijms20153609.

Gatius, S. *et al.* (2011) 'FGFR2 alterations in endometrial carcinoma', *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 24(11), pp. 1500–1510. doi:10.1038/modpathol.2011.110.

Gaubitz, C. *et al.* (2016) 'TORC2 Structure and Function', *Trends in Biochemical Sciences*, 41(6), pp. 532–545. doi:10.1016/j.tibs.2016.04.001.

Gaudet, P. and Dessimoz, C. (2017) 'Gene Ontology: Pitfalls, Biases, and Remedies', in Dessimoz, C. and Škunca, N. (eds) *The Gene Ontology Handbook*. New York, NY: Springer New York, pp. 189–205. doi:10.1007/978-1-4939-3743-1_14.

Gaullier, J.-M. *et al.* (1998) 'FYVE fingers bind PtdIns(3)P', *Nature*, 394(6692), pp. 432–433. doi:10.1038/28767.

Geiger, T. *et al.* (2010) 'Super-SILAC mix for quantitative proteomics of human tumor tissue', *Nature Methods*, 7(5), pp. 383–385. doi:10.1038/nmeth.1446.

Geiger, T. *et al.* (2011) 'Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics', *Nature Protocols*, 6(2), p. 147. doi:10.1038/nprot.2010.192.

Geistlinger, L. *et al.* (2021) 'Toward a gold standard for benchmarking gene set enrichment analysis', *Briefings in Bioinformatics*, 22(1), pp. 545–556. doi:10.1093/bib/bbz158.

The Gene Ontology Consortium (2018) 'The Gene Ontology Resource: 20 years and still GOing strong', *Nucleic Acids Research*, 47(D1), pp. D330–D338. doi:10.1093/nar/gky1055.

Gillet, L.C. *et al.* (2012) 'Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis', *Molecular & Cellular Proteomics : MCP*, 11(6). doi:10.1074/mcp.O111.016717.

Gillis, J., Ballouz, S. and Pavlidis, P. (2014) 'Bias tradeoffs in the creation and analysis of protein–protein interaction networks', *Journal of Proteomics*, 100, pp. 44–54. doi:10.1016/j.jprot.2014.01.020.

Gillis, J. and Pavlidis, P. (2012) '"Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks', *PLoS Computational Biology*. Edited by A. Rzhetsky, 8(3), p. e1002444. doi:10.1371/journal.pcbi.1002444.

Gingras, A.-C., Abe, K.T. and Raught, B. (2019) 'Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles', *Current Opinion in Chemical Biology*, 48, pp. 44–54. doi:10.1016/j.cbpa.2018.10.017.

Giulianelli, S. *et al.* (2008) 'Carcinoma-associated fibroblasts activate progesterone receptors and induce hormone independent mammary tumor growth: A role for the FGF-2/FGFR-2 axis', *International Journal of Cancer*, 123(11), pp. 2518–2531. doi:10.1002/ijc.23802.

Gjerga, E. *et al.* (2020) 'Converting networks to predictive logic models from perturbation signalling data with CellNOpt', *Bioinformatics*, 36(16), pp. 4523–4524. doi:10.1093/bioinformatics/btaa561.

Gjerga, E. *et al.* (2021) 'PHONEMeS: Efficient Modeling of Signaling Networks Derived from Large-Scale Mass Spectrometry Data', *Journal of Proteome Research* [Preprint]. doi:10.1021/acs.jproteome.0c00958.

Go, C.D. *et al.* (2021) 'A proximity-dependent biotinylation map of a human cell', *Nature*, pp. 1–5. doi:10.1038/s41586-021-03592-2.

Godovac-Zimmermann, J. and Brown, L.R. (2001) 'Perspectives for mass spectrometry and functional proteomics', *Mass Spectrometry Reviews*, 20(1), pp. 1–57. doi:10.1002/1098-2787(2001)20:1<1::AID-MAS1001>3.0.CO;2-J.

Goh, L.K. and Sorkin, A. (2013) 'Endocytosis of Receptor Tyrosine Kinases', *Cold Spring Harbor Perspectives in Biology*, 5(5), p. a017459. doi:10.1101/cshperspect.a017459.

Gonzalez, F.A., Raden, D.L. and Davis, R.J. (1991) 'Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases.', *Journal of Biological Chemistry*, 266(33), pp. 22159–22163. doi:10.1016/S0021-9258(18)54548-8.

Gordon, A.D. (1987) 'A Review of Hierarchical Classification', *Journal of the Royal Statistical Society: Series A (General)*, 150(2), pp. 119–137. doi:10.2307/2981629.

Gotoh, N. (2008) 'Regulation of growth factor signaling by FRS2 family docking/scaffold adaptor proteins', *Cancer Science*, 99(7), pp. 1319–1325. doi:10.1111/j.1349-7006.2008.00840.x.

Grandal, M.V. *et al.* (2007) 'EGFRvIII escapes down-regulation due to impaired internalization and sorting to lysosomes', *Carcinogenesis*, 28(7), pp. 1408–1417. doi:10.1093/carcin/bgm058.

Gray, V.E. and Kumar, S. (2011) 'Rampant purifying selection conserves positions with posttranslational modifications in human proteins', *Molecular Biology and Evolution*, 28(5), pp. 1565–1568. doi:10.1093/molbev/msr013.

Greenman, C. *et al.* (2007) 'Patterns of somatic mutation in human cancer genomes', *Nature*, 446(7132), pp. 153–158. doi:10.1038/nature05610.

Griffiths, G. (1989) 'The structure and function of a mannose 6-phosphate receptor-enriched, pre-lysosomal compartment in animal cells', *Journal of Cell Science. Supplement*, 11, pp. 139–147. doi:10.1242/jcs.1989.supplement_11.11.

Gross, S.M. and Rotwein, P. (2016) 'Mapping growth-factor-modulated Akt signaling dynamics', *J Cell Sci*, 129(10), pp. 2052–2063. doi:10.1242/JCS.183764.

Grossmann, S. *et al.* (2007) 'Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis', *Bioinformatics*, 23(22), pp. 3024–3031. doi:10.1093/bioinformatics/btm440.

Gu, H. and Neel, B.G. (2003) 'The "Gab" in signal transduction', *Trends in Cell Biology*, 13(3), pp. 122–130. doi:10.1016/s0962-8924(03)00002-3.

Guimond, S. *et al.* (1993) 'Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4', *The Journal of Biological Chemistry*, 268(32), pp. 23906–23914.

Guo, L., Degenstein, L. and Fuchs, E. (1996) 'Keratinocyte growth factor is required for hair development but not for wound healing.', *Genes & Development*, 10(2), pp. 165–175. doi:10.1101/gad.10.2.165.

Guo, M. *et al.* (2012) 'FGFR2 isoforms support epithelial-stromal interactions in thyroid cancer progression', *Cancer Research*, 72(8), pp. 2017–2027. doi:10.1158/0008-5472.CAN-11-3985.

Guthals, A. and Bandeira, N. (2012) 'Peptide Identification by Tandem Mass Spectrometry with Alternate Fragmentation Modes', *Molecular & Cellular Proteomics : MCP*, 11(9), pp. 550–557. doi:10.1074/mcp.R112.018556.

Guy, G.R. *et al.* (2009) 'Sprouty proteins: modified modulators, matchmakers or missing links?', *Journal of Endocrinology*, 203(2), pp. 191–202. doi:10.1677/JOE-09-0110.

Gygi, S.P. *et al.* (1999) 'Correlation between Protein and mRNA Abundance in Yeast', *Molecular and Cellular Biology*, 19(3), pp. 1720–1730. doi:10.1128/mcb.19.3.1720.

H. B. Mann and D. R. Whitney (1947) 'On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other', *The Annals of Mathematical Statistics*, 18(1), pp. 50–60. doi:10.1214/aoms/1177730491.

Haag, A.M. (2016) 'Mass Analyzers and Mass Spectrometers', in Mirzaei, H. and Carrasco, M. (eds) *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*. Cham: Springer International Publishing (Advances in Experimental Medicine and Biology), pp. 157–169. doi:10.1007/978-3-319-41448-5_7.

Han, H. *et al.* (2018) 'TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions', *Nucleic Acids Research*, 46(D1), pp. D380–D386. doi:10.1093/nar/gkx1013.

Han, S. *et al.* (2017) 'Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells', *Cell Chemical Biology*, 24(3), pp. 404–414. doi:10.1016/j.chembiol.2017.02.002.

Han, Y. *et al.* (2019) 'Directed Evolution of Split APEX2 Peroxidase', *ACS chemical biology*, 14(4), pp. 619–635. doi:10.1021/acschembio.8b00919.

Hanafusa, H. *et al.* (2002) 'Sprouty1 and Sprouty2 provide a control mechanism for the Ras/MAPK signalling pathway', *Nature Cell Biology*, 4(11), pp. 850–858. doi:10.1038/ncb867.

Handler, D. *et al.* (2018) 'The Art of validating quantitative proteomics data', *PROTEOMICS*, p. 1800222. doi:10.1002/pmic.201800222.

Hara, K. *et al.* (2002) 'Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action', *Cell*, 110(2), pp. 177–189. doi:10.1016/s0092-8674(02)00833-4.

Harris, R.C., Chung, E. and Coffey, R.J. (2003) 'EGF receptor ligands', *Experimental Cell Research*, 284(1), pp. 2–13. doi:https://doi.org/10.1016/S0014-4827(02)00105-2.

Harris, T.E. *et al.* (2006) 'mTOR-dependent stimulation of the association of eIF4G and eIF3 by insulin', *The EMBO Journal*, 25(8), pp. 1659–1668. doi:10.1038/sj.emboj.7601047.

Hartigan, J.A. and Wong, M.A. (1979) 'Algorithm AS 136: A K-Means Clustering Algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), pp. 100–108. doi:10.2307/2346830.

Hartwell, L.H. *et al.* (1999) 'From molecular to modular cell biology', *Nature*, 402(6761supp), pp. C47–C52. doi:10.1038/35011540.

Haugsten, E.M. *et al.* (2005) 'Different intracellular trafficking of FGF1 endocytosed by the four homologous FGF receptors.', *Journal of cell science*, 118(Pt 17), pp. 3869–81. doi:10.1242/jcs.02509.

Haugsten, E.M. *et al.* (2011) 'Clathrin- and Dynamin-Independent Endocytosis of FGFR3 – Implications for Signalling', *PLoS ONE*, 6(7). doi:10.1371/journal.pone.0021708.

Haugsten, E.M. *et al.* (2014) 'Photoactivation Approaches Reveal a Role for Rab11 in FGFR4 Recycling and Signalling', *Traffic*, 15(6), pp. 665–683. doi:10.1111/tra.12168.

Haugsten, E.M. *et al.* (2016) 'Proximity Labeling Reveals Molecular Determinants of FGFR4 Endosomal Transport', *Journal of Proteome Research*, 15(10), pp. 3841–3855. doi:10.1021/acs.jproteome.6b00652.

Haynes, W.A., Tomczak, A. and Khatri, P. (2018) 'Gene annotation bias impedes biomedical research', *Scientific Reports*, 8(1), p. 1362. doi:10.1038/s41598-018-19333-x.

Hehnly, H. *et al.* (2012) 'The centrosome regulates the Rab11- dependent recycling endosome pathway at appendages of the mother centriole', *Current biology: CB*, 22(20), pp. 1944–1950. doi:10.1016/j.cub.2012.08.022.

Hemalatha, A. and Mayor, S. (2019) 'Recent advances in clathrin-independent endocytosis', *F1000Research*, 8, p. 138. doi:10.12688/f1000research.16549.1.

Hermida, M.A., Dinesh Kumar, J. and Leslie, N.R. (2017) 'GSK3 and its interactions with the PI3K/AKT/mTOR signalling network', *Advances in Biological Regulation*, 65, pp. 5–15. doi:10.1016/j.jbior.2017.06.003.

Herrero, A. *et al.* (2015) 'Small Molecule Inhibition of ERK Dimerization Prevents Tumorigenesis by RAS-ERK Pathway Oncogenes', *Cancer Cell*, 28(2), pp. 170–182. doi:10.1016/j.ccell.2015.07.001.

Higuchi, M. *et al.* (2008) 'Scaffolding function of PAK in the PDK1–Akt pathway', *Nature Cell Biology*, 10(11), pp. 1356–1364. doi:10.1038/ncb1795.

Hiramatsu, M. *et al.* (1988) 'Influence of age on epidermal growth factor receptor level in the rat brain', *Experientia*, 44(1), pp. 23–25. doi:10.1007/BF01960230.

Hitchinson, B. *et al.* (2018) 'Biased antagonism of CXCR4 avoids antagonist tolerance.', *Science signaling*, 11(552), p. eaat2214. doi:10.1126/scisignal.aat2214.

Hogrebe, A. *et al.* (2018) 'Benchmarking common quantification strategies for large-scale phosphoproteomics', *Nature Communications*, 9(1), p. 1045. doi:10.1038/s41467-018-03309-6.

Hollenhorst, P.C. *et al.* (2011) 'Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells', *Genes & Development*, 25(20), pp. 2147–2157. doi:10.1101/gad.17546311.

Hollenhorst, P.C., McIntosh, L.P. and Graves, B.J. (2011) 'Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors', *Annual Review of Biochemistry*, 80(1), pp. 437–471. doi:10.1146/annurev.biochem.79.081507.103945.

Hopkins, C.R., Miller, K. and Beardmore, J.M. (1985) 'Receptor-Mediated Endocytosis of Transferrin and Epidermal Growth Factor Receptors: A Comparison of Constitutive and Ligand-Induced Uptake', *Journal of Cell Science*, 1985(Supplement 3), pp. 173–186. doi:10.1242/jcs.1985.Supplement_3.17.

Horgan, C.P. *et al.* (2010) 'Rab11-FIP3 links the Rab11 GTPase and cytoplasmic dynein to mediate transport to the endosomal-recycling compartment', *Journal of Cell Science*, 123(2), pp. 181–191. doi:10.1242/jcs.052670.

Horn, H. *et al.* (2014) 'KinomeXplorer: an integrated platform for kinome biology studies', *Nature Methods*, 11(6), pp. 603–604. doi:10.1038/nmeth.2968.

Hornbeck, P.V. *et al.* (2015) 'PhosphoSitePlus, 2014: mutations, PTMs and recalibrations', *Nucleic Acids Research*, 43(D1), pp. D512–D520. doi:10.1093/nar/gku1267.

Hosokawa, N. *et al.* (2009) 'Nutrient-dependent mTORC1 association with the ULK1-Atg13-FIP200 complex required for autophagy', *Molecular Biology of the Cell*, 20(7), pp. 1981–1991. doi:10.1091/mbc.e08-12-1248.

Hou, C. *et al.* (2020) 'Structural Insight into the DNA Binding Function of Transcription Factor ERF', *Biochemistry*, 59(47), pp. 4499–4506. doi:10.1021/acs.biochem.0c00774.

Houles, T. *et al.* (2018) 'RSK Regulates PFK-2 Activity to Promote Metabolic Rewiring in Melanoma', *Cancer Research*, 78(9), pp. 2191–2204. doi:10.1158/0008-5472.CAN-17-2215.

House, S.L. *et al.* (2005) 'Cardioprotection induced by cardiac-specific overexpression of fibroblast growth factor-2 is mediated by the MAPK cascade', *American Journal of Physiology. Heart and Circulatory Physiology*, 289(5), pp. H2167-2175. doi:10.1152/ajpheart.00392.2005.

Hsu, V.W., Bai, M. and Li, J. (2012) 'Getting active: protein sorting in endocytic recycling', *Nature Reviews. Molecular Cell Biology*, 13(5), pp. 323–328. doi:10.1038/nrm3332.

Hu, A., Noble, W.S. and Wolf-Yadlin, A. (2016) 'Technical advances in proteomics: new developments in data-independent acquisition', *F1000Research*, 5. doi:10.12688/F1000RESEARCH.7042.1.

Hu, Q. *et al.* (2005) 'The Orbitrap: a new mass spectrometer', *Journal of mass spectrometry: JMS*, 40(4), pp. 430–443. doi:10.1002/jms.856.

Huang, F. *et al.* (2006) 'Differential Regulation of EGF Receptor Internalization and Degradation by Multiubiquitination within the Kinase Domain', *Molecular Cell*, 21(6), pp. 737–748. doi:10.1016/j.molcel.2006.02.018.

Huang, K.-Y. *et al.* (2019) 'dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications', *Nucleic Acids Research*, 47(D1), pp. D298–D308. doi:10.1093/nar/gky1074.

Huang, Z. *et al.* (2016) 'Two FGF Receptor Kinase Molecules Act in Concert to Recruit and Transphosphorylate Phospholipase Cγ', *Molecular Cell*, 61(1), pp. 98–110. doi:10.1016/j.molcel.2015.11.010.

Huang, Z. *et al.* (2017) 'Uncoupling the Mitogenic and Metabolic Functions of FGF1 by Tuning FGF1-FGF Receptor Dimer Stability', *Cell Reports*, 20(7), pp. 1717–1728. doi:10.1016/j.celrep.2017.06.063.

Hubbard, S.R. (2004) 'Juxtamembrane autoinhibition in receptor tyrosine kinases', *Nature Reviews Molecular Cell Biology*, 5(6), pp. 464–471. doi:10.1038/nrm1399.

Huber, W. *et al.* (2015) 'Orchestrating high-throughput genomic analysis with Bioconductor', *Nature Methods*, 12(2), pp. 115–121. doi:10.1038/nmeth.3252.

Hung, M.-C. and Link, W. (2011) 'Protein localization in disease and therapy', *Journal of Cell Science*, 124(Pt 20), pp. 3381–3392. doi:10.1242/jcs.089110.

Hung, V. *et al.* (2014) 'Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging', *Molecular Cell*, 55(2), pp. 332–341. doi:10.1016/j.molcel.2014.06.003.

Hung, V. *et al.* (2016) 'Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2', *Nature Protocols*, 11(3), pp. 456–475. doi:10.1038/nprot.2016.018.

Hunter, T. (1995) 'Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling', *Cell*, 80(2), pp. 225–236. doi:10.1016/0092-8674(95)90405-0.

Huotari, J. and Helenius, A. (2011) 'Endosome maturation', *The EMBO Journal*, 30(17), pp. 3481–3500. doi:10.1038/emboj.2011.286.

Hur, J.Y. *et al.* (2020) 'High-level FGFR2 amplification is associated with poor prognosis and Lower response to chemotherapy in gastric cancers', *Pathology - Research and Practice*, 216(4), p. 152878. doi:10.1016/j.prp.2020.152878.

Huttlin, E.L. *et al.* (2015) 'The BioPlex Network: A Systematic Exploration of the Human Interactome', *Cell*, 162(2), pp. 425–440. doi:10.1016/j.cell.2015.06.043.

Huttlin, E.L. *et al.* (2017) 'Architecture of the human interactome defines protein communities and disease networks', *Nature*, 545(7655), pp. 505–509. doi:10.1038/nature22366.

Huttlin, E.L. *et al.* (2021) 'Dual proteome-scale networks reveal cell-specific remodeling of the human interactome', *Cell*, 184(11), pp. 3022-3040.e28. doi:10.1016/j.cell.2021.04.011.

Iakoucheva, L.M. *et al.* (2004) 'The importance of intrinsic disorder for protein phosphorylation.', *Nucleic acids research*, 32(3), pp. 1037–49. doi:10.1093/nar/gkh253.

Imamura, H. *et al.* (2010) 'Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data', *BMC Bioinformatics*, 11(1), p. 232. doi:10.1186/1471-2105-11-232.

Inoki, K. *et al.* (2002) 'TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling', *Nature Cell Biology*, 4(9), pp. 648–657. doi:10.1038/ncb839.

Inoki, K. *et al.* (2003) 'Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling', *Genes & Development*, 17(15), pp. 1829–1834. doi:10.1101/gad.1110003.

Inpanathan, S. and Botelho, R.J. (2019) 'The Lysosome Signaling Platform: Adapting With the Times', *Frontiers in Cell and Developmental Biology*, 7, p. 113. doi:10.3389/fcell.2019.00113.

Ioannou, M.S. *et al.* (2017) 'Intersectin-s interaction with DENND2B facilitates recycling of epidermal growth factor receptor', *EMBO Reports*, 18(12), pp. 2119–2130. doi:10.15252/embr.201744034.

Jacobs, B.L. *et al.* (2017) 'Identification of mechanically regulated phosphorylation sites on tuberin (TSC2) that control mechanistic target of rapamycin (mTOR) signaling', *The Journal of Biological Chemistry*, 292(17), pp. 6987–6997. doi:10.1074/jbc.M117.777805.

Janssen-Heininger, Y.M.W. *et al.* (2008) 'REDOX-BASED REGULATION OF SIGNAL TRANSDUCTION: PRINCIPLES, PITFALLS, AND PROMISES', *Free radical biology & medicine*, 45(1), pp. 1–17. doi:10.1016/j.freeradbiomed.2008.03.011.

Jeong, H. *et al.* (2001) 'Lethality and centrality in protein networks', *Nature*, 411(6833), pp. 41–42. doi:10.1038/35075138.

Jiang, R. (2015) 'Walking on multiple disease-gene networks to prioritize candidate genes', *JOURNAL OF MOLECULAR CELL BIOLOGY*, 7(3, SI), pp. 214–230. doi:10.1093/jmcb/mjv008.

Jiang, T., Pan, C.Q. and Low, B.C. (2017) 'BPGAP1 spatially integrates JNK/ERK signaling crosstalk in oncogenesis', *Oncogene*, 36(22), pp. 3178–3192. doi:10.1038/onc.2016.466.

Jiang, X. and Sorkin, A. (2003) 'Epidermal growth factor receptor internalization through clathrin-coated pits requires Cbl RING finger and proline-rich domains but not receptor polyubiquitylation', *Traffic (Copenhagen, Denmark)*, 4(8), pp. 529–543. doi:10.1034/j.1600-0854.2003.t01-1-00109.x.

Jin, F. and Gräter, F. (2021) 'How multisite phosphorylation impacts the conformations of intrinsically disordered proteins', *PLOS Computational Biology*, 17(5), p. e1008939. doi:10.1371/journal.pcbi.1008939.

Jin, L. *et al.* (2021) 'A comparative study of evaluating missing value imputation methods in label-free proteomics', *Scientific Reports*, 11, p. 1760. doi:10.1038/s41598-021-81279-4.

Joensuu, M. *et al.* (2017) 'Visualizing endocytic recycling and trafficking in live neurons by subdiffractional tracking of internalized molecules', *Nature Protocols*, 12(12), pp. 2590–2622. doi:10.1038/nprot.2017.116.

Johannessen, L.E. *et al.* (2006) 'Activation of the Epidermal Growth Factor (EGF) Receptor Induces Formation of EGF Receptor- and Grb2-Containing Clathrin-Coated Pits', *Molecular and Cellular Biology*, 26(2), pp. 389–401. doi:10.1128/MCB.26.2.389-401.2006.

Jones, S.W. *et al.* (1988) 'A Xenopus ribosomal protein S6 kinase has two apparent kinase domains that are each similar to distinct protein kinases', *Proceedings of the National Academy of Sciences*, 85(10), pp. 3377–3381. doi:10.1073/pnas.85.10.3377.

Joung, J. *et al.* (2017) 'Genome-scale CRISPR-Cas9 Knockout and Transcriptional Activation Screening', *Nature protocols*, 12(4), pp. 828–863. doi:10.1038/nprot.2017.016.

Jung, C.H. *et al.* (2009) 'ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery', *Molecular Biology of the Cell*, 20(7), pp. 1992–2003. doi:10.1091/mbc.e08-12-1249.

Kaidanovich-Beilin, O. and Woodgett, J.R. (2011) 'GSK-3: Functional Insights from Cell Biology and Animal Models', *Frontiers in Molecular Neuroscience*, 4. doi:10.3389/fnmol.2011.00040.

Kammers, K. *et al.* (2015) 'Detecting significant changes in protein abundance', *EuPA Open Proteomics*, 7, pp. 11–19. doi:10.1016/j.euprot.2015.02.002.

Kane, M.T., Morgan, P.M. and Coonan, C. (1997) 'Peptide growth factors and preimplantation development', *Human Reproduction Update*, 3(2), pp. 137–157. doi:10.1093/humupd/3.2.137.

Kanehisa, M. and Goto, S. (2000) 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic Acids Research*, 28(1), pp. 27–30. doi:10.1093/nar/28.1.27.

Kanodia, J. *et al.* (2014) 'Deciphering the mechanism behind Fibroblast Growth Factor (FGF) induced biphasic signal-response profiles', *Cell Communication and Signaling : CCS*, 12, p. 34. doi:10.1186/1478-811X-12-34.

Kardami, E. *et al.* (2007) 'Fibroblast growth factor-2 and cardioprotection', *Heart Failure Reviews*, 12(3–4), pp. 267–277. doi:10.1007/s10741-007-9027-0.

Karl, K. *et al.* (2020) 'Ligand bias in receptor tyrosine kinase signaling', *The Journal of Biological Chemistry* [Preprint]. doi:10.1074/jbc.REV120.015190.

Karlsson, M. *et al.* (2004) 'Both Nuclear-Cytoplasmic Shuttling of the Dual Specificity Phosphatase MKP-3 and Its Ability to Anchor MAP Kinase in the Cytoplasm Are Mediated by a Conserved Nuclear Export Signal *', *Journal of Biological Chemistry*, 279(40), pp. 41882–41891. doi:10.1074/jbc.M406720200.

Karp, N.A. *et al.* (2010) 'Addressing accuracy and precision issues in iTRAQ quantitation', *Molecular & cellular proteomics: MCP*, 9(9), pp. 1885–1897. doi:10.1074/mcp.M900628-MCP200.

Karpievitch, Y. *et al.* (2009) 'A statistical framework for protein quantitation in bottom-up MS-based proteomics', *Bioinformatics (Oxford, England)*, 25(16), pp. 2028–2034. doi:10.1093/bioinformatics/btp362.

Kazazic, M. *et al.* (2009) 'Epsin 1 is involved in recruitment of ubiquitinated EGF receptors into clathrin-coated pits', *Traffic (Copenhagen, Denmark)*, 10(2), pp. 235–245. doi:10.1111/j.1600-0854.2008.00858.x.

Keeling, D.M. *et al.* (2019) 'The meanings of "function" in biology and the problematic case of de novo gene emergence', *eLife*, 8, p. e47014. doi:10.7554/eLife.47014.

Kelly, K.L., Ruderman, N.B. and Chen, K.S. (1992) 'Phosphatidylinositol-3-kinase in isolated rat adipocytes. Activation by insulin and subcellular distribution.', *Journal of Biological Chemistry*, 267(5), pp. 3423–3428. doi:10.1016/S0021-9258(19)50748-7.

Kessler, D. *et al.* (2021) 'Targeting Son of Sevenless 1: The pacemaker of KRAS', *Current Opinion in Chemical Biology*, 62, pp. 109–118. doi:10.1016/j.cbpa.2021.02.014.

Ketel, K. *et al.* (2016) 'A phosphoinositide conversion mechanism for exit from endosomes', *Nature*, 529(7586), pp. 408–412. doi:10.1038/nature16516.

Khan, I., Gril, B. and Steeg, P.S. (2019) 'Metastasis Suppressors NME1 and NME2 Promote Dynamin 2 Oligomerization and Regulate Tumor Cell Endocytosis, Motility, and Metastasis', *Cancer Research*, 79(18), pp. 4689–4702. doi:10.1158/0008-5472.CAN-19-0492.

Khan, I. and Steeg, P.S. (2021) 'Endocytosis: a pivotal pathway for regulating metastasis', *British Journal of Cancer*, 124(1), pp. 66–75. doi:10.1038/s41416-020-01179-8.

Kholodenko, B.N. (2009) 'Spatially distributed cell signalling', *FEBS Letters*, 583(24), pp. 4006–4012. doi:https://doi.org/10.1016/j.febslet.2009.09.045.

Kiang, M.Y. (2001) 'Extending the Kohonen self-organizing map networks for clustering analysis', *Computational Statistics & Data Analysis*, 38(2), pp. 161–180. doi:10.1016/S0167-9473(01)00040-8.

Kim, D.-H. *et al.* (2002) 'mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery', *Cell*, 110(2), pp. 163–175. doi:10.1016/s0092-8674(02)00808-5.

Kim, D.-H. *et al.* (2003) 'GbetaL, a positive regulator of the rapamycin-sensitive pathway required for the nutrient-sensitive interaction between raptor and mTOR', *Molecular Cell*, 11(4), pp. 895–904. doi:10.1016/s1097-2765(03)00114-x.

Kim, D.I. *et al.* (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), pp. E2453-2461. doi:10.1073/pnas.1406459111.

Kim, D.I. *et al.* (2016) 'An improved smaller biotin ligase for BioID proximity labeling', *Molecular Biology of the Cell*, 27(8), pp. 1188–1196. doi:10.1091/mbc.E15-12-0844.

Kim, E. *et al.* (2008) 'Regulation of TORC1 by Rag GTPases in nutrient response', *Nature Cell Biology*, 10(8), pp. 935–945. doi:10.1038/ncb1753.

Kim, J. *et al.* (2011) 'AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1', *Nature Cell Biology*, 13(2), pp. 132–141. doi:10.1038/ncb2152.

Kim, M.J. *et al.* (2000) 'Direct Interaction of SOS1 Ras Exchange Protein with the SH3 Domain of Phospholipase C-γ1', *Biochemistry*, 39(29), pp. 8674–8682. doi:10.1021/bi992558t.

Kim, M.-S., Zhong, J. and Pandey, A. (2016) 'Common errors in mass spectrometry-based analysis of post-translational modifications', *PROTEOMICS*, 16(5), pp. 700–714. doi:10.1002/pmic.201500355.

Kim, Y.-M. *et al.* (2012) 'SH3BP4 is a negative regulator of amino acid-Rag GTPase-mTORC1 signaling', *Molecular Cell*, 46(6), pp. 833–846. doi:10.1016/j.molcel.2012.04.007.

Kitano, H. (2007) 'Towards a theory of biological robustness', *Molecular Systems Biology*, 3(1), p. 137. doi:10.1038/msb4100179.

Koenig, T. *et al.* (2008) 'Robust Prediction of the MASCOT Score for an Improved Quality Assessment in Mass Spectrometric Proteomics', *Journal of Proteome Research*, 7(9), pp. 3708–3717. doi:10.1021/pr700859x.

Köhler, S. *et al.* (2008) 'Walking the Interactome for Prioritization of Candidate Disease Genes', *The American Journal of Human Genetics*, 82(4), pp. 949–958. doi:10.1016/J.AJHG.2008.02.013.

Kokla, M. *et al.* (2019) 'Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study', *BMC Bioinformatics*, 20(1), pp. 1–11. doi:10.1186/s12859-019-3110-0.

Köksal, A.S. *et al.* (2018) 'Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data', *Cell Reports*, 24(13), pp. 3607–3618. doi:10.1016/j.celrep.2018.08.085.

Kong, Q., Majeska, R.J. and Vazquez, M. (2011) 'Migration of connective tissue-derived cells is mediated by ultra-low concentration gradient fields of EGF', *Experimental Cell Research*, 317(11), pp. 1491–1502. doi:10.1016/j.yexcr.2011.04.003.

Konturek, J.W. *et al.* (1989) 'Distribution and release of epidermal growth factor in man.', *Gut*, 30(9), pp. 1194–1200. doi:10.1136/gut.30.9.1194.

Korolchuk, V.I. *et al.* (2011) 'Lysosomal positioning coordinates cellular nutrient responses', *Nature Cell Biology*, 13(4), pp. 453–460. doi:10.1038/ncb2204.

Koseska, A. and Bastiaens, P.I.H. (2020) 'Processing Temporal Growth Factor Patterns by an Epidermal Growth Factor Receptor Network Dynamically Established in Space', *Annual Review of Cell and Developmental Biology*, 36(1), pp. 359–383. doi:10.1146/annurev-cellbio-013020-103810.

Kostas, M. *et al.* (2018) 'Translocation of Exogenous FGF1 and FGF2 Protects the Cell against Apoptosis Independently of Receptor Activation', *Journal of Molecular Biology*, 430(21), pp. 4087–4101. doi:10.1016/J.JMB.2018.08.004.

Kouhara, H. *et al.* (1997) 'A Lipid-Anchored Grb2-Binding Protein That Links FGF-Receptor Activation to the Ras/MAPK Signaling Pathway', *Cell*, 89(5), pp. 693–702. doi:10.1016/S0092-8674(00)80252-4.

Kouranti, I. *et al.* (2006) 'Rab35 regulates an endocytic recycling pathway essential for the terminal steps of cytokinesis', *Current biology: CB*, 16(17), pp. 1719–1725. doi:10.1016/j.cub.2006.07.020.

Kovacs, E. *et al.* (2015) 'A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor', *Annual Review of Biochemistry*, 84(1), pp. 739–764. doi:10.1146/annurev-biochem-060614-034402.

Kovács, I.A. *et al.* (2019) 'Network-based prediction of protein interactions', *Nature Communications*, 10(1), p. 1240. doi:10.1038/s41467-019-09177-y.

Kovalenko, D. *et al.* (2003) 'Sef Inhibits Fibroblast Growth Factor Signaling by Inhibiting FGFR1 Tyrosine Phosphorylation and Subsequent ERK Activation *', *Journal of Biological Chemistry*, 278(16), pp. 14087–14091. doi:10.1074/jbc.C200606200.

Kratchmarova, I. *et al.* (2005) 'Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation', *Science (New York, N.Y.)*, 308(5727), pp. 1472–1477. doi:10.1126/science.1107627.

Kreitman, M., Noronha, A. and Yarden, Y. (2018) 'Irreversible modifications of receptor tyrosine kinases', *FEBS Letters*, 592(13), pp. 2199–2212. doi:https://doi.org/10.1002/1873-3468.13095.

Kreuger, J. *et al.* (2005) 'Fibroblast growth factors share binding sites in heparan sulphate', *The Biochemical Journal*, 389(Pt 1), pp. 145–150. doi:10.1042/BJ20042129.

Krishna, M. and Narang, H. (2008) 'The complexity of mitogen-activated protein kinases (MAPKs) made simple', *Cellular and Molecular Life Sciences*, 65(22), pp. 3525–3544. doi:10.1007/s00018-008-8170-7.

Krug, K. *et al.* (2019) 'A Curated Resource for Phosphosite-specific Signature Analysis.', *Molecular & cellular proteomics : MCP*, 18(3), pp. 576–593. doi:10.1074/mcp.TIR118.000943.

Kuleshov, M.V. *et al.* (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic Acids Research*, 44(W1), pp. W90–W97. doi:10.1093/nar/gkw377.

Kunii, K. *et al.* (2008) 'FGFR2-Amplified Gastric Cancer Cell Lines Require FGFR2 and Erbb3 Signaling for Growth and Survival', *Cancer Research*, 68(7), pp. 2340–2348. doi:10.1158/0008-5472.CAN-07-5229.

Kvainickas, A. *et al.* (2019) 'Retromer and TBC1D5 maintain late endosomal RAB7 domains to enable amino acid-induced mTORC1 signaling', *The Journal of Cell Biology*, 218(9), pp. 3019–3038. doi:10.1083/jcb.201812110.

Kwak, S.S. *et al.* (2016) 'Amino acid-dependent NPRL2 interaction with Raptor determines mTOR Complex 1 activation', *Cellular Signalling*, 28(2), pp. 32–41. doi:10.1016/j.cellsig.2015.11.008.

Lachmann, A. *et al.* (2010) 'ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments', *Bioinformatics (Oxford, England)*, 26(19), pp. 2438–2444. doi:10.1093/bioinformatics/btq466.

Lachmann, A. and Ma'ayan, A. (2009) 'KEA: kinase enrichment analysis', *Bioinformatics (Oxford, England)*, 25(5), pp. 684–686. doi:10.1093/bioinformatics/btp026.

Lai, S. and Pelech, S. (2016) 'Regulatory roles of conserved phosphorylation sites in the activation T-loop of the MAP kinase ERK1', *Molecular Biology of the Cell*, 27(6), pp. 1040–1050. doi:10.1091/mbc.E15-07-0527.

Laiouar, S. *et al.* (2020) 'RabX1 Organizes a Late Endosomal Compartment that Forms Tubular Connections to Lysosomes Consistent with a "Kiss and Run" Mechanism', *Current Biology*, 30(7), pp. 1177-1188.e5. doi:10.1016/j.cub.2020.01.048.

Lam, I., Pickering, C.M. and Mac Gabhann, F. (2019) 'Context-Dependent Regulation of Receptor Tyrosine Kinases: Insights from Systems Biology Approaches', *Wiley interdisciplinary reviews. Systems biology and medicine*, 11(2), p. e1437. doi:10.1002/wsbm.1437.

Lamming, D.W. and Bar-Peled, L. (2019) 'Lysosome: The metabolic signaling hub', *Traffic*, 20(1), pp. 27–38. doi:https://doi.org/10.1111/tra.12617.

Lampada, A. *et al.* (2017) 'mTORC1-independent autophagy regulates receptor tyrosine kinase phosphorylation in colorectal cancer cells via an mTORC2-mediated mechanism', *Cell Death and Differentiation*, 24(6), pp. 1045–1062. doi:10.1038/cdd.2017.41.

Landry, C.R., Levy, E.D. and Michnick, S.W. (2009) 'Weak functional constraints on phosphoproteomes', *Trends in Genetics*, 25(5), pp. 193–197. doi:10.1016/j.tig.2009.03.003.

Lanzetti, L. and Di Fiore, P.P. (2017) 'Behind the Scenes: Endo/Exocytosis in the Acquisition of Metastatic Traits', *Cancer Research*, 77(8), pp. 1813–1817. doi:10.1158/0008-5472.CAN-16-3403.

Lao, D.-H. *et al.* (2007) 'Direct Binding of PP2A to Sprouty2 and Phosphorylation Changes Are a Prerequisite for ERK Inhibition Downstream of Fibroblast Growth Factor Receptor Stimulation *', *Journal of Biological Chemistry*, 282(12), pp. 9117–9126. doi:10.1074/jbc.M607563200.

Larsson, C. (2006) 'Protein kinase C and the regulation of the actin cytoskeleton', *Cellular Signalling*, 18(3), pp. 276–284. doi:10.1016/j.cellsig.2005.07.010.

Larsson, H. *et al.* (1999) 'Fibroblast growth factor receptor-1-mediated endothelial cell proliferation is dependent on the Src homology (SH) 2/SH3 domain-containing adaptor protein Crk', *The Journal of Biological Chemistry*, 274(36), pp. 25726–25734. doi:10.1074/jbc.274.36.25726.

Lavoie, H., Gagnon, J. and Therrien, M. (2020) 'ERK signalling: a master regulator of cell behaviour, life and fate', *Nature Reviews Molecular Cell Biology*, 21(10), pp. 607–632. doi:10.1038/s41580-020-0255-7.

Lavoie, J.N. *et al.* (1996) 'Cyclin D1 Expression Is Regulated Positively by the p42/p44MAPK and Negatively by the p38/HOGMAPK Pathway *', *Journal of Biological Chemistry*, 271(34), pp. 20608–20616. doi:10.1074/jbc.271.34.20608.

Lawrence, M.C. and Ward, C.W. (2015) 'Structural Features of the Receptor Tyrosine Kinase Ectodomains', in Wheeler, D.L. and Yarden, Y. (eds) *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease*. New York, NY: Springer, pp. 163–193. doi:10.1007/978-1-4939-2053-2_8.

Lazar, C. *et al.* (2016) 'Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies', *Journal of Proteome Research*, 15(4), pp. 1116–1125. doi:10.1021/acs.jproteome.5b00981.

Lee, B. *et al.* (2020) 'Heterogeneous Multi-Layered Network Model for Omics Data Integration and Analysis', *Frontiers in Genetics*, 10, p. 1381. doi:10.3389/fgene.2019.01381.

Lee, D.C.H., Jones, A.R. and Hubbard, S.J. (2015) 'Computational phosphoproteomics: From identification to localization', *PROTEOMICS*, 15(5–6), pp. 950–963. doi:10.1002/pmic.201400372.

van der Lee, R. *et al.* (2014) 'Classification of intrinsically disordered regions and proteins.', *Chemical reviews*, 114(13), pp. 6589–631. doi:10.1021/cr400525m.

Lee, S. *et al.* (2015) 'Transport through recycling endosomes requires EHD1 recruitment by a phosphatidylserine translocase', *The EMBO Journal*, 34(5), pp. 669–688. doi:10.15252/embj.201489703.

Lee, T. *et al.* (2004) 'Docking Motif Interactions in MAP Kinases Revealed by Hydrogen Exchange Mass Spectrometry', *Molecular Cell*, 14(1), pp. 43–55. doi:10.1016/S1097-2765(04)00161-3.

Leeming, M.G. *et al.* (2020) 'Phosphomatics: interactive interrogation of substrate–kinase networks in global phosphoproteomics datasets', *Bioinformatics* [Preprint], (btaa916). doi:10.1093/bioinformatics/btaa916.

LeFebvre, A.K. *et al.* (2006) 'Translation Initiation Factor eIF4G-1 Binds to eIF3 through the eIF3e Subunit', *The Journal of biological chemistry*, 281(32), pp. 22917–22932. doi:10.1074/jbc.M605418200.

Lemmon, M. and Schlessinger, J. (2011) 'Cell signaling by receptor-tyrosine kinases', *Biochemistry*, 141(7), pp. 1117–1134. doi:10.1016/j.cell.2010.06.011.Cell.

Lemmon, M.A., Schlessinger, J. and Ferguson, K.M. (2014) 'The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases', *Cold Spring Harbor Perspectives in Biology*, 6(4), p. a020768. doi:10.1101/cshperspect.a020768.

Leutert, M., Entwisle, S.W. and Villén, J. (2021) 'Decoding post translational modification crosstalk with proteomics', *Molecular & Cellular Proteomics*, 0(0). doi:10.1016/j.mcpro.2021.100129.

Levkowitz, G. *et al.* (1999) 'Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1', *Molecular Cell*, 4(6), pp. 1029–1040. doi:10.1016/s1097-2765(00)80231-2.

Levy, M.J., Washburn, M.P. and Florens, L. (2018) 'Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background', *Journal of Proteome Research*, 17(10), pp. 3586–3592. doi:10.1021/acs.jproteome.8b00269.

Lew, E.D. *et al.* (2009) 'The Precise Sequence of FGF Receptor Autophosphorylation Is Kinetically Driven and Is Disrupted by Oncogenic Mutations', *Science Signaling*, 2(58), pp. ra6–ra6. doi:10.1126/scisignal.2000021.

Li, C. *et al.* (2007) 'Dusp6(Mkp3) is a negative feedback regulator of FGF stimulated ERK signaling during mouse development', *Development (Cambridge, England)*, 134(1), pp. 167–176. doi:10.1242/dev.02701.

Li, S. *et al.* (2001) 'Tyrosine phosphorylation of Grb2 by Bcr/Abl and epidermal growth factor receptor: a novel regulatory mechanism for tyrosine kinase signaling', *The EMBO journal*, 20(23), pp. 6793–6804. doi:10.1093/emboj/20.23.6793.

Li, X. and Gao, T. (2014) 'mTORC2 phosphorylates protein kinase Cζ to regulate its stability and activity', *EMBO reports*, 15(2), pp. 191–198. doi:10.1002/embr.201338119.

Li, Y. and Patra, J.C. (2010) 'Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network', *Bioinformatics*, 26(9), pp. 1219–1224. doi:10.1093/bioinformatics/btq108.

Li, Z. *et al.* (2012) 'Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos', *Journal of Proteome Research*, 11(3), pp. 1582–1590. doi:10.1021/pr200748h.

Li, Z. *et al.* (2016) 'Autophagic flux is highly active in early mitosis and differentially regulated throughout the cell cycle', *Oncotarget*, 7(26), pp. 39705–39718. doi:10.18632/oncotarget.9451.

Li, Z. *et al.* (2020) 'ULK1-ATG13 and their mitotic phospho-regulation by CDK1 connect autophagy to cell cycle', *PLoS biology*, 18(6), p. e3000288. doi:10.1371/journal.pbio.3000288.

Liao, S. *et al.* (2007) 'The cardioprotective effect of the low molecular weight isoform of fibroblast growth factor-2: the role of JNK signaling', *Journal of Molecular and Cellular Cardiology*, 42(1), pp. 106–120. doi:10.1016/j.yjmcc.2006.10.005.

Liao, Y. *et al.* (2019) 'WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs', *Nucleic Acids Research*, 47(W1), pp. W199–W205. doi:10.1093/nar/gkz401.

Liberzon, A. *et al.* (2011) 'Molecular signatures database (MSigDB) 3.0', *Bioinformatics*, 27(12), pp. 1739–1740. doi:10.1093/bioinformatics/btr260.

Lill, N.L. and Sever, N.I. (2012) 'Where EGF Receptors Transmit Their Signals', *Science Signaling*, 5(243), pp. pe41–pe41. doi:10.1126/scisignal.2003341.

Lim, M.Y., Paulo, J.A. and Gygi, S.P. (2019) 'Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model', *Journal of Proteome Research*, 18(11), pp. 4020–4026. doi:10.1021/acs.jproteome.9b00492.

Linding, R. *et al.* (2008) 'NetworKIN: A resource for exploring cellular phosphorylation networks', *Nucleic Acids Research* [Preprint]. doi:10.1093/nar/gkm902.

Liu, J. *et al.* (2021) 'Stable and EGF-Induced Temporal Interactome Profiling of CBL and CBLB Highlights Their Signaling Complex Diversity', *Journal of Proteome Research* [Preprint]. doi:10.1021/acs.jproteome.1c00284.

Liu, P. *et al.* (2015) 'PtdIns(3,4,5)P3-Dependent Activation of the mTORC2 Kinase Complex', *Cancer Discovery*, 5(11), pp. 1194–1209. doi:10.1158/2159-8290.CD-15-0460.

Liu, S.-L. *et al.* (2018) 'Quantitative Lipid Imaging Reveals a New Signaling Function of Phosphatidylinositol-3,4-Bisphophate: Isoform- and Site-Specific Activation of Akt', *Molecular Cell*, 71(6), pp. 1092-1104.e5. doi:10.1016/j.molcel.2018.07.035.

Liu, X. *et al.* (2020) 'Combined proximity labeling and affinity purification−mass spectrometry workflow for mapping and visualizing protein interaction networks', *Nature Protocols*, 15(10), pp. 3182–3211. doi:10.1038/s41596-020-0365-x.

Lobingier, B.T. *et al.* (2017) 'An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells', *Cell*, 169(2), pp. 350-360.e12. doi:10.1016/j.cell.2017.03.022.

Locard-Paulet, M. *et al.* (2020) 'Comparing 22 Popular Phosphoproteomics Pipelines for Peptide Identification and Site Localization', *Journal of Proteome Research*, 19(3), pp. 1338–1345. doi:10.1021/acs.jproteome.9b00679.

Locasale, J.W. and Wolf-Yadlin, A. (2009) 'Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data', *PLoS ONE*, 4(8), p. e6522. doi:10.1371/journal.pone.0006522.

Locascio, L.E. and Donoghue, D.J. (2013) 'KIDs rule: regulatory phosphorylation of RTKs', *Trends in Biochemical Sciences*, 38(2), pp. 75–84. doi:10.1016/j.tibs.2012.12.001.

Long, X. *et al.* (2005) 'Rheb binds and regulates the mTOR kinase', *Current biology: CB*, 15(8), pp. 702–713. doi:10.1016/j.cub.2005.02.053.

Longatti, A. *et al.* (2012) 'TBC1D14 regulates autophagosome formation via Rab11- and ULK1-positive recycling endosomes', *The Journal of Cell Biology*, 197(5), pp. 659–675. doi:10.1083/jcb.201111079.

Lonic, A. *et al.* (2021) 'Phosphorylation of PKCδ by FER tips the balance from EGFR degradation to recycling', *Journal of Cell Biology*, 220(e201902073). doi:10.1083/jcb.201902073.

Lonze, B.E. and Ginty, D.D. (2002) 'Function and Regulation of CREB Family Transcription Factors in the Nervous System', *Neuron*, 35(4), pp. 605–623. doi:10.1016/S0896-6273(02)00828-0.

Lorenz, K. *et al.* (2009) 'A new type of ERK1/2 autophosphorylation causes cardiac hypertrophy', *Nature Medicine*, 15(1), pp. 75–83. doi:10.1038/nm.1893.

Lorenzen, E. *et al.* (2018) 'G protein subtype-specific signaling bias in a series of CCR5 chemokine analogs.', *Science signaling*, 11(552), p. eaao6152. doi:10.1126/scisignal.aao6152.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. doi:10.1186/s13059-014-0550-8.

Low, T.Y. *et al.* (2021) 'WIDENING THE BOTTLENECK OF PHOSPHOPROTEOMICS: EVOLVING STRATEGIES FOR PHOSPHOPEPTIDE ENRICHMENT', *Mass Spectrometry Reviews*, 40(4), pp. 309–333. doi:10.1002/mas.21636.

Lu, C. *et al.* (2010) 'Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor', *Molecular and Cellular Biology*, 30(22), pp. 5432–5443. doi:10.1128/MCB.00742-10.

Lu, C. *et al.* (2012) 'Mechanisms for Kinase-mediated Dimerization of the Epidermal Growth Factor Receptor *', *Journal of Biological Chemistry*, 287(45), pp. 38244–38253. doi:10.1074/jbc.M112.414391.

Lu, H. *et al.* (2020) 'Resistance to allosteric SHP2 inhibition in FGFR-driven cancers through rapid feedback activation of FGFR', *Oncotarget*, 11(3), pp. 265–281. doi:10.18632/oncotarget.27435.

Lundberg, E. *et al.* (2010) 'Defining the transcriptome and proteome in three functionally different human cell lines', *Molecular Systems Biology*, 6(1), p. 450. doi:10.1038/msb.2010.106.

Lundberg, E. and Borner, G.H.H.H. (2019) *Spatial proteomics: a powerful discovery tool for cell biology*, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. doi:10.1038/s41580-018-0094-y.

Lundby, A. *et al.* (2019) 'Oncogenic Mutations Rewire Signaling Pathways by Switching Protein Recruitment to Phosphotyrosine Sites', *Cell*, 179(2), pp. 543-560.e26. doi:10.1016/J.CELL.2019.09.008.

Lynch, T.J. *et al.* (2004) 'Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib', *The New England Journal of Medicine*, 350(21), pp. 2129–2139. doi:10.1056/NEJMoa040938.

Macia, E. *et al.* (2006) 'Dynasore, a cell-permeable inhibitor of dynamin', *Developmental Cell*, 10(6), pp. 839–850. doi:10.1016/j.devcel.2006.04.002.

Madshus, I.H. and Stang, E. (2009) 'Internalization and intracellular sorting of the EGF receptor: a model for understanding the mechanisms of receptor trafficking', *Journal of Cell Science*, 122(19), pp. 3433–3439. doi:10.1242/jcs.050260.

Maertens, A. *et al.* (2020) 'Functionally Enigmatic Genes in Cancer: Using TCGA Data to Map the Limitations of Annotations.', *Scientific reports*, 10(1), p. 4106. doi:10.1038/s41598-020-60456-x.

Maher, S., Jjunju, F.P.M. and Taylor, S. (2015) 'Colloquium: 100 years of mass spectrometry: Perspectives and future trends', *Reviews of Modern Physics*, 87(1), pp. 113–135. doi:10.1103/RevModPhys.87.113.

Mailman, R.B. (2007) 'GPCR functional selectivity has therapeutic impact', *Trends in Pharmacological Sciences*, 28(8), pp. 390–396. doi:10.1016/j.tips.2007.06.002.

Makarenkova, H.P. *et al.* (2009) 'Differential interactions of FGFs with heparan sulfate control gradient formation and branching morphogenesis', *Science Signaling*, 2(88), p. ra55. doi:10.1126/scisignal.2000304.

Małecki, J. *et al.* (2004) 'Translocation of FGF-1 and FGF-2 across Vesicular Membranes Occurs during G_1-Phase by a Common Mechanism', *Molecular Biology of the Cell*, 15(2), pp. 801–814. doi:10.1091/mbc.e03-08-0589.

Maleki, F. *et al.* (2020) 'Gene Set Analysis: Challenges, Opportunities, and Future Research', *Frontiers in Genetics*, 11, p. 654. doi:10.3389/fgene.2020.00654.

Manadas, B. *et al.* (2010) 'Peptide fractionation in proteomics approaches', *Expert Review of Proteomics*, 7(5), pp. 655–663. doi:10.1586/epr.10.46.

Manning, B.D. and Toker, A. (2017) 'AKT/PKB Signaling: Navigating the Network', *Cell*, 169(3), pp. 381–405. doi:10.1016/j.cell.2017.04.001.

Mantini, G. *et al.* (2021) 'Computational Analysis of Phosphoproteomics Data in Multi-Omics Cancer Studies', *PROTEOMICS*, 21(3–4), p. 1900312. doi:10.1002/pmic.201900312.

Marshall, C.J. (1995) 'Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation', *Cell*, 80(2), pp. 179–185. doi:10.1016/0092-8674(95)90401-8.

Martina, J.A. *et al.* (2012) 'MTORC1 functions as a transcriptional regulator of autophagy by preventing nuclear transport of TFEB', *Autophagy*, 8(6), pp. 903–914. doi:10.4161/auto.19653.

Martinez-Lopez, N. *et al.* (2013) 'Autophagy proteins regulate ERK phosphorylation', *Nature Communications*, 4, p. 2799. doi:10.1038/ncomms3799.

Martinez-Val, A., Bekker-Jensen, Dorte Breinholdt, *et al.* (2021) 'Data Processing and Analysis for DIA-Based PhosphoproteomicsPhosphoproteomics Using SpectronautSpectronaut', in Cecconi, D. (ed.) *Proteomics Data Analysis*. New York, NY: Springer US (Methods in Molecular Biology), pp. 95–107. doi:10.1007/978-1-0716-1641-3_6.

Martinez-Val, A., Bekker-Jensen, Dorte B., *et al.* (2021) 'Spatial-proteomics reveal in-vivo phospho-signaling dynamics at subcellular resolution', *BioRXiv* [Preprint]. doi:10.1101/2021.02.02.425898.

Masuda, H. *et al.* (2012) 'Role of Epidermal Growth Factor Receptor in Breast Cancer', *Breast cancer research and treatment*, 136(2), p. 10.1007/s10549-012-2289–9. doi:10.1007/s10549-012-2289-9.

Matallanas, D. *et al.* (2006) 'Distinct Utilization of Effectors and Biological Outcomes Resulting from Site-Specific Ras Activation: Ras Functions in Lipid Rafts and Golgi Complex Are Dispensable for Proliferation and Transformation', *Molecular and Cellular Biology*, 26(1), pp. 100–116. doi:10.1128/MCB.26.1.100-116.2006.

Mathiassen, S.G., De Zio, D. and Cecconi, F. (2017) 'Autophagy and the Cell Cycle: A Complex Landscape', *Frontiers in Oncology*, 7, p. 51. doi:10.3389/fonc.2017.00051.

Mathur, R. *et al.* (2018) 'Gene set analysis methods: a systematic comparison', *BioData Mining*, 11(1), pp. 1–19. doi:10.1186/s13040-018-0166-8.

Matsui, T. and Fukuda, M. (2011) 'Small GTPase Rab12 regulates transferrin receptor degradation: Implications for a novel membrane trafficking pathway from recycling endosomes to lysosomes', *Cellular Logistics*, 1(4), pp. 155–158. doi:10.4161/cl.1.4.18152.

Matsui, T. and Fukuda, M. (2013) 'Rab12 regulates mTORC1 activity and autophagy through controlling the degradation of amino-acid transporter PAT4', *EMBO Reports*, 14(5), pp. 450–457. doi:10.1038/embor.2013.32.

Matsui, T., Itoh, T. and Fukuda, M. (2011) 'Small GTPase Rab12 regulates constitutive degradation of transferrin receptor', *Traffic (Copenhagen, Denmark)*, 12(10), pp. 1432–1443. doi:10.1111/j.1600-0854.2011.01240.x.

Matzke, M.M. *et al.* (2013) 'A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments', *Proteomics*, 13(0), pp. 493–503. doi:10.1002/pmic.201200269.

Maxfield, F.R. and McGraw, T.E. (2004) 'Endocytic recycling', *Nature Reviews Molecular Cell Biology*, 5(2), pp. 121–132. doi:10.1038/nrm1315.

McAlister, G.C. *et al.* (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.

McNally, K.E. *et al.* (2017) 'Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling', *Nature Cell Biology*, 19(10), pp. 1214–1225. doi:10.1038/ncb3610.

Mellis, I.A. *et al.* (2021) 'Responsiveness to perturbations is a hallmark of transcription factors that maintain cell identity in vitro', *Cell Systems*, 0(0). doi:10.1016/j.cels.2021.07.003.

Ménard, L., Parker, P.J. and Kermorgant, S. (2014) 'Receptor tyrosine kinase c-Met controls the cytoskeleton from different endosomes via different pathways', *Nature Communications*, 5(1), p. 3907. doi:10.1038/ncomms4907.

Meng, X.-M., Nikolic-Paterson, D.J. and Lan, H.Y. (2016) 'TGF-β: the master regulator of fibrosis', *Nature Reviews. Nephrology*, 12(6), pp. 325–338. doi:10.1038/nrneph.2016.48.

Menon, S. *et al.* (2014) 'Spatial Control of the TSC Complex Integrates Insulin and Nutrient Regulation of mTORC1 at the Lysosome', *Cell*, 156(4), pp. 771–785. doi:10.1016/j.cell.2013.11.049.

Mettlen, M. *et al.* (2018) 'Regulation of Clathrin-Mediated Endocytosis', *Annual review of biochemistry*, 87, pp. 871–896. doi:10.1146/annurev-biochem-062917-012644.

Meyer, J.G. and Schilling, B. (2017) 'Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques', *Expert review of proteomics*, 14(5), pp. 419–429. doi:10.1080/14789450.2017.1322904.

Meyer-Baese, A. *et al.* (2014) 'Data analysis techniques in phosphoproteomics', *ELECTROPHORESIS*, 35(24), pp. 3452–3462. doi:10.1002/elps.201400219.

Mi, L.-Z. *et al.* (2011) 'Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor', *Nature Structural & Molecular Biology*, 18(9), pp. 984–989. doi:10.1038/nsmb.2092.

Miaczynska, M. (2013) 'Effects of Membrane Trafficking on Signaling by Receptor Tyrosine Kinases', *Cold Spring Harbor Perspectives in Biology*, 5(11). doi:10.1101/cshperspect.a009035.

Miettinen, P.J. *et al.* (1995) 'Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor', *Nature*, 376(6538), pp. 337–341. doi:10.1038/376337a0.

van der Mijn, J.C. *et al.* (2015) 'Evaluation of different phospho-tyrosine antibodies for label-free phosphoproteomics', *Journal of Proteomics*, 127, pp. 259–263. doi:10.1016/j.jprot.2015.04.006.

Miki, T. *et al.* (1992) 'Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene.', *Proceedings of the National Academy of Sciences of the United States of America*, 89(1), pp. 246–50. doi:10.1073/pnas.89.1.246.

Miller, D.L. *et al.* (2000) 'Compensation by Fibroblast Growth Factor 1 (FGF1) Does Not Account for the Mild Phenotypic Defects Observed in FGF2 Null Mice', *Molecular and Cellular Biology*, 20(6), pp. 2260–2268. doi:10.1128/MCB.20.6.2260-2268.2000.

Miller, K. *et al.* (1986) 'Localization of the epidermal growth factor (EGF) receptor within the endosome of EGF-stimulated epidermoid carcinoma (A431) cells', *The Journal of Cell Biology*, 102(2), pp. 500–509. doi:10.1083/jcb.102.2.500.

Miller, M.A., Sullivan, R.J. and Lauffenburger, D.A. (2017) 'Molecular Pathways: Receptor Ectodomain Shedding in Treatment, Resistance, and Monitoring of Cancer', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 23(3), pp. 623–629. doi:10.1158/1078-0432.CCR-16-0869.

Miller, M.L. *et al.* (2008) 'Linear motif atlas for phosphorylation-dependent signaling', *Science Signaling*, 1(35), p. ra2. doi:10.1126/scisignal.1159433.

Min, K.-W., Lee, S.-H. and Baek, S.J. (2016) 'Moonlighting proteins in cancer', *Cancer Letters*, 370(1), pp. 108–116. doi:10.1016/j.canlet.2015.09.022.

Mingoti, S.A. and Lima, J.O. (2006) 'Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms', *European Journal of Operational Research*, 174(3), pp. 1742–1759. doi:10.1016/j.ejor.2005.03.039.

Mischnik, M. *et al.* (2016) 'IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data', *Bioinformatics*, 32(3), pp. 424–431. doi:10.1093/bioinformatics/btv699.

Mizukami, Y. *et al.* (1997) 'A novel mechanism of JNK1 activation. Nuclear translocation and activation of JNK1 during ischemia and reperfusion', *The Journal of Biological Chemistry*, 272(26), pp. 16657–16662. doi:10.1074/jbc.272.26.16657.

Moasser, M.M. (2007) 'The oncogene HER2; Its signaling and transforming functions and its role in human cancer pathogenesis', *Oncogene*, 26(45), pp. 6469–6487. doi:10.1038/sj.onc.1210477.

Mohammadi, M. *et al.* (1991) 'A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Flg) is a binding site for the SH2 domain of phospholipase C-gamma 1', *Molecular and Cellular Biology*, 11(10), pp. 5068–5078. doi:10.1128/mcb.11.10.5068.

Mukherji, M. *et al.* (2006) 'A phosphoproteomic analysis of the ErbB2 receptor tyrosine kinase signaling pathways', *Biochemistry*, 45(51), pp. 15529–15540. doi:10.1021/bi060971c.

Müller, A.-K., Meyer, M. and Werner, S. (2012) 'The roles of receptor tyrosine kinases and their ligands in the wound repair process', *Seminars in Cell & Developmental Biology*, 23(9), pp. 963–970. doi:10.1016/j.semcdb.2012.09.015.

Mulvey, C.M. *et al.* (2017) 'Using hyperLOPIT to perform high-resolution mapping of the spatial proteome', *Nature Protocols*, 12(6), pp. 1110–1135. doi:10.1038/nprot.2017.026.

Munk, S., Refsgaard, J.C. and Olsen, J.V. (2016) 'Systems Analysis for Interpretation of Phosphoproteomics Data', in von Stechow, L. (ed.) *Phospho-Proteomics: Methods and Protocols*. New York, NY: Springer New York, pp. 341–360. doi:10.1007/978-1-4939-3049-4_23.

Murphy, R.F., Powers, S. and Cantor, C.R. (1984) 'Endosome pH measured in single cells by dual fluorescence flow cytometry: rapid acidification of insulin to pH 6.', *Journal of Cell Biology*, 98(5), pp. 1757–1762. doi:10.1083/jcb.98.5.1757.

Myllymäki, S.-M. *et al.* (2021) 'Proximity-Dependent Biotinylation (BioID) of Integrin Interaction Partners', *Methods in Molecular Biology (Clifton, N.J.)*, 2217, pp. 57–69. doi:10.1007/978-1-0716-0962-0_6.

Mylona, A. *et al.* (2016) 'Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation', *Science*, 354(6309), pp. 233–237. doi:10.1126/science.aad1872.

Nadratowska-Wesolowska, B. *et al.* (2014) 'RSK2 regulates endocytosis of FGF receptor 1 by phosphorylation on serine 789', *Oncogene*, 33(40), pp. 4823–4836. doi:10.1038/onc.2013.425.

Naeem, H. *et al.* (2012) 'Rigorous assessment of gene set enrichment tests', *Bioinformatics*, 28(11), pp. 1480–1486. doi:10.1093/bioinformatics/bts164.

Nag, S. *et al.* (2018) 'Rab4A organizes endosomal domains for sorting cargo to lysosome-related organelles', *Journal of Cell Science*, 131(18). doi:10.1242/jcs.216226.

Naguib, A. *et al.* (2015) 'PTEN functions by recruitment to cytoplasmic vesicles', *Molecular Cell*, 58(2), pp. 255–268. doi:10.1016/j.molcel.2015.03.011.

Nakakuki, T. *et al.* (2010) 'Ligand-Specific c-Fos Expression Emerges from the Spatiotemporal Control of ErbB Network Dynamics', *Cell*, 141(5), pp. 884–896. doi:10.1016/j.cell.2010.03.054.

Nakamura, Y. and Fukami, K. (2017) 'Regulation and physiological functions of mammalian phospholipase C', *The Journal of Biochemistry*, 161(4), pp. 315–321. doi:10.1093/jb/mvw094.

Nakayama, M. *et al.* (2013) 'Spatial regulation of VEGF receptor endocytosis in angiogenesis', *Nature Cell Biology*, 15(3), pp. 249–260. doi:10.1038/ncb2679.

Naldi, A. *et al.* (2017) 'Reconstruction and signal propagation analysis of the Syk signaling network in breast cancer cells', *PLOS Computational Biology*. Edited by L. You, 13(3), p. e1005432. doi:10.1371/journal.pcbi.1005432.

Nanni, M., Ranieri, D., Raffa, S., *et al.* (2018) 'Interplay between FGFR2b-induced autophagy and phagocytosis: role of PLCγ-mediated signalling', *Journal of Cellular and Molecular Medicine*, 22(1), pp. 668–683. doi:10.1111/jcmm.13352.

Nanni, M., Ranieri, D., Rosato, B., *et al.* (2018) 'Role of Fibroblast Growth Factor Receptor 2b in the Cross Talk between Autophagy and Differentiation: Involvement of Jun N-Terminal Protein Kinase Signaling', *Molecular and Cellular Biology*, 38(13), pp. e00119-18. doi:10.1128/MCB.00119-18.

Naro, C. and Sette, C. (2013) 'Phosphorylation-mediated regulation of alternative splicing in cancer', *International Journal of Cell Biology*, 2013. doi:10.1155/2013/151839.

Narushima, Y. *et al.* (2016) 'Quantitative phosphoproteomics-based molecular network description for high-resolution kinase-substrate interactome analysis', *Bioinformatics*, 32(14), pp. 2083–2088. doi:10.1093/bioinformatics/btw164.

Naslavsky, N. and Caplan, S. (2018) 'The enigmatic endosome - Sorting the ins and outs of endocytic trafficking', *Journal of Cell Science*, 131(13). doi:10.1242/jcs.216499.

Needham, E.J. *et al.* (2019) 'Illuminating the dark phosphoproteome', *Science Signaling*, 12(565), p. eaau8645. doi:10.1126/scisignal.aau8645.

Neukamm, S.S. *et al.* (2012) 'Identification of the Amino acids 300-600 of IRS-2 as 14-3-3 binding region with the importance of IGF-1/insulin-regulated phosphorylation of Ser-573', *PLoS ONE*, 7(8). doi:10.1371/journal.pone.0043296.

Nguyen, L.H. and Holmes, S. (2019) 'Ten quick tips for effective dimensionality reduction', *PLOS Computational Biology*. Edited by F. Ouellette, 15(6), p. e1006907. doi:10.1371/journal.pcbi.1006907.

Nightingale, D.J. *et al.* (2019) 'The subcellular organisation of Saccharomyces cerevisiae', *Current Opinion in Chemical Biology*, 48, pp. 86–95. doi:10.1016/j.cbpa.2018.10.026.

Nojima, H. *et al.* (2003) 'The mammalian target of rapamycin (mTOR) partner, raptor, binds the mTOR substrates p70 S6 kinase and 4E-BP1 through their TOR signaling (TOS) motif', *The Journal of Biological Chemistry*, 278(18), pp. 15461–15464. doi:10.1074/jbc.C200665200.

Nola, S. *et al.* (2008) 'Scrib regulates PAK activity during the cell migration process', *Human Molecular Genetics*, 17(22), pp. 3552–3565. doi:10.1093/hmg/ddn248.

Norris, A. and Grant, B.D. (2020) 'Endosomal microdomains: Formation and function', *Current Opinion in Cell Biology*, 65, pp. 86–95. doi:https://doi.org/10.1016/j.ceb.2020.02.018.

Nowosad, A. *et al.* (2020) 'p27 controls Ragulator and mTOR activity in amino acid-deprived cells to regulate the autophagy-lysosomal pathway and coordinate cell cycle and cell growth', *Nature Cell Biology*, 22(9), pp. 1076–1090. doi:10.1038/s41556-020-0554-4.

Oakley, F.D. *et al.* (2009) 'Signaling Components of Redox Active Endosomes: The Redoxosomes', *Antioxidants & Redox Signaling*, 11(6), pp. 1313–1333. doi:10.1089/ars.2008.2363.

Oberg, A.L. and Mahoney, D.W. (2012) 'Statistical methods for quantitative mass spectrometry proteomic experiments with labeling.', *BMC bioinformatics*, 13 Suppl 1(Suppl 16), p. S7. doi:10.1186/1471-2105-13-S16-S7.

Ochoa, D. *et al.* (2019) 'The functional landscape of the human phosphoproteome', *Nature Biotechnology*, 38(3), pp. 365–373. doi:10.1038/s41587-019-0344-3.

Oda, K. *et al.* (2005) 'A comprehensive pathway map of epidermal growth factor receptor signaling', *Molecular Systems Biology*, 1(1), p. 2005.0010. doi:10.1038/msb4100014.

Odle, R.I. *et al.* (2020) 'An mTORC1-to-CDK1 Switch Maintains Autophagy Suppression during Mitosis', *Molecular Cell*, 77(2), pp. 228-240.e7. doi:10.1016/j.molcel.2019.10.016.

Ohbayashi, N. *et al.* (2002) 'FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis', *Genes & Development*, 16(7), pp. 870–879. doi:10.1101/gad.965702.

Ohuchi, H. *et al.* (2000) 'FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development', *Biochemical and Biophysical Research Communications*, 277(3), pp. 643–649. doi:10.1006/bbrc.2000.3721.

Olsen, J.V. *et al.* (2006) 'Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.', *Cell*, 127(3), pp. 635–48. doi:10.1016/j.cell.2006.09.026.

Olsen, J.V. *et al.* (2010) 'Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis', *Science Signaling*, 3(104), p. ra3.

Olsen, S.K. *et al.* (2004) 'Insights into the molecular basis for fibroblast growth factor receptor autoinhibition and ligand-binding promiscuity.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(4), pp. 935–40. doi:10.1073/pnas.0307287101.

Ong, S.-E. *et al.* (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. doi:10.1074/MCP.M200025-MCP200.

Ong, S.-E. and Mann, M. (2007) 'A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)', *Nature Protocols*, 1(6), pp. 2650–2660. doi:10.1038/nprot.2006.427.

Ong, S.H. *et al.* (2000) 'FRS2 Proteins Recruit Intracellular Signaling Pathways by Binding to Diverse Targets on Fibroblast Growth Factor and Nerve Growth Factor Receptors', *Molecular and Cellular Biology*, 20(3), pp. 979–989. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC85215/ (Accessed: 15 March 2021).

Orchard, S. *et al.* (2014) 'The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases', *Nucleic acids research*, 42(Database issue), pp. D358-63. doi:10.1093/nar/gkt1115.

Ornitz, D.M. *et al.* (1996) 'Receptor specificity of the fibroblast growth factor family.', *The Journal of biological chemistry*, 271(25), pp. 15292–7. doi:10.1074/jbc.271.25.15292.

Ornitz, D.M. and Itoh, N. (2015) 'The Fibroblast Growth Factor signaling pathway.', *Wiley interdisciplinary reviews. Developmental biology*, 4(3), pp. 215–66. doi:10.1002/wdev.176.

Orre, L.M. *et al.* (2019) 'SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization', *Molecular Cell*, 73(1), pp. 166-182.e7. doi:10.1016/j.molcel.2018.11.035.

O'Sullivan, M.J. and Lindsay, A.J. (2020) 'The Endosomal Recycling Pathway—At the Crossroads of the Cell', *International Journal of Molecular Sciences*, 21(17), p. 6074. doi:10.3390/ijms21176074.

Padul, V. *et al.* (2015) 'ETV/Pea3 family transcription factor-encoding genes are overexpressed in CIC-mutant oligodendrogliomas', *Genes, Chromosomes and Cancer*, 54(12), pp. 725–733. doi:10.1002/gcc.22283.

Paek, J. *et al.* (2017) 'Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling', *Cell*, 169(2), pp. 338-349.e11. doi:10.1016/j.cell.2017.03.028.

Palmeri, A. *et al.* (2014) 'A proteome-wide domain-centric perspective on protein phosphorylation', *Molecular and Cellular Proteomics*, 13(9), pp. 2198–2212. doi:10.1074/mcp.M114.039990.

Pang, C.N.I. and Wilkins, M.R. (2019) 'Protein Post-Translational Modification Prediction', in Ranganathan, S. et al. (eds) *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press, pp. 15–27. doi:10.1016/B978-0-12-809633-8.20265-3.

Patino-Lopez, G. *et al.* (2008) 'Rab35 and its GAP EPI64C in T cells regulate receptor recycling and immunological synapse formation', *The Journal of Biological Chemistry*, 283(26), pp. 18323–18330. doi:10.1074/jbc.M800056200.

Pawson, T. (2004) 'Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems', *Cell*, 116(2), pp. 191–203. doi:10.1016/S0092-8674(03)01077-8.

Pellegrini, L. *et al.* (2000) 'Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin', *Nature*, 407(6807), pp. 1029–1034. doi:10.1038/35039551.

Pennock, S. and Wang, Z. (2008) 'A Tale of Two Cbls: Interplay of c-Cbl and Cbl-b in Epidermal Growth Factor Receptor Downregulation', *Molecular and Cellular Biology*, 28(9), pp. 3020–3037. doi:10.1128/mcb.01809-07.

Peters, K.G. *et al.* (1992) 'Point mutation of an FGF receptor abolishes phosphatidylinositol turnover and Ca 2+ flux but not mitogenesis', *Nature*, 358(6388), pp. 678–681. doi:10.1038/358678a0.

Peterson, T.R. *et al.* (2009) 'DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival', *Cell*, 137(5), pp. 873–886. doi:10.1016/j.cell.2009.03.046.

Petsalaki, E. *et al.* (2015) 'SELPHI: correlation-based identification of kinase-associated networks from global phospho-proteomics data sets', *Nucleic Acids Research*, 43(W1), pp. W276–W282. doi:10.1093/nar/gkv459.

Phillips, B.P., Gomez-Navarro, N. and Miller, E.A. (2020) 'Protein quality control in the endoplasmic reticulum', *Current Opinion in Cell Biology*, 65, pp. 96–102. doi:10.1016/j.ceb.2020.04.002.

Pinho, J.P.C. *et al.* (2017) 'Spike-In SILAC Approach for Proteomic Analysis of Ex Vivo Microglia', *Methods in Molecular Biology (Clifton, N.J.)*, 1598, pp. 295–312. doi:10.1007/978-1-4939-6952-4_15.

Plotnikov, A.N. *et al.* (2000) 'Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity', *Cell*, 101(4), pp. 413–424. doi:10.1016/s0092-8674(00)80851-x.

Pope, S.D. and Medzhitov, R. (2018) 'Emerging Principles of Gene Expression Programs and Their Regulation', *Molecular Cell*, 71(3), pp. 389–397. doi:10.1016/j.molcel.2018.07.017.

Porębska, N. *et al.* (2018) 'Targeting Cellular Trafficking of Fibroblast Growth Factor Receptors as a Strategy for Selective Cancer Treatment', *Journal of Clinical Medicine*, 8(1), p. E7. doi:10.3390/jcm8010007.

Posewitz, M.C. and Tempst, P. (1999) 'Immobilized gallium(III) affinity chromatography of phosphopeptides', *Analytical Chemistry*, 71(14), pp. 2883–2892. doi:10.1021/ac981409y.

Prentzell, M.T. *et al.* (2021) 'G3BPs tether the TSC complex to lysosomes and suppress mTORC1 signaling', *Cell*, 0(0). doi:10.1016/j.cell.2020.12.024.

Prior, I.A. and Hancock, J.F. (2012) 'Ras trafficking, localization and compartmentalized signalling', *Seminars in Cell and Developmental Biology*, 23(2), pp. 145–153. doi:10.1016/j.semcdb.2011.09.002.

Protter, D.S.W. and Parker, R. (2016) 'Principles and Properties of Stress Granules', *Trends in Cell Biology*, 26(9), pp. 668–679. doi:10.1016/j.tcb.2016.05.004.

Pryor, P.R. and Luzio, J.P. (2009) 'Delivery of endocytosed membrane proteins to the lysosome', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(4), pp. 615–624. doi:10.1016/j.bbamcr.2008.12.022.

Qiao, D. *et al.* (2003) 'Heparan Sulfate Proteoglycans as Regulators of Fibroblast Growth Factor-2 Signaling in Brain Endothelial Cells: SPECIFIC ROLE FOR GLYPICAN-1 IN GLIOMA ANGIOGENESIS *', *Journal of Biological Chemistry*, 278(18), pp. 16045–16053. doi:10.1074/jbc.M211259200.

Qiao, J. *et al.* (1999) 'FGF-7 modulates ureteric bud growth and nephron number in the developing kidney', *Development (Cambridge, England)*, 126(3), pp. 547–554.

Qin, W. *et al.* (2021) 'Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery', *Nature Communications*, 12(1), p. 4980. doi:10.1038/s41467-021-25259-2.

Qiu, B. *et al.* (2019) 'Sprouty4 correlates with favorable prognosis in perihilar cholangiocarcinoma by blocking the FGFR-ERK signaling pathway and arresting the cell cycle', *EBioMedicine*, 50, pp. 166–177. doi:10.1016/j.ebiom.2019.11.021.

Rabanal-Ruiz, Y. *et al.* (2021) 'mTORC1 activity is supported by spatial association with focal adhesions', *Journal of Cell Biology*, 220(e202004010). doi:10.1083/jcb.202004010.

Ranieri, D. *et al.* (2018) 'Expression of the FGFR2c mesenchymal splicing variant in human keratinocytes inhibits differentiation and promotes invasion', *Molecular Carcinogenesis*, 57(2), pp. 272–283. doi:10.1002/mc.22754.

Rauniyar, N. and Yates, J.R. (2014) 'Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics', *Journal of Proteome Research*, 13(12), pp. 5293–5309. doi:10.1021/pr500880b.

Reddy, R.J. *et al.* (2016) 'Early signaling dynamics of the epidermal growth factor receptor', *Proceedings of the National Academy of Sciences of the United States of America*, 113(11), pp. 3114–3119. doi:10.1073/pnas.1521288113.

Redpath, G.M.I. *et al.* (2019) 'Flotillins promote T cell receptor sorting through a fast Rab5-Rab11 endocytic recycling axis', *Nature Communications*, 10(1), p. 4392. doi:10.1038/s41467-019-12352-w.

Redpath, G.M.I. *et al.* (2020) 'Membrane Heterogeneity Controls Cellular Endocytic Trafficking', *Frontiers in Cell and Developmental Biology*, 8. doi:10.3389/fcell.2020.00757.

Reid, S.E. *et al.* (2017) 'Tumor matrix stiffness promotes metastatic cancer cell interaction with the endothelium', *The EMBO Journal*, 36(16), pp. 2373–2389. doi:10.15252/embj.201694912.

Reilly, J. and Maher, P. (2001) 'Importin β–Mediated Nuclear Import of Fibroblast Growth Factor Receptor', *The Journal of Cell Biology*, 152(6), pp. 1307–1312. Available at: https://www.semanticscholar.org/paper/Importin-%CE%B2%E2%80%93Mediated-Nuclear-Import-of-Fibroblast-Reilly-Maher/f226dcdd876df94962fc98a59e74782b79757be2 (Accessed: 8 September 2021).

Reyland, M.E. (2009) 'Protein kinase C isoforms: Multi-functional regulators of cell life and death', *Frontiers in bioscience (Landmark edition)*, 14, pp. 2386–2399. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5204454/ (Accessed: 3 March 2021).

Rhee, H.-W. *et al.* (2013) 'Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging', *Science*, 339(6125), pp. 1328–1331. doi:10.1126/science.1230593.

Riley, N.M. and Coon, J.J. (2016) 'Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling', *Analytical Chemistry*, 88(1), pp. 74–94. doi:10.1021/acs.analchem.5b04123.

Rink, J. *et al.* (2005) 'Rab conversion as a mechanism of progression from early to late endosomes', *Cell*, 122(5), pp. 735–749. doi:10.1016/j.cell.2005.06.043.

Ritchie, M.E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), pp. e47–e47. doi:10.1093/nar/gkv007.

Rivera, K.D. *et al.* (2021) 'Automating UbiFast for High-throughput and Multiplexed Ubiquitin Enrichment', *Molecular & cellular proteomics: MCP*, 20, p. 100154. doi:10.1016/j.mcpro.2021.100154.

Rivera, V.M. *et al.* (1993) 'A growth factor-induced kinase phosphorylates the serum response factor at a site that regulates its DNA-binding activity', *Molecular and Cellular Biology*, 13(10), pp. 6260–6273. doi:10.1128/MCB.13.10.6260.

Roczniak-Ferguson, A. *et al.* (2012) 'The Transcription Factor TFEB Links mTORC1 Signaling to Transcriptional Control of Lysosome Homeostasis', *Science Signaling*, 5(228), pp. ra42–ra42. doi:10.1126/scisignal.2002790.

Roepstorff, K. *et al.* (2009) 'Differential effects of EGFR ligands on endocytic sorting of the receptor', *Traffic*, 10(8), pp. 1115–1127. doi:10.1111/j.1600-0854.2009.00943.x.

Roskoski, R. (2014) 'The ErbB/HER family of protein-tyrosine kinases and cancer', *Pharmacological Research*, 79, pp. 34–74. doi:https://doi.org/10.1016/j.phrs.2013.11.002.

Roskoski, R. (2020) 'Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update', *Pharmacological Research*, 152, p. 104609. doi:10.1016/j.phrs.2019.104609.

Ross, P.L. *et al.* (2004) 'Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents', *Molecular & cellular proteomics: MCP*, 3(12), pp. 1154–1169. doi:10.1074/mcp.M400129-MCP200.

Rosse, C. *et al.* (2010) 'PKC and the control of localized signal dynamics', *Nature Reviews Molecular Cell Biology*, 11(2), pp. 103–112. doi:10.1038/nrm2847.

Rousseau, A. and Bertolotti, A. (2018) 'Regulation of proteasome assembly and activity in health and disease', *Nature Reviews Molecular Cell Biology*, 19(11), pp. 697–712. doi:10.1038/s41580-018-0040-z.

Rousseeuw, P.J. (1987) 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics*, 20, pp. 53–65. doi:10.1016/0377-0427(87)90125-7.

Roux, K.J. *et al.* (2012) 'A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells', *The Journal of Cell Biology*, 196(6), pp. 801–810. doi:10.1083/jcb.201112098.

Roux, P.P. *et al.* (2004) 'Tumor-promoting phorbol esters and activated Ras inactivate the tuberous sclerosis tumor suppressor complex via p90 ribosomal S6 kinase', *Proceedings of the National Academy of Sciences*, 101(37), pp. 13489–13494. doi:10.1073/pnas.0405659101.

Roux, P.P. and Thibault, P. (2013) 'The Coming of Age of Phosphoproteomics—from Large Data Sets to Inference of Protein Functions', *Molecular & Cellular Proteomics*, 12(12), pp. 3453–3464. doi:10.1074/mcp.R113.032862.

Rubin, D.B. (1976) 'Inference and missing data', *Biometrika*, 63(3), pp. 581–592. doi:10.1093/biomet/63.3.581.

Rudolph, J.D. *et al.* (2016) 'Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks.', *Cell systems*, 3(6), pp. 585-593.e3. doi:10.1016/j.cels.2016.11.005.

Ruprecht, B. *et al.* (2017) 'Lapatinib resistance in breast cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis', *Cancer Research*, 77(8), pp. 1842–1853. doi:10.1158/0008-5472.CAN-16-2976.

Rush, J. *et al.* (2005) 'Immunoaffinity profiling of tyrosine phosphorylation in cancer cells', *Nature Biotechnology*, 23(1), pp. 94–101. doi:10.1038/nbt1046.

Ruvinsky, I. and Meyuhas, O. (2006) 'Ribosomal protein S6 phosphorylation: from protein synthesis to cell size', *Trends in Biochemical Sciences*, 31(6), pp. 342–348. doi:10.1016/j.tibs.2006.04.003.

Sabapathy, K. *et al.* (2004) 'Distinct roles for JNK1 and JNK2 in regulating JNK activity and c-Jun-dependent cell proliferation', *Molecular Cell*, 15(5), pp. 713–725. doi:10.1016/j.molcel.2004.08.028.

Sacco, F., Perfetto, L. and Cesareni, G. (2018) 'Combining Phosphoproteomics Datasets and Literature Information to Reveal the Functional Connections in a Cell Phosphorylation Network', *PROTEOMICS*, 18(5–6), p. 1700311. doi:10.1002/pmic.201700311.

Saffi, G.T. and Botelho, R.J. (2019) 'Lysosome Fission: Planning for an Exit', *Trends in Cell Biology*, 29(8), pp. 635–646. doi:10.1016/j.tcb.2019.05.003.

Sahin, U. *et al.* (2004) 'Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands', *The Journal of Cell Biology*, 164(5), pp. 769–779. doi:10.1083/jcb.200307137.

Samavarchi-Tehrani, P., Samson, R. and Gingras, A.-C. (2020) 'Proximity dependent biotinylation: key enzymes and adaptation to proteomics approaches', *Molecular & Cellular Proteomics*, 19(5), p. mcp.R120.001941. doi:10.1074/mcp.r120.001941.

Sampattavanich, S. *et al.* (2018) 'Encoding Growth Factor Identity in the Temporal Dynamics of FOXO3 under the Combinatorial Control of ERK and AKT Kinases', *Cell Systems*, 6(6), pp. 664-678.e9. doi:10.1016/j.cels.2018.05.004.

Sancak, Y. *et al.* (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.

Sandilands, E. *et al.* (2007) 'Src kinase modulates the activation, transport and signalling dynamics of fibroblast growth factor receptors', *EMBO reports*, 8(12), pp. 1162–1169. doi:10.1038/sj.embor.7401097.

Santra, T. *et al.* (2019) 'An Integrated Global Analysis of Compartmentalized HRAS Signaling.', *Cell reports*, 26(11), pp. 3100-3115.e7. doi:10.1016/j.celrep.2019.02.038.

Sarabipour, S. and Hristova, K. (2016) 'Mechanism of FGF receptor dimerization and activation', *Nature Communications*, 7(1), p. 10262. doi:10.1038/ncomms10262.

Sarbassov, D.D. *et al.* (2004) 'Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton', *Current biology: CB*, 14(14), pp. 1296–1302. doi:10.1016/j.cub.2004.06.054.

Sarbassov, D.D. *et al.* (2005) 'Phosphorylation and Regulation of Akt/PKB by the Rictor-mTOR Complex', *Science*, 307(5712), pp. 1098–1101. doi:10.1126/science.1106148.

Sato, K. *et al.* (2003) 'Reconstitution of Src-dependent Phospholipase Cγ Phosphorylation and Transient Calcium Release by Using Membrane Rafts and Cell-free Extracts from Xenopus Eggs*', *Journal of Biological Chemistry*, 278(40), pp. 38413–38420. doi:10.1074/jbc.M302617200.

Sato, M. *et al.* (2003) 'Production of PtdInsP 3 at endomembranes is triggered by receptor endocytosis', *Nature Cell Biology*, 5(11), pp. 1016–1022. doi:10.1038/ncb1054.

Savini, M., Zhao, Q. and Wang, M.C. (2019) 'Lysosome: Signaling Hub for Metabolic Sensing and Longevity', *Trends in cell biology*, 29(11), pp. 876–887. doi:10.1016/j.tcb.2019.08.008.

Saxton, R.A. and Sabatini, D.M. (2017) 'mTOR Signaling in Growth, Metabolism, and Disease', *Cell*, 168(6), pp. 960–976. doi:10.1016/j.cell.2017.02.004.

Schaaf, M.B.E. *et al.* (2016) 'LC3/GABARAP family proteins: autophagy-(un)related functions', *The FASEB Journal*, 30(12), pp. 3961–3978. doi:10.1096/fj.201600698R.

Schäfer, A. *et al.* (2019) 'Elucidating essential kinases of endothelin signalling by logic modelling of phosphoproteomics data', *Molecular Systems Biology*, 15(8), p. e8828. doi:10.15252/msb.20198828.

Scheltema, R.A. *et al.* (2014) 'The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer', *Molecular & Cellular Proteomics : MCP*, 13(12), pp. 3698–3708. doi:10.1074/mcp.M114.043489.

Schindler, C. *et al.* (2015) 'EARP is a multisubunit tethering complex involved in endocytic recycling', *Nature Cell Biology*, 17(5), pp. 639–650. doi:10.1038/ncb3129.

Schlessinger, J. (2000) 'Cell Signaling by Receptor Tyrosine Kinases', *Cell*, 103(2), pp. 211–225. doi:10.1016/S0092-8674(00)00114-8.

Schlessinger, J. and Lemmon, M.A. (2003) 'SH2 and PTB Domains in Tyrosine Kinase Signaling', *Science's STKE*, 2003(191), pp. re12–re12. doi:10.1126/stke.2003.191.re12.

Schmid, S.L. (2017) 'Reciprocal regulation of signaling and endocytosis: Implications for the evolving cancer cell', *The Journal of Cell Biology*, 216(9), pp. 2623–2632. doi:10.1083/jcb.201705017.

Schmidt, K.A., Dall, S.R.X. and Gils, J.A.V. (2010) 'The ecology of information: an overview on the ecological significance of making informed decisions', *Oikos*, 119(2), pp. 304–316. doi:10.1111/j.1600-0706.2009.17573.x.

Schmidt-Glenewinkel, H. *et al.* (2008) 'An ultrasensitive sorting mechanism for EGF Receptor Endocytosis', *BMC Systems Biology*, 2(1), p. 32. doi:10.1186/1752-0509-2-32.

Schneider, M.R. and Wolf, E. (2009) 'The epidermal growth factor receptor ligands at a glance', *Journal of Cellular Physiology*, 218(3), pp. 460–466. doi:10.1002/jcp.21635.

Schuler, M. (2020) 'Inhibiting fibroblast growth factor receptors in cancer – the next generation', *Annals of Oncology*, 31(10), pp. 1285–1286. doi:10.1016/j.annonc.2020.07.007.

Schwämmle, V. and Jensen, O.N. (2010) 'A simple and fast method to determine the parameters for fuzzy c–means cluster analysis', *Bioinformatics*, 26(22), pp. 2841–2848. doi:10.1093/bioinformatics/btq534.

Schwämmle, V., León, I.R. and Jensen, O.N. (2013) 'Assessment and Improvement of Statistical Tools for Comparative Proteomics Analysis of Sparse Data Sets with Few Experimental Replicates', *Journal of Proteome Research*, 12(9), pp. 3874–3883. doi:10.1021/pr400045u.

Schwämmle, V., Verano-Braga, T. and Roepstorff, P. (2015) 'Computational and statistical methods for high-throughput analysis of post-translational modifications of proteins', *Journal of Proteomics*, 129, pp. 3–15. doi:10.1016/j.jprot.2015.07.016.

Scott, C.C., Vacca, F. and Gruenberg, J. (2014) 'Endosome maturation, transport and functions', *Seminars in Cell & Developmental Biology*, 31, pp. 2–10. doi:10.1016/j.semcdb.2014.03.034.

Semenova, I. *et al.* (2014) 'Regulation of microtubule-based transport by MAP4', *Molecular Biology of the Cell*, 25(20), pp. 3119–3132. doi:10.1091/mbc.E14-01-0022.

Settembre, C. *et al.* (2012) 'A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB', *The EMBO journal*, 31(5), pp. 1095–1108. doi:10.1038/emboj.2012.32.

Sgouras, D.N. *et al.* (1995) 'ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress ets-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation.', *The EMBO Journal*, 14(19), pp. 4781–4793. doi:10.1002/j.1460-2075.1995.tb00160.x.

Shan, Y. *et al.* (2012) 'Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization', *Cell*, 149(4), pp. 860–870. doi:10.1016/j.cell.2012.02.063.

Sharma, K. *et al.* (2014) 'Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling', *Cell Reports*, 8(5), pp. 1583–1594. doi:10.1016/j.celrep.2014.07.036.

Shaulian, E. and Karin, M. (2001) 'AP-1 in cell proliferation and survival', *Oncogene*, 20(19), pp. 2390–2400. doi:10.1038/sj.onc.1204383.

Shaw, J.D. *et al.* (2003) 'PtdIns(3,5)P2 is required for delivery of endocytic cargo into the multivesicular body', *Traffic (Copenhagen, Denmark)*, 4(7), pp. 479–490. doi:10.1034/j.1600-0854.2003.t01-1-00106.x.

Shee, K. *et al.* (2018) 'Therapeutically targeting tumor microenvironment-mediated drug resistance in estrogen receptor-positive breast cancer', *The Journal of Experimental Medicine*, 215(3), pp. 895–910. doi:10.1084/jem.20171818.

Shevchenko, A. *et al.* (2006) 'In-gel digestion for mass spectrometric characterization of proteins and proteomes', *Nature Protocols*, 1(6), pp. 2856–2860. doi:10.1038/nprot.2006.468.

Shi, T. *et al.* (2015) 'Sensitive targeted quantification of ERK phosphorylation dynamics and stoichiometry in human cells without affinity enrichment.', *Analytical chemistry*, 87(2), pp. 1103–10. doi:10.1021/ac503797x.

Shieh, G.S. (1998) 'A weighted Kendall's tau statistic', *Statistics & Probability Letters*, 39(1), pp. 17–24. doi:10.1016/S0167-7152(98)00006-6.

Shin, J.H. *et al.* (2018) 'NNMT depletion contributes to liver cancer cell survival by enhancing autophagy under nutrient starvation', *Oncogenesis*, 7(8), p. 58. doi:10.1038/s41389-018-0064-4.

Shishkova, E., Hebert, A.S. and Coon, J.J. (2016) 'Now, more than ever, proteomics needs better chromatography', *Cell systems*, 3(4), pp. 321–324. doi:10.1016/j.cels.2016.10.007.

Sibilia, M. *et al.* (1998) 'A strain-independent postnatal neurodegeneration in mice lacking the EGF receptor', *The EMBO journal*, 17(3), pp. 719–731. doi:10.1093/emboj/17.3.719.

Sidoli, S., Kulej, K. and Garcia, B.A. (2017) 'Why proteomics is not the new genomics and the future of mass spectrometry in cell biology.', *The Journal of cell biology*, 216(1), pp. 21–24. doi:10.1083/jcb.201612010.

Sigismund, S. *et al.* (2005) 'Clathrin-independent endocytosis of ubiquitinated cargos', *Proceedings of the National Academy of Sciences*, 102(8), pp. 2760–2765. doi:10.1073/pnas.0409817102.

Sigismund, S. *et al.* (2008) 'Clathrin-Mediated Internalization Is Essential for Sustained EGFR Signaling but Dispensable for Degradation', *Developmental Cell*, 15(2), pp. 209–219. doi:10.1016/J.DEVCEL.2008.06.012.

Sigismund, S. *et al.* (2012) 'Endocytosis and Signaling: Cell Logistics Shape the Eukaryotic Cell Plan', *Physiological reviews*, 92(1), pp. 273–366. doi:10.1152/physrev.00005.2011.

Sigismund, S. *et al.* (2013) 'Threshold-controlled ubiquitination of the EGFR directs receptor fate', *The EMBO Journal*, 32(15), pp. 2140–2157. doi:10.1038/emboj.2013.149.

Sigismund, S. *et al.* (2021) 'Endocytosis in the context-dependent regulation of individual and collective cell properties', *Nature Reviews Molecular Cell Biology*, pp. 1–19. doi:10.1038/s41580-021-00375-5.

Simon, A.J. *et al.* (2014) 'Intrinsic disorder as a generalizable strategy for the rational design of highly responsive, allosterically cooperative receptors', *Proceedings of the National Academy of Sciences*, 111(42), pp. 15048–15053. doi:10.1073/pnas.1410796111.

Singh, B., Carpenter, G. and Coffey, R.J. (2016) 'EGF receptor ligands: recent advances', *F1000Research*, 5. doi:10.12688/f1000research.9025.1.

Smith, J.S. *et al.* (2018) 'Biased agonists of the chemokine receptor CXCR3 differentially control chemotaxis and inflammation', *Science Signaling*, 11(555), p. eaaq1075. doi:10.1126/scisignal.aaq1075.

Smith, M.P. *et al.* (2021) 'Reciprocal priming between receptor tyrosine kinases at recycling endosomes orchestrates cellular signalling outputs', *The EMBO Journal*, 40, p. e107182. doi:10.15252/embj.2020107182.

Smyth, G.K. (2004) 'Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments', *Statistical Applications in Genetics and Molecular Biology*, 3(1). doi:10.2202/1544-6115.1027.

Solari, F.A. *et al.* (2015) 'Why phosphoproteomics is still a challenge', *Mol. BioSyst.*, 11(6), pp. 1487–1493. doi:10.1039/C5MB00024F.

Solinger, J.A. *et al.* (2020) 'FERARI is required for Rab11-dependent endocytic recycling', *Nature Cell Biology*, 22(2), pp. 213–224. doi:10.1038/s41556-019-0456-5.

Sönnichsen, B. *et al.* (2000) 'Distinct Membrane Domains on Endosomes in the Recycling Pathway Visualized by Multicolor Imaging of Rab4, Rab5, and Rab11', *Journal of Cell Biology*, 149(4), pp. 901–914. doi:10.1083/jcb.149.4.901.

Sørensen, V. *et al.* (2006) 'Different abilities of the four FGFRs to mediate FGF-1 translocation are linked to differences in the receptor C-terminal tail', *Journal of Cell Science*, 119(20), pp. 4332–4341. doi:10.1242/JCS.03209.

Sørensen, V. *et al.* (2008) 'Phosphorylation of Fibroblast Growth Factor (FGF) Receptor 1 at Ser777 by p38 Mitogen-Activated Protein Kinase Regulates Translocation of Exogenous FGF1 to the Cytosol and Nucleus', *Molecular and Cellular Biology*, 28(12), pp. 4129–4141. doi:10.1128/MCB.02117-07.

Sorkin, A. and Carpenter, G. (1993) 'Interaction of activated EGF receptors with coated pit adaptins', *Science*, 261(5121), pp. 612–615. doi:10.1126/science.8342026.

Sorkin, A. and Goh, L.K. (2008) 'Endocytosis and intracellular trafficking of ErbBs', *Experimental cell research*, 314(17), pp. 3093–3106. doi:10.1016/j.yexcr.2008.07.029.

Sorokin, A. *et al.* (1994) 'Internalization of fibroblast growth factor receptor is inhibited by a point mutation at tyrosine 766', *The Journal of Biological Chemistry*, 269(25), pp. 17056–17061.

Soto, A.M. and Sonnenschein, C. (2011) 'The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory', *BioEssays*, 33(5), pp. 332–340. doi:10.1002/bies.201100025.

Soul, J. *et al.* (2015) 'PhenomeExpress: A refined network analysis of expression datasets by inclusion of known disease phenotypes', *Scientific Reports*, 5(1), p. 8117. doi:10.1038/srep08117.

Sousa, L.P. *et al.* (2012) 'Suppression of EGFR endocytosis by dynamin depletion reveals that EGFR signaling occurs primarily at the plasma membrane', *Proceedings of the National Academy of Sciences*, 109(12), pp. 4419–4424. doi:10.1073/pnas.1200164109.

Sowa, M.E. *et al.* (2009) 'Defining the Human Deubiquitinating Enzyme Interaction Landscape', *Cell*, 138(2), pp. 389–403. doi:10.1016/j.cell.2009.04.042.

Sparrow, L.G. *et al.* (1997) 'The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain', *The Journal of Biological Chemistry*, 272(47), pp. 29460–29467. doi:10.1074/jbc.272.47.29460.

Sparta, B. *et al.* (2015) 'Receptor Level Mechanisms Are Required for Epidermal Growth Factor (EGF)-stimulated Extracellular Signal-regulated Kinase (ERK) Activity Pulses', *Journal of Biological Chemistry*, 290(41), pp. 24784–24792. doi:10.1074/jbc.M115.662247.

Stachowiak, J.C. *et al.* (2012) 'Membrane bending by protein–protein crowding', *Nature Cell Biology*, 14(9), pp. 944–949. doi:10.1038/ncb2561.

Stanoev, A. *et al.* (2018) 'Interdependence between EGFR and Phosphatases Spatially Established by Vesicular Dynamics Generates a Growth Factor Sensing and Responding Network', *Cell Systems*, 7(3), pp. 295-309.e11. doi:10.1016/j.cels.2018.06.006.

Stasyk, T. and Huber, L.A. (2016) 'Spatio-Temporal Parameters of Endosomal Signaling in Cancer: Implications for New Treatment Options', *Journal of Cellular Biochemistry*, 117(4), pp. 836–843. doi:10.1002/jcb.25418.

Steen, H. and Mann, M. (2004) 'The abc's (and xyz's) of peptide sequencing', *Nature Reviews Molecular Cell Biology*, 5(9), pp. 699–711. doi:10.1038/nrm1468.

Stegmeier, F. *et al.* (2007) 'Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities', *Nature*, 446(7138), pp. 876–881. doi:10.1038/nature05694.

Stenmark, H. (2009) 'Rab GTPases as coordinators of vesicle traffic', *Nature Reviews Molecular Cell Biology*, 10(8), pp. 513–525. doi:10.1038/nrm2728.

Stern, K.A., Place, T.L. and Lill, N.L. (2008) 'EGF and amphiregulin differentially regulate Cbl recruitment to endosomes and EGF receptor fate', *The Biochemical Journal*, 410(3), pp. 585–594. doi:10.1042/BJ20071505.

Stewart, E.L. *et al.* (2015) 'Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations-a review', *Translational Lung Cancer Research*, 4(1), pp. 67–81. doi:10.3978/j.issn.2218-6751.2014.11.06.

Stoney, R.A. *et al.* (2015) 'Disentangling the multigenic and pleiotropic nature of molecular function.', *BMC systems biology*, 9 Suppl 6(Suppl 6), p. S3. doi:10.1186/1752-0509-9-S6-S3.

Stramucci, L., Pranteda, A. and Bossi, G. (2018) 'Insights of Crosstalk between p53 Protein and the MKK3/MKK6/p38 MAPK Signaling Pathway in Cancer', *Cancers*, 10(5), p. 131. doi:10.3390/cancers10050131.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–50. doi:10.1073/pnas.0506580102.

Sugiyama, M.G., Fairn, G.D. and Antonescu, C.N. (2019) 'Akt-ing Up Just About Everywhere: Compartment-Specific Akt Activation and Function in Receptor Tyrosine Kinase Signaling', *Frontiers in Cell and Developmental Biology*, 7. doi:10.3389/fcell.2019.00070.

Sugiyama, N., Imamura, H. and Ishihama, Y. (2019) 'Large-scale Discovery of Substrates of the Human Kinome', *Scientific Reports*, 9. doi:10.1038/s41598-019-46385-4.

Sun, C. *et al.* (2016) 'Selectivity in glycosaminoglycan binding dictates the distribution and diffusion of fibroblast growth factors in the pericellular matrix', *Open Biology*, 6(3), p. 150277. doi:10.1098/rsob.150277.

Sun, Y., Goderie, S.K. and Temple, S. (2005) 'Asymmetric distribution of EGFR receptor during mitosis generates diverse CNS progenitor cells', *Neuron*, 45(6), pp. 873–886. doi:10.1016/j.neuron.2005.01.045.

Swaney, D.L., Wenger, C.D. and Coon, J.J. (2010) 'Value of using multiple proteases for large-scale mass spectrometry-based proteomics', *Journal of Proteome Research*, 9(3), pp. 1323–1329. doi:10.1021/pr900863u.

Szklarczyk, D. *et al.* (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.

Szybowska, P. *et al.* (2021) 'Negative Regulation of FGFR (Fibroblast Growth Factor Receptor) Signaling', *Cells*, 10(6), p. 1342. doi:10.3390/cells10061342.

Takahashi, S. *et al.* (2012) 'Rab11 regulates exocytosis of recycling vesicles at the plasma membrane', *Journal of Cell Science*, 125(Pt 17), pp. 4049–4057. doi:10.1242/jcs.102913.

Takeda, M. *et al.* (2007) 'AZD2171 shows potent antitumor activity against gastric cancer over-expressing fibroblast growth factor receptor 2/keratinocyte growth factor receptor', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 13(10), pp. 3051–3057. doi:10.1158/1078-0432.CCR-06-2743.

Tang, Q.-Y. and Kaneko, K. (2020) 'Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis', *PLOS Computational Biology*. Edited by B.L. de Groot, 16(2), p. e1007670. doi:10.1371/journal.pcbi.1007670.

Tang, Y. *et al.* (2019) 'DRUM: Inference of Disease-Associated m6A RNA Methylation Sites From a Multi-Layer Heterogeneous Network', *Frontiers in Genetics*, 10, p. 266. doi:10.3389/fgene.2019.00266.

Tanimura, S. *et al.* (2016) 'ERK signaling promotes cell motility by inducing the localization of myosin 1E to lamellipodial tips', *Journal of Cell Biology*, 214(4), pp. 475–489. doi:10.1083/jcb.201503123.

Tee, A.R. *et al.* (2003) 'Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb', *Current biology: CB*, 13(15), pp. 1259–1268. doi:10.1016/s0960-9822(03)00506-2.

Tefft, D. *et al.* (2002) 'mSprouty2 inhibits FGF10-activated MAP kinase by differentially binding to upstream target proteins', *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 283(4), pp. L700–L706. doi:10.1152/ajplung.00372.2001.

Teis, D., Wunderlich, W. and Huber, L.A. (2002) 'Localization of the MP1-MAPK Scaffold Complex to Endosomes Is Mediated by p14 and Required for Signal Transduction', *Developmental Cell*, 3(6), pp. 803–814. doi:10.1016/S1534-5807(02)00364-7.

Teo, G. *et al.* (2014) 'SAINTexpress: improvements and additional features in Significance Analysis of Interactome software', *Journal of proteomics*, 100, pp. 37–43. doi:10.1016/j.jprot.2013.10.023.

Teo, G. *et al.* (2016) 'SAINTq: Scoring protein-protein interactions in affinity purification - mass spectrometry experiments with fragment or peptide intensity data', *Proteomics*, 16(15–16), pp. 2238–2245. doi:10.1002/pmic.201500499.

Terfve, C. *et al.* (2012) 'CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms', *BMC Systems Biology*, 6, p. 133. doi:10.1186/1752-0509-6-133.

Terfve, C. *et al.* (2015) 'Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data', *Nature Communications*, 6, p. 8033. doi:10.1038/ncomms9033.

Thapa, N. *et al.* (2020) 'Phosphatidylinositol 3-kinase signalling is spatially organized at endosomal compartments by microtubule-associated protein 4', *Nature Cell Biology*, 22(11), pp. 1357–1370. doi:10.1038/s41556-020-00596-4.

The UniProt Consortium (2021) 'UniProt: the universal protein knowledgebase in 2021', *Nucleic Acids Research*, 49(D1), pp. D480–D489. doi:10.1093/nar/gkaa1100.

Thingholm, T.E. *et al.* (2006) 'Highly selective enrichment of phosphorylated peptides using titanium dioxide', *Nature Protocols*, 1(4), pp. 1929–1935. doi:10.1038/nprot.2006.185.

Thomas, P.D. (2017) 'The Gene Ontology and the meaning of biological function', *Methods in molecular biology (Clifton, N.J.)*, 1446, pp. 15–24. doi:10.1007/978-1-4939-3743-1_2.

Thompson, A. *et al.* (2003) 'Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS', *Analytical Chemistry*, 75(8), pp. 1895–1904. doi:10.1021/ac0262560.

Thoms, H.C., Dunlop, M.G. and Stark, L.A. (2007) 'p38-Mediated Inactivation of Cyclin D1/Cyclin-Dependent Kinase 4 Stimulates Nucleolar Translocation of RelA and Apoptosis in Colorectal Cancer Cells', *Cancer Research*, 67(4), pp. 1660–1669. doi:10.1158/0008-5472.CAN-06-1038.

Thoreen, C.C. *et al.* (2012) 'A unifying model for mTORC1-mediated regulation of mRNA translation', *Nature*, 485(7396), pp. 109–113. doi:10.1038/nature11083.

Thorndike, R.L. (1953) 'Who belongs in the family?', *Psychometrika*, 18(4), pp. 267–276. doi:10.1007/BF02289263.

Threadgill, D.W. *et al.* (1995) 'Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype', *Science (New York, N.Y.)*, 269(5221), pp. 230–234. doi:10.1126/science.7618084.

Timp, W. and Timp, G. (2020) 'Beyond mass spectrometry, the next step in proteomics', *Science Advances*, 6(2), p. eaax8978. doi:10.1126/sciadv.aax8978.

Ting, L. *et al.* (2009) 'Normalization and Statistical Analysis of Quantitative Proteomics Data Generated by Metabolic Labeling', *Molecular & Cellular Proteomics : MCP*, 8(10), pp. 2227–2242. doi:10.1074/mcp.M800462-MCP200.

Tomas, A., Futter, C.E. and Eden, E.R. (2014) 'EGF receptor trafficking: consequences for signaling and cancer', *Trends in Cell Biology*, 24(1), pp. 26–34. doi:10.1016/J.TCB.2013.11.002.

Tomczak, A. *et al.* (2018) 'Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations.', *Scientific reports*, 8(1), p. 5115. doi:10.1038/s41598-018-23395-2.

Tomlinson, V. *et al.* (2010) 'JNK phosphorylates Yes-associated protein (YAP) to regulate apoptosis', *Cell Death & Disease*, 1, p. e29. doi:10.1038/cddis.2010.7.

Torii, S. *et al.* (2004) 'Sef Is a Spatial Regulator for Ras/MAP Kinase Signaling', *Developmental Cell*, 7(1), pp. 33–44. doi:10.1016/j.devcel.2004.05.019.

Traer, C.J. *et al.* (2007) 'SNX4 coordinates endosomal sorting of TfnR with dynein-mediated transport into the endocytic recycling compartment', *Nature Cell Biology*, 9(12), pp. 1370–1380. doi:10.1038/ncb1656.

Traub, L.M. (2009) 'Tickets to ride: selecting cargo for clathrin-regulated internalization', *Nature Reviews. Molecular Cell Biology*, 10(9), pp. 583–596. doi:10.1038/nrm2751.

Trinkle-Mulcahy, L. (2019) 'Recent advances in proximity-based labeling methods for interactome mapping [version 1; referees: 2 approved]', *F1000Research*, 8. doi:10.12688/f1000research.16903.1.

Tsang, M. *et al.* (2002) 'Identification of Sef, a novel modulator of FGF signalling', *Nature Cell Biology*, 4(2), pp. 165–169. doi:10.1038/ncb749.

Tubbesing, K. *et al.* (2020) 'Complex Rab4-Mediated Regulation of Endosomal Size and EGFR Activation', *Molecular Cancer Research*, 18(5), pp. 757–773. doi:10.1158/1541-7786.MCR-19-0052.

Turner, N. and Grose, R. (2010) 'Fibroblast growth factor signalling: from development to cancer', *Nature Reviews Cancer*, 10(2), pp. 116–129. doi:10.1038/nrc2780.

Tusher, V.G., Tibshirani, R. and Chu, G. (2001) 'Significance analysis of microarrays applied to the ionizing radiation response', *Proceedings of the National Academy of Sciences*, 98(9), pp. 5116–5121.

Tyanova, S. *et al.* (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nature Methods*, 13(9), pp. 731–740. doi:10.1038/nmeth.3901.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', *Nature Protocols*, 11(12), pp. 2301–2319. doi:10.1038/nprot.2016.136.

Tzeng, H.-T. and Wang, Y.-C. (2016) 'Rab-mediated vesicle trafficking in cancer', *Journal of Biomedical Science*, 23(1), p. 70. doi:10.1186/s12929-016-0287-7.

Uchida, Y. *et al.* (2011) 'Intracellular phosphatidylserine is essential for retrograde membrane traffic through endosomes', *Proceedings of the National Academy of Sciences*, 108(38), p. 15846. doi:10.1073/pnas.1109101108.

Uchikawa, E. *et al.* (2019) 'Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor–ligand complex', *eLife*. Edited by A.T. Brunger, P.A. Cole, and N. Grigorieff, 8, p. e48630. doi:10.7554/eLife.48630.

Uhlitz, F. *et al.* (2017) 'An immediate–late gene expression module decodes ERK signal duration', *Molecular Systems Biology*, 13(5). doi:10.15252/msb.20177554.

Ullrich, O. *et al.* (1996) 'Rab11 regulates recycling through the pericentriolar recycling endosome', *The Journal of Cell Biology*, 135(4), pp. 913–924. doi:10.1083/jcb.135.4.913.

Ünal, E.B., Uhlitz, F. and Blüthgen, N. (2017) 'A compendium of ERK targets', *FEBS Letters*, 591(17), pp. 2607–2615. doi:https://doi.org/10.1002/1873-3468.12740.

Vaites, L.P. *et al.* (2017) 'Systematic Analysis of Human Cells Lacking ATG8 Proteins Uncovers Roles for GABARAPs and the CCZ1/MON1 Regulator C18orf8/RMC1 in Macroautophagic and Selective Autophagic Flux', *Molecular and Cellular Biology*, 38(1), pp. e00392-17. doi:10.1128/MCB.00392-17.

Vanhaesebroeck, B. *et al.* (2021) 'PI3K inhibitors are finally coming of age', *Nature Reviews. Drug Discovery* [Preprint]. doi:10.1038/s41573-021-00209-1.

Vanhaesebroeck, B., Stephens, L. and Hawkins, P. (2012) 'PI3K signalling: the path to discovery and understanding', *Nature Reviews Molecular Cell Biology*, 13(3), pp. 195–203. doi:10.1038/nrm3290.

Venerando, A., Cesaro, L. and Pinna, L.A. (2017) 'From phosphoproteins to phosphoproteomes: a historical account', *The FEBS Journal*, 284(13), pp. 1936–1951. doi:10.1111/febs.14014.

Verger, A. *et al.* (2001) 'Identification of Amino Acid Residues in the ETS Transcription Factor Erg That Mediate Erg-Jun/Fos-DNA Ternary Complex Formation *', *Journal of Biological Chemistry*, 276(20), pp. 17181–17189. doi:10.1074/jbc.M010208200.

Vidal, M., Cusick, M.E. and Barabási, A.-L. (2011) 'Interactome networks and human disease.', *Cell*, 144(6), pp. 986–98. doi:10.1016/j.cell.2011.02.016.

Vieira, A.V., Lamaze, C. and Schmid, S.L. (1996) 'Control of EGF Receptor Signaling by Clathrin-Mediated Endocytosis', *Science*, 274(5295), p. 2086. doi:10.1126/science.274.5295.2086.

Villaseñor, R. *et al.* (2015) 'Regulation of EGFR signal transduction by analogue-to-digital conversion in endosomes', *eLife*. Edited by S.R. Pfeffer, 4, p. e06156. doi:10.7554/eLife.06156.

Villaseñor, R., Kalaidzidis, Y. and Zerial, M. (2016) 'Signal processing by the endosomal system', *Current Opinion in Cell Biology*, 39, pp. 53–60. doi:10.1016/j.ceb.2016.02.002.

Vyse, S., Desmond, H. and Huang, P.H. (2017) 'Advances in mass spectrometry based strategies to study receptor tyrosine kinases.', *IUCrJ*, 4(Pt 2), pp. 119–130. doi:10.1107/S2052252516020546.

Wagner, E.F. and Nebreda, Á.R. (2009) 'Signal integration by JNK and p38 MAPK pathways in cancer development', *Nature Reviews Cancer*, 9(8), pp. 537–549. doi:10.1038/nrc2694.

Wan, Y.Y. *et al.* (2006) 'The kinase TAK1 integrates antigen and cytokine receptor signaling for T cell development, survival and function', *Nature Immunology*, 7(8), pp. 851–858. doi:10.1038/ni1355.

Wang, S.-J. *et al.* (2004) 'Differential effects of EGF gradient profiles on MDA-MB-231 breast cancer cell chemotaxis', *Experimental Cell Research*, 300(1), pp. 180–189. doi:10.1016/j.yexcr.2004.06.030.

Wang, W., Bian, J. and Li, Z. (2021) 'Internalized Activation of Membrane Receptors: From Phenomenon to Theory', *Trends in Cell Biology*, 31(6), pp. 428–431. doi:10.1016/j.tcb.2021.03.008.

Wang, Y. *et al.* (2002) 'Endosomal signaling of epidermal growth factor receptor stimulates signal transduction pathways leading to cell survival', *Molecular and Cellular Biology*, 22(20), pp. 7279–7290. doi:10.1128/MCB.22.20.7279-7290.2002.

Wang, Y., Wu, J. and Wang, Z. (2006) 'Akt binds to and phosphorylates phospholipase C-gamma1 in response to epidermal growth factor', *Molecular Biology of the Cell*, 17(5), pp. 2267–2277. doi:10.1091/mbc.e05-10-0918.

Wang, Y.-N. *et al.* (2012) 'Membrane-bound Trafficking Regulates Nuclear Transport of Integral Epidermal Growth Factor Receptor (EGFR) and ErbB-2', *Journal of Biological Chemistry*, 287(20), pp. 16869–16879. doi:10.1074/jbc.M111.314799.

Ward, M.D. and Leahy, D.J. (2015) 'Kinase Activator-Receiver Preference in ErbB Heterodimers Is Determined by Intracellular Regions and Is Not Coupled to Extracellular Asymmetry', *Journal of Biological Chemistry*, 290(3), pp. 1570–1579. doi:10.1074/jbc.M114.612085.

Watson, J. and Francavilla, C. (2018) 'Regulation of FGF10 Signaling in Development and Disease', *Frontiers in Genetics*, 9, p. 500. doi:10.3389/fgene.2018.00500.

Watson, J., Schwartz, J.-M. and Francavilla, C. (2021) 'Using Multilayer Heterogeneous Networks to Infer Functions of Phosphorylated Sites', *Journal of Proteome Research*, 20(7), pp. 3532–3548. doi:10.1021/acs.jproteome.1c00150.

Webb, A.E. and Brunet, A. (2014) 'FOXO transcription factors: key regulators of cellular quality control', *Trends in Biochemical Sciences*, 39(4), pp. 159–169. doi:10.1016/j.tibs.2014.02.003.

Webb-Robertson, B.-J.M. *et al.* (2015) 'Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics', *Journal of Proteome Research*, 14(5), pp. 1993–2001. doi:10.1021/pr501138h.

Wee, P. and Wang, Z. (2018) 'Regulation of EGFR Endocytosis by CBL During Mitosis', *Cells*, 7(12). doi:10.3390/cells7120257.

van Weert, A.W. *et al.* (2000) 'Primaquine interferes with membrane recycling from endosomes to the plasma membrane through a direct interaction with endosomes which does not involve neutralisation of endosomal pH nor osmotic swelling of endosomes', *European Journal of Cell Biology*, 79(6), pp. 394–399. doi:10.1078/0171-9335-00062.

Wei, X. *et al.* (2020) 'Detecting Gene Ontology misannotations using taxon-specific rate ratio comparisons', *Bioinformatics*, 36(16), pp. 4383–4388. doi:10.1093/bioinformatics/btaa548.

Wei, Y. *et al.* (2008) 'JNK1-mediated phosphorylation of Bcl-2 regulates starvation-induced autophagy', *Molecular Cell*, 30(6), pp. 678–688. doi:10.1016/j.molcel.2008.06.001.

Weidemüller, P. *et al.* (2021) 'Transcription factors: Bridge between cell signaling and gene regulation', *PROTEOMICS*, 2000034. doi:10.1002/pmic.202000034.

Weistuch, C. *et al.* (2020) 'Inferring a network from dynamical signals at its nodes', *PLOS Computational Biology*, 16(11), p. e1008435. doi:10.1371/journal.pcbi.1008435.

Wells, A. *et al.* (1990) 'Ligand-induced transformation by a noninternalizing epidermal growth factor receptor', *Science (New York, N.Y.)*, 247(4945), pp. 962–964. doi:10.1126/science.2305263.

Werner, S. *et al.* (1992) 'Differential splicing in the extracellular region of fibroblast growth factor receptor 1 generates receptor variants with different ligand-binding specificities', *Molecular and Cellular Biology*, 12(1), pp. 82–88. doi:10.1128/mcb.12.1.82.

Werner, T. *et al.* (2014) 'Ion coalescence of neutron encoded TMT 10-plex reporter ions', *Analytical Chemistry*, 86(7), pp. 3594–3601. doi:10.1021/ac500140s.

Widmann, C. *et al.* (1999) 'Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module From Yeast to Human', *Physiological Reviews*, 79(1), pp. 143–180. doi:10.1152/physrev.1999.79.1.143.

Więdłocha, A. *et al.* (2005) 'Phosphorylation-regulated Nucleocytoplasmic Trafficking of Internalized Fibroblast Growth Factor-1', *Molecular Biology of the Cell*, 16(2), pp. 794–810. doi:10.1091/mbc.E04-05-0389.

Wiley, H.S. (1988) 'Anomalous binding of epidermal growth factor to A431 cells is due to the effect of high receptor densities and a saturable endocytic system', *The Journal of Cell Biology*, 107(2), pp. 801–810. doi:10.1083/jcb.107.2.801.

Wilm, M. (2011) 'Principles of electrospray ionization', *Molecular & cellular proteomics: MCP*, 10(7), p. M111.009407. doi:10.1074/mcp.M111.009407.

Wintheiser, G.A. and Silberstein, P. (2021) 'Physiology, Tyrosine Kinase Receptors', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK538532/ (Accessed: 29 June 2021).

Wiredja, D.D., Koyutürk, M. and Chance, M.R. (2017) 'The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics', *Bioinformatics*, 33(21), pp. 3489–3491. doi:10.1093/bioinformatics/btx415.

Wiśniewski, J.R. and Mann, M. (2012) 'Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis', *Analytical Chemistry*, 84(6), pp. 2631–2637. doi:10.1021/ac300006b.

Wong, A. *et al.* (2002) 'FRS2α attenuates FGF receptor signaling by Grb2- mediated recruitment of the ubiquitin ligase Cbl', *Proceedings of the National Academy of Sciences*, 99(10), pp. 6684–6689. doi:10.1073/pnas.052138899.

Woo, M.S. *et al.* (2004) 'Ribosomal S6 Kinase (RSK) Regulates Phosphorylation of Filamin A on an Important Regulatory Site', *Molecular and Cellular Biology*, 24(7), pp. 3025–3035. doi:10.1128/MCB.24.7.3025-3035.2004.

Wybenga-Groot, L.E. *et al.* (2001) 'Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region', *Cell*, 106(6), pp. 745–757. doi:10.1016/s0092-8674(01)00496-2.

Xie, J. and Proud, C.G. (2014) 'Signaling crosstalk between the mTOR complexes', *Translation*, 2(1), p. e28174. doi:10.4161/trla.28174.

Xie, S. *et al.* (2016) 'The endocytic recycling compartment maintains cargo segregation acquired upon exit from the sorting endosome', *Molecular biology of the cell*, 27(1), pp. 108–126. doi:10.1091/mbc.e15-07-0514.

Xiong, S. *et al.* (2003) 'hSef Inhibits PC-12 Cell Differentiation by Interfering with Ras-Mitogen-activated Protein Kinase MAPK Signaling *', *Journal of Biological Chemistry*, 278(50), pp. 50273–50282. doi:10.1074/jbc.M306936200.

Xu, X. and Wang, M. (2016) 'Inferring Disease Associated Phosphorylation Sites via Random Walk on Multi-Layer Heterogeneous Network', *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(5), pp. 836–844. doi:10.1109/TCBB.2015.2498548.

Yadav, L. *et al.* (2017) 'Systematic Analysis of Human Protein Phosphatase Interactions and Dynamics', *Cell Systems*, 4(4), pp. 430-444.e5. doi:10.1016/j.cels.2017.02.011.

Yaffe, M.B. (2013) 'The scientific drunk and the lamppost: massive sequencing efforts in cancer discovery and treatment.', *Science signaling*, 6(269), p. pe13. doi:10.1126/scisignal.2003684.

Yaffe, M.B. (2019) 'Why geneticists stole cancer research even though cancer is primarily a signaling disease', *Sci. Signal.*, 12(565), p. eaaw3483. doi:10.1126/SCISIGNAL.AAW3483.

Yamasaki, A., Jin, Y. and Ohsumi, Y. (2020) 'Mitotic phosphorylation of the ULK complex regulates cell cycle progression', *PLoS Biology*, 18(6), p. e3000718. doi:10.1371/journal.pbio.3000718.

Yang, G. *et al.* (2015) 'A Positive Feedback Loop between Akt and mTORC2 via SIN1 Phosphorylation', *Cell Reports*, 12(6), pp. 937–943. doi:10.1016/j.celrep.2015.07.016.

Yang, H. *et al.* (2013) 'mTOR kinase structure, mechanism and regulation', *Nature*, 497(7448), pp. 217–223. doi:10.1038/nature12122.

Yang, P. *et al.* (2015) 'Knowledge-Based Analysis for Detecting Key Signaling Events from Time-Series Phosphoproteomics Data', *PLOS Computational Biology*. Edited by L.M. Iakoucheva, 11(8), p. e1004403. doi:10.1371/journal.pcbi.1004403.

Yang, R.-B. *et al.* (2003) 'A Novel Interleukin-17 Receptor-like Protein Identified in Human Umbilical Vein Endothelial Cells Antagonizes Basic Fibroblast Growth Factor-induced Signaling *', *Journal of Biological Chemistry*, 278(35), pp. 33232–33238. doi:10.1074/jbc.M305022200.

Yang, Y. *et al.* (2020) 'In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics', *Nature Communications*, 11(1), p. 146. doi:10.1038/s41467-019-13866-z.

Yarwood, R. *et al.* (2020) 'Membrane trafficking in health and disease', *Disease Models & Mechanisms*, 13(4), p. dmm043448. doi:10.1242/dmm.043448.

Yeh, B.K. *et al.* (2003) 'Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors', *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), pp. 2266–2271. doi:10.1073/pnas.0436500100.

Yu, G. *et al.* (2010) 'GOSemSim: an R package for measuring semantic similarity among GO terms and gene products', *Bioinformatics*, 26(7), pp. 976–978. doi:10.1093/bioinformatics/btq064.

Yuan, W. and Song, C. (2020) 'The Emerging Role of Rab5 in Membrane Receptor Trafficking and Signaling Pathways', *Biochemistry Research International*, 2020, p. 4186308. doi:10.1155/2020/4186308.

Yusoff, P. *et al.* (2002) 'Sprouty2 Inhibits the Ras/MAP Kinase Pathway by Inhibiting the Activation of Raf *', *Journal of Biological Chemistry*, 277(5), pp. 3195–3201. doi:10.1074/jbc.M108368200.

Zachari, M. and Ganley, I.G. (2017) 'The mammalian ULK1 complex and autophagy initiation', *Essays in Biochemistry*. Edited by J.D. Lane, V.I. Korolchuk, and J.T. Murray, 61(6), pp. 585–596. doi:10.1042/EBC20170021.

Zakrzewska, M. *et al.* (2013) 'ERK-Mediated Phosphorylation of Fibroblast Growth Factor Receptor 1 on Ser777 Inhibits Signaling', *Science Signaling*, 6(262), pp. ra11–ra11. doi:10.1126/scisignal.2003087.

Zakrzewska, M. *et al.* (2019) 'Crosstalk between p38 and Erk 1/2 in Downregulation of FGF1-Induced Signaling', *International Journal of Molecular Sciences*, 20(8). doi:10.3390/ijms20081826.

Zambelli, A.E. (2016) 'A data-driven approach to estimating the number of clusters in hierarchical clustering', *F1000Research*, 5, p. 2809. doi:10.12688/f1000research.10103.1.

Zanivan, S., Krueger, M. and Mann, M. (2012) 'In vivo quantitative proteomics: the SILAC mouse', *Methods in Molecular Biology (Clifton, N.J.)*, 757, pp. 435–450. doi:10.1007/978-1-61779-166-6_25.

Zellner, S., Schifferer, M. and Behrends, C. (2021) 'Systematically defining selective autophagy receptor-specific cargo using autophagosome content profiling', *Molecular Cell*, 81(6), pp. 1337-1354.e8. doi:10.1016/j.molcel.2021.01.009.

Zhang, Xuewu *et al.* (2006) 'An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor', *Cell*, 125(6), pp. 1137–1149. doi:10.1016/j.cell.2006.05.013.

Zhang, Xiuqin *et al.* (2006) 'Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family.', *The Journal of biological chemistry*, 281(23), pp. 15694–700. doi:10.1074/jbc.M601252200.

Zhang, Y. *et al.* (2005) 'Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules', *Molecular & cellular proteomics: MCP*, 4(9), pp. 1240–1250. doi:10.1074/mcp.M500089-MCP200.

Zhao, L. *et al.* (2020) 'Comparative evaluation of label-free quantification strategies', *Journal of Proteomics*, 215, p. 103669. doi:10.1016/j.jprot.2020.103669.

Zhen, Y. *et al.* (2018) 'Proximity Labeling by a Recombinant APEX2-FGF1 Fusion Protein Reveals Interaction of FGF1 with the Proteoglycans CD44 and CSPG4', *Biochemistry*, 57(26), pp. 3807–3816. doi:10.1021/acs.biochem.8b00120.

Zheng, Y. *et al.* (2013) 'Temporal regulation of EGF signalling networks by the scaffold protein Shc1', *Nature*, 499(7457), pp. 166–171. doi:10.1038/nature12308.

Zhou, H. *et al.* (2013) 'Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography', *Nature Protocols*, 8(3), pp. 461–480. doi:10.1038/nprot.2013.010.

Zhu, Y. *et al.* (2020) 'DEqMS: a method for accurate variance estimation in differential protein expression analysis', *Molecular & Cellular Proteomics*, 19(6), p. mcp.TIR119.001646. doi:10.1074/mcp.tir119.001646.

Zinkle, A. and Mohammadi, M. (2019) 'Structural Biology of the FGF7 Subfamily', *Frontiers in Genetics*, 10, p. 102. doi:10.3389/fgene.2019.00102.

Zoncu, R. *et al.* (2011) 'mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase', *Science (New York, N.Y.)*, 334(6056), pp. 678–683. doi:10.1126/science.1207056.

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.

Appendix 1

Supplementary Information for Watson et al. 2021 (Chapter 2).

Appendix 1 Figure 1. Optimization of RWHN parameters using the validation dataset. 254

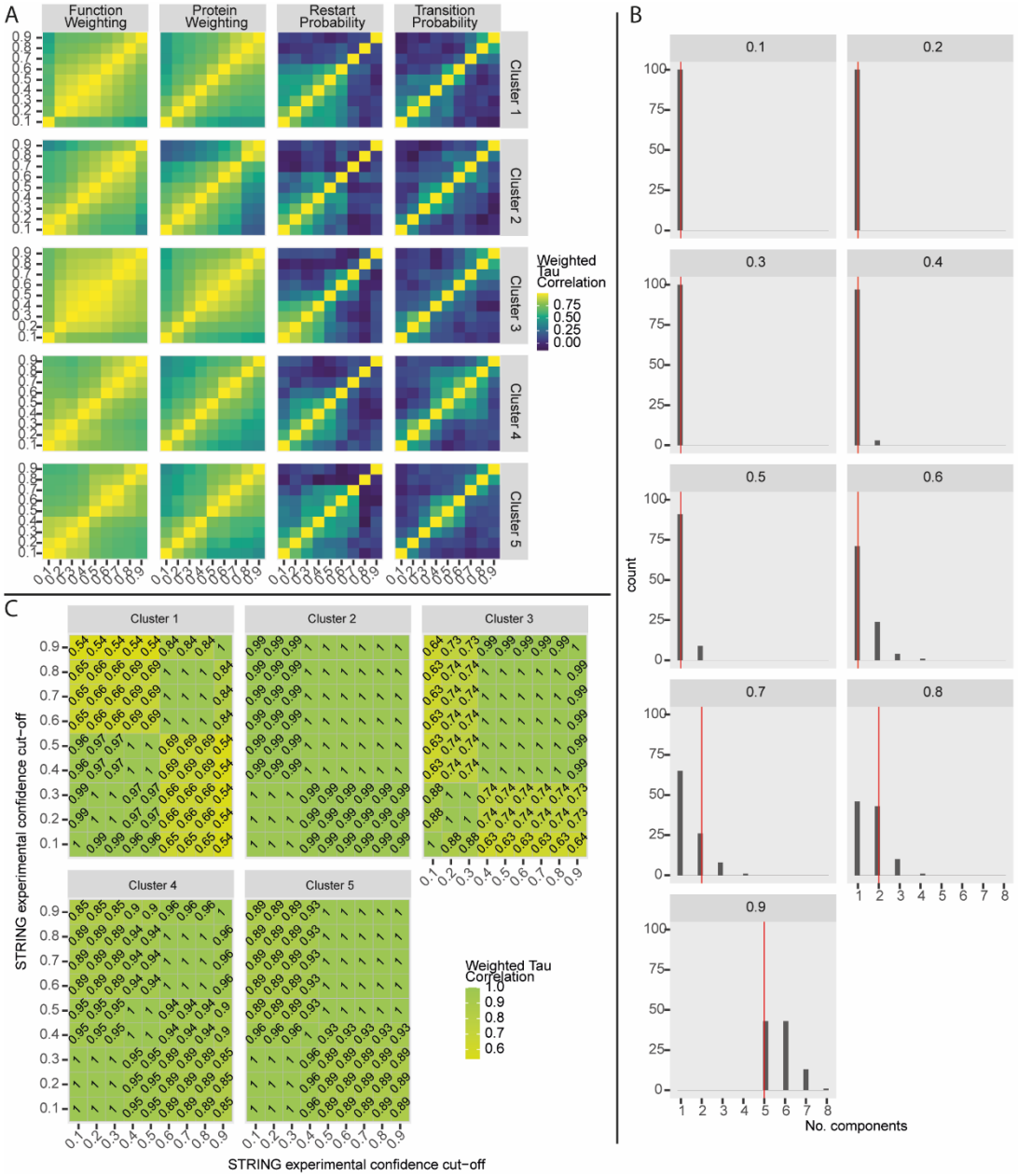
Appendix 1 Figure 2. Assessment of RWHN performance using randomly permuted validation network. 255

Appendix 1 Figure 3. Phosphorylation levels over time of phosphorylated sites within each of the 6 clusters from the Francavilla et al. dataset that were regulated by EGF or TGF-α stimulation. 257

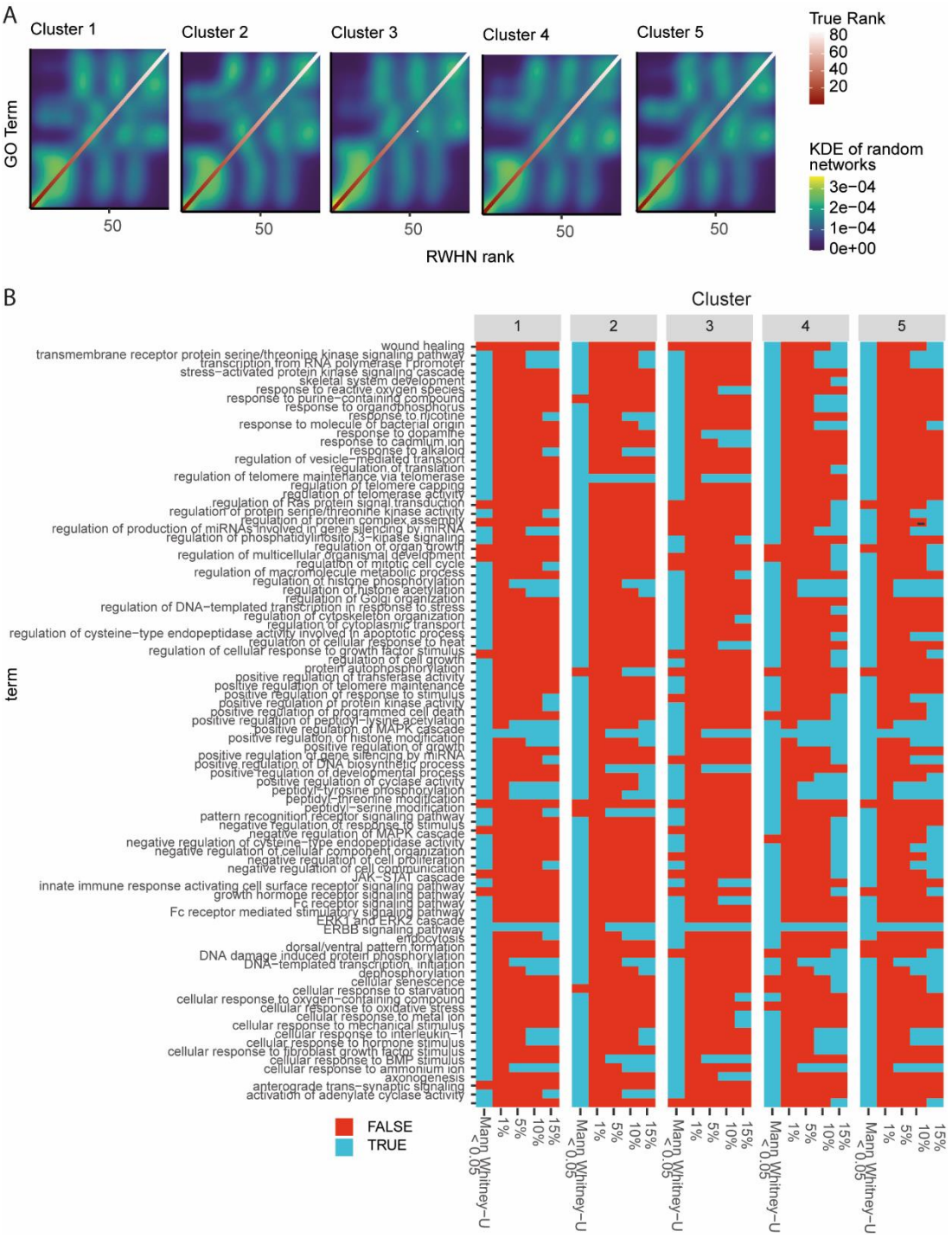
Appendix 1 Figure 4. Clustering of regulated phosphorylated sites in the Ruprecht et al. dataset using k-means clustering. 257

Appendix 1 Figure 5. Clustering of regulated phosphorylated sites from the Santra et al. dataset using k-means clustering. 259

Appendix 1 Table 1. PSP ON_PROCESS annotations, that describe biological function of sites, mapped to most closely related GOBP term 260



Appendix 1 Figure 1. Optimization of RWHN parameters using the validation dataset. ***A.*** *The parameters in the iterative RWHN calculation (Equation 8) – transition probability (λ), restart probability (r) – and the tunable weighting applied to the protein (η_P) or function (η_F) nodes of the multilayer heterogenous network were tested at values within the range of 0.1 – 0.9, changing one parameter whilst setting all others to 0.5. This was repeated with seed nodes set to all five clusters found in the validation dataset. We then calculated how frequently each GO term would be ranked at the same position when the value of each parameter was altered, indicating stability over the range of potential values. Each cell in the heatmaps indicates a GO Term at a rank, with the color indicating the percentage of times that term was given that rank over the range of parameters.* ***B.*** *Connectivity of randomly permuted networks based on the STRING PPI that forms the protein layer, with different confidence thresholds. As the confidence threshold increases above 0.7, the likelihood of the PPI being disconnected increased. The red line indicates the "true" PPI based on the validation data.* ***C.*** *Weighted Tau correlation of RWHN results with STRING PPI constructed with a range of confidence thresholds.*

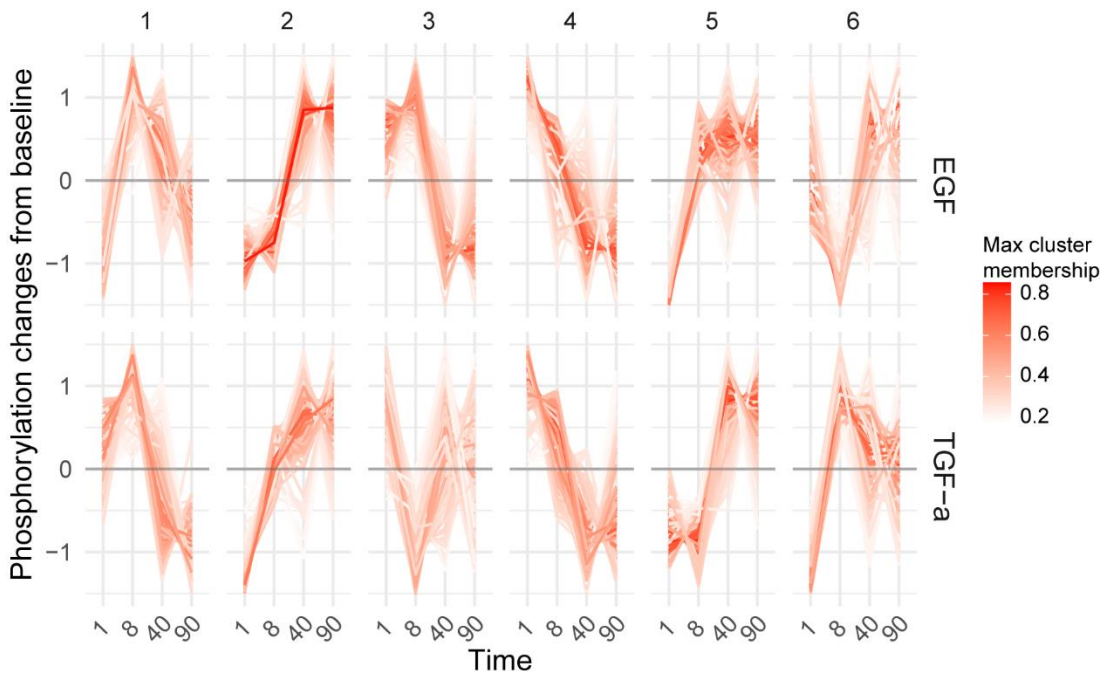


Appendix 1 Figure 2. Assessment of RWHN performance using randomly permuted validation network.

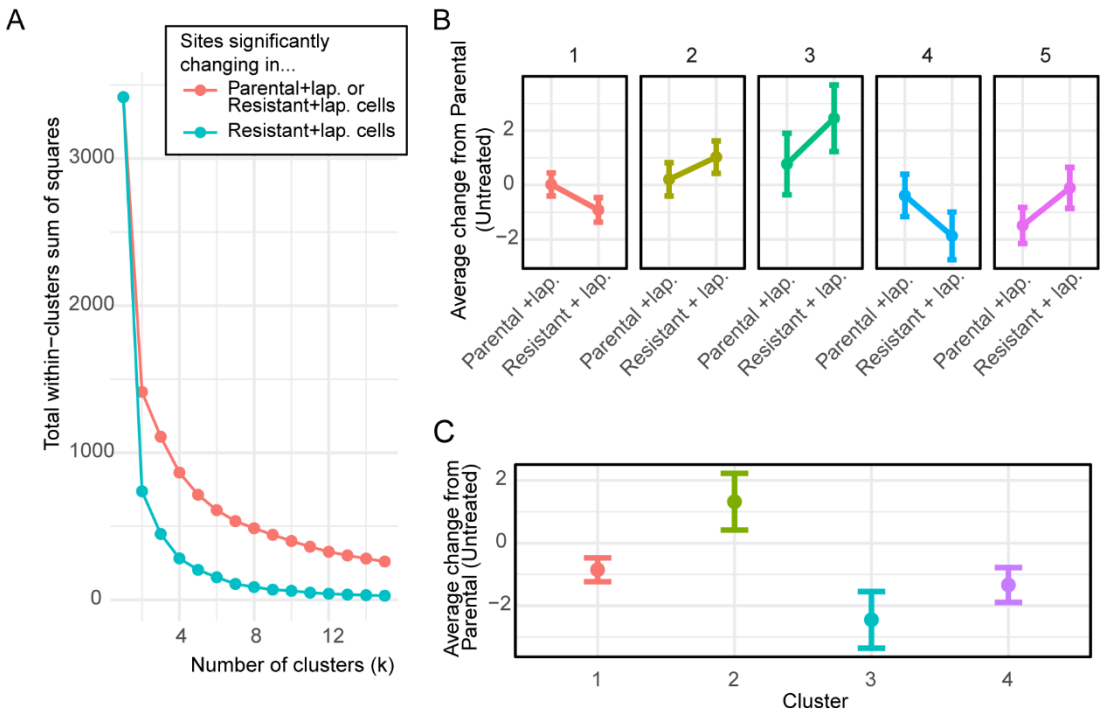
The validation network was permuted 100 times and RWHN run over each permutation with seed nodes set to sites found in clusters 1 to 5.

***A**. KDE was calculated to estimate how frequently each term appeared at each rank in the random networks. The y-axes are reordered based on the "true" ranking of the GO terms (I.e., the ranking in the non-permuted network) for that cluster. If the random networks generated the same result as the actual network, we would expect the KDE plots to closely resemble the trend of the true rank. **B.** The probability of each term in the "true" result (I.e., RWHN results on the non-randomly permuted data) being ranked in that position for each cluster compared*

to the rankings from the randomly permuted networks was calculated using the Mann-Whitney U test ($p < 0.05$, with Benjamini & Hochberg correction). This was then compared to the RWHN output with different thresholds applied – the top 1%, 5%, 10% or 15% of terms. Those terms whose rankings did not differ from random are highlighted with a black outline.

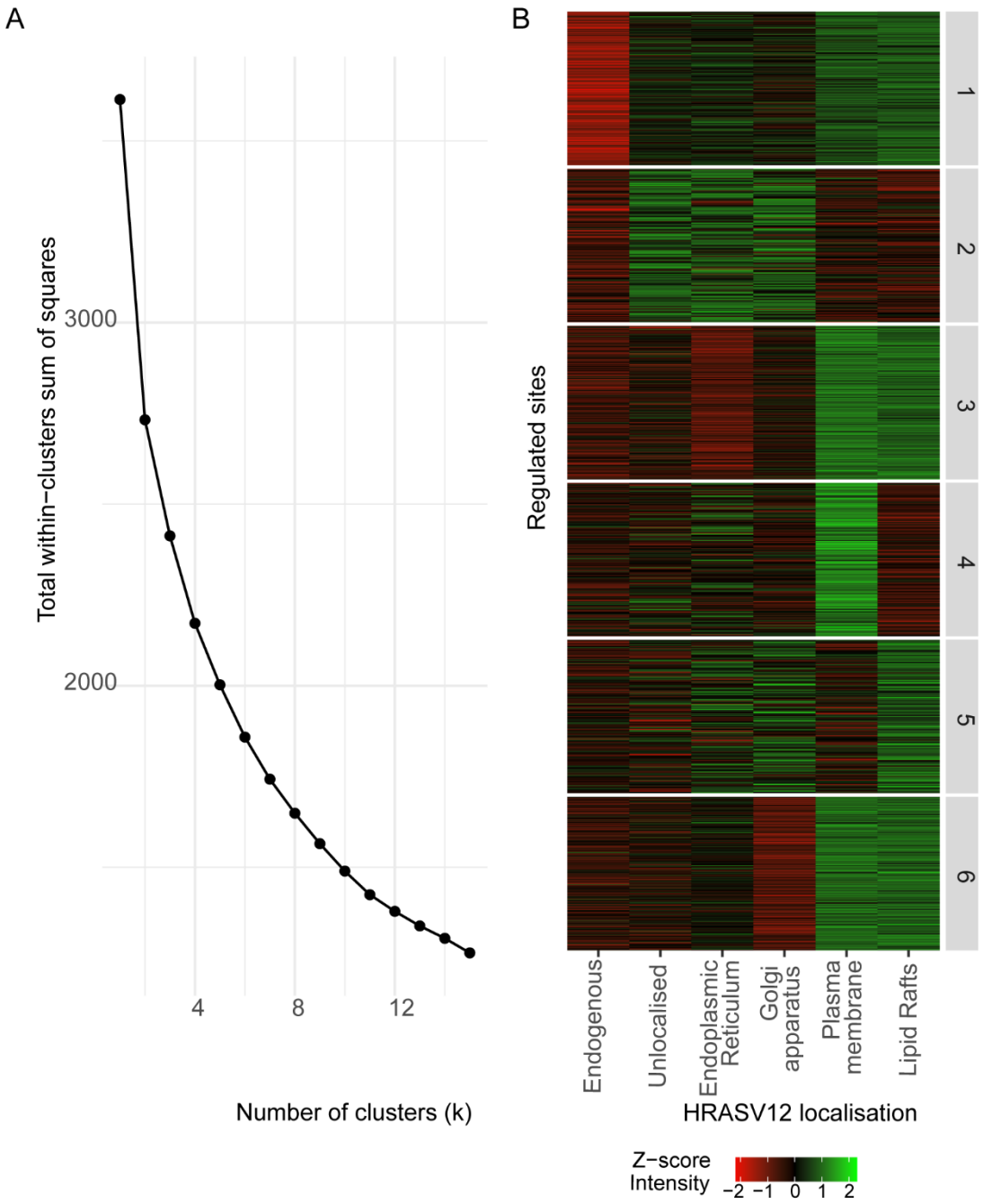


Appendix 1 Figure 3. Phosphorylation levels over time of phosphorylated sites within each of the 6 clusters from the Francavilla et al. dataset that were regulated by EGF or TGF-α stimulation.



Appendix 1 Figure 4. Clustering of regulated phosphorylated sites in the Ruprecht et al. dataset using k-means clustering.
A. *Elbow plot was used to determine the number of clusters (k) to group data into. The pink line represents sites regulated in lapatinib treated parental cells or resistant cells, whilst the blue line represents sites regulated in lapatinib-treated resistant cells only.* ***B.*** *Clustering of*

all lapatinib regulated sites in lapatinib-resistant and parental cell lines. C) Clustering of significantly changed sites in lapatinib-resistant cells.



Appendix 1 Figure 5. Clustering of regulated phosphorylated sites from the Santra et al. dataset using k-means clustering.
A. *Elbow plot used to determine the number of clusters (k) to group the data into.* ***B.*** *Clustering of regulated sites into the selected 6 clusters*

Appendix 1 Table 1. PSP ON_PROCESS annotations, that describe biological function of sites, mapped to most closely related GOBP term

PSP "ON_PROCESS" annotation	Mapped GOBP term	GOID
apoptosis	apoptotic process	GO:0006915
autophagy	autophagy	GO:0006914
cell adhesion	cell adhesion	GO:0007155
cell cycle regulation	cell cycle	GO:0007049
cell differentiation	cell differentiation	GO:0030154
cell growth	cell growth	GO:0016049
cell motility	cell motility	GO:0048870
chromatin organization	chromatin organization	GO:0006325
cytoskeletal reorganization	cytoskeletal reorganization	GO:0007010
DNA repair	DNA repair	GO:0006281
endocytosis	endocytosis	GO:0006897
exocytosis	exocytosis	GO:0006887
RNA splicing	RNA splicing	GO:0008380
RNA stability	regulation of RNA stability	GO:0043487
signaling pathway regulation	signal transduction	GO:0007165
transcription	Gene expression	GO:0010467
translation	translation	GO:0006412

Appendix 2

Appendix 2 Figure 1. Quality assessment of experimental setup and MS-based phosphoproteomics. 262

Appendix 2 Figure 2. Quality assessment of T47D dominant negative dynamin (DnDNM2) and dominant negative RAB11a (DnRAB11) MS-based phosphoproteomics, as an example system with endogenous FGFR2b expression. 264

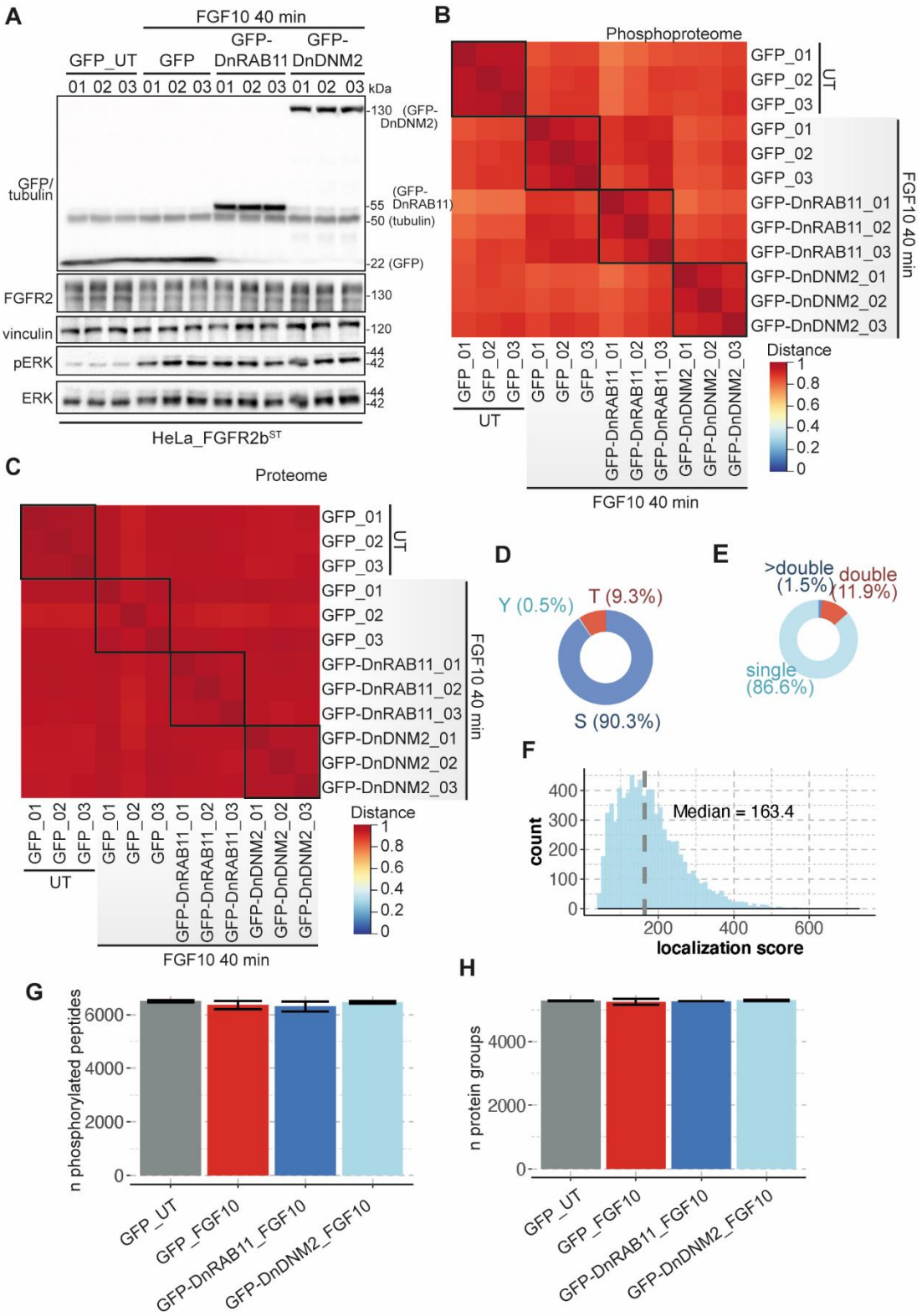
Appendix 2 Figure 3. Quality control of the experiments based on APEX2-tagged proteins. 266

Appendix 2 Figure 4. Quality assessment of proximal and global proteomics and phosphoproteomics. 268

Appendix 2 Figure 5. Role of FGFR2b and mTOR downstream of FGFR2b in autophagy regulation. 270

Appendix 2 Figure 6. FGFR2b regulates mTOR and ULK1 signalling from the REs. 272

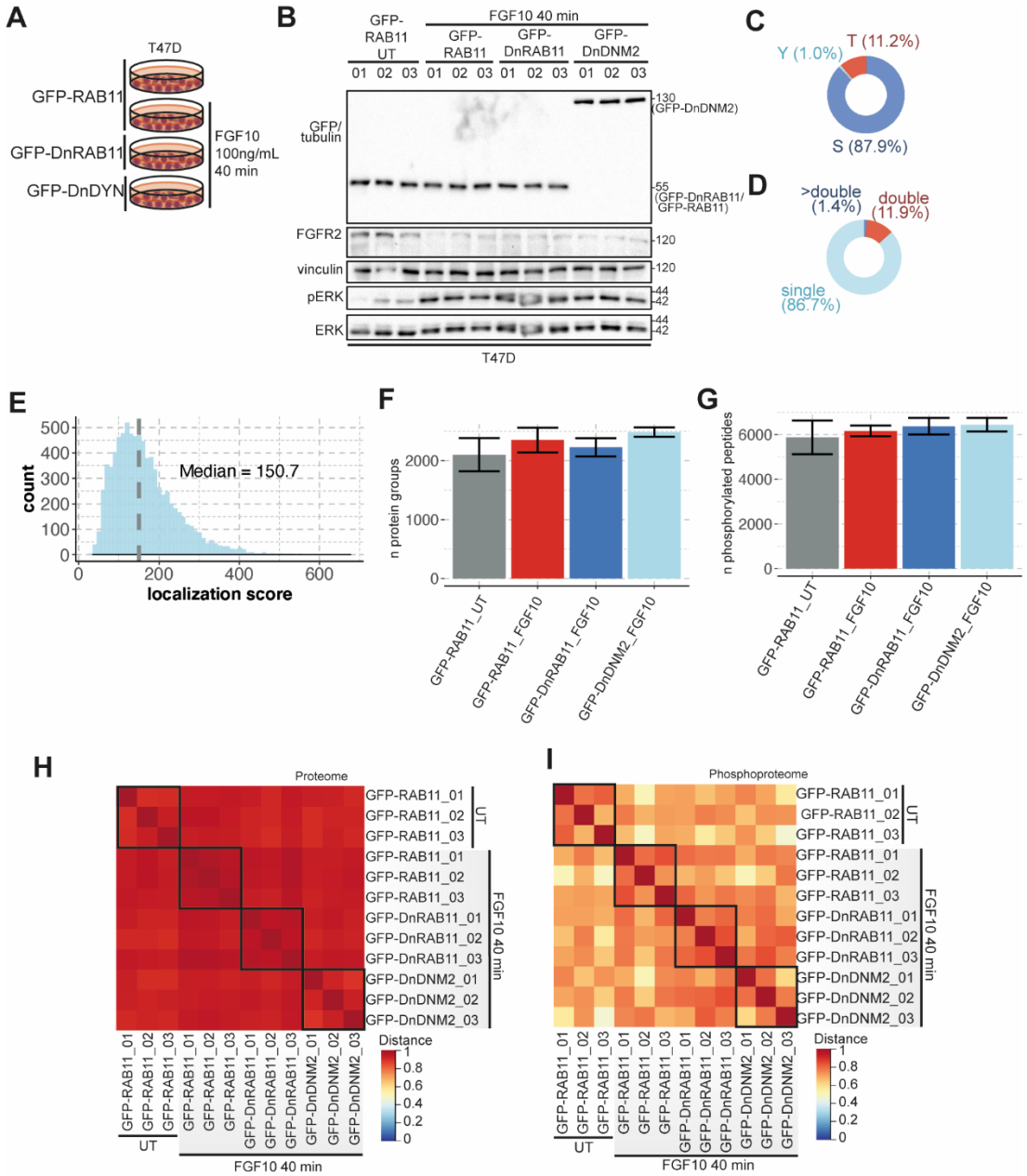
Appendix 2 Figure 7. Inhibiting FGFR2b recycling leads to dysregulated autophagy and an altered balance of proliferation and cell death. 274



Appendix 2 Figure 1. Quality assessment of experimental setup and MS-based phosphoproteomics.

***A**. Immunoblot of input lysates collected for the experiment described in Figure 1D, briefly, HeLa-FGFR2b cells expressing GFP untreated (UT) and either GFP, GFP-DnRAB11 or GFP-DnDNM2, were treated with FGF10 (100 ng/mL) for 40 min (n = 3). Immunostaining for GFP*

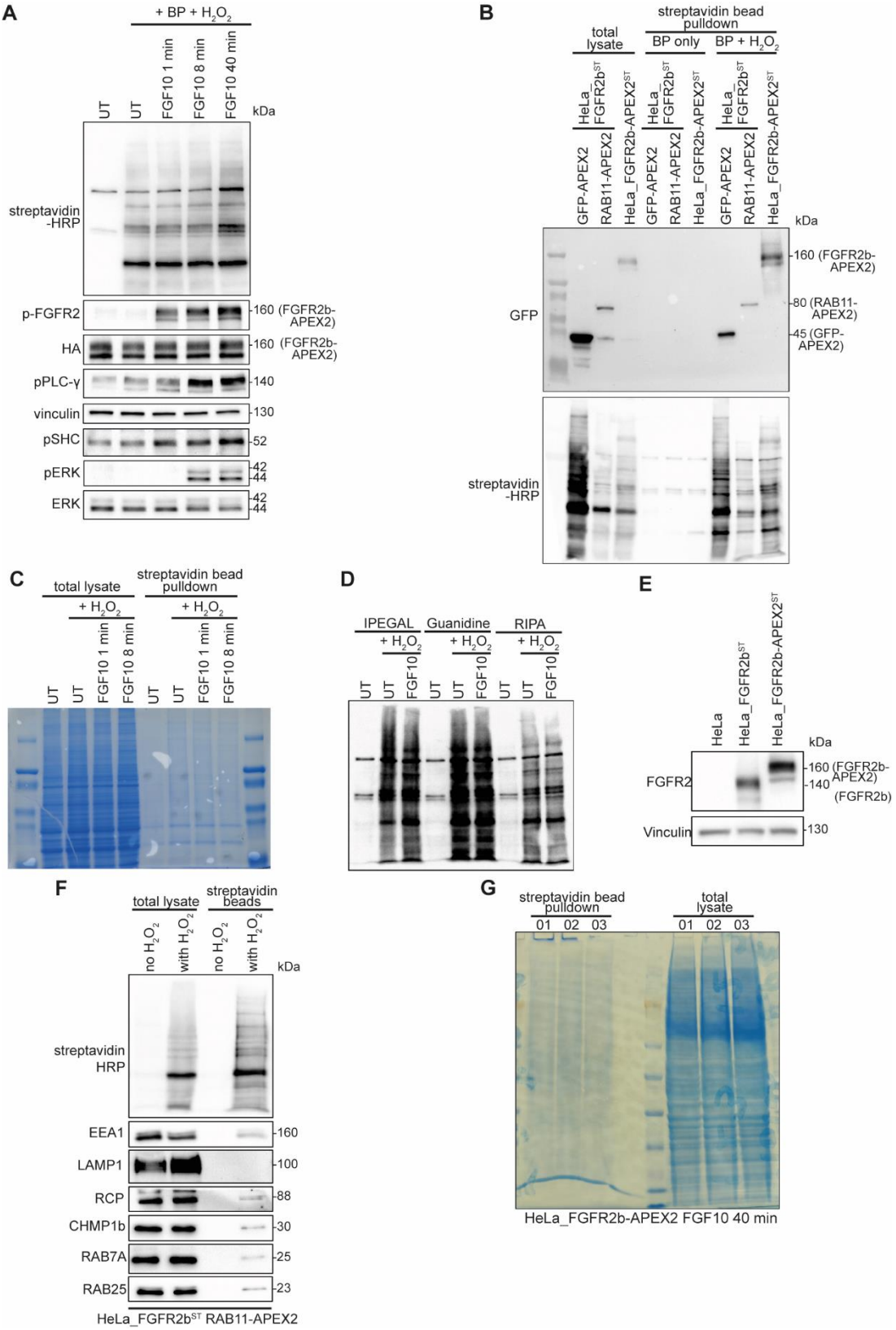
shows consistent transfection efficiency across conditions and replicates. Phosphorylation of ERK1/2 is increased in all FGF10 treated conditions, indicating consistent FGFR2b activation in all three FGF10 treated conditions. B. Number (n) of phosphorylated sites in each condition following QC filtering; removal of potential contaminants, reverse sequences and a localization probability < 0.75. Values represent mean ± SD of n = 3. C. Distribution of peptides containing phosphorylated Serine (S), Threonine (T) and Tyrosine (Y) residues (%). ***B-C****. Heatmap of Pearson correlation of the proteome (B) phosphoproteome(C) show good reproducibility and variation within experimental conditions. Transient expression (< 72h) of dominant negative constructs does not noticeably change the background proteome of HeLa cells.* ***D.*** *Distribution of peptides containing phosphorylated Serine (S), Threonine (T) and Tyrosine (Y) residues (%).* ***E.*** *Distribution of peptides containing a single, double of more than double modification (%).* ***F.*** *Distribution of phosphorylated peptides localization score with a median of 163.4H.* ***G-H.*** *Number (n) of proteins (G) and phosphorylated sites (H) quantified in each condition following QC filtering: removal of potential contaminants, reverse sequences. Proteins were also filtered for razor and unique peptides > 1, unique and razor sequence coverage ≥ 5%. Phosphorylated sites were also filtered for localization probability > 0.75. Values represent mean +/- SD of n = 3.*



Appendix 2 Figure 2. Quality assessment of T47D dominant negative dynamin (DnDNM2) and dominant negative RAB11a (DnRAB11) MS-based phosphoproteomics, as an example system with endogenous FGFR2b expression.

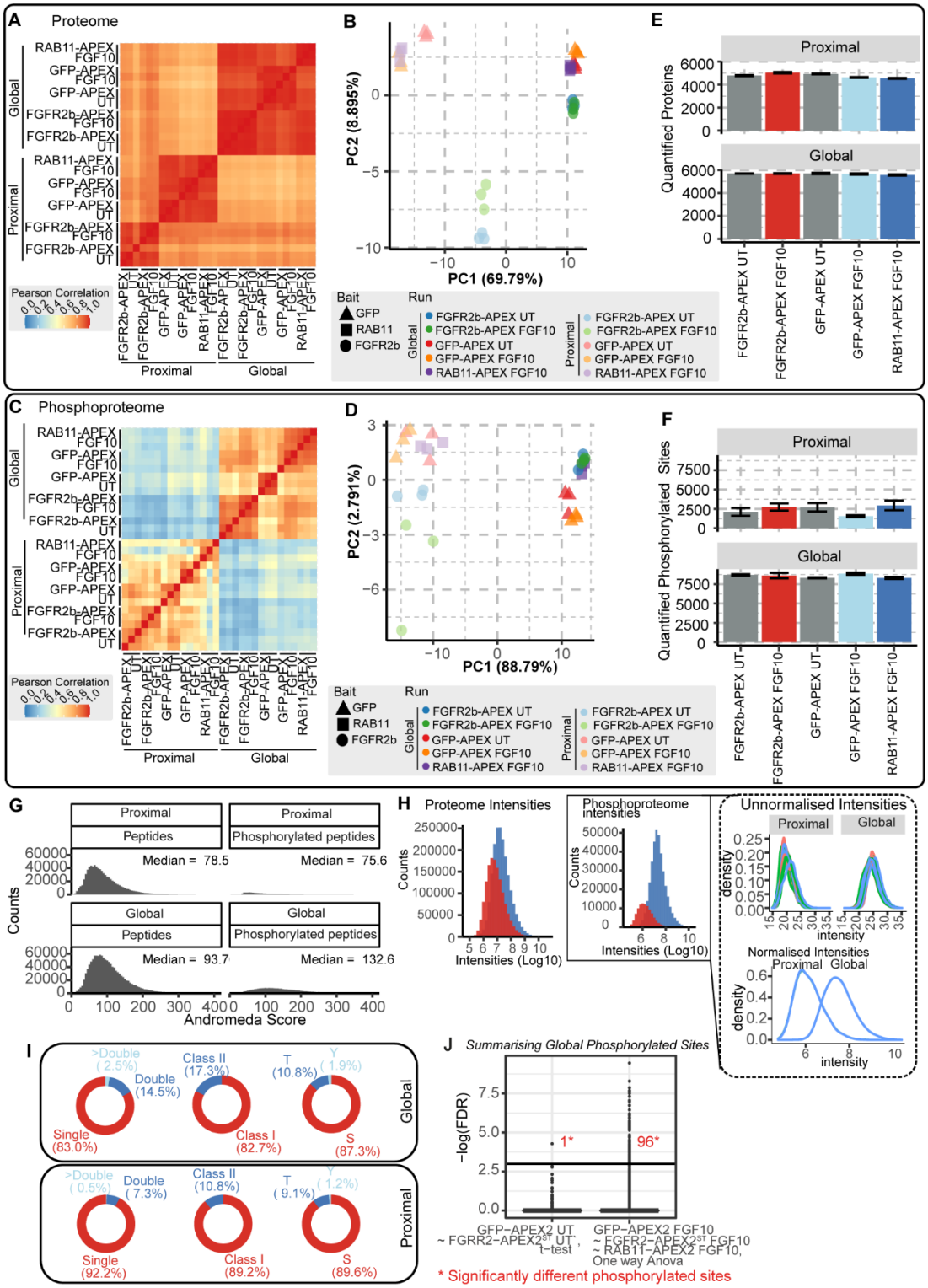
A. *Conditions for label-free MS-based phosphoproteomics in T47D breast cancer cell line.* ***B.*** *Immunoblot of input lysates collected for the experiment described in above, briefly, HeLa-FGFR2b cells expressing GFP-RAB11 untreated (UT) and either GFP-RAB11, GFP-DnRAB11 or GFP-DnDNM2, were treated with FGF10 (100 ng/mL) for 40 min (n = 3). Immunostaining for GFP shows consistent transfection efficiency across conditions and replicates. Phosphorylation of ERK1/2 is increased in all FGF10 treated conditions, indicating consistent FGFR2b activation in all three FGF10 treated conditions.* ***C.*** *Distribution (%) of peptides containing phosphorylated Serine (S), Threonine (T) and Tyrosine (Y) residues.* ***D.*** *Distribution (%) of peptides containing a single, double of > double modification.* ***E.*** *Distribution of phosphorylated peptide localization scores with a median of 150.7.* ***F-G.*** *Number (n) of proteins (F) and phosphorylated sites (G) quantified in each condition following QC filtering: removal of potential contaminants, reverse sequences. Proteins were also filtered for razor*

and unique peptides > 1, unique and razor sequence coverage ≥ 5%. Phosphorylated sites were also filtered for localization probability > 0.75. Values represent mean +/- SD of n = 3. ***H-I****. Heatmap of Pearson correlation of the total T47D proteome (H) and phosphoproteome (I) shows good reproducibility and variation within experimental conditions*



Appendix 2 Figure 3. Quality control of the experiments based on APEX2-tagged proteins. A. Immunoblot analysis with the indicated antibodies illustrating the expression of APEX2 constructs in the HeLa-FGFR2b-APEX2ST and HeLa-FGFR2b. B. Immunoblot analysis with the indicated antibodies of HeLa cells transfected with GFP-RAB11-APEX2 or GFP-APEX2 or HA-FGFR2b-APEX2. Total lysates (total) and the pulldown following enrichment of biotinylated

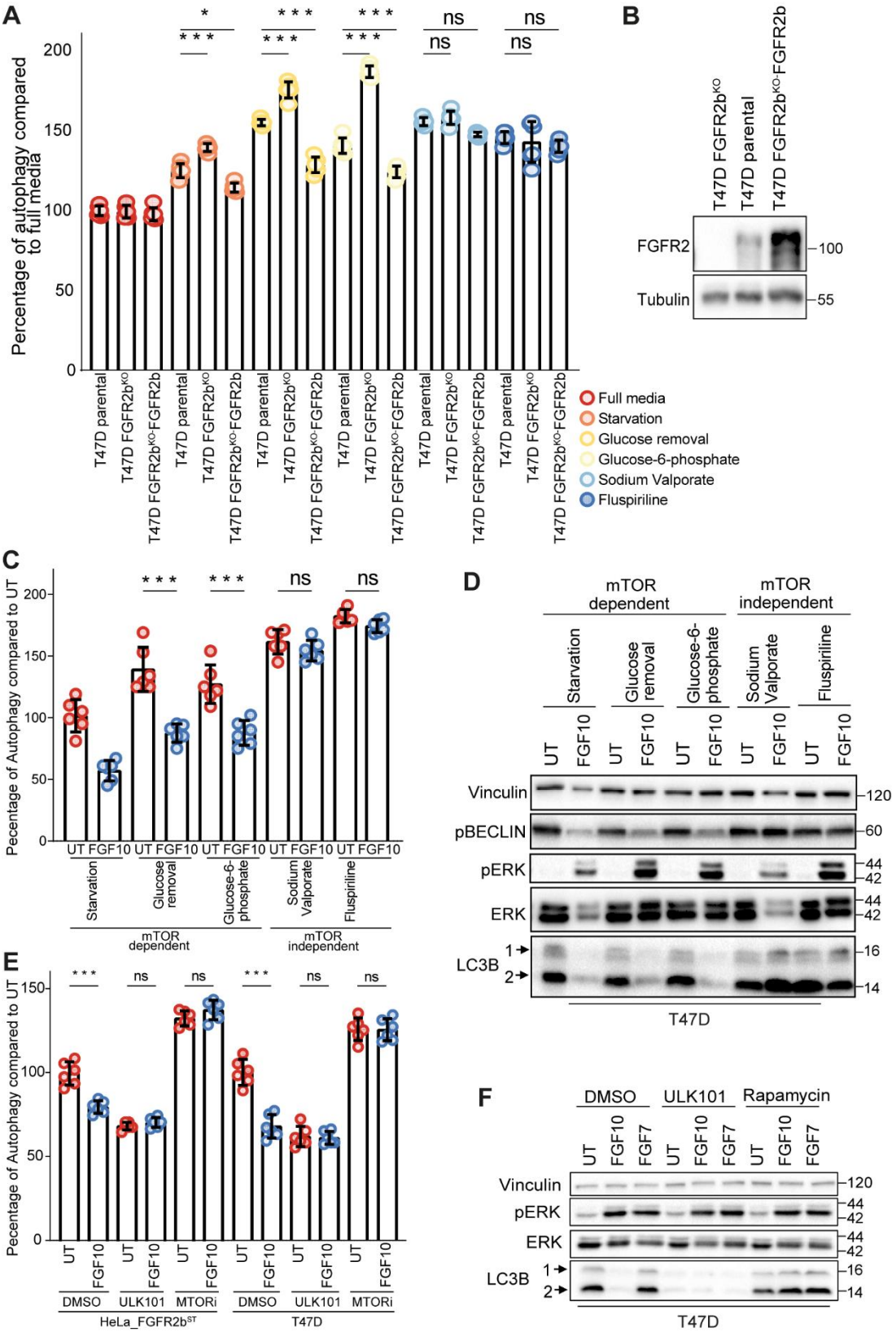
samples with streptavidin beads are shown. BP is the source of biotin for biotinylation by APEX2 and H2O2 is required for catalysis of biotinylation. C. Western blot of indicated proteins including Strep-HRP as control on lysates from HeLa-FGFR2b-APEX2 cells; biotinylation was performed in indicated samples and cells stimulated with FGF10 for indicated time points. D. Lysates from HeLa-FGFR2b-APEX2 cells ran on an 8% SDS PAGE gel and stained with Coomassie substitute; biotinylation was performed in indicated samples and cells stimulated with FGF10 for indicated time points; streptavidin bead pulldown, running the control samples against, lysates extracted off the beads. E. Western blot of Strep-HRP indicating recovered protein efficacy from lysates extracted off the same volume of beads following biotinylation of HeLa-FGFR2b-APEX2 and control conditions. F. Western blot of indicated proteins in HeLa-FGFR2b transfected with GFP-RAB11-APEX2 (48hours); biotinylation was performed followed by streptavidin bead pulldown, running the control samples against, lysates extracted off the beads and including Strep-HRP as control.



Appendix 2 Figure 4. Quality assessment of proximal and global proteomics and phosphoproteomics.

A. *Heat map showing the Pearson correlation of the global and proximal proteome of HeLa_FGFR2b-APEX2ST GFP-RAB11, HeLa_FGFR2ST GFP-APEX2 and HeLa_FGFR2ST RAB11-APEX2 treated with FGF10 or untreated, showed very good reproducibility among biological replicates (Pearson correlation coefficient higher than or equal to 0.9). Differences were seen among UT and FGF10 stimulated cells expressing different APEX2 constructs in the proximal samples, between UT and FGF10 stimulated cells in the global samples, and between the total*

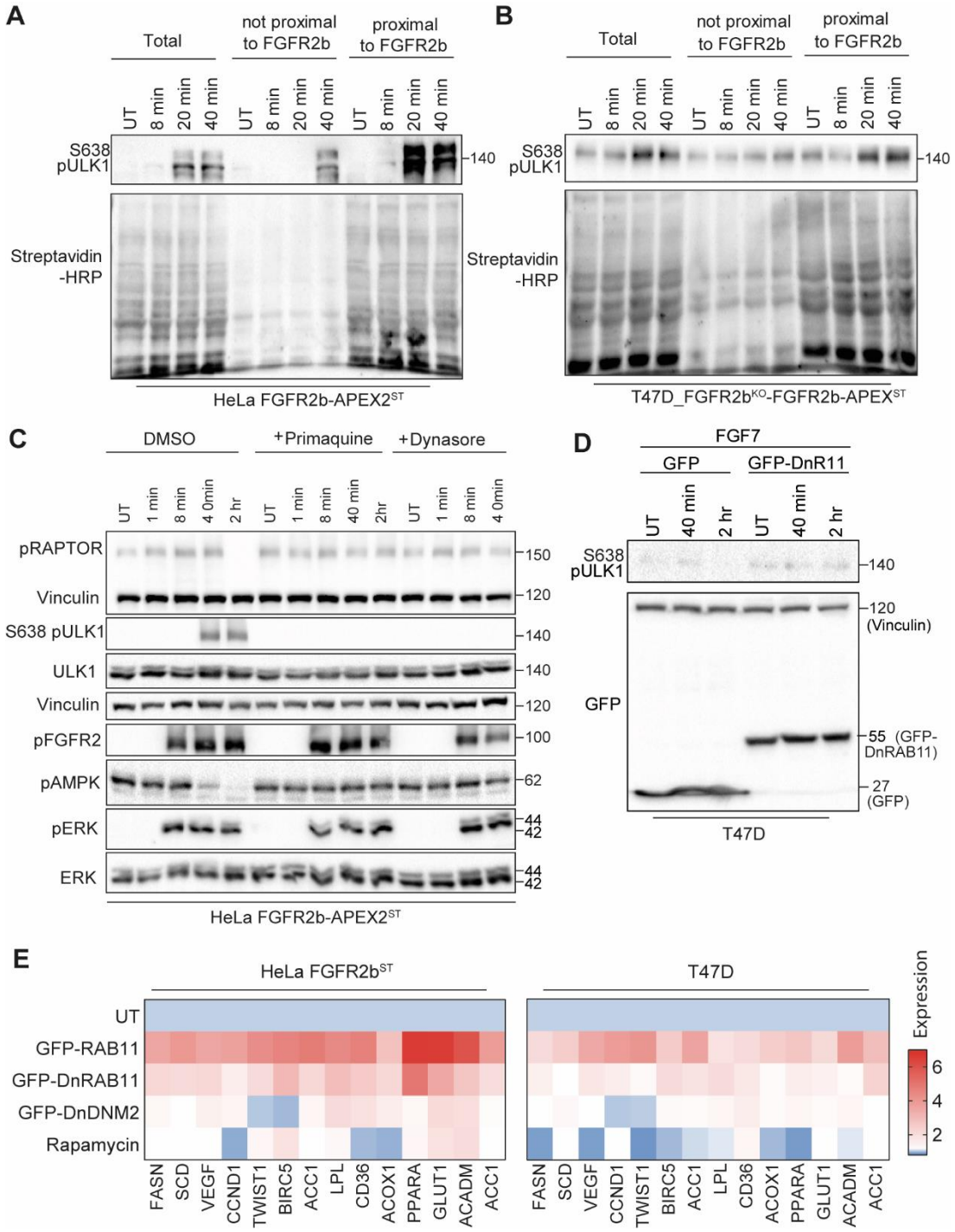
and proximal samples more broadly (Pearson correlation coefficient smaller than 0.9). ***B.*** *PCA of the proteome samples.* ***C.*** *Heat map showing the Pearson correlation of the phosphoproteome of HeLa_FGFR2b-APEX2ST GFP-RAB11, HeLa_FGFR2STGFP-APEX2 and HeLa_FGFR2ST RAB11-APEX2 cells, treated with FGF10 or untreated, showed very good reproducibility among biological replicates (Pearson correlation coefficient higher than or equal to 0.74). Differences were seen among UT and FGF10 stimulated cells expressing different APEX2 constructs in the proximal samples, between UT and FGF10 stimulated cells in the global samples, and between the total and proximal samples more broadly.* ***D****. PCA of the phosphoproteome samples.* ***E.*** *Number (n) of protein groups quantified in each condition. Values represent mean ± SD of N = 3.* ***F.*** *Number (n) of phosphorylated sites quantified in each condition. Values represent mean ± SD of n = 3.* ***G****. The distribution of peptides and phosphorylated peptides Andromeda score showed that most of the peptides were identified with relatively high confidence (median score > 75).* ***H.*** *Distribution of Log10 intensities of the proteome and phosphoproteome from proximal or global experimental conditions. Inset: densities representing the unnormalized and normalized distributions of intensities. Global and proximal proteome samples were normalized together, whilst phosphoproteome samples were normalized separately.* ***I.*** *Visualisation of phosphorylated site characteristics: phosphorylated sites containing a single, double of more than double modification; percentage of the phosphorylated sites with a localisation probability ≥ 0.75 (Class I) (%); and phosphorylation a modification on Serine (S), Threonine (T) and Tyrosine (Y) residues (%). Separated by phosphorylated sites quantified in global or proximal.* ***J.*** *To assess whether APEX2 tags altered the quantification of the phosphoproteome, statistical tests were performed prior to combining the data from the separate mass spectrometry runs (Figures 3A and 4A). A t-test was performed between GFP-APEX2 and FGFR2b-APEX2 UT samples, and a one-way ANOVA between GFP-APEX2, FGFR2b-APEX2 and RAB11-APEX2 treated with 100 ng/ml FGF10 for 40 minutes. Multiple hypothesis testing was accounted for by calculating the false discovery rate (FDR). In each the UT and FGF10 treated samples, 1 and 96 phosphorylated sites were found to be statistically different due to the APEX2 tags. Given that this was within the 5% chance of statistical error, we concluded that the APEX2 tags were not affecting global phosphoproteome quantification and the global UT and FGF10-treated quantitative values could be treated as replicates.*



Appendix 2 Figure 5. Role of FGFR2b and mTOR downstream of FGFR2b in autophagy regulation.

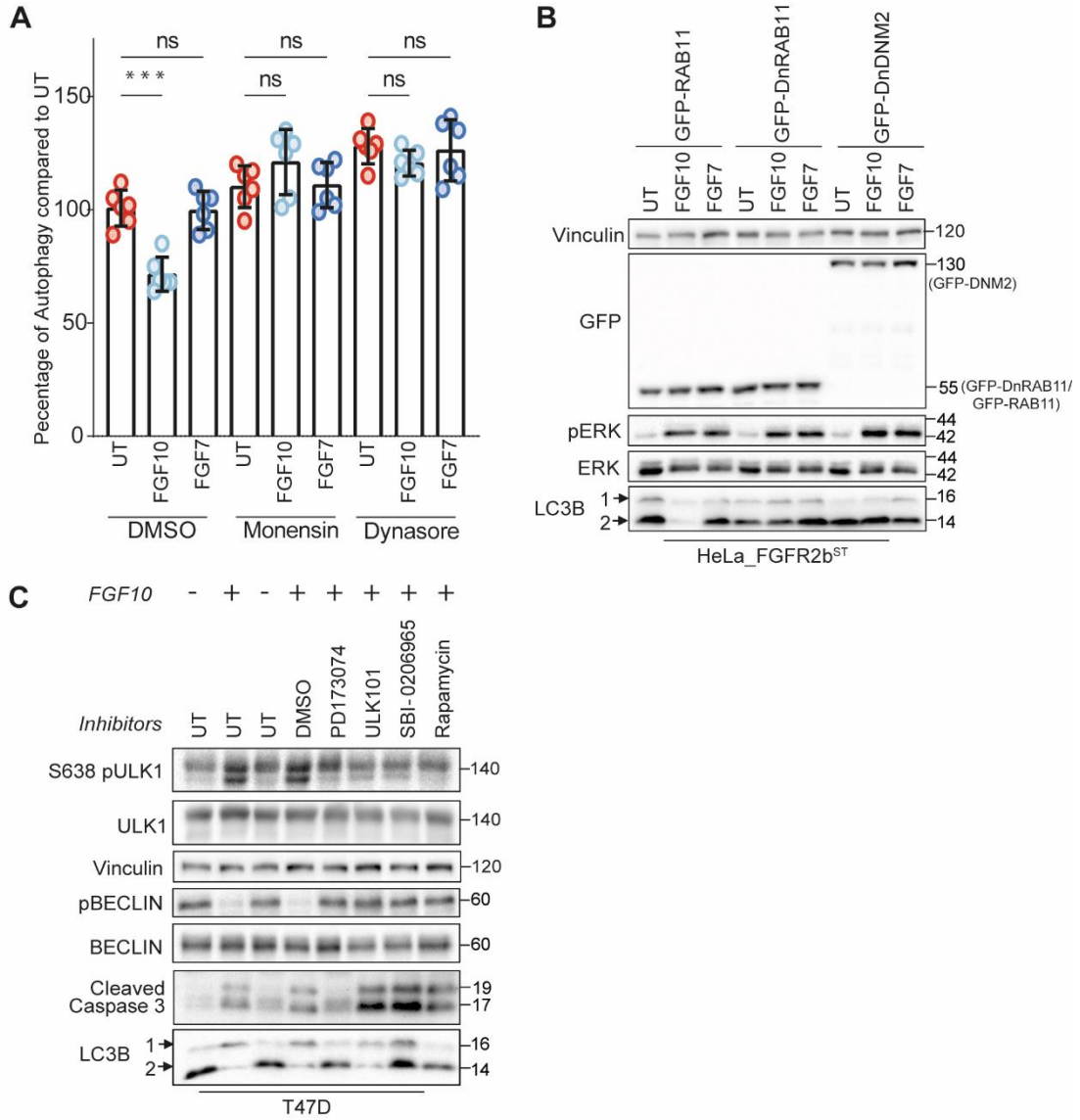
A**. Autophagy measured by acridine orange after 24 hours treatment with the indicated conditions. N = 6, P =< 0.001 (one-way ANOVA with Tukey test). **B**. Immunoblot analysis of FGFR2 expression in T47D, T47D_FGFR2b^{KO} and lentiviral driven. T47D_FGFR2^{KO}_FGFRb.**C**.*

*Autophagy measured by acridine orange after 2 hours treatment of T47D cells with the indicated conditions. N = 6, P =< 0.001*** (one-way ANOVA with Tukey test).* ***D****. Immunoblot analysis with the indicated antibodies of T47D treated for 2 hours with the indicated conditions.* ***E****. Autophagy measured by acridine orange after 2 hours treatment of HeLa_FGFR2bST and T47D cells with FGF10 or FGF7, and ULK1010 and Rapamycin inhibitors (See Figure 7D inset table for inhibitor details). N = 6, P =< 0.001*** (one-way ANOVA with Tukey test).* ***F****. Immunoblot analysis with the indicated antibodies of T47D cells treated for 2 hours as indicated.*



Appendix 2 Figure 6. FGFR2b regulates mTOR and ULK1 signalling from the REs.
*Immunoblot analysis with the indicated antibodies of HeLa FGFR2b-APEX2ST (**A**) and T47D_FGFR2b^{KO}-FGFR2b-APEX (**B**). Non proximal and proximal samples represent the supernatant and the pulldown following enrichment of biotinylated samples with streptavidin beads, respectively, and run against total lysates (total). **C**. Immunoblot analysis with the indicated antibodies of HeLa_FGFR2bST cells pre-treated with primaquine or Dynasore for 2 hours followed by stimulation with FGF10 for the indicated time points. **D**. Immunoblot analysis with the indicated antibodies of T47D cells transfected with either GFP or GFP-DnRAB11 and stimulated with FGF7 for the indicated time points. **E**. qPCR fold expression of indicated genes in HeLa FGFR2b or T47D transfected with GFP-RAB11, GFP-DnRAB11 or GFP-*

DnDNM2 or pre-incubated with rapamycin for 2hours followed by stimulation with FGF10 for 4 hours.



Appendix 2 Figure 7. Inhibiting FGFR2b recycling leads to dysregulated autophagy and an altered balance of proliferation and cell death.

A**. Autophagy measured by acridine orange after treatment of HeLa_FGFR2bST cells in the indicated conditions. Cells were incubated with Monensin or Dynasore for 2 hours before incubation with FGF7 or FGF10 for 2hours. N = 6, P =< 0.001 (one-way ANOVA with Tukey test). **B**. Immunoblot analysis with the indicated antibodies of T47D cells transfected with GFP-RAB11 or GFP-DnRAB11 or GFP-DnDNM2 and stimulated with FGF10 or FGF7 for 2 hours. **C**. Immunoblot analysis with the indicated antibodies of T47D cells pre-treated for 2 hours with PD173074, ULK101, SBI0206965, or Rapamycin (targets of inhibition specified in Figure 3-16D inset table) and stimulated with FGF10 for 2 hours.*