Mechanistic Characterization of RTK Signaling Networks Using Phosphoproteomic Approaches

by

Jacqueline S. Gerritsen

M.S. Biopharmaceutical Sciences
Leiden University, 2016

B.S. Pharmaceutical Sciences
Vrije Universiteit Amsterdam, 2014

SUBMITTED TO THE DEPARTMENT OF BIOLOGICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2022

Signature of Author: _____
Jacqueline S. Gerritsen
Department of Biological Engineering
April 29, 2022

Certified by: _____
Forest M. White, PhD
Ned C. and Janet Bemis Rice Professor of Biological Engineering
Thesis Supervisor

Accepted by: _____
Katharina Ribbeck, PhD
Hyman Career Development Professor
Graduate Program Committee Chair

This doctoral thesis has been examined by a Committee of the Department of Biological Engineering as follows:

Douglas A. Lauffenburger, PhD
Ford Professor of Biological Engineering, Chemical Engineering, and Biology
Massachusetts Institute of Technology
Thesis Committee Chair

Forest M. White, PhD
Ned C. and Janet Bemis Rice Professor of Biological Engineering
Massachusetts Institute of Technology
Thesis Supervisor

Sebastian Lourido, PhD
Associate professor of Biology
Massachusetts Institute of Technology
Committee Member

Mechanistic Characterization of RTK Signaling Networks Using Phosphoproteomic Approaches

by

Jacqueline S. Gerritsen

Submitted to the Department of Biological Engineering on

April 29th

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

ABSTRACT

Cancer is a complex disease, which often stems from aberrant gene expression and protein signaling. In order to improve development of novel therapeutics, underlying mechanisms employed by malignant cells to maintain their function need to be deeply understood. The field of phosphoproteomics has advanced over the past few decades to allow analysis of phosphorylated proteins with increased sensitivity and accuracy. Using these methods in combination with a systematic mutational strategy that evaluates the network effects of loss of phosphorylatable tyrosines that the protein of interest seemingly depends on, allows for an unbiased mechanistic characterization of the signaling network. When implemented with various computational tools, these data can provide a predictive model that can inform future targeting strategies in disease.

Here, I present an overview of the value phosphoproteomics adds to cancer research, and the insights we can gain when combined with Y-to-F mutational approaches to gain mechanistic understanding underlying protein function. We successfully applied these approaches in characterizing EGFR, a protein often dysregulated in cancer. Furthermore, we experimentally evaluated the use of fluorophores such as GFP in these signaling studies. We also applied these approaches in other settings, when evaluating function of AXL, another RTK that has been associated with acquired resistance to EGFR-targeting tyrosine kinase inhibitors. We applied phosphoproteomic analysis to explore the regulation of EGFR by phosphatase PTPRJ as a potential indirect mechanism of modulation. Beyond cancer, we applied our phosphoproteomic approach to validate a novel biomarker in Alzheimer's disease.

Together, these findings highlight how these approaches can result in invaluable mechanistic insight that can propel future drug development efforts.

Thesis supervisor: Forest M. White, PhD
Title: Ned C. and Janet Bemis Rice Professor of Biological Engineering

~

I am dedicating this thesis to my parents Jan Willem Gerritsen and Josephine Gerritsen-Schoonis, one of them in heaven, as they have always empowered me and continue to inspire me to be the best version of myself.

~

Table of Contents

Abstract **3**

Table of Contents **5**

CHAPTER 1: General Introduction **8**

CHAPTER 2 Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells ... **19**

Abstract 20

Introduction: Phosphoproteomics, signaling and cancer 21

Sample preparation and phosphopeptide enrichment 23

Enrichment of subsets of the phosphoproteome 29

Quantitative phosphoproteomic methods 32

MS-based analysis of the phosphoproteome: applications and challenges 36

Computational analysis of phosphoproteomic data 41

Future of phosphoproteomics 45

Expert Opinion 47

References 51

CHAPTER 3: Elucidating phosphorylation-mediated signaling networks in receptor tyrosine kinases through Y-to-F mutational studies **61**

Abstract 62

Receptor Tyrosine Kinases in Disease 63

RTK structure and Function 64

RTK Signaling Network Models and Y-to-F Mutational Studies 65

Consideration of choosing tyrosine-replacing amino-acid 69

Y-to-F interactome studies 70

Y-to-F studies in disease models 72

Site-specific roles beyond downstream signaling conveyance 72

Omics based Y-to-F studies 74

A panel of Y-to-F mutant cell liens to model RTK-specific responses 74

Computational tools to analyze and interpret phosphoproteomic signaling networks 75

Establishing associations between cell phenotypes and RTK downstream signaling 79

Validation experiments to demonstrate model predictions 81

Forward-looking perspective 83

References 85

CHAPTER 4: Predictive data-driven modeling of C-terminal tyrosine function in the EGFR signaling network **93**
Abstract 94
Introduction 95
Results 98
Development and initial characterization of EGFR-expression system 98
NR6 cells expressing wild-type exhibit signaling network response to EGF stimulation 101
EGFR network is largely resilient to mutation of C-terminal tyrosine residues 104
Downstream network analysis reveals network rewiring for Y992F isoform involving PLCγ-1 and other RTKs 106
Phenotypic characterization confirms known biology and reveals interesting trends 108
Partial Least Squares Regression model reveals Epsin 8- and PLCɣ-1-mediated mechanisms associated with cellular phenotype 111
Discussion 115
Material and Methods 119
References 128
Supplemental Figures 132
CHAPTER 5: Using Green Fluorescent Protein in Signaling Studies **139**
Abstract 140
Introduction 141
Results 142
Development of EGFR-expression system 142
NR6 cells expressing wild-type and lacking GFP exhibit stronger signaling network response to EGF stimulation 145
Mutant EGFR isoforms exhibit stronger signaling network in absence of GFP 147
Phenotypes are differentially altered between GFP/non-GFP lines. 148
Discussion 150
Material and Methods 151
References 158
CHAPTER 6: Mapping of AXL signaling networks driving resistance to EGFR inhibition in lung cancer **159**
Abstract 160
Introduction 161
Results 163

Discussion 170
Material and Methods 172
References 179
Supplemental Figures 180
CHAPTER 7: Regulation of EGFR signaling by disrupting PTPRJ dimerization – exploring potential mechanisms 186
Abstract 187
Introduction 188
Results 190
Discussion 196
Material and Methods 197
References 201
CHAPTER 8: RNA-binding protein ELAVL4/HuD ameliorates Alzheimer's disease-related molecular changes in human iPSC-derived neurons 202
Abstract 203
Background 204
Results 208
Generation of ELAVL4/HuD knock out, overexpression and rescue induced neurons 208
ELAVL4 expression affects APP phenotypes in induced neurons 210
ELAVL4 expression affects tau phenotypes in induced neurons 212
Canonical pathway and upstream regulator analysis results 217
Discussion 223
Conclusion 229
Material and Methods 230
References 238
Supplemental Figures 245
CHAPTER 9: Conclusions & Future Perspectives 254
Conclusions & Future Perspectives 256
Summary 259
List of Publications 261
Bibliography 262
Acknowledgements 263

CHAPTER 1

Introduction

Cancer is a signaling disease

Cancer is driven by genetic and epigenetic alterations in oncogenes or tumor suppressor genes that leads to uncontrolled cell proliferation, migration, angiogenesis and invasion, among others.[1] Despite the root of cancer usually being genetic or epigenetic in nature, cancer progression is generally associated with complex mechanisms in and between tumor cells, healthy cells and the tumor microenvironment. Many of the phenotypic outcomes that are common amongst cancer cells, are the result of dysregulated cellular signal transduction pathways.[2]

Despite the many research efforts and therapies that have hit the market, success rates in the clinic have often been dampened or plateaud due to heterogeneity of tumors as well as acquired resistance to chemotherapies.[3,4] Heterogeneous responses to drugs or resistance mechanisms employed by cancer cells have motivated the field of personalized medicine as well as targeted strategies. However, even these highly efficacious inhibitors targeting tyrosine kinases as well as immune therapies were eventually met with resistance, highlighting the complexity of this disease.[5] Given that cancer displays itself as a signaling disease, it is no surprise that nearly all molecularly targeted therapeutics are directed against signaling molecules. If we consider the fundamentals of biology, in which DNA transcribes to RNA, which is translated to protein and can be modified by various post translational modifications (PTMs), which then control biological activity, it seems obvious to consider the closest node to biological activity as a point of intervention if our aim is to regain control of aberrant biological activity in cancer: protein PTMs (Fig 1).[6] PTMs are essential for a number of activities, such

regulation of gene expression, protein stability, protein-protein interactions and activation/deactivation of enzymes. PTMs may provide key information about the derangements in cellular programming in disease conditions and serve as the major targets for therapeutics.[7]

Figure 1: Central dogma of biology: DNA transcribed to RNA, translated to protein, modified with post-translational modifications, to control biological activity in the cell.

Phosphorylation as a measure of protein activation

One of the most extensively studied PTMs is phosphorylation, and with good reason. Phosphorylation regulates a number of critical cellular functions including migration, proliferation, apoptosis and differentiation, similar to the functions dysregulated in cancer cells. [8] Phosphorylation is a reversible modification that occurs involves the addition of a phosphate group to the polar group of various amino acids, most notably serine (~90%), threonine(~10%) and tyrosine (~0.1-1%). The donation of this phosphate group occurs as ATP transfers their phosphoryl group to the hydroxyl group. Proteins are dephosphorylated by phosphatases (Fig 2).[9]

As phosphorylation is the most common mechanism of regulation protein function and transmitting signals throughout the cells, the level of phosphorylation is often taken as a

measure of protein activity in the cell. Phosphorylated residues can be recognized by specific binding domains, most commonly Src homolgy 2 (SH2) or Protein Tyrosine Binding (PTB) domains, which opens the door for downstream signal relay. Rather than linear signal transduction, researchers acknowledge the fact that these stimuli often elicit signaling casacades that can amplify the initial stimulus for large-scale or global cellular responses such as proliferation.[10] Depending on the stimuli and corresponding conformational changes of the protein, pathways can be differentially activated and crosstalk between signaling axes often the norm rather than exception.

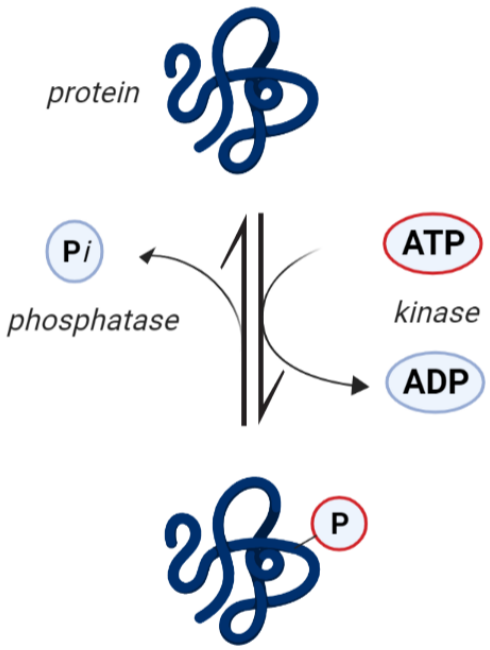


Figure 2: Phosphorylation is a reversible modification, mediated by kinases and phosphates.

Although phosphotyrosine is not as prevalent as phosphoserine or phosphothreonine, tyrosine phosphorylation is at the forefront of biomedical research because of its relation to human disease, mainly through the dysregulation of receptor tyrosine kinases(RTKs).[11]

RTKs and the EGFR signaling network

The RTK family and their downstream tyrosine kinase targets encompass a large number of oncoproteins. RTKs are single-pass membrane proteins with an extracellular ligand-binding domain and intracellular kinase domain. As their name indicates, they rely on tyrosines and more specifically phosphorylation of tyrosines for signal transduction.[12] Extracellular stimulation, e.g. by binding of a ligand, causes dimerization in the plasma membrane that also induces a conformational change of the protein that is necessary for functional maturity.[13] Once the RTK have assumed this “active” conformation, cross- and auto-phosphorylation can occur on the C-terminal residues, thereby initiating a cascade of signaling events ultimately resulting in one or multiple of the cellular phenotypes described earlier (Fig 3).

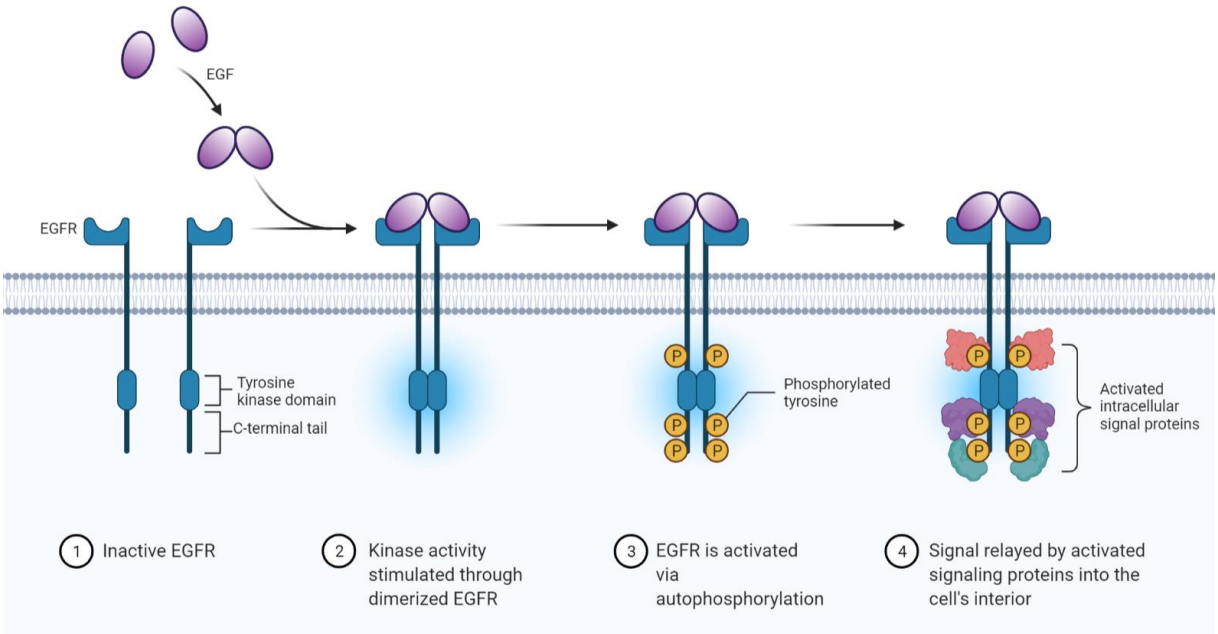


Figure 3: RTK(EGFR) activation upon ligand binding. Monomeric (inactive) EGFR binds ligand, prompting dimerization with another RTK. The subsequent conformational change induces autophosphorylation of C-terminal tyrosines, allowing for signal transduction.

Signaling pathways often engaged by RTKs include the MAPK, PI3K-AKT and JAK-STAT signaling axes.[14] Strong evidence exists that associates altered phosphorylation of these pathways with cancer and other pathologies, making RTKs attractive therapeutic targets.

One of the most well studied RTKs is the Epidermal Growth Factor Receptor (EGFR) which is a member of the ErbB subfamily (EGFR, HER2, HER3, HER4).[15] EGFR is found to be frequently overexpressed, mutated or otherwise dysregulated in a large fraction of human cancers. Despite many studies characterizing EGFR and its associated protein-protein interactions (Fig 4), the underlying mechanisms explaining the individual functions of the C-terminal tyrosines are still poorly understood. [16]

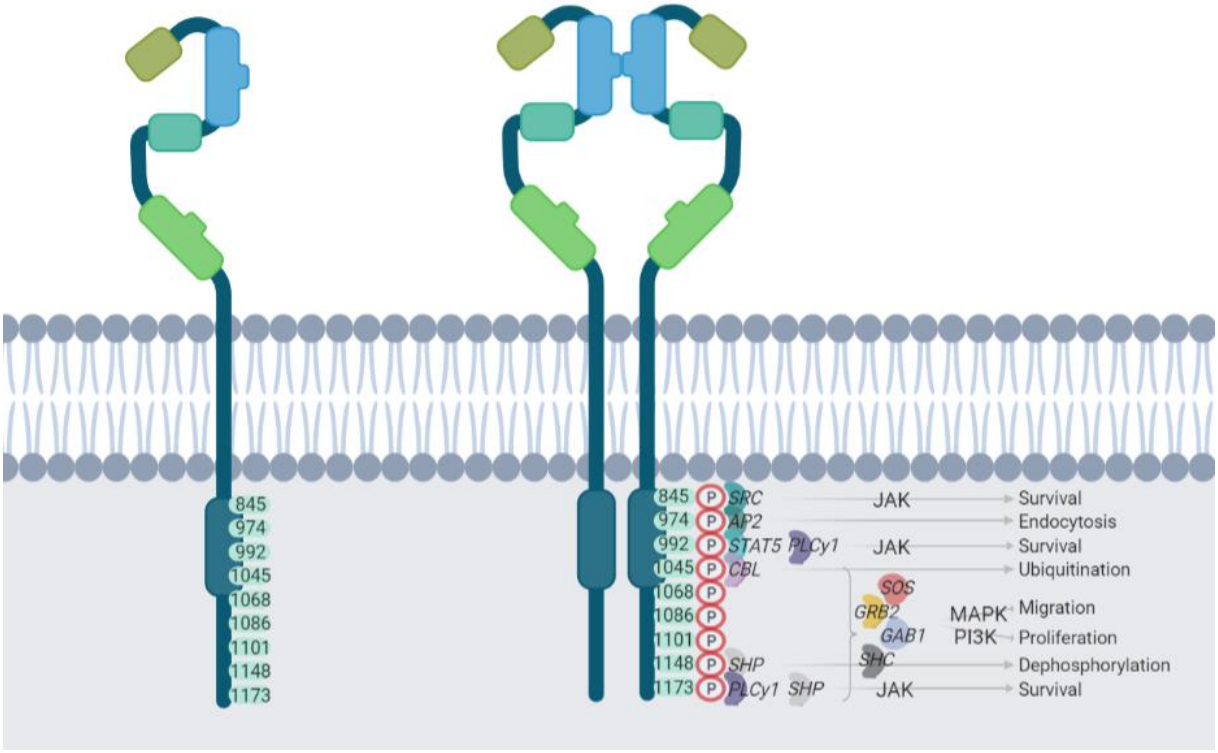


Figure 4: EGFR with most prominent autophosphorylation sites and reported binders.

Where most studies focus on EGFR as either an active or inactive protein, we need to admit to the fact that EGFR signaling is a lot more complex than that, if we want to explain the variety of downstream responses EGFR can regulate based on a variety of inputs. This point is further emphasized by the apparent overlap in binding domains for a variety of adapter molecules, such as Shc and Grb2, which also suggests a built-in adaptative mechanism is in place should one of these sites be unable to participate in signal transduction.[17]

If our goal is to identify novel biomarkers and develop more effective therapies that overcome the resistance and other hurdles currently holding us back in the clinic, we need to analyze the phosphorylation-mediated signaling networks using a systems biology approach, that is

quantitative, physiologically relevant and has great network coverage. On top of that, we need to mechanistically understand how these interactions control downstream signaling and phenotypic responses.

The power of phosphoproteomics

One of the challenges with measuring tyrosine phosphorylation of proteins remains identification and quantification of low abundance pTyr peptides, especially when sample input material is limited.[18] However, advances in mass spectrometry tools and methods have advanced tremendously over the past two decades and now enables gathering of deep biological insight. Phosphoproteomics, the large-scale analysis of protein phosphorylation sites, in particular has emerged as a powerful tool to define signaling network regulation and dysregulation.[19] Optimizations in sample preparation and enrichment have aided as well in increasing the feasibility of using mass spectrometry to measure protein phosphorylation. Chapter 2 will discuss this in greater detail.

This thesis

Aim

The overarching goal of the projects described in this thesis is to use phosphoproteomics in combination with mutational approaches to gain mechanistic understanding of signaling networks. These data and the resulting predictive models provide novel insights regarding RTK

signaling and suggest potential therapeutic targets to modulate cell biological response in disease.

Outline

Chapter 2 will detail the advancements in phosphoproteomics and its value in cancer research. Chapter 3 will assess the significant contributions Y-to-F mutational approaches have made in discovering novel biomarkers and gathering mechanistic insight, and discuss the potential of combining this approach with phosphoproteomics to obtain a deeper understanding of the signaling networks in question. Chapter 4 will describe the main project of this thesis, where previously described approaches are applied to reveal novel insights into the EGFR signaling pathway. Chapter 5 takes a little segue to discuss the use of fluorophores such as GFP in these types of signaling studies, comparing resulting signaling data. Chapter 6 applies an approach similar to that in chapter 4, but to better understand the AXL signaling pathway, in particular as it pertains to AXL-mediated resistance to EGFR tyrosine kinase inhibitors (TKIs). Chapter 7 is a mechanistic evaluation of PTPRJ, a phosphatase that regulates EGFR activity, as a potential therapeutic target to indirectly modulate EGFR activity. Chapter 8 highlights ELAVL4, a novel biomarker in Alzheimer's disease, and the use of phosphoproteomics as a tool to better understand the mechanisms underlying its therapeutic potential. Finally, chapter 9 concludes and summarizes this thesis as well as highlights areas for future exploration and innovation.

References

1 Sever R, Brugge JS. Signal Transduction in Cancer. Cold Spring Harb Perspect Med Internet. 2015 cited 2022 Mar 23;5. Available from: /pmc/articles/PMC4382731/.

2 Devreotes P, Horwitz AR. Signaling networks that regulate cell migration. Cold Spring Harb Perspect Biol. 2015;7.

3 Bonadonna G. Combination chemotherapy as an adjuvant treatment in operable breast cancer. N Engl J Med. 1976;294:405–410.

4 W Hryniuk HB. The importance of dose intensity in chemotherapy of metastatic breast cancer. J Clin Oncol. 1984;2:1281–1288.

5 Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. FEBS J Internet. 2010 cited 2022 Feb 4;277:301–308. Available from: https://pubmed.ncbi.nlm.nih.gov/19922469/.

6 Singh V, Ram M, Kumar R, et al. Phosphorylation: Implications in Cancer Internet. Protein J. Springer New York LLC; 2017 cited 2021 Jul 5. Available from: https://pubmed.ncbi.nlm.nih.gov/28108801/.

7 Swietlik JJ, Sinha A, Meissner F. Dissecting intercellular signaling with mass spectrometry–based proteomics Internet. Curr. Opin. Cell Biol. Elsevier Ltd; 2020 cited 2021 May 19. p. 20–30. Available from: https://pubmed.ncbi.nlm.nih.gov/31927463/.

8 Ardito F, Giuliani M, Perrone D, et al. The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review) Internet. Int. J. Mol. Med. Spandidos Publications; 2017 cited 2021 Jun 6. p. 271–280. Available from: /pmc/articles/PMC5500920/.

9 Walters RW, Shukla AK, Kovacs JJ, et al. beta-Arrestin1 mediates nicotinic acid-induced flushing, but not its antilipolytic effect, in mice. J Clin Invest Internet. 2009 cited 2015 Apr 28;119:1312–1321. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2673863&tool=pmcentrez&rendertype=abstract.

10 PH H, ER M, AM X, et al. Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth. Mol Biosyst Internet. 2010 cited 2021 Jul 31;6:1227–1237. Available from: https://pubmed.ncbi.nlm.nih.gov/20461251/.

11 Miller MA, Meyer AS, Beste MT, et al. ADAM-10 and -17 regulate endometriotic cell migration via concerted ligand and receptor shedding feedback on kinase signaling. Proc Natl Acad Sci U S A Internet. 2013 cited 2015 Feb 10;110:E2074-83. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3670354&tool=pmcentrez&rendertype=abstract.

12 Adrain C, Freeman M. Regulation of receptor tyrosine kinase ligand processing. Cold Spring Harb Perspect Biol Internet. 2014 cited 2015 Sep 16;6. Available from: http://cshperspectives.cshlp.org/content/6/1/a008995.abstract.

13 KM N, FM W, DA L, et al. Robust co-regulation of tyrosine phosphorylation sites on proteins

reveals novel protein interactions. Mol Biosyst Internet. 2012 cited 2021 Jul 31;8:2771–2782. Available from: https://pubmed.ncbi.nlm.nih.gov/22851037/.

14 Lin Y, Yang Z, Xu A, et al. PIK3R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3β/CTNNB1 signaling pathway. Sci Rep Internet. 2015 cited 2017 Apr 16;5:8997. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25757764.

15 Macdonald-Obermann JL, Pike LJ. Different Epidermal Growth Factor (EGF) Receptor Ligands Show Distinct Kinetics and Biased or Partial Agonism for Homodimer and Heterodimer Formation. J Biol Chem Internet. 2014 cited 2017 Jan 5;289:26178–26188. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25086039.

16 Brewer MR, Yun CH, Lai D, et al. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. Proc Natl Acad Sci U S A Internet. 2013 cited 2022 Feb 9;110. Available from: https://pubmed.ncbi.nlm.nih.gov/24019492/.

17 Daub H, Weiss FU, Wallasch C, et al. Role of transactivation of the EGF receptor in signalling by G-protein-coupled receptors. Nature Internet. 1996 cited 2015 Dec 10;379:557–560. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8596637.

18 PJ B, LY F, VM D, et al. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol Cell Proteomics Internet. 2010 cited 2021 Jul 30;9:84–99. Available from: https://pubmed.ncbi.nlm.nih.gov/19770167/.

19 Gerritsen JS, White FM. Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells. Expert Rev Proteomics [Internet]. 2021 [cited 2022 Feb 24];18:661–674. Available from: https://pubmed.ncbi.nlm.nih.gov/34468274/.

CHAPTER 2

REVIEW

Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells

Jacqueline Gerritsen, M.Sc.

Forest M. White, Ph.D.

Expert Review of Proteomics, ***2021*** *18(8): 661-674*

Abstract

Background: Protein phosphorylation is crucial in regulating a plethora of cellular signaling processes. Many pathologies, including cancer, have been associated with aberrant phosphorylation-mediated signaling networks that drive altered cell proliferation, migration, metabolic regulation, and can lead to systemic inflammation. Phosphoproteomics, the large-scale analysis of protein phosphorylation sites, has emerged as a powerful tool to define signaling network regulation and dysregulation in normal and pathological conditions.

Areas covered: We provide an overview of methodology for global phosphoproteomics as well as enrichment of specific subsets of the phosphoproteome, including phospho-motif enrichment of kinase substrates. We review quantitative methods, advantages and limitations of different mass spectrometry acquisition formats, and computational approaches to extract biological insight from phosphoproteomics data. Throughout, we discuss various applications and their challenges in implementation.

Expert Opinion: Over the past 20 years the field of phosphoproteomics has advanced to now enable deep biological, and in some cases, clinical insight through the quantitative analysis of signaling networks. Future areas of development include Clinical Laboratory Improvement Amendments (CLIA)-approved methods for analysis of clinical samples, continued improvements in sensitivity to enable analysis of small numbers of rare cells and tissue microarrays, and computational methods to integrate data resulting from multiple systems-level quantitative analytical methods.

Introduction: Phosphoproteomics, signaling and cancer

Phosphorylation is a reversible post-translational modification(PTM) that is critical for regulating inter- and intracellular signaling networks [1]. In mammalian cells, phosphorylation typically occurs on the side chains of three amino acids; serine, threonine and tyrosine, although phosphorylation of other residues has been reported [2,3]. The addition of a phosphate group can have a variety of effects, as phosphorylated proteins may have altered conformation, stability, activity, subcellular localization, or protein-protein interactions. Due to its ability to modify protein function like a modular switch mechanism, phosphorylation is crucial to the regulation of signal transduction pathways [4].

Most major cellular processes, including proliferation, migration, apoptosis, and others, are regulated by protein phosphorylation-mediated signaling cascades that are also critical for relaying information about the external cell micro-environment and internal cell state. Dysregulated protein phosphorylation signaling due to aberrant kinase or phosphatase activity has been associated with a host of human pathologies [5–7]. Indeed, many hallmarks of cancer, including sustained proliferation, resistance to cell death, angiogenesis, avoiding growth suppression and invasion and metastasis can be linked to dysregulated signaling pathways and inappropriate kinase activity [8,9].

The transformative potential of multiple constitutively activated kinases, as well as the role of protein phosphorylation in regulating other aspects of biology, has fueled a deep interest in protein phosphorylation, including studies at the single protein level, protein complexes, enzyme-substrate relationships, or at the level of the phosphoproteome, the compendium of

protein phosphorylation sites in a given biological sample. Phosphoproteomics, the large-scale analysis of protein phosphorylation sites, was pioneered in 2002 by Ficarro *et al.*, and has developed rapidly over the past few decades [10]. Although phosphoproteomics may be performed using a variety of instruments and can encompass both targeted and discovery analyses, phosphoproteomics-based mapping of phosphorylation events in a large-scale, relatively unbiased manner mainly relies on mass spectrometry (MS)-based approaches [11,12]. Alternative techniques to measure protein phosphorylation include immunofluorescence / immunohistochemistry, phospho-flow, reverse-phase protein microarrays, and multiple different forms of western blotting. Although these techniques are widely used, they are dependent on antibody availability and specificity, and can be limited in the number of phosphorylation sites monitored per analysis [13,14]. By comparison, MS-based methods require minimal a priori knowledge, can identify and quantify >10,000 phosphorylation sites in a given sample, and provide high specificity by directly sequencing the site of protein phosphorylation.

MS-based phosphoproteomics has the potential to uncover activated signaling networks and novel targets in cancer cells, yet there are some inherent challenges. For instance, phosphorylation is a reversible modification that can be highly dynamic on the seconds-to minutes time scale [15]. Additionally, the phosphoproteome comprises approximately 0.1% of the proteome, and low-level phosphorylation events such as phosphotyrosine comprise only 0.1-1% of the phosphoproteome [15]; in many cases these ultra-low-abundance phosphorylation events are critical to decipher cellular signaling networks mediating oncogenic initiation and progression. Thus, MS-based phosphoproteomics must be able to identify and

quantify ultra-low level, dynamic phosphorylation events. At the same time, some highly abundant proteins are phosphorylated at high stoichiometry, thus MS-based phosphoproteomics must also be able to handle a large dynamic range of phosphorylation.

Despite these challenges, MS-based phosphoproteomics has already proven to be capable of generating valuable phosphorylation data leading to biological insight [16,17]. It has been used to detect and validate potential biomarkers and drug targets, for example in identifying kinases and their phosphorylation sites not previously known to be involved in cellular signaling in multiple disease states [18–20]. Recently, progress was made in the mechanistic understanding of the hedgehog pathway; Scheidt *et al.* showed that aberrant signaling in this pathway is correlated with various cancers [21]. Phosphoproteomics may lead to new biomarkers for drug development, and also can elucidate resistance mechanisms and other mechanisms of action in disease [22–24]. For example, as cancer therapy development becomes increasingly focused on personalized medicine, phosphoproteomics has been used to uncover cancer cell signaling networks in patient tissues and signaling signatures of response to tyrosine kinase inhibitors [25–28]. Here we will focus on MS-based phosphoproteomics, including technical aspects of the analysis, challenges in implementation, and applications where phosphoproteomics has been used to uncover novel information in cancer signaling networks.

Sample preparation and phosphopeptide enrichment

While MS is the method of choice for large-scale analysis of phosphorylated proteins, high-yield sample preparation with minimal losses of specific classes of proteins is important, especially for analysis of ultra-low level phosphorylation sites (Figure 1). The first steps in MS-based

sample preparation are cell lysis and protein extraction. Many protocols now utilize a urea-based buffer to lyse cells, as it rapidly denatures proteins to preserve the physiological modification state of proteins, and is easily removed by desalting in later steps. If heated, urea-based buffers can lead to the carbamylation of proteins; higher temperatures for longer times can exacerbate this effect. As an alternative, guanidine hydrochloride (Gnd-HCl) can be used, as this buffer allows for heating to high temperature for better solubilization, recovery and denaturation. However, Gnd-HCl can negatively impact digestion efficiency, so additional dilution needs to be performed prior to adding proteolytic enzymes. Although sodium dodecyl sulfate (SDS) is a powerful chaotropic agent that is widely used in molecular biology, western blotting, and protein arrays, it is challenging to remove in sample processing steps for MS analysis and can suppress protease activity and MS signal. Recent clean-up strategies like SP3 and S-trap have been developed that can overcome this hurdle, although complete removal of SDS can still be challenging [29,30]. Acid-labile surfactant is an MS-friendly alternative to SDS, yet due to the relatively high cost of this reagent and the additional clean-up steps required, most groups favor urea-based cell lysis.

Following cell lysis, denatured proteins are typically reduced and alkylated, most commonly using dithiothreitol (DTT) and iodoacetamide (IAA), respectively.[18] After chemical modification, digestion of proteins into peptides is performed in bottom-up proteomics. Trypsin is the most widely used protease, as it cuts C-terminal to arginine and lysine residues; due to their frequency in the proteome, trypsin tends to produce peptide lengths that are compatible with MS analysis [31]. Additional benefits include the wide availability and high specificity of trypsin. Peptides generated by trypsin proteolysis tend to have improved ionization and

fragmentation due to the basicity of the arginine and lysine residues at the C-terminus, thereby enhancing sequencing by MS. As an alternative choice, Lys-C cleaves C-terminal to lysine and is compatible with a range of different buffer conditions. Lys-C is commonly used in combination with trypsin in a double-digestion protocol, with the goal of reduced missed cleavages [32]. While trypsin is the most commonly preferred protease, it is worth noting that many phosphorylation sites fall in regions of the protein that are not amenable to MS-analysis following trypsin digestion. To gain deeper coverage of the phosphoproteome, some researchers combine data resulting from proteolysis with multiple enzymes (typically a single enzyme at a time), including GluC (V8 protease), Chymotrypsin and Elastase [33–35]. Regardless of the protease used, it must be ensured that the lysis buffer used is compatible with the protease or diluted sufficiently as not to inhibit protease activity [36].

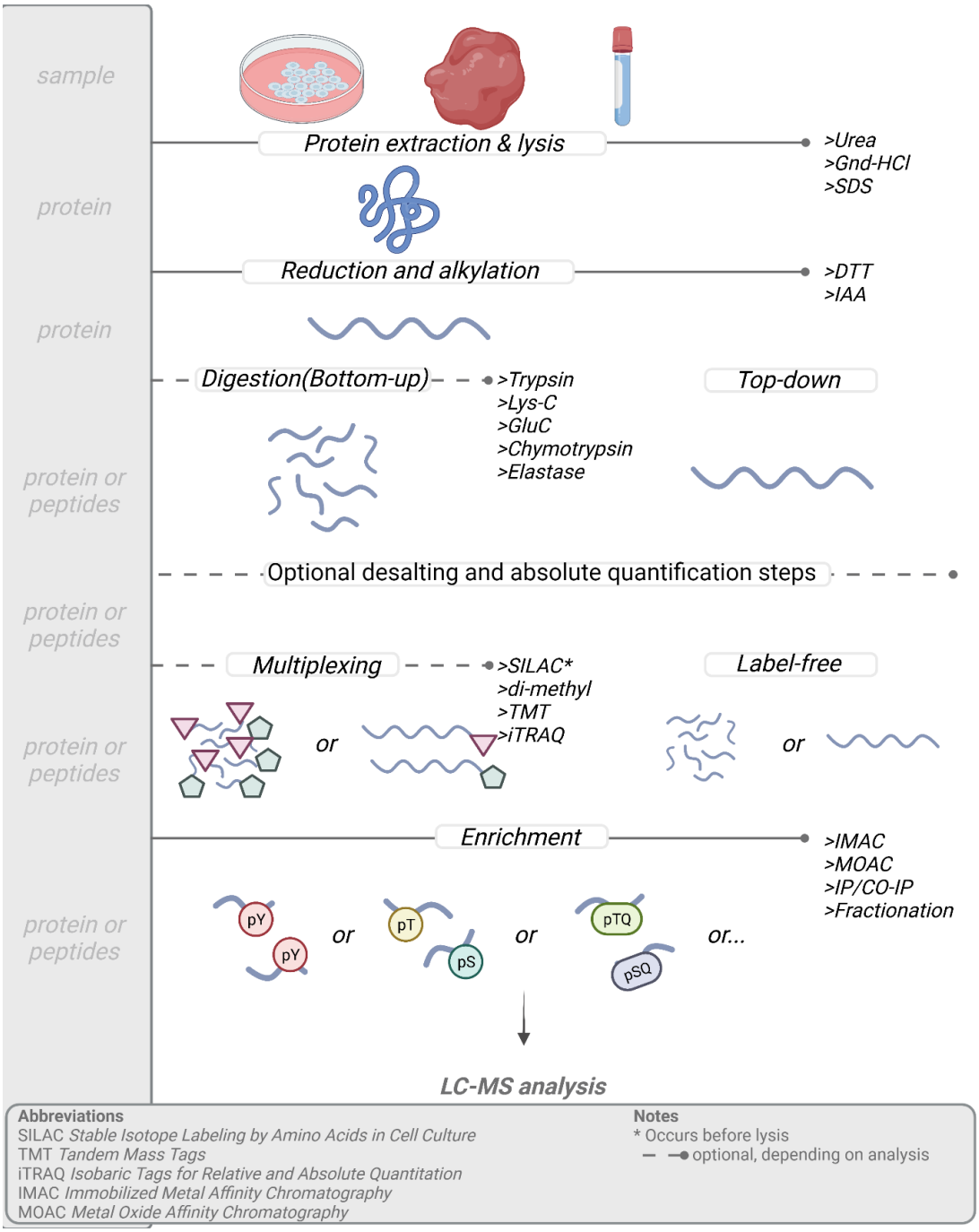


Figure 5: General sample preparation protocol for mass-spec based phosphoproteomic analysis. Protein is extracted from in vitro cell cultures or in vivo tissue or liquid samples, followed by chemical modification. In bottom-up proteomics, proteins are digested into peptides. Quantification can be performed by multiple methods, including non-isobaric and isobaric labeling, many of which are applied after digestion and before enrichment. Phosphopeptide enrichment purification can be performed at the global or subset-enrichment level depending on the biological question or experimental goals. Enrichment steps are critical in ensuring detection of these low-abundant peptides before analysis. Created with biorender.com.

Following digestion, phosphoproteomic workflows can take multiple directions. Some groups choose to desalt peptides prior to chemical modification (e.g., non-isobaric or isobaric labeling, see below), while other applications may skip these steps and move straight to phosphopeptide enrichment. It is worth noting that the optimal workflow for a given application may not follow either of these approaches. We recommend developing a simple protocol with minimal unnecessary steps that could lead to increased peptide loss, especially for enrichment and analysis of low-level phosphopeptides from small amounts of starting material.

Due to the low stoichiometry of phosphoproteins compared to non-phosphorylated proteins, enrichment is a key step towards successful detection of phosphorylated peptides [37]. For large-scale phosphoproteomic analysis, phosphopeptide enrichment is typically performed by either immobilized metal affinity chromatography (IMAC) or by metal-oxide affinity chromatography (MOAC). Both of these techniques depend on the affinity and coordination of negatively charged phosphate groups towards positively charged metal ions such as Fe^{3+} (IMAC) and Ti^{4+} (MOAC). [38] Since binding to the metal occurs through the phosphate moiety, phosphorylated serine, threonine, and tyrosine appear to be enriched equivalently, although phosphorylated serine (pSer) and phosphorylated threonine (pThr) constitute approximately 90% and ~9-10% of the phosphoproteome, respectively, with phosphorylated tyrosine comprising the remaining 0.1-1%. Since the selectivity of most metal ions for phosphorylated peptides is often imperfect, nonspecific binding of non-phosphorylated peptides can confound the analysis, especially for low-level samples [10,39]. Non-specific binding can be mitigated through a variety of techniques, including addition of organic acids to MOAC-based enrichment, chemical modification of carboxylate groups, or by using a nitrilotriacetic acid (NTA)-based resin

for Fe^{3+}-IMAC. While several studies have pointed to the complementary nature of IMAC and MOAC in phosphopeptide enrichment, [40] using an excess of either material enables similar degree of enrichment, and thus selecting either IMAC or MOAC should suffice for most large-scale phosphoproteomic analyses [41]. For most applications, sufficient depth of coverage can be achieved with a simple one-step enrichment protocol (e.g., NTA-Fe^{3+} IMAC spin columns), and additional depth can be afforded by fractionation of the sample prior to phosphopeptide enrichment of each fraction [42].

Using any of the above strategies to enrich phosphopeptides from mammalian cell lysate for analysis on a high-resolution mass spectrometer, it is fairly straightforward to identify thousands of high confidence phosphopeptides from a single sample. While large-scale data generation has become relatively easy, unfortunately, gaining biological understanding from this data remains challenging, largely due to the complexity of the phosphoproteome. Moreover, the overwhelming abundance of pSer sites are generally static (e.g., unaltered by a given biological perturbation), and their role in regulating the biology of the system is poorly understood. For instance, in one of the early large scale phosphoproteomics analysis, Olsen *et al.* identified and quantified over 6,000 phosphosites in HeLa cells stimulated with EGF at several time points [43]. Despite this massive data set, ~85% of the phosphosites were unaffected by the stimulation conditions. As might be expected given their overall low abundance in the phosphoproteome only ~103 pTyr sites out of 6,600 total phosphorylation sites (~1.6%) were identified in this analysis, despite stimulation of the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase (RTK) that is highly expressed in these cells. Similar results have been seen for a range of other systems, including recent large-scale

phosphoproteomic studies of human tumors, where ~34,000 total phosphopeptides were identified and quantified, yet only ~2% of these were pTyr sites [44,45]. Since aberrantly activated tyrosine kinases are known to be oncogenic and tyrosine phosphorylation sites on kinase activation loops are known to be critical regulators of kinase activity, the paucity of tyrosine phosphorylation sites in these data sets would suggest that some of the potentially most informative signals might be missed in these large-scale studies [46–48].

Enrichment of subsets of the phosphoproteome

In order to access low-abundance phosphorylation sites within the complex, high dynamic range phosphoproteome, it is often necessary to perform an additional enrichment step. For instance, if the goal is to identify and quantify RTK-driven signaling networks in a given cell line or tissue specimen, specific enrichment for pTyr-containing proteins or peptides will significantly improve the analysis [49–51]. Perhaps the most common method for pTyr-enrichment involves using pan-specific anti-pTyr antibodies to immunoprecipitate pTyr-containing peptides following tryptic digestion [49,50]. There are multiple pan-specific anti-pTyr antibodies available commercially; while each one has affinity for a range of pTyr sites, each tends to have some degree of bias. We have found that mixing several antibodies together can give greater overall coverage of the pTyr sites in a given biological sample [52,53]. As with most enrichment methods, the amount of non-specific binding in pTyr-immunoprecipitations tends to increase with decreasing sample amount. To address this issue, a second stage of enrichment (e.g., with an Fe^{3+}-NTA IMAC spin column) can be used to reduce the level of non-phosphorylated peptides present in the pTyr IP. Alternative strategies to enrich pTyr peptides

or proteins have been explored. In cell signaling networks, Src-homology 2 (SH2) domains bind to pTyr sites within a protein or on a different protein to regulate activity and protein-protein interactions (i.e., in the context of recruitment of adaptor proteins to an activated RTK) [54]. Taking advantage of their binding affinity, several groups, including by Bian *et al.,* have utilized SH2 domains for enrichment of pTyr containing peptides, and have also engineered SH2 domains for increased binding affinity [55,56]. These "SH2 superbinders" have enabled enrichment of thousands of pTyr sites, yet different SH2 domains appear to have different specificity, thus sequential analysis or mixing SH2 domains for a given analysis might provide increased depth of coverage [57].

The strategy of enriching phosphorylation site subsets for deeper analysis can be applied to a range of biological applications requiring cellular signaling network analyses. For instance, cells respond to many environmental contaminants and cytotoxic chemotherapies by initiating a DNA damage response; in both contexts this response is critical for halting the cell cycle, fixing DNA lesions / adducts, and enabling the cell to survive the damage. To identify critical regulators of the DNA damage response, cells were treated with ionizing radiation and phosphopeptides were enriched using a combination of phosphorylation site specific antibodies [58]. In this study, the non-specificity of the antibodies resulted in enrichment and identification of hundreds of phosphorylation sites potentially involved in DNA damage response signaling networks. An alternative approach is to use phospho-motif specific antibodies that recognize phosphorylation sites within a particular amino acid sequence. For instance, ATM and ATR kinases tend to phosphorylate serines or threonines followed by a glutamine; ATM/ATR substrate phospho-motif antibodies therefore enrich pSer or pThr

followed by glutamine (e.g., pSQ/pTQ) [59]. Using these antibodies for peptide IP from proteolyzed cell lysate allows for the selective enrichment of hundreds of pSQ/pTQ containing peptides in a single analysis. Similarly, the ERK 1/2 mitogen activated kinases can regulate cell proliferation and tend to phosphorylate S or T residues in the context of a proline in the -2 position and in the +1 position. ERK 1/2 substrate motif antibodies can therefore be used to IP hundreds of peptides containing a PXXpSP or PXXpTP motif, where X represents any amino acid. In both of these cases, as with multiple other available phospho-motif specific antibodies, it is critical to note that all peptides that match these phosphorylation motifs are not necessarily substrates of the given kinase [60]. Indeed, many peptides that happen to have the motif of interest are quantitatively unaffected by activation or inhibition of the kinase of interest. As with pTyr enrichment, the complexity of the phosphoproteome tends to obscure many of these phosphomotif-containing peptides; recent large-scale phosphoproteomics data sets contain tens of pSQ/pTQ peptides, while specific enrichment using phospho-motif antibodies would likely provide over five hundred pSQ/pTQ-conatining peptides from the same sample [44,59]. It is worth noting that phosphorylation subset enrichments can be performed serially on the same sample to gain increased depth of coverage on multiple biological pathways [61]. For instance, cancer cells are often driven by aberrantly activated RTKs that signal through the ERK 1/2 MAP kinases to drive proliferation while also activating protein kinase B (AKT) to promote cell survival. Proliferating cells incur DNA damage during replication and therefore activate a DNA damage response. Each of these networks could be interrogated in a given biological sample, or

across multiple biological samples, by serial IP using pan-specific and phospho-motif specific antibodies, followed by global phosphoproteomics on the resulting supernatant.

Quantitative phosphoproteomics methods

As mentioned above, with either subset enrichment strategies or with global phosphoproteomics analyses, it is now possible to identify thousands to well over 10,000 phosphorylation sites from a given biological sample, respectively. Despite this plethora of data, gaining biological knowledge from these analyses is still quite challenging, as identification of a phosphorylation site on a protein in a given sample does not necessarily mean that the site is regulatory of the biological processes in that sample. To infer biological knowledge from phosphoproteomics, quantitative data comparing phosphorylation levels across different conditions (i.e., cell stimulation with a mitogen, treatment with a given kinase inhibitor, or treatment with a chemotherapy agent, among others) is often required. Multiple label-free or label-based quantification strategies can be employed to determine the relative abundance of phosphorylation sites between various conditions of a biological system [62,63]. The simplest of these approaches is label-free quantification, in which the amount of a given phosphorylation site, typically estimated by the area under the curve (AUC) of the chromatographic elution profile of the precursor ion, is compared across phosphoproteomic analyses of different biological conditions. This approach is straightforward, can be highly quantitative, does not require additional sample handling, and can be used to compare across hundreds of samples without requiring a normalization channel. Moreover, since quantification typically occurs from the precursor signal intensity, dynamic range signal compression, e.g., from MS/MS-based

quantification techniques, see below, may be less of an issue. On the other hand, the vast number of phosphopeptides present in global phosphoproteomics experiments can lead to highly complex full scan mass spectra (MS1 spectra), and may result in overlap of isotope envelopes and inaccurate quantification, especially on mass spectrometers with lower resolving power. Additionally, sample handing and chromatographic reproducibility are critically important for accurate label-free quantification. More complex workflows, including phosphopeptide IP's, can adversely affect the accuracy of label-free quantification. Early versions of label-free quantification relied on spectral counting, in which the number of MS/MS spectra for a given precursor were compared across different conditions. This approach is most accurate when applied to highly abundant peptides and can be confounded by low-level peptides (typically less than 5 spectral counts), where stochasticity between runs can confound quantitative accuracy.

As an alternative to label-free quantification, multiplexed, label-based strategies have been developed. These approaches fall into two categories: non-isobaric labels (MS1 quantification) and isobaric labels (MS/MS quantification). Non-isobaric labels, as the name implies, utilize different numbers of heavy isotopes on each tag to generate labels for each sample that differ in mass. Non-isobaric labels can be applied at multiple steps in sample generation or sample processing. Stable Isotope Labeling with Amino Acid in Cell Culture (SILAC), as described by Mann's group, is one of the most commonly used non-isobaric multiplexing methods [64]. In SILAC, cells from one condition are cultured in media that contains one or more heavy-isotope labeled amino acids, while cells from another condition are cultured in media with corresponding light-isotope labeled amino acids. Lysine or arginine are most commonly used

for labeling to ensure that tryptic peptides will be quantifiable. Although SILAC labeling can be expensive, especially for in vivo applications [65,66], the protocol is relatively easy to implement, and since labeling occurs during cell culture (or organism growth), samples can be mixed early in sample processing and quantification is relatively unaffected by variation in sample handling. However, SILAC is typically limited to multiplexing of 2 or 3 conditions and results in increased complexity of full scan mass spectra that can adversely affect quantification dynamic range [67]. SILAC has frequently been used in combination with phosphoproteomics to quantify cancer cell signaling networks [68–71]. For instance, Zhang *et al.* combined pTyr IP with SILAC to quantify the adaptive response of lung adenocarcinoma cells to EGFR inhibition [72], and Cunningham *et al.* combined SILAC with global phosphoproteomics to define fibroblast growth factor receptor (FGFR) signaling networks in triple negative breast cancer [73].

Beyond SILAC, non-isobaric chemical labels have also frequently been used for quantitative analysis of phosphorylation-mediated signaling networks. In this approach, peptides from each sample are chemically modified after proteolytic digestion, with each sample being tagged with the same label with different number of heavy isotopes. For instance, di-methyl labels, with one sample labeled with 12C-H2 formaldehyde and other samples labeled with 12C-D2 or 13C-D2 formaldehyde offer an inexpensive alternative to SILAC and have been used to elucidate phosphorylation changes following cell stimulation in a variety of contexts [74,75]. This approach is similarly limited to multiplexing of 2 or 3 samples and can be affected by complexity of MS1 spectra due to the combination of multiple non-isobarically tagged samples.

Isobaric tags label all samples with a tag of the same mass; quantification occurs in MS/MS through production and detection of reporter ions generated during fragmentation [76,77]. Isobaric tags enable multiplexing of up to 18 samples [78] in a single analysis with relatively minimal increase in the complexity of MS1 spectra. This high degree of multiplexing reduces MS analysis time while also reducing inter-analysis variability, including chromatographic differences and precursor selection for MS2. While inter-analysis irreproducibility in discovery-mode analyses can still be a problem for larger sample sets [27], 16- or 18-plex isobaric tags significantly reduce this potential problem. Since the *m/z* ratio for each tagged peptide is identical from all samples, isobaric tags have been used for 'boost' experiments for phosphoproteomics, among other applications, wherein one of the samples is present in much higher amount, improving the signal-to-noise ratio in full scan mass spectra and driving selection of peaks for MS/MS. This approach can improve sensitivity, enabling analysis of smaller amounts of samples in the 'non-boost' channel, but may lead to decreased quantitative accuracy [79]. Dynamic range compression, one of the major potential problems with isobaric labels, is thought to be due to co-isolation of multiple peptides for MS/MS. This problem is exacerbated in complex mixtures, including global phosphoproteomics, due to the massive number of peptides in the sample. Although narrowing the isolation window can reduce dynamic range compression, MS3 may provide more accurate quantification as demonstrated by McAllister *et al.* [80], but may not be applicable to low-abundance peptides including pTyr and phospho-motif enriched peptides [62,79]. Enriched subsets of the phosphoproteome tend to be much less complex, and thus dynamic range compression is less of an issue, although it can still adversely affect quantitative accuracy. In addition to dynamic range compression,

isobaric labels tend to be similar in cost to SILAC, and introduce at least one additional step in the workflow, thereby risking sample loss. Even with these issues, isobaric labeling has enabled a wide range of quantitative phosphoproteomic analyses, including quantification of pTyr dynamics following growth factor stimulation [81,82], quantification of pTyr signaling networks from formalin fixed parrafin embedded (FFPE) sections of human tumor specimens by Kohale *et al.* [26], and quantification of adaptive response to therapy in cancer models and tumor tissues [63,83,84], among many others.

MS-based analysis of the phosphoproteome: applications and challenges

Proteomics experiments can be generally divided into top-down / middle down and bottom-up proteomics. In top-down proteomics, intact proteins are ionized and analyzed by mass spectrometry (MS) in the absence of a digestion step [85]. In middle-down proteomics, typically a single digestion step, e.g., Cyanogen bromide (CNBr)-digestion, is used to produce a few large protein fragments which are then ionized for MS analysis [86]. Both of these approaches enable identification of proteoforms: the post-translation modification 'code' present on a given protein, although middle-down approaches can suffer from data integration requirements to reconstitute the intact protein [87,88]. Although top-down and middle-down proteomics have been extensively used to analyze the post-translational modification code of histones, to date there are only a few applications of these approaches to the phosphoproteome [89–92], with no in-depth analysis of signaling network alterations between conditions. The dearth of top-down phosphoproteomics experiments might be due to the inherent challenges of top-down proteomics, as characterization of intact proteins, and especially PTMs on intact proteins, can

require significant time, effort, and expertise. Additionally, dynamic PTMs such as phosphorylation are highly challenging, as each new phosphorylation site results in a new proteoform, and thus a dozen or more sites on a given protein may lead to a combinatorial explosion of protein states. With that said, top-down and middle-down phosphoproteomics are needed to define the associations between phosphorylation sites across a given protein and to answer fundamental questions, including whether EGFR or other RTKs are phosphorylated on multiple sites on the C-terminal tail in a given isoform, or whether each of the multiple sites are mostly exclusive to a given isoform. Almost all phosphoproteomics experiments have been performed by bottom-up proteomics, where proteins are proteolyzed to peptides, typically using enzymes such as trypsin (see above). In general, bottom-up proteomics experiments are much easier compared to top- or middle-down experiments, but this ease of analysis comes at the cost of information that may be critical for understanding protein function and signaling network mechanisms. For instance, can phosphorylation of an inhibitory site, e.g., the C-terminal phosphorylation site on Src-family kinases, co-occur with phosphorylation of the activation loop? Since these sites are separated by ~100 residues, this information is only available by top-down or middle-down analyses.

Bottom-up proteomics can be sub-divided into three subcategories based on the data acquisition method: data-dependent acquisition (DDA), data-independent acquisition (DIA) and targeted phosphoproteomics [93](Figure 2). These methods mainly differentiate from each other by the manner in which specific ions are chosen for fragmentation. DDA, also known as discovery mode, generally selects the most abundant peptides for isolation and fragmentation. When coupled with fractionation, DDA experiments can provide deep characterization of the

phosphoproteome [61,94,95], and are still the most common approach to signaling network analyses. Although the total number of detected peptides is generally higher with DDA compared to DIA or targeted approaches, DDA tends to suffer from run-to-run irreproducibility of identified and quantified peptides, due to the stochastic nature of peptide sampling. DDA-based quantification of cancer signaling networks in cell lines and tumors can suffer from the 'sparse-matrix' problem, where only a subset of phosphorylation sites are quantified across all samples, even when coupled with multiplexed isobaric reagents [27,96,97]. In DIA, pre-set windows of m/z ratios are sequentially isolated, fragmented, and analyzed, with the goal of covering the full m/z ratio rapidly enough to fragment and analyze all peptides present in a given sample. Since peak selection is absent in DIA, run-to-run reproducibility tends to be much greater compared to DDA.[98] However, depending on the width of the isolation window and the sample complexity being analyzed, DIA can be adversely affected by dynamic range, e.g., loss of the lowest abundance fragments in the presence of highly abundant fragments. Due to the complex MS/MS spectra, DIA is also best applied with a library of spectra, typically obtained from multiple DDA experiments, and is suboptimal for discovery. However, more recent DIA experiments have utilized smaller isolation windows that are on-par with DDA experiments; these new settings suggest a powerful combination of high reproducibility with the potential for discovery of novel signaling components. Quantification with DIA tends to occur through label-free, as the duty cycle for full scan MS is suboptimal for non-isobaric labels, and the mixed MS/MS spectra are suboptimal for isobaric labels. As mentioned above, label-free quantification for DIA can be problematic when working with more complex workflows involving IP of pTyr or phospho-motifs, both of which are highly susceptible to slight variations

between sample processing steps. Nevertheless, DIA methods have been applied to analysis of signaling networks for multiple applications [99–101], and new tools are emerging to facilitate these approaches [102].

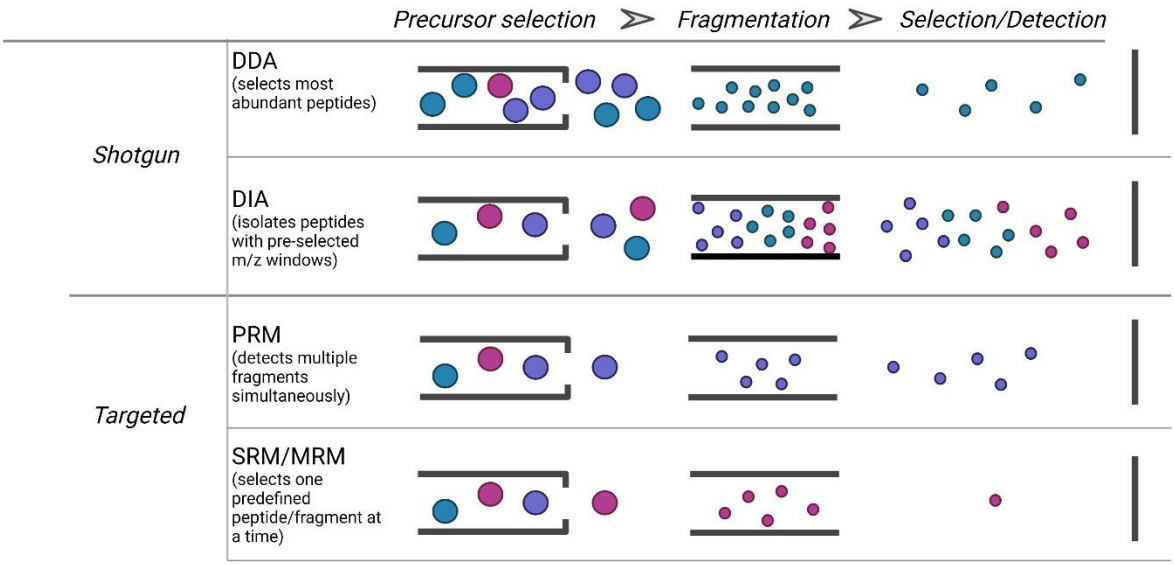


Figure 6: Types of MS data acquisition modes in bottom-up proteomics. data dependent acquisition (DDA) selects the most abundant peptides which are then isolated and fragmented sequentially. Data Independent Acquisition (DIA) isolates and fragments peptides within defined m/z windows. In targeted acquisition mode, the instrument is set up to detect and fragment selected peptides that are defined a priori. In parallel reaction monitoring (PRM), multiple fragments per precursor can be detected at the same time. In SRM or MRM, fragment ions from a given precursor are typically detected sequentially. Created with biorender.com.

Targeted phosphoproteomics provides consistent reproducibility with high accuracy quantification and can provide high sensitivity, especially when coupled with internal standard trigger peptides, but it comes at the cost of discovery and coverage as shown in our research group [27]. As the name implies, targeted methods require prior knowledge about the signals (e.g., phosphopeptides) of interest, as the instrument method is typically constructed to select the precursor m/z ratios of interest for fragmentation and quantification. Targeted methods typically utilize multiple reaction monitoring (MRM) or parallel reaction monitoring (PRM)

based methods to isolate and fragment the precursor of interest. In MRM, specific fragment ions are detected sequentially, typically on a triple quadrupole MS, while in PRM, all fragment ions are detected in the same analysis, on an ion trap or TOF MS. Although targeted methods have historically been used to quantify a few targets per analysis, MRM-based targeted phosphoproteomics have been used to monitor signaling networks. In one example application by Wolf-Yadlin *et al.*, 226 pTyr-containing peptides were quantified by MRM to assess the dynamic signaling network response to EGF stimulation at multiple time points [103]. MRM methods have also been used to quantify the DNA damage response to chemotherapy; in this example, a combination of phospho- and non-phosphopeptides were targeted for quantification [104,105]. In targeted analyses with a large number of targets, elution time windows have allowed for longer detection times while maintaining high duty cycles. More recently, heavy-isotope labeled internal standard (IS) trigger peptides have bypassed the need for elution time windows; in this approach, fragmentation of the heavy labeled IS trigger peptide is coupled with on-the-fly pseudo-spectral matching to a known fragment ion fingerprint [106]. Matching of multiple fragment ions initiates high resolution MS/MS of the endogenous peptide, and quantitative data can be generated by comparing the light:heavy ratio of multiple fragment ion peaks. The SureQuant framework available on selected ThermoFisher MS instruments allows for monitoring of hundreds of IS-peptides. We have recently developed an approach (SureQuant pTyr) to quantify pTyr signaling networks comprised of ~400 pTyr peptides in EGF-stimulated cell lines and across dozens of human colorectal tumor specimens [27]. Although SureQuant pTyr provides quantitative data for each detected peptide in each sample, quantification relies on one-point calibration and assumes a

linear dynamic range. To generate absolute quantification data with internal standard curves for each peptide, we developed Multiplex Absolute Regressed Quantification with Internal Standards (MARQUIS), a multiplexed MRM- or PRM-based targeted approach in which different amounts of heavy-isotope labeled IS phosphopeptides were added to each biological sample. Following the addition of these reference peptides, multiple samples were isobarically labeled and combined for analysis [107]. Reporter ions from MS/MS of the TMT-labeled endogenous peptides could then be regressed against the internal standard curve provided by MS/MS of the TMT-labeled IS peptides. Application of this approach to ~20 pTyr phosphorylation in the EGFR signaling network provided absolute quantification, e.g., copies/cell, for the temporal dynamic profiles for each phosphorylation site and relative stoichiometry information for multiple sites on the EGFR C-terminal tail. In the future, combining a MARQUIS-style approach with SureQuant pTyr may provide absolute quantification of hundreds of nodes in the network, enabling more mechanistic computational modeling of cancer cell signaling networks.

Computational analysis of phosphoproteomic data

While technological and methodological advancements have greatly expanded the capabilities of phosphoproteomics, the ultimate impact depends on the biological knowledge that can be gleaned from the data. In this section, steps in data analysis as well as potential tools to gain biological insight will be discussed (Figure 3).

High quality / high accuracy phosphopeptide identification and quantification can be critical for defining cancer cell signaling networks from phosphoproteomic data [108]. Since each phosphopeptide is typically a ‘one-hit wonder’, high stringency in data filtering is recommended

to remove as many false positive assignments as possible and to localize the phosphorylation site to the correct residue, if possible. While restrictive filters may compromise the ultimate data set size, high quality data facilitates computational data analysis and ultimate biological insight.

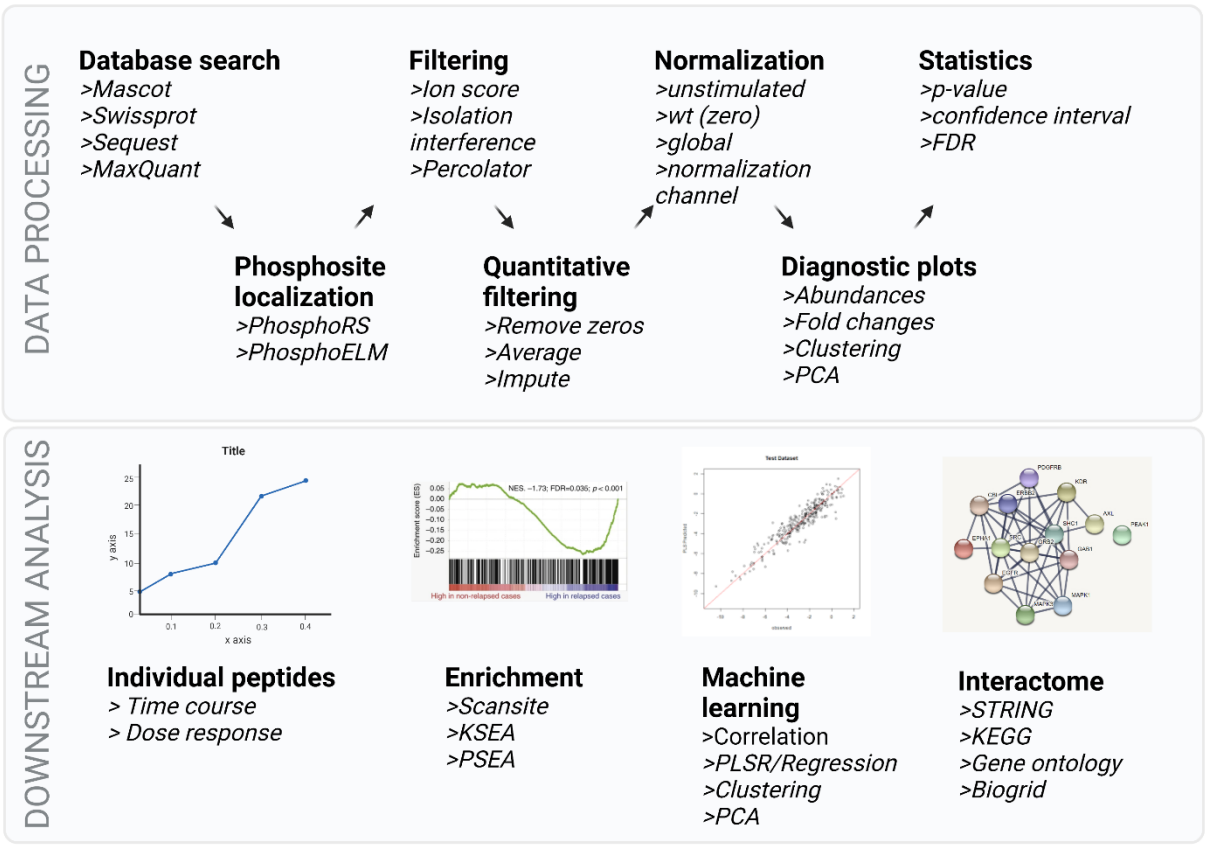


Figure 7: Steps in data processing and downstream analysis tools for phosphoproteomic data. Top: multiple data processing steps are needed in order to identify and quantify phosphorylation sites from MS/MS data, including database searching, site localization and quality filtering for identifications. Depending on the quantitative method, data may go through additional filtering, normalization, clustering, and statistical analysis as a first pass at identifying differentially phosphorylated peptides. Bottom: more nuanced biological information can be gained through additional computational analysis, including temporal analysis, kinase / substrate or pathway enrichment of phosphorylation subsets, and machine learning approaches to identify modules and pathways of phosphorylation-mediated signaling networks. Each of these tools can lead to predicted functions for phosphorylation sites in the data. Validation experiments should be performed to confirm analysis results. Created with BioRender.com.

A vast number of computational methods have been developed to extract cellular signaling network information from phosphoproteomic data [109]; here we will highlight a few general computational data analysis strategies as well as some specific instances where tools have been applied to provide novel biological insight. As a first pass approach, clustering methods, including hierarchical clustering, k-means-clustering, self-organizing maps, and others, have been applied to reveal associations between phosphorylation sites and to thereby define 'modules' of co-regulated sites across different biological conditions. In some cases, clustering has helped to predict the function of unknown phosphorylation sites, for instance when they are co-regulated with multiple sites of known function [43,49,110]. Although different clustering methods can be applied to extract additional information from a given data set, combining the results from multiple clustering methods can reveal consensus modules and provide novel insight. As one example of this strategy, the multiple cluster analysis methodology (MCAM) algorithm was developed by Naegle *et al.* Application of MCAM to dynamic phosphotyrosine data uncovered phospho-site specific interactions, including a novel interaction between EGFR phosphorylation and the PDLIM1 cytoskeletal protein [111,112]. Co-correlation can also be used to identify phosphorylation sites that are co-regulated across multiple biological conditions. This approach was applied to studies of pTyr signaling networks in glioblastoma and non-small cell lung cancer tumor tissues to highlight patient-specific activated signaling networks [26,113,114]. To gain additional insight into signaling networks from large-scale phosphoproteomic datasets, multiple computational approaches have been developed to predict kinase-phosphorylation site associations. These algorithms, e.g., ptmRS and Thesaurus, typically rely on kinase motifs, known substrates, or a combination of motifs,

substrates, and protein-protein interactions to predict kinases that may be responsible for phosphorylating particular sites in the data [115–122]. As with all predictive algorithms, the results from these analyses should be confirmed through additional experiments, e.g., by chemical or genetic perturbation of the kinase.

Integrative 'omics, combining information from several different types of data, can provide additional insight into the structure and function of phosphorylation mediated signaling networks. As one example, proteogenomics, combining proteomic and phosphoproteomic data with genomic and transcriptomic data has been used to classify subtypes of different human tumors and identify putative activated signaling networks and central nodes [44,123–125]. Other integrative 'omics tools have combined proteomics, phosphoproteomics, transcriptional profiling, and metabolomics to attempt to characterize the role of phosphorylation sites in regulating the state of a given system and to identify targeted nodes [44,126–128]. Another computational tool that allows integration of multiple datatypes is partial least squares regression (PLSR), which utilizes input matrices X (e.g., phosphoproteomic dataset) and Y (e.g., phenotypic outcome such as proliferation) to build a predictive model correlating associations between phosphorylation sites and phenotypic outcome. We have previously used this approach to identify pTyr sites associated with migration and proliferation in HER2 overexpressing mammary epithelial cells and to highlight a counter-intuitive role for the ERK 1/2 MAP kinases in EGFRvIII expressing glioma cells [96,97,129]. Frejno *et al.* used a PLSR-based analysis on multi-omics data to determine the landscape of proteome activity in a large set of cancer cell lines, with the goal of predicting drug response and novel functional associations from these networks [130]. Other approaches have been developed over the years to integrate

different types of omics as well as non-omics data, such as immunohistochemical data [83,131]. Computational analysis can be challenging with discovery-mode proteomic or phosphoproteomic datasets, due to the sparse matrix problem associated with missing observations between biological replicates or different biological conditions. From a computational perspective, one option is to disregard all peptides that are not quantified across all conditions; unfortunately this approach radically reduces the dataset size [96]. [An alternative approach is to impute quantitative values for the missing data, typically zeros, or an average of the remaining channels. Although this allows conservation of more peptide information, it risks false hypothesis generation in downstream analysis, as it assumes high similarity between biological conditions [132]. Targeted data acquisition approaches (see above) help to reduce the sparse matrix problem and should enable improved computational analysis.

Future of phosphoproteomics

MS-based phosphoproteomics has made great strides over the past decade in terms of detection limits, speed, accuracy and resolution. Phosphoproteomics has emerged as a powerful tool for analysis of signaling networks in diseased tissues and model systems, both *in vivo* and *in vitro*. Nonetheless, one of the main challenges in MS-based phosphoproteomics is the application to limited sample amounts / low abundance model systems. Progress in this area has enabled successful phosphoproteomic analysis on extracellular vesicles secreted from cancer cells to identify potential biomarkers in glioblastoma-EGFRVIII variant [133]. Additionally, methods have emerged that address the challenge of limited starting material

when working with patient tissue specimens [134], and phosphoproteomic analyses have now been successfully performed on FFPE tissue samples [26,135]. Simultaneously, given the value of patient samples and the importance of multi-omics analysis, methods are being developed that allow for simultaneous extraction of DNA, RNA and protein from samples [136].

Another challenge for phosphoproteomics is spatial analysis, due to the highly dynamic nature of this PTM. With the use of proximity labeling strategies, Liu *et al*. demonstrated the ability to monitor altered phosphorylation patterns due to ER stress in *in vitro* and *in vivo* systems [137]. We previously used phosphoproteomics to characterize the immediate-early signaling dynamics in the EGFR network and proximity ligation assays (PLA) to characterize dynamic recruitment of adaptor proteins to the membrane [82], or total internal reflection fluorescence (TIRF) microscopy to monitor *in vivo* SH2 binding dynamics and binding site kinetics [138]. Being able to directly quantify spatially resolved signaling networks by MS-based phosphoproteomic analysis has yet to be accomplished.

Another fascinating development in the field is single cell proteomics (scProteomics) [139]. This type of analysis has the potential to provide information about co-occurrence of phosphorylation sites and states of individual cells and would therefore enable significantly improved definition of cellular signaling networks. Although progress has been made at the proteome level using nanodroplet sample preparation platforms, phosphopeptide analysis remains challenging, as most signaling nodes are well below detection limits thus far [140].

Continued improvements in MS instrumentation, phosphoproteomic methodology, and computational modeling algorithms will facilitate the application of MS-based

phosphoproteomics to additional systems, including tissue microarrays, needle biopsies, and rare liquid tumors, and the interpretation of this data to define activated signaling networks in these tissues.

Expert opinion

Quantitative MS-based phosphoproteomics has begun to emerge as a powerful technique enabling the identification of activated oncogenic signaling networks from a variety of biological systems, including *in vitro* and *in vivo* model systems as well as patient tumor tissue specimens. While global phosphoproteomics can provide an impressive overview of the phosphorylation state of thousands of proteins within a given sample, this analysis method may fail to identify and quantify low-abundance sites, e.g., pTyr and many oncogenic kinase substrates, that may be critical regulators of biological function, especially in cancer cell signaling. Accessing these low abundance sites requires enrichment of selected subsets of the phosphoproteome, often through immunoprecipitation. These experiments can be more technically challenging compared to global phosphoproteomics, as the low amount of phosphorylated peptides are strongly impacted by sample losses during processing steps. Nonetheless, pTyr phosphoproteomics can provide insight into activated tyrosine kinases regulating a range of oncogenic phenotypes, including cellular migration, proliferation, invasion, therapeutic resistance, among others. Advances in pTyr proteomics, including higher affinity reagents for pTyr enrichment, targeted approaches for monitoring selected pTyr phosphorylation sites, and MS-instrumentation improvements to provide enhanced sensitivity, have begun to make this approach more turn-key and accessible to a wider range of proteomics research labs and core facilities.

The next step in the evolution of phosphoproteomics and especially for pTyr proteomics may be development of a CLIA (Clinical Laboratory Improvement Amendments) approved analysis workflow to provide clinical insight for patient stratification and therapeutic selection based on signaling network activation in clinical samples. Moving to CLIA approval will require standardization of standard operating procedure (SOPs), dedicated instrumentation, and highly reproducible quantification. Although difficult to implement, CLIA approval may be facilitated by targeted approaches such as SureQuant pTyr, where heavy-isotope encoded standard peptides are included in every analysis, yielding accurate quantification even in the context of sub-optimal pTyr enrichment. Recently we implemented SureQuant pTyr to identify activated signaling networks in dozens of human colorectal tumor specimens; importantly, this method was performed with commercially available columns, reagents, and instrumentation, and thus should be accessible to most, if not all, proteomics facilities.

Phosphoproteomics continues to be limited by sample requirements, with many studies requiring relatively large amounts of frozen tissue specimens for in vivo studies, or multiple plates of cells per condition for in vitro studies. It is worth noting that multiple labs have been developing improved methods to enable analysis of signaling networks from smaller amounts of starting material, and continued developments over the next several years should enable the analysis of pTyr signaling networks from tissue microarray (TMA) samples, or from small numbers of rare cell types. While starting material requirements may not rival genomic or transcriptional profiling experiments, both of these technologies have amplification steps that are currently not possible with proteomic approaches. Importantly, without amplification it is challenging to perform phosphoproteomics at the single cell level. While biologically relevant

phosphorylation sites span a large range of copy numbers in different cell types, we can take, as an example, a phosphorylation site that may be present at a reasonably high expression level of 10,000 copies/cell. Detection of this site in a single cell would require low-zeptomole sensitivity (10,000 copies/cell equates to ~$1.7x10^{-20}$ moles/cell (17 zeptomoles/cell), assuming no losses in sample processing and high efficiency ionization. At this point, it seems that single cell phosphoproteomics might be limited to very high expression phosphorylation sites, which may be less informative regarding activated signaling networks. Continued technological development is clearly required to advance this aspect of the field forward. Non-MS-based approaches, including immunohistochemistry, immunofluorescence, and phosphoflow can accomplish single-cell signaling network analysis, albeit with limited and targeted detection of selected phosphorylation sites.

One other challenge for phosphoproteomics is data integration with other systems-level measurements that may be performed on additional aliquots of the same samples. For instance, integration of phosphorylation data with transcript expression (bulk or single cell), metabolomics, methylation, genome sequencing, and other 'omics measurements, will provide significant insight into the co-regulation of these often dynamic regulatory networks. While some methods (see above) have been developed to attempt to integrate these data, our understanding of how transcript expression is related to signaling network regulation, and vice-versa, is still elementary. Development of improved computational algorithms, including artificial intelligence / machine learning approaches, are needed to define the linkages between these often highly complex data sets.

Over the past two decades the field of phosphoproteomics has experienced explosive growth in new tools, technologies, algorithms, and applications that have advanced the field to enable deep biological and clinical insight. Continued development of the field will lead to more stable technologies that can be applied to identify therapeutic targets, track therapeutic efficacy in clinical trials, and identify non-genomic adaptive response / resistance mechanisms that adversely impact patient survival.

References

1 Humphrey SJ, James DE, Mann M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. Trends Endocrinol Metab. 2015. p. 676–687.

2 Fuhs SR, Hunter T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. Curr. Opin. Cell Biol. 2017;45:8–16.

3 Fuhs SR, Meisenhelder J, Aslanian A, et al. Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation. Cell. 2015;162:198–210.

4 Schmelzle K, White FM. Phosphoproteomic approaches to elucidate cellular signaling networks. Current Opinion in Biotechnology. Curr Opin Biotechnol. 2006. p. 406–414.

5 Tsuji S, Ohno Y, Nakamura S, et al. Temozolomide has anti-tumor effects through the phosphorylation of cPLA2 on glioblastoma cells. Brain Res. 2019;1723.

6 Mao L, Zhan Y, Wu B, et al. ULK1 phosphorylates Exo70 to suppress breast cancer metastasis. Nat. Commun. 2020;11.

7 Xiao M, Xie J, Wu Y, et al. The eEF2 kinase-induced STAT3 inactivation inhibits lung cancer cell proliferation by phosphorylation of PKM2. Cell Commun. Signal. 2020;18.

8 Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell; 2011. p. 646–674.

9 Ardito F, Giuliani M, Perrone D, et al. The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). Int. J. Mol. 2017. 271–280.

10 Ficarro SB, McCleland M, Stukenberg PT, et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat. biotechnol. 2002;20:301–305.

*** This manuscript describes a method for radically reducing non-specific binding during phosphopeptide enrichment, thereby enabling large-scale phosphoproteomics analyses.**

11 Carter AM, Tan C, Pozo K, et al. Phosphoprotein-based biomarkers as predictors for cancer therapy. Proc. Natl. Acad. Sci. U.S.A. 2020;117:18401–18411.

12 Bekker-Jensen DB, Bernhardt OM, Hogrebe A, et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat. Commun. 2020;11.

13 Sevecka M, MacBeath G. State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling. Nat. Methods. 2006;3:825–831.

*** In addition to highlighting the potential for analysis of signaling networks with reverse-phase protein arrays (RPPA), this manuscript also showed that most phospho-specific antibodies do not provide high accuracy quantitative data in the RPPA format.**

14 Krutzik PO, Crane JM, Clutter MR, et al. High-content single-cell drug screening with phosphospecific flow cytometry. Nat. Chem. Biol. 2008;4:132–142.

*** Single-cell phosphoproteomics using phospho-specific antibodies and mass cytometry.**

15 Hunter T, Sefton BM. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proc. Natl. Acad. Sci. U.S.A. 1980;77:1311–1315.

*** This manuscript was the first report of tyrosine phosphorylation and also estimated the frequency of tyrosine phosphorylation to be ~0.1-1% of the phosphoproteome.**

16 Raffel S, Klimmeck D, Falcone M, et al. Quantitative proteomics reveals specific metabolic features of acute myeloid leukemia stem cells. Blood. 2020;136:1507–1519.

17 Nelson ME, Parker BL, Burchfield JG, et al. Phosphoproteomics reveals conserved exercise-stimulated signaling and AMPK regulation of store-operated calcium entry. EMBO J. 2019;38.

18 Morshed N, Ralvenius WT, Nott A, et al. Phosphoproteomics identifies microglial Siglec-F inflammatory response during neurodegeneration. Mol Syst Biol. 2020;16.

19 Kubiniok P, Finicle BT, Piffaretti F, et al. Dynamic Phosphoproteomics Uncovers Signaling Pathways Modulated by Anti-oncogenic Sphingolipid Analogs. Mol Cell Proteom. 2019;18:408–422.

20 Hu Y, Sun L, Zhang Y, et al. Phosphoproteomics reveals key regulatory kinases and modulated pathways associated with ovarian cancer tumors. Onco Targets Ther. 2020;13:3595–3605.

21 Scheidt T, Alka O, Gonczarowska-Jorge H, et al. Phosphoproteomics of short-term hedgehog signaling in human medulloblastoma cells. Cell Commun Signal. 2020;18.

22 Malik N, Nirujogi RS, Peltier J, et al. Phosphoproteomics reveals that the hVPS34 regulated SGK3 kinase specifically phosphorylates endosomal proteins including Syntaxin-7, Syntaxin-12, RFIP4 and WDR44. Biochem J. 2019;476:3081–3107.

23 Batth TS, Papetti M, Pfeiffer A, et al. Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling. Cell Rep. 2018;22:2784–2796.

24 Hijazi M, Smith R, Rajeeve V, et al. Reconstructing kinase network topologies from phosphoproteomics data reveals cancer-associated rewiring. Nat Biotechnol. 2020;38:493–502.

25 van Alphen C, Cloos J, Beekhof R, et al. Phosphotyrosine-based Phosphoproteomics for Target Identification and Drug Response Prediction in AML Cell Lines. Mol Cell Proteom. 2020;19:884–899.

26 Kohale IN, Burgenske DM, Mladek AC, et al. Quantitative Analysis of Tyrosine Phosphorylation from FFPE Tissues Reveals Patient-Specific Signaling Networks. Cancer Res. 2021;81:3930–3941.

27 Stopfer LE, Flower CT, Gajadhar AS, et al. High-density, targeted monitoring of tyrosine phosphorylation reveals activated signaling networks in human tumors. Cancer Res. 2021;81:2495-2509.

*** This manuscript describes the targeted analysis of hundreds of tyrosine phosphorylation sites through SureQuant pTyr**

28 Tuncbag N, Milani P, Pokorny JL, et al. Network Modeling Identifies Patient-specific Pathways in Glioblastoma. Sci Rep. 2016;6.

29 Ludwig KR, Schroll MM, Hummon AB. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. J Proteome Res. 2018;17:2480–2490.

30 Hailemariam M, Eguez RV, Singh H, et al. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. J Proteome Res. 2018;17:2917–2924.

31 Bubis JA, Gorshkov V, Gorshkov M V., et al. PhosphoShield: Improving Trypsin Digestion of Phosphoproteins by Shielding the Negatively Charged Phosphate Moiety. JASMS 2020;31:2053–2060.

32 Lundby A, Franciosa G, Emdal KB, et al. Oncogenic Mutations Rewire Signaling Pathways by Switching Protein Recruitment to Phosphotyrosine Sites. Cell. 2019;179:543-560.e26.

33 Giansanti P, Tsiatsiani L, Low TY, et al. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat Protoc. 2016;11:993–1006.

34 van der Laarse SAM, van Gelder CAGH, Bern M, et al. Targeting proline in (phospho)proteomics. FEBS J. 2020;287:2979–2997.

35 Pankow S, Bamberger C, Yates JR. A posttranslational modification code for CFTR maturation is altered in cystic fibrosis. Sci Signal. 2019;12.

36 Gao X, Li Q, Liu Y, et al. Multi-in-One: Multiple-Proteases, One-Hour-Shot Strategy for Fast and High-Coverage Phosphoproteomic Investigation. Anal Chem. 2020;92:8943–8951.

37 Humphrey SJ, Karayel O, James DE, et al. High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. Nat Protoc. 2018;13:1897–1916.

*** Demonstrates a rapid and facile method for deep characterization of the phosphoproteome**

38 Gao Y, Ha YS, Kwon TG, et al. Characterization of Kinase Expression Related to Increased Migration of PC-3M Cells Using Global Comparative Phosphoproteome Analysis. Cancer Genom Proteom. 2020;17:543–553.

39 CM P, MH L, AJR H, et al. Defeating Major Contaminants in Fe 3+- Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. Mol Cell Proteom. 2018;17:1028–1034.

40 Bodenmiller B, Mueller LN, Mueller M, et al. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods. 2007;4:231–237.

41 Ruprecht B, Koch H, Medard G, et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. Mol Cell Proteom. 2015;14:205–215.

42 Yeh T, Ho M, Chen W, et al. Comparison of different fractionation strategies for in-depth phosphoproteomics by liquid chromatography tandem mass spectrometry. Anal Bioanal. Chem. 2019;411:3417–3424.

43 Olsen JV, Blagoev Blagoy, Gnad F, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127:635–648.

*** Among the first very large scale phosphoproteomics papers, reporting quantitative data for over 6,000 phosphorylation sites in cells stimulated with epidermal growth factor.**

44 Wang LB, Karpova A, Gritsenko MA, et al. Proteogenomic and metabolomic characterization of human glioblastoma. Cancer cell. 2021;39:509-528.e20.

45 Clark DJ, Dhanasekaran SM, Petralia F, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell. 2019;179:964-983.e31.

46 Pawson T. Protein modules and signalling networks. Nature. 1995;373:573–580.

*** An excellent review of the fundamentals of cellular signaling networks**

47 Pawson T, Hunter T. Signal transduction and growth control in normal and cancer cells. Current Opin Genet Dev. 1994;4:1–4.

48 Rozakis-Adcock M, Fernley R, Wade J, et al. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. Nature. 1993;363:83–85.

49 Zhang Y, Wolf-Yadlin A, Ross PL, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol Cell Proteom. 2005;4:1240–1250.

50 Rush J, Moritz A, Lee KA, et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nature Biotechnol. 2005;23:94–101.

*** The first report of immunoprecipitation of phosphotyrosine-containing peptides for subsequent proteomic analysis**

51 Blagoev Blagoy, Ong SE, Kratchmarova I, et al. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol. 2004;22:1139–1145.

52 Reddy R, Curran TG, Zhang Y, et al. Measurement of Phosphorylated Peptides with Absolute Quantification. Methods Mol Biol. 2016;1410:281–292.

53 Zhang Y, Wolf-Yadlin A, White FM. Quantitative proteomic analysis of phosphotyrosine-mediated cellular signaling networks. Methods Mol Biol. 2007;359:203–212.

54 Pawson T. SH2 and SH3 domains in signal transduction. Adv Cancer Res. 1994;64:87–110.

55 Bian Y, Li L, Dong M, et al. Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. Nat Chem Biol. 2016;12:959–966.

56 Tong J, Cao B, Martyn G, et al. Protein-phosphotyrosine proteome profiling by superbinder-SH2 domain affinity purification mass spectrometry, sSH2-AP-MS. Proteomics. 2017;17.

57 Li S, Zou Y, Zhao D, et al. Revisiting the phosphotyrosine binding pocket of Fyn SH2 domain led to the identification of novel SH2 superbinders. Protein Sci. 2021;30:558–570.

58 Matsuoka S, Ballif BA, Smogorzewska A, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science. 2007;316:1160–1166.

59 Stokes MP, Rush J, MacNeill J, et al. Profiling of UV-induced ATM/ATR signaling pathways. Proc. Natl. Acad. Sci. U.S.A. 2007;104:19855–19860.

60 Carlson SM, Chouinard CR, Labadorf A, et al. Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. Sci Signal. 2011;4.

61 Mertins P, Qiao J, Patel J, et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods. 2013;10:634–637.

62 Hogrebe A, von Stechow L, Bekker-Jensen DB, et al. Benchmarking common quantification strategies for large-scale phosphoproteomics. Nat Commun. 2018;9:1045.

63 Stepath M, Zulch B, Maghnouj A,, et al. Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi. J Proteome Res. 2020;19:926–937.

64 Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteom. 2002;1:376–386.

*** This manuscript describes the use of stable isotope encoded amino acids for labeling of proteins in cell culture, now one of the most common methods for non-isobaric multiplexing of biological samples.**

65 Westman-Bringmalm A, Abramsson A, Pannee J, et al. SILAC zebrafish for quantitative analysis of protein turnover and tissue regeneration. J Proteomics. 2011;75:425–434.

66 Kruger M, Moser M, Ussar S, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell. 2008;134:353–364.

67 Altelaar A, Frese C, Preisinger C, et al. Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics. 2013;88:14–26.

68 Osinalde N, Moss H, Arrizabalaga O, et al. Interleukin-2 signaling pathway analysis by quantitative phosphoproteomics. J Proteomics. 2011;75:177–191.

69 Hammond DE, Hyde R, Kratchmarova I, et al. Quantitative analysis of HGF and EGF-dependent phosphotyrosine signaling networks. J Proteome Res. 2010;9:2734–2742.

70 Zhang G, Neubert TA. Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation. Methods Mol Biol. 2009;527.

71 Dengjel J, Aimov V, Olsen JV, et al. Quantitative proteomic assessment of very early cellular signaling events. Nat Biotechnol. 2007;25:566–568.

72 Zhang X, Maity T, Kashyap MK, et al. Quantitative Tyrosine Phosphoproteomics of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor-treated Lung Adenocarcinoma Cells

Reveals Potential Novel Biomarkers of Therapeutic Response. Mol Cell Proteom. 2017;16:891–910.

73 Cunningham DL, Sarhan AR, Creese AJ, et al. Differential responses to kinase inhibition in FGFR2-addicted triple negative breast cancer cells: a quantitative phosphoproteomics study. Sci Rep. 2020;10:7950.

74 Boersema PJ, Foong LY, Ding VMY, et al. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol Cell Proteom. 2010;9:84–99.

75 Polat AN, Kraiczek K, Heck AJR, et al. Fully automated isotopic dimethyl labeling and phosphopeptide enrichment using a microfluidic HPLC phosphochip. AnalBioanal Chem. 2012;404:2507–2512.

76 Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteom. 2004;3:1154–1169.

*** Describes the multiplexed analysis of samples using isobaric tags; this manuscript generated a great deal of interest in multiplexed analyses.**

77 Thompson A, Schafer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003;75:1895–1904.

78 Li J, Cai Z, Bomgarden RD, et al. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. J Proteome Res. 2021;20:2964–2972.

79 Stopfer LE, Conage-Pough JE, White FM. Quantitative Consequences of Protein Carriers in Immunopeptidomics and Tyrosine Phosphorylation MS 2 Analyses. Mol Cell Proteom. 2021;20:100104.

80 McAlister GC, Nusinow DP, Jedrychowski MP, et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem. 2014;86:7150–7158.

*** Dynamic range compression can adversely affect the quantitative accuracy of multiplexed analyses using isobaric tags. This manuscript describes the development and application of multinotch MS3 to improve sensitivity and quantitative accuracy for these samples.**

81 Schmelzle K, Kane S, Gridley S, et al. Temporal dynamics of tyrosine phosphorylation in insulin signaling. Diabetes. 2006;55:2171–2179.

82 Reddy RJ, Gajadhar AS, Swenson EJ, et al. Early signaling dynamics of the epidermal growth factor receptor. Proc. Natl. Acad. Sci. U.S.A. 2016;113:3114–3119.

*** Describes the use of quantitative pTyr phosphoproteomics to quantify signaling network dynamics with second-scale temporal resolution, highlighting the rapid activation of EGFR signaling.**

83 Randall EC, Emdal KB, Laramy JK, et al. Integrated mapping of pharmacokinetics and pharmacodynamics in a patient-derived xenograft model of glioblastoma. Nat Commun. 2018;9.

84 Shen L, Li Z, Shen L. Quantitative Tyrosine Phosphoproteomic Analysis of Resistance to Radiotherapy in Nasopharyngeal Carcinoma Cells. Cancer Manag Res. 2020;12:12667–12678.

85 Donnelly DP, Rawlins CM, DeHart CJ, et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nat Methods. 2019;16:587–594.

86 Sidoli S, Garcia BA. Middle-down proteomics: a still unexploited resource for chromatin biology. Expert Rev Proteom. 2017;14:617–626.

87 Zhen Y, Huang X, Kelleher NL. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. Curr Opin Chem Biol. 2016;33:142–150.

88 Lu C, Coradin M, Porter EG, et al. Accelerating the Field of Epigenetic Histone Modification Through Mass Spectrometry-Based Approaches. Mol Cell Proteom. 2020;20.

89 Roberts DS, Chen B, Tiambeng TN, et al. Reproducible Large-Scale Synthesis of Surface Silanized Nanoparticles as an Enabling Nanoproteomics Platform: Enrichment of the Human Heart Phosphoproteome. Nano Res. 2019;12:1473–1481.

90 Chen B, Hwang L, Ochowicz W, et al. Coupling functionalized cobalt ferrite nanoparticle enrichment with online LC/MS/MS for top-down phosphoproteomics. Chem Sci. 2017;8:4306–4311.

91 Riley NM, Hebert AS, Durnberger G, et al. Phosphoproteomics with Activated Ion Electron Transfer Dissociation. Anal Chem. 2017;89:6367–6376.

92 Wu S, Yang F, Zhao R, et al. Integrated workflow for characterizing intact phosphoproteins from complex mixtures. Anal Chem. 2009;81:4210–4219.

93 White FM, Wolf-Yadlin A. Methods for the Analysis of Protein Phosphorylation-Mediated Cellular Signaling Networks. Annu Rev Anal Chem (Palo Alto, Calif). 2016;9:295–315.

94 Sharma K, D'Souza RCJ, Tyanova S, et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep. 2014;8:1583–1594.

95 Batth TS, Olsen JV. Offline High pH Reversed-Phase Peptide Fractionation for Deep Phosphoproteome Coverage. Methods Mol Biol. 2016;1355:179–192.

96 Kumar N, Wolf-Yadlin A, White FM, et al. Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. PLoS Comput Biol. 2007;3:0035–0048.

97 Wolf-Yadlin A, Kumar N, Zhang Y, et al. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Mol Syst Biol. 2006;2.

***Describes the use of partial least squares regression (PLSR) to integrate quantitative phosphoproteomics data with quantitative phenotypic data from cells under the same conditions to highlight network nodes associated with phenotype.**

98 Kitata RB, Choong W-K, Tsai C-F, et al. A data-independent acquisition-based global phosphoproteomics system enables deep profiling. Nat Commun. 2021 12:1. 2021;12:1–14.

99 Gao E, Li W, Wu C, et al. Data-independent acquisition-based proteome and phosphoproteome profiling across six melanoma cell lines reveals determinants of proteotypes. Mol Omics. 2021;17:413–425.

100 Hu Y, Wang M, Ren S, et al. Quantitative proteomics and phosphoproteomic analyses of mouse livers after tick-borne Babesia microti infection. Int J Parasitol. 2021;51:167–182.

101 Parker BL, Yang G, Humphrey SJ, et al. Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. Sci Signal. 2015;8.

102 Martinez-Val A, Bekker-Jensen DB, Hogrebe A, et al. Data Processing and Analysis for DIA-Based Phosphoproteomics Using Spectronaut. Methods Mol Biol. 2021;2361:95–107.

103 Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, et al. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proc. Natl. Acad. Sci. U.S.A. 2007;104:5860–5865.

***The first report of MRM-based targeted pTyr phosphoproteomics on hundreds of phosphorylation sites in the EGFR signaling network.**

104 Whiteaker JR, Zhao L, Saul R, et al. A Multiplexed Mass Spectrometry-Based Assay for Robust Quantification of Phosphosignaling in Response to DNA Damage. Radiat Res. 2018;189:505–518.

105 Whiteaker JR, Zhao L, Schoenherr RM, et al. Peptide Immunoaffinity Enrichment with Targeted Mass Spectrometry: Application to Quantification of ATM Kinase Phospho-Signaling. Methods Mol Biol. 2017;1599:197–213.

106 Gallien S, Kim SY, Domon B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). Mol Cell Proteom.. 2015;14:1630–1644.

107 Curran TG, Zhang Y, Ma DJ, et al. MARQUIS: A multiplex method for absolute quantification of peptides and posttranslational modifications. Nat Commun. 2015;6:5924.

108 White FM. The potential cost of high-throughput proteomics. Sci Signal. 2011;4.

109 Savage SR, Zhang B. Using phosphoproteomics data to understand cellular signaling: A comprehensive guide to bioinformatics resources. Clin Proteom. 2020. p. 27.

110 Emdal KB, Dittmann A, Reddy RJ, et al. Characterization of in vivo resistance to osimertinib and JNJ-61186372, an EGFR/Met bispecific antibody, reveals unique and consensus mechanisms of resistance. Mol Cancer Ther. 2017;16:2572–2585.

111 Naegle KM, White FM, Lauffenburger DA, et al. Robust co-regulation of tyrosine phosphorylation sites on proteins reveals novel protein interactions. Mol Biosyst. 2012;8:2771–2782.

112 Naegle KM, Welsch RE, Yaffe MB, et al. MCAM: multiple clustering analysis methodology for deriving hypotheses and insights from high-throughput proteomic datasets. PLoS Comput. Biol. 2011;7.

113 Archer TC, Ehrenberger T, Mundt F, et al. Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups. Cancer cell. 2018;34:396-410.e8.

***Combines multi-omic analysis with computational modeling to identify subtypes within medulloblastoma and highlights some of the markers important for each subtype.**

114 Johnson H, White FM. Quantitative analysis of signaling networks across differentially embedded tumors highlights interpatient heterogeneity in human glioblastoma. J Proteome Res. 2014;13:4581–4593.

115 Linding R, Jensen LJ, Ostheimer GJ, et al. Systematic discovery of in vivo phosphorylation networks. Cell. 2007;129:1415–1426.

116 Kuleshov MV, Xie Z, London ABK, et al. KEA3: improved kinase enrichment analysis via data integration. Nucleic Acids Res. 2021;49.

117 Strasser SD, Ghazi PC, Starchenko A, et al. Substrate-based kinase activity inference identifies MK2 as driver of colitis.Integr Biol(Camb). 2019;11:301–314.

118 Krug K, Mertins P, Zhang B, et al. A Curated Resource for Phosphosite-specific Signature Analysis. Mol Cell Proteom.2019;18:576–593.

119 Storey AJ, Naceanceno KS, Lan RS, et al. ProteoViz: A tool for the analysis and interactive visualization of phosphoproteomics data. Mol Omics. 2020;16:316–326.

120 Pedersen AK, Pfeiffer A, Karemore G, et al. Proteomic investigation of Cbl and Cbl-b in neuroblastoma cell differentiation highlights roles for SHP-2 and CDK16. iScience. 2021;24.

121 Wirbel J, Cutillas P, Saez-Rodriguez J. Phosphoproteomics-based profiling of kinase activities in cancer cells. Methods Mol Biol. 2018:103–132.

122 Searle BC, Lawrence RT, MacCoss MJ, et al. Thesaurus: quantifying phosphopeptide positional isomers. Nat Methods. 2019;16:703–706.

123 Krug K, Jaehnig EJ, Satpathy S, et al. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell. 2020;183:1436-1456.e31.

124 Vasaikar S, Huang C, Wang X, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell. 2019;177:1035-1049.e19.

125 Stewart PA, Welsh EA, Slebos RJC, et al. Proteogenomic landscape of squamous cell lung cancer. Nat Commun. 2019;10.

126 Kedaigle A, Fraenkel E. Turning omics data into therapeutic insights. Curr Opin Pharmacol. 2018;42:95–101.

127 Soltis AR, Kennedy NJ, Xin X, et al. Hepatic Dysfunction Caused by Consumption of a High-Fat Diet. Cell Rep. 2017;21:3317–3328.

128 Pirhaji L, Milani P, Leidl M, et al. Revealing disease-associated pathways by network integration of untargeted metabolomics. Nat Methods. 2016;13:770–776.

129 Huang PH, Miraldi ER, Xu AM, et al. Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth. Mol Biosyst. 2010;6:1227–1237.

130 Frejno M, Meng C, Ruprecht B, et al. Proteome activity landscapes of tumor cell lines determine drug responses. Nat Commun. 2020 11:1. 2020;11:1–12.

131 Patel-Murray NL, Adam M, Huynh N, et al. A Multi-Omics Interpretable Machine Learning Model Reveals Modes of Action of Small Molecules. Sci Rep. 2020;10:1–14.

132 Wang S, Li W, Hu L, et al. NAguideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses. Nucleic Acids Res. 2020;48:e83–e83.

133 Bijnsdorp I V., Schelfhorst T, Luinenburg M, et al. Feasibility of phosphoproteomics to uncover oncogenic signalling in secreted extracellular vesicles using glioblastoma-EGFRVIII cells as a model. J Proteomics. 2021;232.

134 Murilo JR, Kuras M, Rezeli M, et al. Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. PloS One. 2018;13.

135 Coscia F, Doll S, Bech JM, et al. A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis. J Pathol. 2020;251:100–112.

136 Sanchez-Quiles V, Shi MJ, Dingli F, et al. Triple extraction method enables high quality mass spectrometry-based proteomics and phospho-proteomics for eventual multi-omics integration studies. Proteomics. 2021;2000303.

137 Liu Y, Zeng R, Wang R, et al. Spatiotemporally resolved subcellular phosphoproteomics. Proc. Natl. Acad. Sci. U.S.A. 2021;118:e2025299118.

138 Jadwin JA, Oh D, Curran TG, et al. Time-resolved multimodal analysis of Src Homology 2 (SH2) domain binding in signaling by receptor tyrosine kinases. ELife. 2016;5.

139 Budnik B, Levy E, Harmange G, et al. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 2018;19.

***Describes the use of single-cell proteomics to distinguish different cell states**

140 Dou M, Clair G, Tsai CF, et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. Anal Chem. 2019;91:13119–13127.

CHAPTER 3

REVIEW

Elucidating phosphorylation-mediated signaling networks in receptor tyrosine kinases through Y-to-F mutational studies

Jacqueline Gerritsen, M.Sc.*

Marc Creixell*

Forest M. White, Ph.D.

Aaron S Meyer, Ph.D.

*These authors contributed equally to this work

Manuscript in preparation

Abstract

Receptor tyrosine kinases (RTKs) play a crucial role in cellular processes through complex phosphorylation-mediated signaling networks. Their aberrant activity in disease has resulted in many efforts to understand the underlying mechanisms that explain their contribution to disease. However, RTK crosstalk and downstream signaling redundancy hinder the identification of signaling components triggered by individual RTK members. One powerful approach that allows for mechanistic characterization of RTKs is the use of tyrosine-to-phenylalanine (Y-to-F) mutational studies, in which intracellular tyrosine residues are systematically mutated to non-phosphorylatable residues. Different Y-to-F mutations induce subtle but distinct RTK-specific signaling and phenotypic perturbations in a cellular environment. Both signaling and cellular responses driven by an RTK can then be functionally linked by coupling proteomic experiments with computational tools. Here, we review the mechanistic and translatable insights gained from these types of studies, as well as discuss current approaches to collect and analyze this type of data. Finally, we offer a future perspective discussing the potential value current advances in our tools and knowledge can add to the level of biological insight we can gain from this mechanistic research, especially as throughput and sensitivity of methods continues to improve.

Receptor Tyrosine Kinases in Disease

Phosphorylation Receptor tyrosine kinases (RTKs) are key regulators of a plethora of cellular processes including cell proliferation, migration, differentiation, and metabolism[1,2]. Regulation occurs as RTKs transfer and translate signals from the extracellular environment to an intracellular response. Given their critical cellular function, aberrant RTK signaling has been associated with a variety of human diseases, including many cancers[3]. Large-scale omics studies have revealed various mechanisms through which abnormal RTK activation can occur, including gene amplification, autocrine activation, chromosomal rearrangements, and gain-of-function mutations[4–7]. As a result, many tyrosine kinase inhibitors (TKIs) and monoclonal antibodies (MAbs) have been developed and approved over the past few decades that target various pathologies, e.g., cetuximab in lung and head and neck cancer, panitumumab in colon cancer and trastuzumab and pertuzumab in breast cancer[8].

Despite these advancements, acquired resistance to these targeted therapies has often dampened success in the clinic, which can occur through several mechanisms, including driver mutations, drug-target interaction interference and RTK bypass mechanisms. One well-known example of acquired resistance to EGFR-TKIs is the T790M mutation in non-small cell lung cancer (NSCLC). The Threonine in position 790 regulates inhibitor specificity in the ATP binding pocket. The mutation confers resistance through enhancement of ATP affinity, thereby successfully competing with TKIs[9,10]. Drug treatment may lead to acquired mutations that modify drug metabolism, and hence inducing resistance, e.g., by changes in drug efflux and uptake[1]. Bypass resistance involves the hyperactivation of a receptor not targeted by therapy.

For instance, MET and AXL have been shown to provide bypass resistance to EGFR-targeted therapies in lung carcinoma while HER3 drives resistance to HER2-inhibition in breast cancer[11–15]. Lack of treatment response or cancer growth after remission due to acquired resistance or cancer heterogeneity emphasizes the complexity of the signaling networks employed by RTKs to propagate cellular signals. Understanding the receptor physiology and mechanisms underlying their signaling function is critical to understand why and how RTKs contribute to disease and has important implications for selection of anti-cancer therapies.

RTK Structure and Function

All RTKs (58 known in humans) are grouped into 20 subfamilies based on their domain architecture and share a similar protein structure that includes an extracellular ligand binding domain, a single transmembrane helix, an intracellular region that contains a juxtamembrane regulatory region, a tyrosine kinase domain (TKD) and a carboxyl (C-) terminal tail[16]. Stimulation of these receptors, e.g., by binding of a ligand, results in conformational changes and receptor dimerization, allowing the kinase domain to assume an active conformation. At this time, these changes allow trans-autophosphorylation of tyrosine residues located on the TKD and C-terminal[17]. Autophosphorylation leads to recruitment and activation of a wide variety of downstream signaling proteins that contain Src homology-2 (SH2) or phosphotyrosine-binding (PTB) domains. This increased kinase activity in turn causes association with several cellular proteins due to recognition of the phosphorylated residues by their SH2 and SH3 binding domains, i.e., PLCɣ-1 and PI3K. These proteins and other adapter

proteins become phosphorylated and produce second messengers that trigger intracellular changes that aid in the receptor's biological signal relay and activate major signaling pathways including, but not limited to, the activation of the RAS/MAPK, PI3K/AKT and JAK/STAT pathways[18,19](Figure 1A). The interaction between activated growth factor receptors and signal transduction molecules relies on (1) the RTK being tyrosine phosphorylated for binding motif recognition of adapter proteins containing an appropriate binding domain, (2) the subsequent adapter-mediated recruitment of signaling molecules, and (3) the direct RTK phosphorylation of downstream targets. The importance of these events for signal relay is further illustrated by the fact that mutant receptors lacking kinase activity fail to signal[20].

RTK Signaling Network Models and Y-to-F Mutational Studies

Over the past few decades, numerous models have been generated that attempt to explain RTK activation and predict their signaling response[21–23]. Many of the experimental models are based on bait experiments that helped identify phosphorylation-dependent interactions between RTKs and intracellular components[24–26]. Unfortunately, most of these studies do not account for the signaling dynamics, competition or recruitment of other kinases and phosphatases. More importantly, they are not capable of isolating the proximal interactors and downstream signaling pathways specific to individual RTK members. This continuous challenge in investigating RTK biology is further complicated by the redundancy in downstream signaling nodes and converging pathways, which hinder the ability to functionally link an RTK input to a given phenotypic output.

Generating a panel of Y-to-F RTK mutant cell lines represents a powerful approach to distinctly perturb both RTK-driven signaling pathways and phenotypic consequences. Genetic editing of tyrosine phosphosites to phenylalanine—a non-phosphorylatable tyrosine mimetic—induces signaling changes due to their impaired ability to engage in signal transduction (**Figure 1B**).

Besides affecting downstream signaling pathways, Y-to-F mutations can influence the overall activation state of the RTK. As described, ligand induced RTK activation involves RTK monomer dimerization and autophosphorylation of several tyrosines within the kinase domain. The absolute amount of intracellular tyrosine phosphorylation is typically used as a proxy for RTK activation[27]. However, biochemical assays have provided evidence of phosphorylation events inhibiting the catalytic activity of protein kinases. For instance, the enzymatic inhibition of CDK1 through Y15 phosphorylation leads to exit from mitosis and locks CDK1 in its inactive state in G1 phase[28,29]. Thus, while Y-to-F mutations are generally shown to cause loss-of-function effects, it is possible that mutation to phenylalanine of an inhibitory phosphosite leads to increased RTK activation (**Figure 1B**).

Finally, there is evidence of Y-to-F mutations affecting the specificity of RTKs for their downstream substrates. Moreover, phosphorylation at conserved tyrosine residues in PTB domains has been shown to lead to an increased preference for a particular set of substrates due to a shift in kinase specificity[30,31]. This suggests that the lack of tyrosine phosphorylation near the RTK binding pocket might affect the receptor's kinase domain specificity. In such case, the signaling differences observed in RTK Y-to-F mutants would not only be due to the interaction with adapter molecules and overall kinase activation but also due

to differences in complex formation with a distinct set of recognized downstream substrates (**Figure 1C**).

In this review, we will describe valuable insights gained from site-directed mutagenesis methods in various RTKs subfamilies through Y-F mutations. We provide an overview of experimental and computational methods to measure, analyze, and integrate the signaling and phenotypic consequences of RTK Y-to-F mutant cell lines, describe ways in which model predictions of downstream drivers can be validated, and provide a forward-looking perspective

of the utility of this system, the potential impact and challenges that lie ahead.

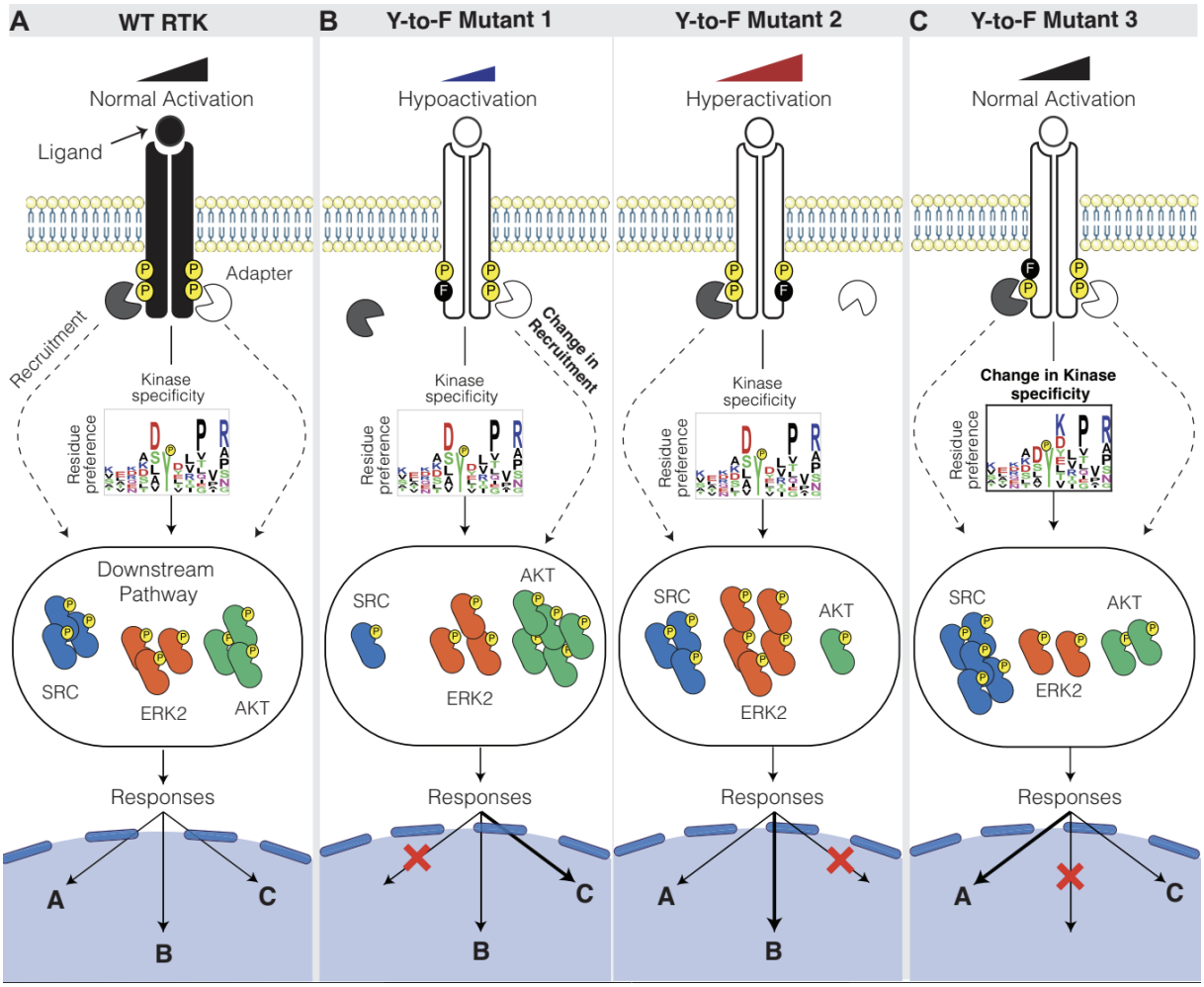


Figure 8: Y to F mutational studies to distinguish individual contributions of tyrosine residues to signaling network. A. Under wild-type conditions, tyrosine residues are phosphorylated and recognized by adapter molecules for signal relay. B. Loss of tyrosine residue can lead to decreased recruitment and subsequent activation of downstream signaling pathways. Alternatively, loss of tyrosine residue causes increase of recruitment by other residues leading to increased activation of certain downstream network nodes. C. Loss of tyrosine may not affect pathways or responses measured and therefore display a response similar to wild-type.

Consideration of choosing tyrosine-replacing amino-acid

Because each phosphorylated tyrosine and their surrounding amino acids selectively recognize certain SH2 or PTB domains, many studies leverage gene editing techniques to identify the pathways regulated by these tyrosines[26]. Phenylalanine is the replacement residue of choice as it most closely replicates tyrosine structure and is still able to make π-π and hydrophobic interactions.[32] Phenylalanine lacks the hydroxyl group that could lead to modifications including phosphorylation, thereby functionally knocking out tyrosine phosphorylation[33]. Other residues, including alanine, have been used as well, but are structurally quite different and may confound signaling data as it can cause changes in protein folding[34].

Counterintuitively, Zisch *et al*. found that replacing tyrosines 605 and 611 in EphB2 with a glutamic acid prevented binding of SH2 domains without abrogating kinase activity and biological responses, while mutating these sites to tyrosine impaired receptor kinase activity, complicating analysis of their function in terms of identifying binding partners and signaling pathways[35]. This approach may prove useful in establishing the role of other autophosphorylation sites in systems where mutation to phenylalanine would affect kinase activity.

C-terminal deletions of the EGFR receptor to investigate the ability of tyrosines to autophosphorylate have yielded replicable results, although caution must be exerted given the risk of more extensive effects to protein folding and dimerization[36,37].

Y-to-F interactome studies

Computational simulations as well as experimental studies have shown that the EGFR pathway contains regions of functional redundancy in the upstream region. Results suggest that this redundancy helps to maintain functional robustness in the event of low EGF stimulus or partial system failure. Downstream signaling nodes, like ERK, have fewer phosphosite redundancies, more often resulting in an attenuated signal response[23]. A study by Gill *et al* found that EGFR containing just one functional tyrosine on the C- terminal tail was still signaling-competent[38]. Given the multiplicity of binding sites, i.e. the ability of adapter molecules to bind to multiple phosphosites, Schlesssinger *et al* utilized Y-to-F mutant cell lines to determine the hierarchy of binding sites for Grb2 and Shc1 on EGFR using co-immunoprecipitation experiments. Although at least 5 tyrosines have been reported to bind Grb2 and Shc1 on EGFR, their data demonstrated a preference for Grb2 binding to Y1068 and Y1173, and that Y1173 is the preferred binding site for Shc1[39]. Y992 was identified as a main binding domain for PLCɣ-1 and has been connected to various phenotypic outcomes[38,40]. As a primary binder of Abl, Y1173 plays a major role in ABL-mediated endocytosis making it a potential target in human tumors[41]. This binding multiplicity was demonstrated in other RTKs as well. In IGF1R, Y1150 and Y1151 were responsible for ~80% of autophosphorylation, while no single Y-to-F mutation was able to abolish kinase activation[42].

In HER3, Y1325 was identified as the major binding site for Shc1 as well as the responsible residue for inducing heregulin (HRG)-dependent ERK activation[43]. Furthermore, Y-to-F studies have demonstrated cooperativity between residues on HER3 as well as requirement co-activation of HER2 and HER3 for PI3K activity[44]. **Figure 2** summarizes the efforts that involved

mutation of selected C-terminal phosphorylation sites that revealed loss of a given adapter binding or confirmed the multiplicity of a binding interaction44–56.

Instead of mutating the residues on the RTK, others have aimed to evaluate RTK function through determining the preferred interaction domain on the adapter proteins[45,46]. In the case of Shc1, for example, the PTB domain interaction with EGFR was shown to be the dominant interaction compared to the SH2 domain interaction[47,48]. Other studies have evaluated the effect of loss of a tyrosine site by measuring a phenotypic response, as a decrease in response is often expected as a result of decreased phosphorylation[49]. However, the VEGFR2 Y1006F mutant showed an increased ability to stimulate cell proliferation compared to wildtype[50].

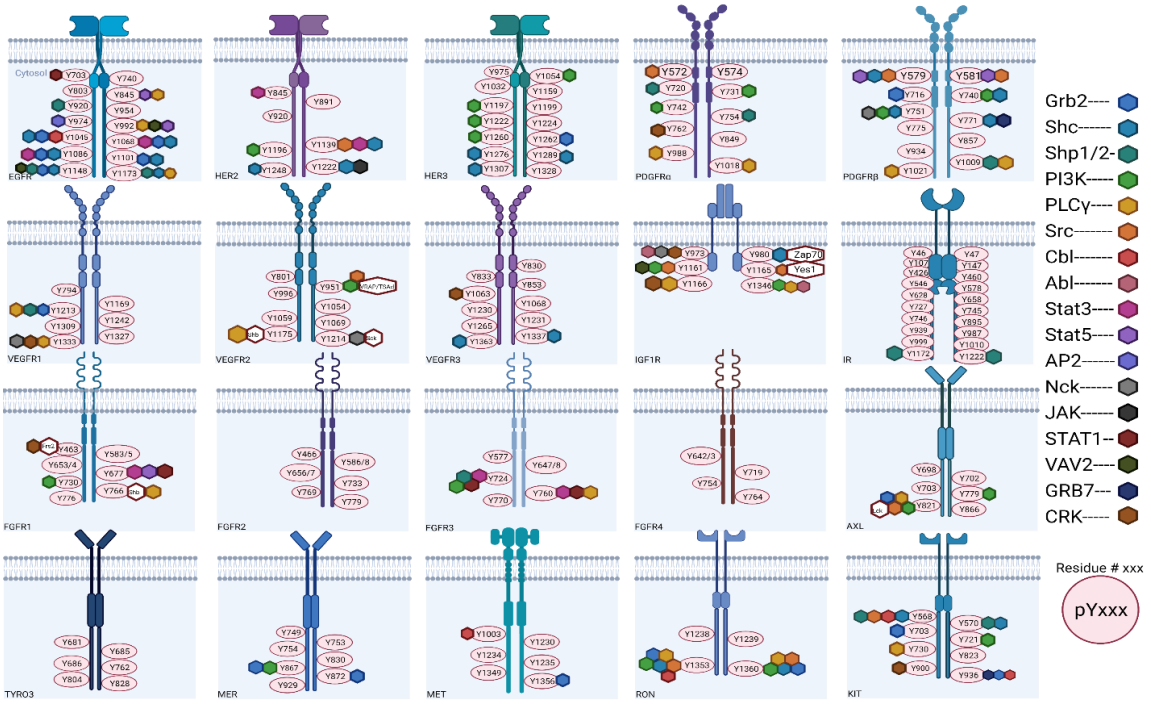


Figure 9: Known binding interactions between tyrosine residues of a subset of RTKs and secondary adaptor molecules. Numbers indicating residue location on protein. Proteins shown: EGFR[89,90], HER2[91], HER3[92], PDGFRa/b[93], VEGFR1-3[94,95], IGF1R[96], IR[97], FGFR1-4[98], AXL/TYRO3/MER[99,100], MET[101], RON[102], KIT[93]. Several are still poorly characterized in terms of binding partners and function.

Y-to-F studies in disease models

One major application of these mechanistic studies has been the identification of novel therapeutic targets. Given a strong correlation between HER2 activation and constitutively active Stat3a in a variety of human tumors, Ren *et al* found that Y1139 to be the main residue responsible for Stat3a activation through JAK2 and SRC-dependent mechanisms, and provided multiple novel targets for HER2-overexpressing tumors[51]. In EGFR-VIII mutants, site-specific mutational experiments have led to the identification of a set of phosphoproteins as potential candidates for future drug development in Glioblastoma Multiforme[52]. Using a similar approach, PDGFRα was identified to be important in development of cardiac and brain cells, providing a model to study aberrant NCC development[53]. In constitutively active mutants of FGFR3, Y724 was found to be a critical node in activation of multiple signaling pathways[54,55]. Finally, AXL-Y821 was found to mediate resistance to cetuximab through c-ABL in a head and neck squamous carcinoma (HNSCC) model, proposing ABL inhibition as a potential therapeutic strategy to resensitize tumor cells to cetuximab treatment[56].

Site-specific roles beyond downstream signaling conveyance

Besides their role in signal transduction as binding domains for adapter molecules, certain tyrosine residues have also been found to have functions related to receptor dimerization or substrate specificity. For example, Y762 was found to bind Crk as well as suggested to be partly responsible for the signaling divergence between the two PDGFR receptors[57]. Furthermore,

Y857 on PDGFRβ was identified as a critical site in transactivation of EGFR, providing mechanistic clues of crosstalk mechanisms between RTKs[58]. Interestingly, another study found this same tyrosine to affect cell proliferation, but not migration, while EGFR is known to affect both phenotypes[59]. Finally, while an IGFR1 Y1136F mutation only minimally affects its overall kinase activity, it was found to substantially reduce the phosphorylation of various substrates[60]. This seeming discrepancy highlights the complexity of the network, but it should also be noted that these experiments are often performed in a single cell line. Depending on the cell context, the concentration of the adapter protein may vary and be a limiting factor in certain model systems and yield different functional outcomes. In FGFR1, Y677 was found to function as a stabilizer of the active conformation[61]. Certain tyrosine residues on FGFR2 were found responsible in determining substrate specificity of the receptor[24,62]. Jiang *et al.* demonstrated IGFR's dependence on Y950 as a mediator of substrate binding and showed Y1136 dramatically reduces the transforming ability and overall substrate phosphorylation[60]. Finally, these mutational approaches have been used to understand molecular-level specificity in transphosphorylation. For example, Bae *et al.* showed that asymmetric receptor contact is required for autophosphorylation of FGFR1 and transphosphorylation of Y583[63]. Even for the same cancer types, the driving signaling pathways and targets identified across studies are largely context-dependent and thus cross-referencing insights generated in different systems can often be challenging and misleading. RTKs are responsible for the regulation of several oncogenic processes. For example, the RTK AXL has been shown to promote cell proliferation, epithelial-to-mesenchymal transition, metastasis, macropinocytosis, metabolic oxidation, DNA damage response (DDR), and immunosuppression in various types of cancer[64–67]. All these

phenotypes affect disease progression and response to targeted therapy against RTK-driven cancers.

Omics based Y-to-F studies

The studies described above have helped identify binders and function to specific tyrosines on the TKD and C-terminal tails or measured phenotypic variation because of functional tyrosine knock-out. To fully understand the mechanisms underlying binding and phenotype, however, it is critical to understand to functionally link these together. Recent advances in omics-based approaches have allowed for improved mechanistic interrogation of the entire network in model systems that recapitulate the complexity and competition of the cellular signaling networks. The next section of this review will focus on how Y-to-F mutational model systems and omics-tools can be combined to gain valuable biological insight. The ability to link functional consequences to specific tyrosine phosphorylation events, has important implications in understanding acquired resistance mechanisms in disease and identifying novel biomarkers for drug development.

A panel of Y-to-F mutant cell lines to model RTK-specific responses

We propose that the use of a panel of RTK mutant cell lines provides a unique opportunity to link specific phosphoproteomic downstream components to various cell phenotypes that collectively drive a complex multifactorial response (**Figure 3A**). For instance, as reviewed by Baselga *et al*, drug resistance should be viewed and investigated as a multifaceted problem driven by the development of simultaneous “collateral” malignant phenotypes that

coordinately promote tumor growth and metastasis in the presence of therapy[68]. Through coupling signaling and phenotypic experiments with computational modeling, a panel of Y-to-F RTK cell lines represents an ideal system to separate the signaling components driving specific phenotypes in the same RTK-driven cellular context. Below, we provide an overview of experimental and computational methods to measure, analyze, and integrate the signaling and phenotypic consequences of RTK Y-to-F mutant cell lines, describe ways in which model predictions of downstream drivers can be validated, and provide a forward-looking perspective of the utility of this system.

Computational tools to analyze and interpret phosphoproteomic signaling networks

Phosphoproteomics is a powerful approach that allows gathering of quantitative, network-level, dynamic signaling data, making it well suited for interrogating site-specific tyrosine functions. However, challenges inherent to phosphoproteomic data hinder the identification of relevant proteomic alterations; namely measuring substrates rather than kinases, the incomplete and stochastic coverage of the phosphoproteome across experiments and its high-content but low-sample throughput. In the last decade there has been substantial advancements in the development of computational methods tailored to overcome these challenges and improve our capabilities to disentangle and reconstruct signaling networks. Depending on the question at hand, we can categorize computational tools tailored to signaling data of Y-to-F RTK mutant cell lines in two main groups: Univariate and Multivariate methods (**Figure 3B**).

Univariate methods are those that allow one to profile the phosphoproteome of an individual sample or RTK mutant cell line. Kinome enrichment tools such as kinase-substrate enrichment analysis (KSEA) and Integrative Inferred Kinase Activity (INKA) identify hyperactive upstream kinases using the phosphoproteomic quantification of a sample. KSEA averages the signals of groups of known kinase substrates to infer enriched pathways in biological samples[69], while INKA ranks kinases based on their inferred activity by integrating kinases' overall and activation loop phosphorylation with the phosphorylation abundance of its known substrates[70]. These methods require making assumptions about kinase-substrate relationships for which there is experimental evidence in the literature or are predicted by kinase prediction algorithms such as Scansite or KinomeXplorer. Scansite makes predictions using kinase specificity profiles generated from oriented peptide library scanning experiments while KinomeXplorer uses sequence motif and protein-protein network information[71,72]. Furthermore, ranked gene identifiers based on the signaling data of a sample can be used to run Gene Set Enrichment Analysis (GSEA) or STRING analysis to identify and visualize overrepresented biological processes and protein interaction network maps, respectively. Thus, univariate methods can be used to find enriched signaling components and their associated biological processes in specific Y-to-F RTK perturbations.

Widely used multivariate methods such as clustering or dimensionality reduction algorithms to analyze phosphoproteomic data are extremely useful to enhance our interpretation of a dramatically underdetermined system comprised of thousands of peptides—or variables—and only a few observations. Recently, the Meyer lab constructed an expectation-maximization (EM) algorithm, Dual Data and Motif Clustering (DDMC) that can be viewed as a

multivariate form of KSEA and INKA. DDMC generates clusters displaying similar phosphorylation patterns and sequence motif features. Incorporating the sequence information into the clustering criterion allows one to make predictions about the upstream kinases regulating clusters by comparing the clusters' sequence features with previously reported kinase specificity profiling data[73]. These univariate and multivariate methods, often in combination, can help reconstruct signaling networks from phosphoproteomic data. Another clustering approach, ENSEMBLE, generates and incorporates multiple unique clustering solutions utilizing different combinations of data transformations (e.g., untransformed, z-score, mean-centering), clustering algorithms (k-means, hierarchical clustering, GMM…), and parameters (number of clusters, distance metrics, linkage methods, etc.). Samples are then

grouped based on how often they cluster together[74].

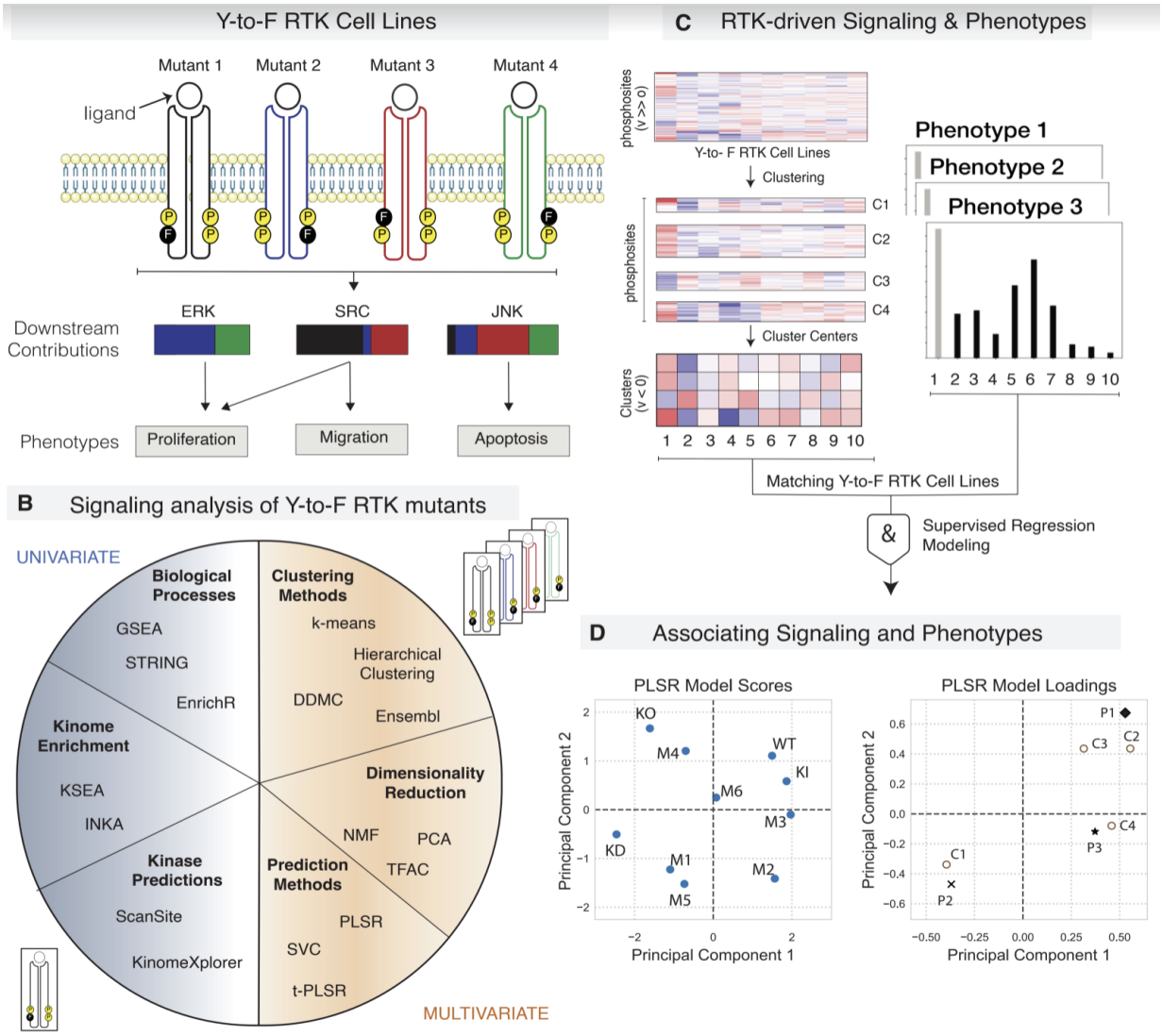


Figure 10: Combining Y-to-F mutational data and computational analysis to gain biological insight. A. Collecting signaling and phenotypic data of a panel of isoform mutants. B. Categories of Univariate or Multivariate analyses. C. Supervised learning algorithms to increase prediction of data-driven models. D. PLSR as an example to functionally link phenotypic and signaling data together to generate meaningful insight.

Thus, multivariate methods such as the described clustering algorithms or other common dimensionality reduction techniques such as principal component analysis (PCA) or non-negative matrix factorization (NMF) find behavioral patterns across RTK mutant samples which allow the systematic identification of RTK-driven signaling pathways and processes.

Establishing associations between cell phenotypes and RTK downstream signaling

RTKs are involved in a myriad of cellular processes, and data-driven multivariate modeling of a panel of Y-to-F mutants can help map the downstream signaling pathways that correlate with different phenotypes. To build a regression model linking both information sources, the phosphoproteomic data—matrix X—is used to explain the cellular responses—matrix Y. In both matrices, the observations consist of the RTK mutant cell lines treated under the exact conditions and the variables are phosphopeptides and phenotypes, respectively (**Figure 3A/C**). Practically any cell phenotype data of interest such as cell proliferation, apoptosis, migration, or genetic signature scores can be incorporated in the Y matrix if the cell culture conditions match the ones utilized to generate the signaling data. Importantly, the signaling and phenotypic consequences that Y-to-F mutations might induce could be influenced by an impairment of RTK regulation at different levels. Thus, trafficking and degradation rates, gene and protein expression, as well as dynamic measurements of downstream signaling of the different RTKs are crucial factors that should be measured to obtain a holistic view of how these mutations induce the observed responses.

Y-to-F mutational studies can provide a unique perspective on the extent to which RTK point mutations affect distinct phenotypes through specific downstream signaling components. For instance, RTK mutant 1 might largely activate ERK to promote proliferation whereas mutant 2 mainly induces JNK activity and apoptosis (**Figure 3A**). However, the benefit of systematically investigating the signaling changes across mutants via multivariate modeling provides the opportunity of more generally attributing RTK-specific signaling effects to phenotypic consequences. As pointed above, MS-global phosphoproteomic data sets are hugely underdetermined given its high-content but low-sample throughput. Previous efforts in regressing signaling measurements against phenotypic data were based on the great ability of partial least squares (PLSR) models to robustly handle prediction in the presence of high-dimensional and correlated data[75]. This is an excellent strategy with moderately large phosphoproteomic data sets. However, as the number of phosphosites—or variables—increases, the system becomes increasingly underdetermined, and the prediction power diminishes. In addition, while these models can generally be predictive with such data, they are not easily interpretable. Therefore, clustering large-scale MS measurements based on biologically meaningful features and utilizing the cluster averages to fit PLSR could enhance the predictive performance of the model while providing highly interpretable results wherein clusters constitute signaling nodes distinctly correlated with cell phenotypes (**Figure 3C/D**).

Moreover, The RTK Y-to-F cell signaling data can also be visualized with PCA which identifies a low-rank structure within a space of possible outcomes. The resulting PCA scores and loadings can then be effectively visualized to establish associations between cell lines and signaling molecules. Interestingly, the analysis of omics data can be integrated using higher-

mode generations of PCA to predict phenotypes. These methods, broadly referred to as tensor decomposition techniques, can be remarkably effective at dimensionality reduction even beyond PCA while associating omics-based signaling information to phenotypic outcomes. For instance, researchers used tensor partial least squares regression to predict the transcriptional signatures from signaling induced by different stimuli over time. The model predicted that an early-phase, AKT-associated signal downstream of insulin repressed a set of transcripts induced by TNF[76].

Here, we propose two multivariate modeling strategies, namely (i) fitting of cluster centers to a PLSR model to predict phenotypes or (ii) tensor factorization to associate signaling responses with phenotypes. The resulting model predictions facilitate the generation of hypothesis about key signaling pathway components driving cell responses which in turn informs the design of downstream of validation experiments.

Validation experiments to demonstrate model predictions

The predicted RTK-driven molecular pathways influencing cellular response can be effectively validated through several experimental methods. Typical biochemical assays such as western blot and proximity ligation assay (PLA) or immunoprecipitation (IP) can be very useful to demonstrate a change in protein activation or interaction, respectively. However, here we highlight high-throughput methodologies capable of simultaneously quantifying the activation of several proteins or detecting RTK-protein interactions, as well as biochemical assays that enable the characterization of the specificity profile of a kinase of interest.

To validate RTK-protein or downstream protein-protein interactions, BioID interaction mapping facilitates a global view of the interactome of any protein, including RTKs. This system utilizes a proximity-dependent labeling strategy wherein a bait protein is fused to a mutant form of biotin ligase (BirA*) that biotinylates interacting proteins within a 10-nm radius of the protein of interest[77]. Such data can serve as direct evidence of adapter molecules, kinases, phosphatases and other signaling molecules distinctly interacting with RTK Y-to-F mutants which can then be subsequently associated with downstream signaling consequences.

Biochemical kinase specificity assays such as position-specific peptide library scanning (PSPL) enable the specificity profiling of a kinase of interest which in turn can help elucidate downstream substrates displaying the kinase's optimal target motif. In this technique, a kinase of interest is individually incubated with each of 180 different peptide libraries in which each library contains a central phosphoacceptor residue (S/T or Y), a second fixed amino acid located any of the peptide residues spanning positions 5 throughout +4 relative to the phosphorylation site, and a degenerate mixture containing all natural amino acids at all other positions. The kinase and peptide libraries are incubated in the presence of radioactive ATP, which allows the quantification of phosphorylation abundance per residue and position and the identification of the kinase's ''optimal'' substrate motif. The accumulation of PSPL kinase profilings facilitate the association of putative upstream kinases regulating specific substrates or groups of phosphosites displaying similar kinase motifs[78–81]. Since they do not provide docking information, a major limitation of PSPL experiments is that the real affinity between the string of identified peptide residues as key determinants of specificity of a sequence motif and the interacting kinase domain is unknown. This limitation could compromise kinase-substrate or

kinase-cluster associations. A method combining bacterial surface-display of peptide libraries with next-generation sequences would tackle this limitation by quantifying the specificity of a kinase to virtually all possible motif combinations[82].

Besides any combination of these computational and experimental techniques, the resulting mechanistic knowledge can be further validated by repeating the phenotypic measurements used to build the regression model in the presence of pharmacological or genetic inhibition of the predicted and validated RTK-regulated downstream drivers. In summary, we propose an experimental and computational framework to investigate, associate, and validate RTK-driven signaling pathways and phenotypes.

Forward-looking perspective

In this review, we highlighted the value of using Y-to-F studies to research the underlying mechanisms of signal propagation in RTKs. Being able to link signaling data to functional response in a cellular environment is crucial, whether the goal is improved mechanistic understanding or identification of novel biomarkers or drug targets. Recent advances in experimental and computational tools and methods have great potential to add additional insight we can gather with this type of data.

In terms of model systems, advances in CRISPR methods have made it more feasible to make point mutations in model systems expressing endogenous RTKs, allowing generation of a physiologically more relevant system, in terms of expression levels and cellular environment. Viral-expressing model systems, although easy to use and manipulate, are challenging when it comes to controlling expression levels between cell lines and replicates, due to random

insertion, even in an inducible system.[38] Furthermore, the use of a viral system limits the choice of model system, due to potential confounding signals from endogenous expressing genes. CRISPR point mutations are now also shown to be successfully made in *in vivo* settings, opening the door for potential Y-to-F studies in animal studies (e.g., PDX models).[83]

For data analysis, increased sensitivity and accuracy of methods and instruments have yielded promising results for single-cell and single-molecule phosphoproteomics.[84,85] Automated microfluidic chip LC systems are an important component to allow quantitative identification of already low-abundant phosphorylated tyrosine.[86] These analyses are especially beneficial as they will reveal information about the stoichiometry of phosphorylation of the cell or receptor at a given time, and they could allow profiling of a large panel of Y-to-F perturbations. One potential translational application would be in understanding tumor heterogeneity, which continues to be one of the main obstacles when it comes to cancer treatments.[87] Before this method can gain more traction, though, issues regarding sensitivity and low-throughput will need to be resolved.

Finally, combining targeted phosphoproteomic strategies with absolute quantification methods, such as Surequant, will allow for reproducible pTyr profiling of commonly dysregulated oncogenic signaling proteins.[88] These advancements allow us to build on previously gathered mechanistic insight regarding RTK signal and function and to continue to explore the underlying mechanisms that explain the complex signaling network and responses that these RTKs regulate in health and disease.

References

1 Manning G, Whyte DB, Martinez R, et al. The protein kinase complement of the human genome. Science Internet. 2002 cited 2022 Jan 18;298:1912–1934. Available from: https://pubmed.ncbi.nlm.nih.gov/12471243/.

2 Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. Oncogene Internet. 2000 cited 2016 Jan 28;19:5548–5557. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11114734.

3 Sharma S V., Bell DW, Settleman J, et al. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer Internet. 2007 cited 2022 Jan 18;7:169–181. Available from: https://pubmed.ncbi.nlm.nih.gov/17318210/.

4 Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. Science (80-) Internet. 2003 cited 2022 Feb 9;300:949. Available from: https://www.science.org/doi/abs/10.1126/science.1082596.

5 Brewer MR, Yun CH, Lai D, et al. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. Proc Natl Acad Sci U S A Internet. 2013 cited 2022 Feb 9;110. Available from: https://pubmed.ncbi.nlm.nih.gov/24019492/.

6 Wang TH, Wu CC, Huang KY, et al. Integrated Omics Analysis of Non-Small-Cell Lung Cancer Cells Harboring the EGFR C797S Mutation Reveals the Potential of AXL as a Novel Therapeutic Target in TKI-Resistant Lung Cancer. Cancers 2021, Vol 13, Page 111 Internet. 2020 cited 2022 Feb 9;13:111. Available from: https://www.mdpi.com/2072-6694/13/1/111/htm.

7 Friedlaender A, Subbiah V, Russo A, et al. EGFR and HER2 exon 20 insertions in solid tumours: from biology to treatment. Nat Rev Clin Oncol 2021 191 Internet. 2021 cited 2022 Feb 10;19:51–69. Available from: https://www.nature.com/articles/s41571-021-00558-1.

8 Jänne PA, Engelman JA, Johnson BE. Epidermal growth factor receptor mutations in non-small-cell lung cancer: implications for treatment and tumor biology. J Clin Oncol Internet. 2005 cited 2022 Jan 18;23:3227–3234. Available from: https://pubmed.ncbi.nlm.nih.gov/15886310/.

9 Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med Internet. 2005 cited 2022 Jan 18;352:786–792. Available from: https://pubmed.ncbi.nlm.nih.gov/15728811/.

10 Remon J, Steuer CE, Ramalingam SS, et al. Osimertinib and other third-generation EGFR TKI in EGFR-mutant NSCLC patients. Ann Oncol Internet. 2018 cited 2018 Oct 3;29:i20–i27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29462255.

11 Wilson TR, Fridlyand J, Yan Y, et al. Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. Nature Internet. 2012 cited 2022 Mar 13;487:505–509. Available from: https://pubmed.ncbi.nlm.nih.gov/22763448/.

12 Tao JJ, Castel P, Radosevic-Robin N, et al. Antagonism of EGFR and HER3 enhances the

response to inhibitors of the PI3K-Akt pathway in triple-negative breast cancer. Sci Signal Internet. 2014 cited 2022 Mar 13;7. Available from: https://pubmed.ncbi.nlm.nih.gov/24667376/.

13 Turke AB, Zejnullahu K, Wu YL, et al. Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. Cancer Cell Internet. 2010 cited 2022 Mar 13;17:77–88. Available from: https://pubmed.ncbi.nlm.nih.gov/20129249/.

14 Wilson TR, Lee DY, Berry L, et al. Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. Cancer Cell Internet. 2011 cited 2022 Mar 13;20:158–172. Available from: https://pubmed.ncbi.nlm.nih.gov/21840482/.

15 Zhang Z, Lee JC, Lin L, et al. Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat Genet. 2012;44:852–860.

16 Hubbard SR. Structural analysis of receptor tyrosine kinases. Prog Biophys Mol Biol Internet. 1999 cited 2022 Jan 18;71:343–358. Available from: https://pubmed.ncbi.nlm.nih.gov/10354703/.

17 Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell Internet. 2010 cited 2014 Jul 10;141:1117–1134. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2914105&tool=pmcentrez&rendertype=abstract.

18 Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. Cell Internet. 2007 cited 2022 Jan 18;129:1261–1274. Available from: https://pubmed.ncbi.nlm.nih.gov/17604717/.

19 Patterson RL, Van Rossum DB, Nikolaidis N, et al. Phospholipase C-gamma: diverse roles in receptor-mediated calcium signaling. Trends Biochem Sci Internet. 2005 cited 2022 Jan 18;30:688–697. Available from: https://pubmed.ncbi.nlm.nih.gov/16260143/.

20 Westermark B, Claesson-Welsh L, Heldin C-H. Structural and functional aspects of platelet-derived growth factor and its receptors. Ciba Found Symp Internet. 1990 cited 2022 Jan 18;150:6–22. Available from: https://pubmed.ncbi.nlm.nih.gov/2164910/.

21 Oda K, Matsuoka Y, Funahashi A, et al. A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol Internet. 2005 cited 2022 Feb 10;1:2005.0010. Available from: /pmc/articles/PMC1681468/.

22 Resat H, Ewald JA, Dixon DA, et al. An Integrated Model of Epidermal Growth Factor Receptor Trafficking and Signal Transduction. Biophys J Internet. 2003 cited 2022 Feb 10;85:730. Available from: /pmc/articles/PMC1303198/.

23 Wang DYQ, Cardelli L, Phillips A, et al. Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics. BMC Syst Biol Internet. 2009 cited 2022 Feb 10;3:1–17. Available from: https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-3-118.

24 Hinsby AM, Olsen J V., Mann M. Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4. J Biol Chem Internet. 2004 cited 2022 Jan 21;279:46438–46447. Available from: https://pubmed.ncbi.nlm.nih.gov/15316024/.

25 Schulze WX, Deng L, Mann M. Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol Internet. 2005 cited 2018 Sep 17; Available from: www.molecularsystemsbiology.com.

26 Jones RB, Gordus A, Krall JA, et al. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature Internet. 2006 cited 2022 Feb 10;439:168–174. Available from: https://pubmed.ncbi.nlm.nih.gov/16273093/.

27 Lemmon MA, Freed DM, Schlessinger J, et al. The Dark Side of Cell Signaling: Positive Roles for Negative Regulators. Cell Internet. 2016 cited 2022 Mar 13;164:1172–1184. Available from: https://pubmed.ncbi.nlm.nih.gov/26967284/.

28 Potapova TA, Daum JR, Byrd KS, et al. Fine tuning the cell cycle: activation of the Cdk1 inhibitory phosphorylation pathway during mitotic exit. Mol Biol Cell Internet. 2009 cited 2022 Mar 13;20:1737–1748. Available from: https://pubmed.ncbi.nlm.nih.gov/19158392/.

29 Yamada M, Watanabe K, Mistrik M, et al. ATR–Chk1–APC/CCdh1-dependent stabilization of Cdc7–ASK (Dbf4) kinase is required for DNA lesion bypass under replication stress. Genes Dev Internet. 2013 cited 2022 Mar 13;27:2459–2472. Available from: http://genesdev.cshlp.org/content/27/22/2459.full.

30 Songyang Z, Gish G, Mbamalu G, et al. A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. J Biol Chem Internet. 1995 cited 2022 Feb 18;270:26029–26032. Available from: https://pubmed.ncbi.nlm.nih.gov/7592796/.

31 Jin LL, Wybenga-Groot LE, Tong J, et al. Tyrosine phosphorylation of the Lyn Src homology 2 (SH2) domain modulates its binding affinity and specificity. Mol Cell Proteomics Internet. 2015 cited 2022 Feb 18;14:695–706. Available from: https://pubmed.ncbi.nlm.nih.gov/25587033/.

32 Gallivan JP, Dougherty DA. Cation-π interactions in structural biology. Proc Natl Acad Sci U S A Internet. 1999 cited 2022 Mar 11;96:9459–9464. Available from: www.pnas.org.

33 Colette Daubner S, Fitzpatrick PF. Mutation to phenylalanine of tyrosine 371 in tyrosine hydroxylase increases the affinity for phenylalanine. Biochemistry Internet. 1998 cited 2022 Feb 10;37:16440–16444. Available from: https://pubmed.ncbi.nlm.nih.gov/9819237/.

34 Teng S, Srivastava AK, Schwartz CE, et al. Structural Assessment of the Effects of Amino Acid Substitutions on Protein Stability and Protein-Protein Interaction. Int J Comput Biol Drug Des Internet. 2010 cited 2022 Feb 10;3:334. Available from: /pmc/articles/PMC3319068/.

35 Zisch AH, Pazzagli C, Freeman AL, et al. Replacing two conserved tyrosines of the EphB2 receptor with glutamic acid prevents binding of SH2 domains without abrogating kinase activity and biological responses. Oncogene 2000 192 Internet. 2000 cited 2022 Feb 10;19:177–187. Available from: https://www.nature.com/articles/1203304.

36 G M Walton, W S Chen, M G Rosenfeld GNG. Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced in vivo tyrosine phosphorylation of cell substrates. J Biol Chem. 1990;265:1750–1754.

37 Roskoski R. The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacol Res Internet. 2014 cited 2018 Oct 1;79:34–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24269963.

38 Gill K, Macdonald-Obermann JL, Pike LJ. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. J Biol Chem Internet. 2017 cited 2022 Jan 18;292:20744–20755. Available from: https://pubmed.ncbi.nlm.nih.gov/29074618/.

39 Batzer AG, Rotin D, Ureña JM, et al. Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. Mol Cell Biol Internet. 1994 cited 2022 Jan 18;14:5192–5201. Available from: https://pubmed.ncbi.nlm.nih.gov/7518560/.

40 Nogami M, Yamazaki M, Watanabe H, et al. Requirement of autophosphorylated tyrosine 992 of EGF receptor and its docking protein phospholipase Cγ1 for membrane ruffle formation. FEBS Lett. 2003;536:71–76.

41 Tanos B, Pendergast AM. Abl tyrosine kinase regulates endocytosis of the epidermal growth factor receptor. J Biol Chem Internet. 2006 cited 2022 Feb 4;281:32714–32723. Available from: https://pubmed.ncbi.nlm.nih.gov/16943190/.

42 Smith JE, Sheng zu F, Kallen RG. Effects of tyrosine-->phenylalanine mutations on auto- and trans-phosphorylation reactions catalyzed by the insulin receptor beta-subunit cytoplasmic domain. DNA Cell Biol Internet. 1994 cited 2022 Jan 18;13:593–604. Available from: https://pubmed.ncbi.nlm.nih.gov/8024702/.

43 Vijapurkar U, Cheng K, Koland JG. Mutation of a Shc binding site tyrosine residue in ErbB3/HER3 blocks heregulin-dependent activation of mitogen-activated protein kinase. J Biol Chem Internet. 1998 cited 2022 Feb 10;273:20996–21002. Available from: https://pubmed.ncbi.nlm.nih.gov/9694850/.

44 Hellyer NJ, Kim MS, Koland JG. Heregulin-dependent activation of phosphoinositide 3-kinase and Akt via the ErbB2/ErbB3 co-receptor. J Biol Chem Internet. 2001 cited 2022 Feb 10;276:42153–42161. Available from: https://pubmed.ncbi.nlm.nih.gov/11546794/.

45 N G, M T, M S. Tyrosine phosphorylation sites at amino acids 239 and 240 of Shc are involved in epidermal growth factor-induced mitogenic signaling that is distinct from Ras/mitogen-activated protein kinase activation. Mol Cell Biol Internet. 1997 cited 2022 Feb 10;17:1824–1831. Available from: https://pubmed.ncbi.nlm.nih.gov/9121430/.

46 Salas V, Sánchez-Torres J, Cusidó-Hita DM, et al. Characterisation of tyrosine-phosphorylation-defective calmodulin mutants. Protein Expr Purif Internet. 2005 cited 2022 Feb 10;41:384–392. Available from: https://pubmed.ncbi.nlm.nih.gov/15866726/.

47 Sakaguchi K, Okabayashi Y, Kido Y, et al. Shc phosphotyrosine-binding domain dominantly interacts with epidermal growth factor receptors and mediates Ras activation in intact cells. Mol Endocrinol Internet. 1998 cited 2022 Jan 18;12:536–543. Available from:

https://pubmed.ncbi.nlm.nih.gov/9544989/.

48 van der Geer P, Wiley S, Ka-Man Lai V, et al. A conserved amino-terminal Shc domain binds to phosphotyrosine motifs in activated receptors and phosphopeptides. Curr Biol Internet. 1995 cited 2022 Feb 10;5:404–412. Available from: https://pubmed.ncbi.nlm.nih.gov/7542991/.

49 Blanes MG, Oubaha M, Rautureau Y, et al. Phosphorylation of tyrosine 801 of vascular endothelial growth factor receptor-2 is necessary for Akt-dependent endothelial nitric-oxide synthase activation and nitric oxide release from endothelial cells. J Biol Chem Internet. 2007 cited 2022 Feb 10;282:10660–10669. Available from: https://pubmed.ncbi.nlm.nih.gov/17303569/.

50 Meyer RD, Latz C, Rahimi N. Recruitment and activation of phospholipase Cgamma1 by vascular endothelial growth factor receptor-2 are required for tubulogenesis and differentiation of endothelial cells. J Biol Chem Internet. 2003 cited 2022 Feb 10;278:16347–16355. Available from: https://pubmed.ncbi.nlm.nih.gov/12598525/.

51 Ren Z, Schaefer TS. ErbB-2 activates Stat3 alpha in a Src- and JAK2-dependent manner. J Biol Chem Internet. 2002 cited 2022 Jan 18;277:38486–38493. Available from: https://pubmed.ncbi.nlm.nih.gov/11940572/.

52 Huang PH, Miraldi ER, Xu AM, et al. Analysis of site-specific mutations on EGFRVIII phosphotyrosine signaling network identifies determinants governing U87MG Glioblastoma cell growth. Mol Biosyst Internet. 2010 cited 2022 Feb 10;6:1227. Available from: /pmc/articles/PMC3291333/.

53 Tallquist MD, Soriano P. Cell autonomous requirement for PDGFRalpha in populations of cranial and cardiac neural crest cells. Development Internet. 2003 cited 2022 Feb 10;130:507–518. Available from: https://pubmed.ncbi.nlm.nih.gov/12490557/.

54 Porta R, Borea R, Coelho A, et al. FGFR a promising druggable target in cancer: Molecular biology and new drugs. Crit Rev Oncol Hematol Internet. 2017 cited 2022 Jan 18;113:256–267. Available from: https://pubmed.ncbi.nlm.nih.gov/28427515/.

55 Hart KC, Robertson SC, Donoghue DJ. Identification of Tyrosine Residues in Constitutively Activated Fibroblast Growth Factor Receptor 3 Involved in Mitogenesis, Stat Activation, and Phosphatidylinositol 3-Kinase Activation. Mol Biol Cell Internet. 2001 cited 2022 Feb 10;12:931. Available from: /pmc/articles/PMC32277/.

56 McDaniel NK, Iida M, Nickel KP, et al. AXL Mediates Cetuximab and Radiation Resistance Through Tyrosine 821 and the c-ABL Kinase Pathway in Head and Neck Cancer. Clin Cancer Res Internet. 2020 cited 2022 Feb 10;26:4349–4359. Available from: https://clincancerres.aacrjournals.org/content/26/16/4349.

57 Yokote K, Hellman U, Ekman S, et al. Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. Oncogene Internet. 1998 cited 2022 Feb 10;16:1229–1239. Available from: https://pubmed.ncbi.nlm.nih.gov/9546424/.

58 Kazlauskas A, Durden DL, Cooper JA. Functions of the major tyrosine phosphorylation site of the

PDGF receptor beta subunit. Cell Regul Internet. 1991 cited 2022 Jan 18;2:413–425. Available from: https://pubmed.ncbi.nlm.nih.gov/1653029/.

59 Wardega P, Heldin CH, Lennartsson J. Mutation of tyrosine residue 857 in the PDGF beta-receptor affects cell proliferation but not migration. Cell Signal Internet. 2010 cited 2022 Feb 10;22:1363–1368. Available from: https://pubmed.ncbi.nlm.nih.gov/20494825/.

60 Jiang Y, Chan JLK, Zong CS, et al. Effect of Tyrosine Mutations on the Kinase Activity and Transforming Potential of an Oncogenic Human Insulin-like Growth Factor I Receptor (∗). J Biol Chem Internet. 1996 cited 2022 Feb 10;271:160–167. Available from: http://www.jbc.org/article/S0021925818953184/fulltext.

61 Foehr ED, Raffioni S, Murray-Rust J, et al. The Role of Tyrosine Residues in Fibroblast Growth Factor Receptor 1 Signaling in PC12 Cells: Systematic site-directed mutagenesis in the endodomain *. J Biol Chem Internet. 2001 cited 2022 Feb 10;276:37529–37536. Available from: http://www.jbc.org/article/S002192581956700X/fulltext.

62 Zhang Y, McKeehan K, Lin Y, et al. Fibroblast Growth Factor Receptor 1 (FGFR1) Tyrosine Phosphorylation Regulates Binding of FGFR Substrate 2α (FRS2α) But Not FRS2β to the Receptor. Mol Endocrinol Internet. 2008 cited 2022 Feb 10;22:167–175. Available from: https://academic.oup.com/mend/article/22/1/167/2684090.

63 Bae JH, Boggon TJ, Tomé F, et al. Asymmetric receptor contact is required for tyrosine autophosphorylation of fibroblast growth factor receptor in living cells. Proc Natl Acad Sci Internet. 2010 cited 2022 Feb 10;107:2866–2871. Available from: https://www.pnas.org/content/107/7/2866.

64 Balaji K, Vijayaraghavan S, Diao L, et al. AXL inhibition suppresses the DNA damage response and sensitizes cells to PARP inhibition in multiple cancers. Mol Cancer Res Internet. 2017 cited 2022 Feb 18;15:45–58. Available from: https://aacrjournals.org/mcr/article/15/1/45/285706/AXL-Inhibition-Suppresses-the-DNA-Damage-Response.

65 Revach OY, Sandler O, Samuels Y, et al. Cross-talk between receptor tyrosine kinases AXL and ERBB3 regulates invadopodia formation in melanoma cells. Cancer Res Internet. 2019 cited 2022 Feb 18;79:2634–2648. Available from: https://aacrjournals.org/cancerres/article/79/10/2634/543883/Cross-Talk-between-Receptor-Tyrosine-Kinases-AXL.

66 Hoj JP, Mayro B, Pendergast AM. A TAZ-AXL-ABL2 Feed-Forward Signaling Axis Promotes Lung Adenocarcinoma Brain Metastasis. Cell Rep Internet. 2019 cited 2022 Feb 18;29:3421-3434.e8. Available from: https://pubmed.ncbi.nlm.nih.gov/31825826/.

67 Zhang Z, Lee JC, Lin L, et al. Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat Genet Internet. 2012 cited 2022 Mar 13;44:852–860. Available from: https://pubmed.ncbi.nlm.nih.gov/22751098/.

68 Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature Internet. 2019 cited

2022 Feb 18;575:299–309. Available from: https://pubmed.ncbi.nlm.nih.gov/31723286/.

69 Casado P, Rodriguez-Prados JC, Cosulich SC, et al. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. Sci Signal Internet. 2013 cited 2022 Feb 18;6. Available from: https://pubmed.ncbi.nlm.nih.gov/23532336/.

70 Beekhof R, Alphen C van, Henneman AA, et al. INKA, an integrative data analysis pipeline for phosphoproteomic inference of active kinases. Mol Syst Biol Internet. 2019 cited 2022 Feb 18;15:e8250. Available from: https://onlinelibrary.wiley.com/doi/full/10.15252/msb.20188250.

71 Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res Internet. 2003 cited 2022 Feb 18;31:3635–3641. Available from: https://pubmed.ncbi.nlm.nih.gov/12824383/.

72 Horn H, Schoof EM, Kim J, et al. KinomeXplorer: an integrated platform for kinome biology studies. Nat Methods Internet. 2014 cited 2022 Feb 18;11:603–604. Available from: https://pubmed.ncbi.nlm.nih.gov/24874572/.

73 Creixell M, Meyer AS. Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data. Cell Reports Methods. 2022;2:100167.

74 KE S, VS S, EA W, et al. Ensemble clustering of phosphoproteomic data identifies differences in protein interactions and cell-cell junction integrity of HER2-overexpressing cells. Integr Biol (Camb) Internet. 2017 cited 2021 Jul 10;9:539–547. Available from: https://pubmed.ncbi.nlm.nih.gov/28492659/.

75 Morshed N, Lee MJ, Rodriguez FH, et al. Quantitative phosphoproteomics uncovers dysregulated kinase networks in Alzheimer's disease. Nat Aging 2021 16 Internet. 2021 cited 2022 Mar 13;1:550–565. Available from: https://www.nature.com/articles/s43587-021-00071-1.

76 Chitforoushzadeh Z, Ye Z, Sheng Z, et al. TNF-insulin crosstalk at the transcription factor GATA6 is revealed by a model that links signaling and transcriptomic data tensors. Sci Signal Internet. 2016 cited 2022 Mar 13;9. Available from: https://pubmed.ncbi.nlm.nih.gov/27273097/.

77 Roux KJ, Kim DI, Burke B, et al. BioID: A Screen for Protein-Protein Interactions. Curr Protoc protein Sci Internet. 2018 cited 2022 Mar 13;91:19.23.1. Available from: /pmc/articles/PMC6028010/.

78 Alexander J, Lim D, Joughin BA, et al. Spatial Exclusivity Combined with Positive and Negative Selection of Phosphorylation Motifs Is the Basis for Context-Dependent Mitotic Signaling. Sci Signal Internet. 2011 cited 2022 Mar 13;4:ra42. Available from: /pmc/articles/PMC3939359/.

79 Begley MJ, Yun CH, Gewinner CA, et al. EGF-receptor specificity for phosphotyrosine-primed substrates provides signal integration with Src. Nat Struct Mol Biol Internet. 2015 cited 2022 Mar 13;22:983–990. Available from: https://pubmed.ncbi.nlm.nih.gov/26551075/.

80 Kooij B van de, Creixell P, Vlimmeren A van, et al. Comprehensive substrate specificity profiling of the human nek kinome reveals unexpected signaling outputs. Elife. 2019;8.

81 Obata T, Yaffe MB, Leparc GG, et al. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. J Biol Chem Internet. 2000 cited 2022 Mar 13;275:36108–36115. Available from: https://pubmed.ncbi.nlm.nih.gov/10945990/.

82 Shah NH, Löbel M, Weiss A, et al. Fine-tuning of substrate preferences of the src-family kinase lck revealed through a high-throughput specificity screen. Elife. 2018;7.

83 Levi T, Sloutskin A, Kalifa R, et al. Efficient In Vivo Introduction of Point Mutations Using ssODN and a Co-CRISPR Approach. Biol Proced Online Internet. 2020 cited 2022 Feb 24;22. Available from: https://pubmed.ncbi.nlm.nih.gov/32684853/.

84 Polat AN, Özlü N. Towards single-cell LC-MS phosphoproteomics. Analyst Internet. 2014 cited 2022 Feb 24;139:4733–4749. Available from: https://pubmed.ncbi.nlm.nih.gov/25068147/.

85 Gerritsen JS, White FM. Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells. Expert Rev Proteomics Internet. 2021 cited 2022 Feb 24;18:661–674. Available from: https://pubmed.ncbi.nlm.nih.gov/34468274/.

86 Wei W, Shin YS, Xue M, et al. Single-Cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma. Cancer Cell Internet. 2016 cited 2022 Feb 24;29:563–573. Available from: https://pubmed.ncbi.nlm.nih.gov/27070703/.

87 Lun XK, Bodenmiller B. Profiling Cell Signaling Networks at Single-cell Resolution. Mol Cell Proteomics Internet. 2020 cited 2022 Feb 24;19:744–756. Available from: https://pubmed.ncbi.nlm.nih.gov/32132232/.

88 Stopfer LE, Flower CT, Gajadhar AS, et al. High-density, targeted monitoring of tyrosine phosphorylation reveals activated signaling networks in human tumors. Cancer Res Internet. 2021 cited 2021 May 24;81:canres.3804.2020. Available from: https://pubmed.ncbi.nlm.nih.gov/33509940/.

CHAPTER 4

Predictive data-driven modeling of C-terminal tyrosine function in the EGFR signaling network

Jacqueline S. Gerritsen

Joseph S. Faraguna

Rudy Bonavia

Frank B. Furnari

Forest M. White

Submitted to Molecular Systems Biology, March 202

Abstract

The Epidermal Growth Factor Receptor (EGFR) has been studied extensively due to its critical role in cellular signaling and association with disease. Previous models have elucidated interactions between EGFR and downstream adaptor proteins, or showed phenotypes affected by EGFR. However, the link between EGFR phosphorylation and phenotypic outcome is still poorly understood. Here, we employed a suite of isogenic cells lines expressing site-specific mutations at each of the EGFR C-terminal phosphorylation sites to interrogate their role in signaling network and cell biological response to stimulation. Our results demonstrate the resilience of the EGFR network, which was largely similar even in the context of multiple Y-to-F mutations in the EGFR C-terminal tail, while also revealing nodes in the network that have not previously been linked to EGFR signaling. Our data-driven model highlights signaling network nodes associated with distinct EGF-driven cell responses, including migration, proliferation, and receptor trafficking. Application of this same approach to less studied RTKs should provide a plethora of novel associations that should lead to a much better understanding of these signaling networks.

Introduction

Epidermal Growth Factor Receptor (EGFR) is a receptor tyrosine kinase (RTK) that is overexpressed, mutated, or otherwise dysregulated in a large fraction of human cancers [1]. On ligand binding, EGFR assumes an active conformation, dimerizes, and cross-/ auto-phosphorylation occurs on C-terminal tyrosine residues, thus initiating a cascade of signaling events ultimately resulting in cellular phenotypes such as proliferation and migration, among others [2]. EGFR signaling has been well characterized in a variety of cell lines, including immediate-early signaling with high temporal resolution as well as deep characterization of the network-wide effects of EGFR activation [3–5]. Despite these large scale phosphoproteomics studies, mechanistic insights connecting kinase and phosphatase activity with particular phosphorylation sites and their associated protein-protein interactions have been challenging. To address this issue and gain a better understanding of how EGFR activity regulates cell biology, there have been extensive efforts to define the association between EGFR C-terminal tyrosine phosphorylation sites (e.g. Y845, Y974, Y992, Y1045, Y1068, Y1086, Y1101, Y1148, Y1173) and recruitment of adaptor or effector proteins. Many of these studies have been performed using in vitro assays, including a landmark study using protein microarrays to identify and quantify the affinity of Src-homology 2 (SH2) and phosphotyrosine binding (PTB) domains for peptides containing phosphorylated tyrosine residues from ErbB family members [6], and another using tyrosine phosphorylated peptides from EGFR to identify protein binders in cell lysate [7]. These studies, along with a plethora of cell-based studies, have yielded critical insight for selected EGFR tyrosine phosphorylation sites. For instance, phosphorylated Tyr 992 has been shown to function as a major binding site for 1-phosphatidylinositol 4,5-bisphosphate

phosphodiesterase gamma-1(PLCɣ-1) [8,9], while phosphorylation of Y1045 serves as the primary EGFR binding site for Cbl, an E3 Ubiquitin ligase that mediates receptor ubiquitination, endocytosis, and degradation or recycling [10]. Phosphorylated residues Y1068 and Y1086 bind the growth factor receptor-bound protein 2 (Grb2) adaptor protein [11,12], while phosphorylated Tyr1148 and Tyr1173 can recruit the Shc adaptor protein [13]. Recruitment and subsequent phosphorylation of Shc and/or Grb2 can lead to activation of the Erk MAPK pathway and thereby mediate activated cell proliferation or cell migration. In addition to these well-characterized interactions, several pTyr sites on the receptor have been postulated to bind multiple adaptor or effector proteins, and several adaptors can bind to multiple sites. How these interactions are regulated in the cellular environment and how they control downstream signaling is still poorly understood.

Further complicating this picture, cytosolic tyrosine kinases and tyrosine phosphatases can be recruited to phosphorylated EGFR and function to provide positive and negative feedback to regulate receptor phosphorylation and activity. For instance, in a positive feedback loop, initial activation of EGFR can lead to activation of Src (or Src-family kinases) that can then phosphorylate Y845 in the EGFR activation loop, further increasing EGFR activity [14]. Other cytosolic kinases such as the Abelson tyrosine kinase (Abl) have been shown to phosphorylate other cytosolic residues in the C-terminal tail, including Y1173 [15]. Tyrosine phosphatases, including PTP1B and SHP2, have been postulated to be recruited to the receptor and serve to dephosphorylate C-terminal tyrosine residues, thus serving as negative feedback to limit receptor activity [16]. Inhibition of the phosphatases has been shown to have a strong impact on some auto-phosphorylation sites, even in the immediate-early stages following ligand

stimulation [3]. Finally, crosstalk with other RTKs can add to the complexity, through hetero-dimerization or oligomerization and cross-phosphorylation of tyrosine sites on other RTKs, or potentially mediated by activation of cytosolic tyrosine kinases [17]. Regardless of mechanism, tyrosine phosphorylation of other RTKs may lead to adaptor recruitment, phosphorylation, and additional downstream signaling that could alter phenotypic outcome, including therapeutic resistance [18].

Since tyrosine phosphorylation controls many aspects of cellular and tumor biology, analysis of phosphorylation-mediated signaling networks can provide crucial information on novel biomarkers or resistance mechanisms. However, many approaches fail to provide predictive insights as the resulting models are based on few nodes with data collected from isolated interaction experiments (e.g. protein-protein interaction experiments) that do not consider the complexity of intracellular network interactions [19,20].

To attempt to elucidate the functional roles of individual EGFR C-terminal tyrosine phosphorylation sites of EGFR in mediating cellular signaling and phenotypic outcome, NR6 cells lacking ErbB expression were infected to stably express wild-type EGFR or EGFR mutant isoforms containing one or more site-specific mutations of selected C-terminal tyrosine phosphorylation sites. Phosphoproteomic analysis of EGF-stimulated cellular signaling in each cell line demonstrated the resilience of the EGFR network to loss of selected pTyr sites while also highlighting connections between particular pTyr sites and adaptor or effector proteins. To define associations between loss of selected pTyr sites, signaling network alterations, and cell biological effects, we quantified cell phenotypic responses, including receptor trafficking, cell proliferation, and cell migration. Intriguingly, despite similar overall network response in each

line, the phenotypic outcome was, in many cases, significantly altered across cell lines expressing different mutant isoforms. To gain insight into the connection between our highly dynamic early timepoint phosphoproteomic dataset and the various quantitative phenotypic readouts, we developed a predictive data driven model that highlights the importance of selected signaling nodes in mediating each phenotypic outcome. These data and the predictive model provide novel insights regarding EGFR signaling and suggest potential therapeutic targets to modulate cell biological response to EGFR stimulation.

Results

Development and initial characterization of EGFR-expression system

To evaluate the individual contributions of the C-terminal tyrosines in the EGFR signaling network response, mouse fibroblasts (NR6) lacking ErbB expression were retrovirally transfected to stably express human wild type EGFR (wtEGFR) or mutant EGFR isoforms in which one or more tyrosines in the C-terminal tail were converted to a non-phosphorylatable phenylalanine residue (e.g., Y-to-F mutation) to evaluate loss of function. In total nine cell lines were generated; one expressing wtEGFR, six containing single Y-to-F mutations (Y845F; Y992F; Y1045F; Y1068F; Y1148F; Y173F), and two isoforms with most of the canonical C-terminal tyrosine phosphorylation sites mutated to phenylalanine. The latter two isoforms contain 5 Y-to-F mutations (Y992F, Y1068F, Y1086F, Y1148F, and Y1173F) (DY5), or 6 Y-to-F mutations (Y845F, Y992F, Y1068F, Y1086F, Y1148F, and Y1173F) (DY6) (Fig 1A). The DY5 and DY6 multiple mutant isoforms allow for evaluation of signaling effects in the context of loss of the major

autophosphorylation sites; comparison of DY6 to DY5 also allows for assessment of gain-of-function of phosphorylation of Y845, located in the activation loop.

Following successful transfection, expression levels were determined for all cell lines and replicates by flow cytometry (Supplemental Fig 1). Expression of each mutant isoform led to relatively slight changes in surface level receptor expression for most of the transfected cell lines, with Y845F and Y1045F being the major outliers in terms of surface expression (Fig 1B).

Phosphorylation of the Y1045 site has been associated with binding of the Cbl E3 ubiquitin ligase and subsequent ubiquitination of EGFR. Blocking phosphorylation at this site through Y-to-F mutation should therefore result in altered receptor trafficking, including increased recycling and decreased degradation, providing a potential explanation for the significantly greater cell surface expression relative to wtEGFR.

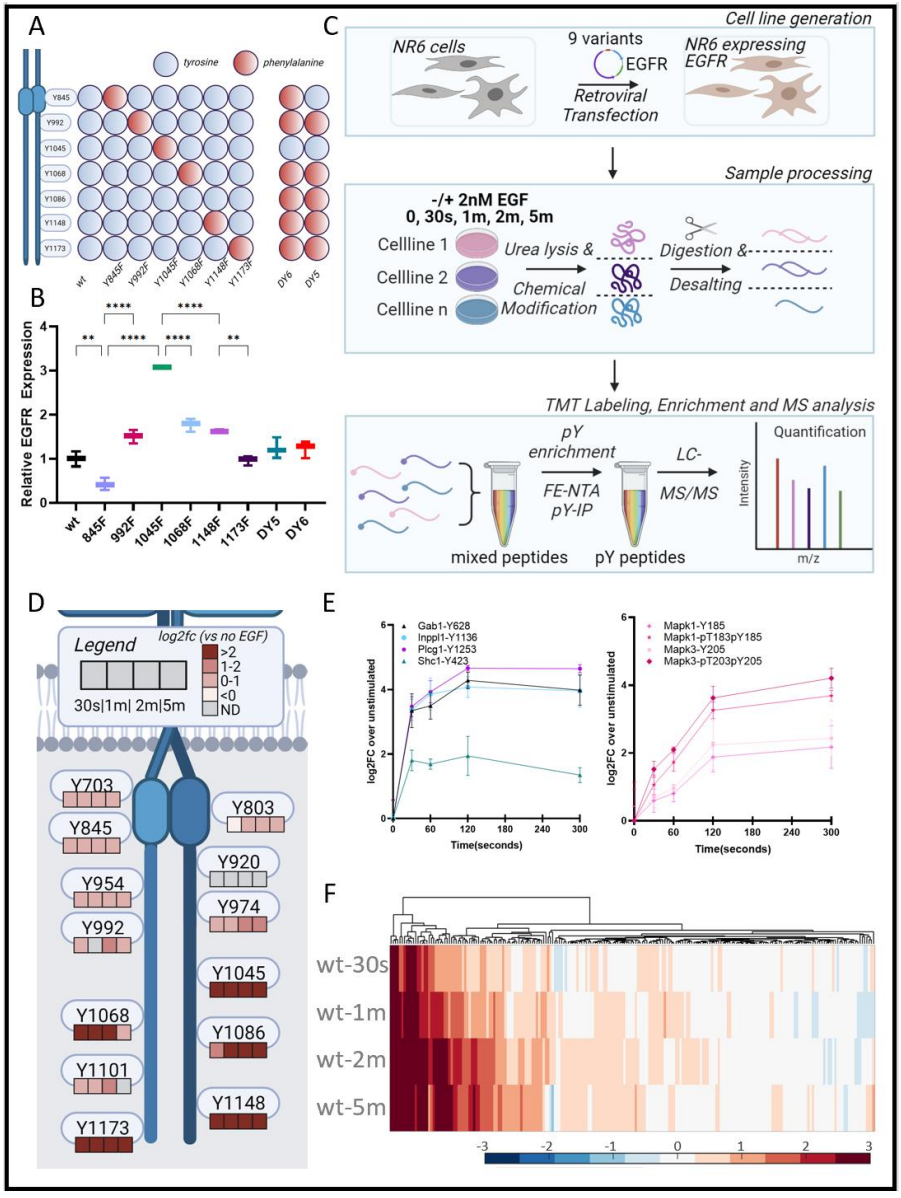


Figure 11: ***EGFR expressing model system exhibits strong EGF response.***
*A. A panel of 9 cell lines that was generated, including wild type, 6 single Y-to-F mutants, and two multiple mutant isoforms lacking 5 or 6 tyrosine residues. Red and Blue dots indicate presence of Phenylalanine or Tyrosine at residue location, respectively. B. EGFR expression levels as determined by flow cytometry with EGF-labeled fluorophore, normalized to wild type, data are presented as mean ± SD. *P≤0.05 (One-way ANOVA), n=3. C. Workflow schematic of NR6 retroviral transfection, cell stimulation and sample processing, followed by TMT labeling and phosphotyrosine enrichment steps for LC-MS/MS analysis. D. EGFR peptide phosphorylation levels in wild type EGFR expressing cells normalized to unstimulated condition, log2 fold change, n=3. E. Temporal dynamics of phosphorylation levels of several downstream nodes in EGFR network, normalized to unstimulated condition, log2 fold change. Data are presented as mean ± SD (student's t-test) n=3. F. Hierarchical clustering of wild type tyrosine phosphorylation data, normalized to unstimulated condition, log2 fold change, data are presented as mean, n=3.*

After confirming that receptor surface expression was roughly similar across most of the mutant isoform expressing cell lines, to assess the effect of each mutation on the cellular signaling network, cells were stimulated with 2nM EGF for 0 seconds, 30 seconds, 1 minute, 2 minutes or 5 minutes. At the appropriate time following stimulation, cells were snap-frozen in liquid nitrogen, lysed in cold 8M Urea to preserve physiological signaling, and proteolytically digested to peptides. To accommodate the large number of samples (5 time points per cell line, 9 total cell lines, and 3 biological replicates per condition), peptides from each cell lysate were labeled with isobaric tandem mass tags and analyzed as 10-plex experiments (sample labeling scheme for each analysis in Supplemental Fig 2) with pooled normalization controls. Labeled samples were subjected to 2-step phosphotyrosine enrichment and subsequent liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis (Fig 1C). These analyses yielded quantitative data for 869 tyrosine phosphorylation (pTyr) containing peptides, on average, for each analysis, with 254 pTyr sites being present across all conditions in at least 2 replicates and 217 pTyr sites present in all conditions and all replicates (quantitative pTyr data for all cell lines and conditions can be found in Supplemental Table 1).

NR6 cells expressing wild-type exhibit signaling network response to EGF stimulation

With this large data set in hand, we initially interrogated pTyr temporal dynamics in wtEGFR-expressing NR6 cells as a starting point to ensure accurate signaling in the NR6 model system. Dynamic pTyr data from these cells displayed expected behavior, with well-characterized autophosphorylation sites in the EGFR C-terminal tail (Y1045, Y1068, Y1086, Y1148, and Y1173)

all displaying strong, significant increase in phosphorylation within 30 seconds following stimulation. Other sites on the receptor (Y703, Y803, Y845, Y954, Y992, Y1101) were detected and also demonstrated significantly increased phosphorylation levels upon EGF induction (Fig 1D). Although NR6 cells do not naturally express EGFR or other ErbB receptors, they are signaling competent, in agreement with multiple previous studies [21–24]. Signaling pathways downstream of the receptor in our wtEGFR-expressing NR6 cells are intact and responsive to EGFR stimulation, as shown by a strong temporal response of several well-characterized proximal adaptor protein phosphorylation sites as well as the ERK 1/2 mitogen activated protein kinases (MAPK1 and MAPK3) (Fig 1E). Signaling effects of EGF stimulation are widespread in these cells, and beyond the canonical EGFR signaling network, we detected an overall increase in pTyr phosphorylation of the signaling network over time (Fig 1F) [7]. Of the 217 pTyr sites detected in these cells, 59% (128/217) were significantly altered by EGF stimulation at one or more timepoints.

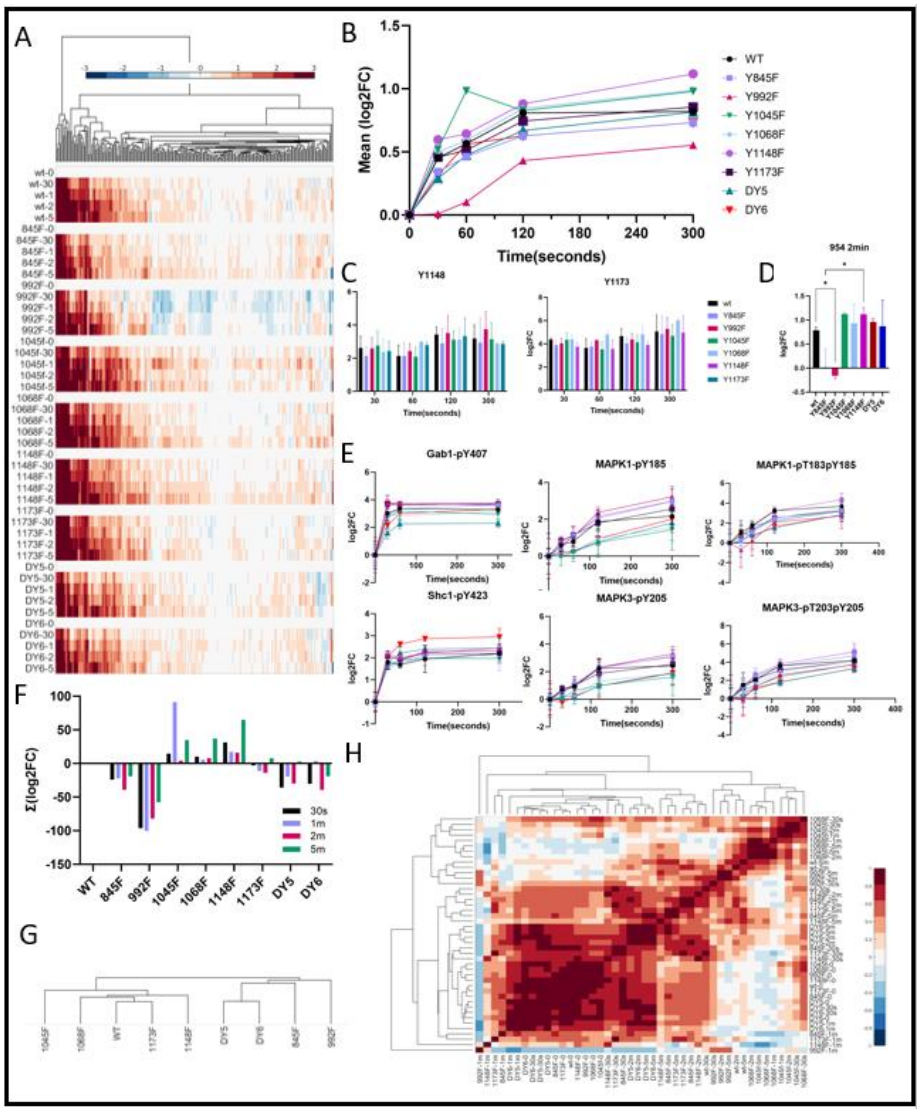


Figure 12: ***EGFR network is largely resilient to loss of pY sites***

*A. Hierarchical clustering of tyrosine phosphorylation data of all EGFR isoforms. Log2 fold change over basal levels, data are presented as mean, n=3. B. Time course plot showing averaged log2 fold change across peptides for each cell line at each timepoint, n=3. C. EGFR-Y1148 and EGFR-Y1173 phosphorylation levels in cell lines (if detected), log2 fold change. No significant differences between conditions. Data are presented as mean ± SD (student's t-test). D. Tyrosine phosphorylation levels of EGFR peptides containing Y954, showing lack of EGF-induced phosphorylation in Y992F isoform. Data are presented as mean ± SD, *P≤0.05 (student's t-test) E. Time course plots of tyrosine phosphorylation levels of several downstream nodes of EGFR network, normalized to unstimulated conditions, log2 fold change. Data are presented as mean ± SD, *P≤0.05 (student's t-test), n=3. F. Sum of log2 fold change for each cell line and timepoint demonstrating overall amount of phosphorylation between timepoints and cell lines. G. Dendrogram depicting how cell lines cluster based on similarity in EGF response, as calculated by Area Under the Curve (AUC). H. Co-clustering analysis of individual cell lines and time points demonstrate order of similarity between conditions. *P≤0.05 (one-way ANOVA), n=3.*

As a negative control, parental, non-transfected NR6 cells were stimulated with EGF and subjected to the same analysis. Although an increase in phosphorylation was observed at the 30 second timepoint, the signaling network nodes responding to stimulation were largely comprised of stress-response signaling, including reactive-oxygen species (ROS) dependent activation of Src-family kinases along with JNK and p38, with no EGFR peptides detected in the analysis. Moreover, the overall trend indicates diminished activation at later timepoints, further suggestive of an EGFR-independent stress response (Supplemental Fig 3). Taken together, these results suggest that NR6 cells transfected to express wtEGFR respond to EGF stimulation in a manner that is highly similar to EGFR signaling in other lines [25], thus establishing this as a viable model system to study the effects of loss of phosphorylation on select C-terminal sites**.**

EGFR network is largely resilient to mutation of C-terminal tyrosine residues

Having confirmed that NR6 cells transfected to express wtEGFR faithfully recapitulate EGFR signaling, we next interrogated the temporal pTyr phosphoproteomic data from NR6 cells expressing each of the mutant isoforms to assess the effect of each mutation on phosphorylation of EGFR, the adaptor proteins, and downstream signaling networks. To account for the different initial signaling states of each line (Supplemental Fig 4), temporal phosphorylation data for each line were normalized to their own unstimulated ("0 sec") time point, thus extracting quantitative changes induced by EGF stimulation at each time point. Interestingly, data from the mutant isoform-expressing NR6 lines indicate largely intact and robust signaling response to EGF stimulation, regardless of site-specific mutation. The overall resilience of EGFR signaling can be seen from the high degree of similarity (quantified in Supplemental Fig 5) between the different cell lines in the heatmap in Fig 2A and average

network phosphorylation levels at each timepoint for each cell line in Fig 2B. Surprisingly, mutating the Y1148 residue did not seem to affect the overall network response compared to wtEGFR, even though this site is reported to be 3-4 times more phosphorylated than other sites on EGFR [3,26]. Moreover, even the cell lines expressing DY5 and DY6 isoforms, lacking 5 or 6 tyrosines, respectively, were able to induce a robust phosphorylation network response.

Correlation clustering of the EGFR phosphopeptide data confirmed high similarity in EGFR activation between isoforms (Supplemental Fig 6). Drilling down further and comparing individual phosphopeptide levels between isoforms, phosphorylation levels of Y1173 and Y1148 were unaffected by mutation of other tyrosine residues, whereas phosphorylation of Y954 was significantly downregulated in isoform Y992F after 2 minutes of EGF treatment (Fig 2C and Fig 2D). This data suggests that dominant phosphorylation sites such as Y1148 and Y1173 phosphorylation may be less dependent on phosphorylation of other residues, whereas Y954 phosphorylation more directly correlates with EGFR's ability to phosphorylate Y992. Network nodes downstream of EGFR demonstrated robustness as well, as shown for selected well-characterized adapters and effectors of EGFR (Fig 2E). Slight but significant differences were observed between wildtype and Y1068F at the 2min time point for the singly phosphorylated MAPK1 and MAPK3, as well as for Y1045F for the doubly phosphorylated variants. Despite similar phosphorylation response for most EGFR peptides, adaptor proteins, and the ERK MAPK effector proteins, decreased phosphorylation in response to stimulation was observed for multiple phosphorylation sites in the Y992F-mutant expressing cells as well as the Y845F-mutant expressing cells (Fig 2F). These differences were recapitulated in clustering the isoforms based on the Area Under the Curve (AUC) for all peptides, which revealed Y992F- and

Y845F-mutant expressing cell lines among the least correlated with wtEGFR (Fig 2G). Co-correlation analysis of the individual timepoints further exposed the differences in response to EGF between these lines and wtEGFR (Fig 2H). This suggests that although the network appears to be highly resilient overall, selected mutations on the receptor can still alter network response to EGF stimulation.

Downstream network analysis reveals network rewiring for Y992F isoform involving PLCɣ-1 and other RTKs

To further interrogate signaling differences between mutant-isoform expressing lines, we performed principal component analysis (PCA). Although most lines and conditions clustered tightly together in principal component (PC) space, stimulation time points from the Y992F-expressing lines occupied a distinct region, strengthening the observation that Y992F is an outlier in terms of EGF network response compared to the other cell lines (Fig 3A). Within PC space, PC1 largely explains the time component, while PC2 appears to depend on the level of phosphorylation in response to EGF (e.g. Y1045F, Y1148F and Y1068F have a stronger overall phosphorylation response compared to Y992F and the multiple mutants).

Consistent with the documented role of phosphorylated Y992 in PLCɣ-1 binding and activation [8], PLCɣ-1 phosphorylation was significantly decreased in the Y992F isoform as well in DY5 and DY6, both of which contain the Y992F mutation, relative to wtEGFR across all timepoints, confirming this site as an important node for PLCɣ-1 activation (Fig 3B). PLCɣ-1 phosphorylation was down in the Y845F isoform as well.

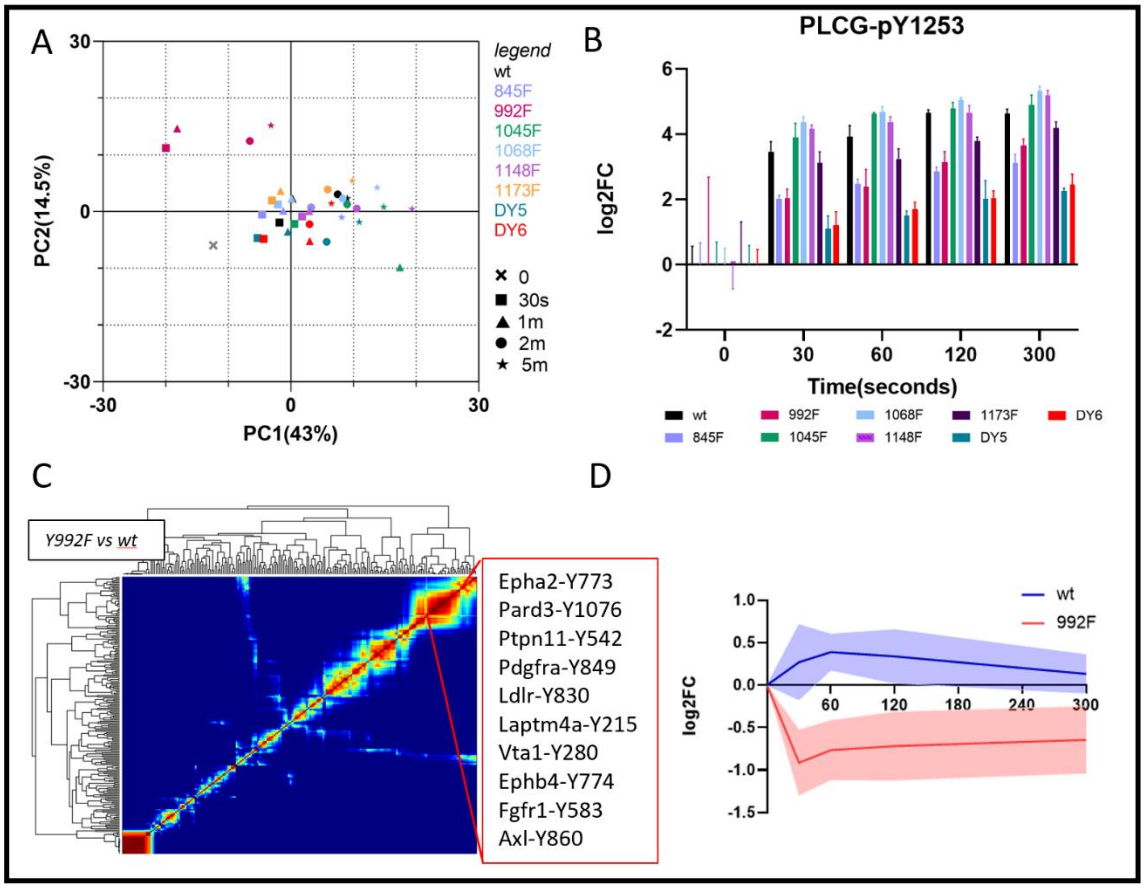


Figure 13: ***Downstream network analysis reveals network rewiring for Y992F isoform involving PLCγ and other RTKs.*** *A. Principle Component Analysis (PCA) explaining 43% (PC1) and 14.5% (PC2) of the variance in the data, data is basal normalized. B. Phosphorylation levels of PLCG-Y1253 at each time point per isoform. Log2fold change compared to basal. Data are presented as mean ± SD, n=3. C. Self-Organizing Map (SOM) clustering to identify similarly differentially regulated peptides between wildtype(wt) and Y992F mutant isoform. Neural network iterations = 1000. D. 95% Confidence interval of peptides from highlighted cluster in panel C, depicting difference in trend between EGFR isoforms. Data presented are mean log2 fold changes and 95% confidence interval, n=3.*

To gain further insight into the earlier observation of decreased phosphorylation on multiple nodes in the Y992F line in the hierarchical clustering analysis (Fig 2A), we generated a self-organizing map (SOM) to identify clusters of co-regulated peptides that were differentially regulated in Y992F compared to wtEGFR (Fig 3C). SOMs utilize neural networks to display similarities in the data, thereby revealing clusters of peptides that are similarly differentially regulated between isoform lines. In the SOM comparing Y992F to wtEGFR signaling response,

one of the large clusters displayed an increase in phosphorylation upon EGF treatment at 30 seconds in the wtEGFR-expressing line (blue line) while these same peptides decrease in phosphorylation upon EGF stimulation in the Y992F-expressing line (red line) (Fig 3D). Interestingly, the majority of peptides in this cluster belonged to other RTKs, including PDGFRα, EPHA2, LDLR, AXL, FGFR1 and EPHB4. This data suggests a potential regulatory function for pY992 in modulating crosstalk with other RTKs or by stabilizing otherwise autoinhibitory intramolecular interactions [27]. Given the role of PLCɣ- as a negative feedback regulator for homodimer formation [28], the increased availability of PLCɣ-1 because of lack of binding opportunity at the 992 residue in the Y992F isoform could have inhibited dimer formation and hence reduced phosphorylation of other RTKs.

Phenotypic characterization confirms known biology and reveals interesting trends

As the phosphoproteomic data reflect overall signaling network similarity as well as some mutant-specific differences, our findings raised the question as to whether and how this variation translates to cellular response. To evaluate the phenotypic consequences of tyrosine loss-of-function, proliferation, migration and internalization rates were measured (Fig 4). Confluency measurements taken over time for each line in the absence or presence of EGF were compared to quantify an EGF-response proliferation rate. EGF-driven proliferation rates were significantly lower for 6 out of 8 mutant isoforms compared to wtEGFR (Fig 4A), with 3 of these lines yielding a counter-intuitive decrease in proliferation in the presence of EGF as compared to the absence of EGF. This result appears to be linked to Y1068F (and may be related to dynamics of the singly phosphorylated ERK activation loop (Fig 2E)) as all three lines contain this mutation.

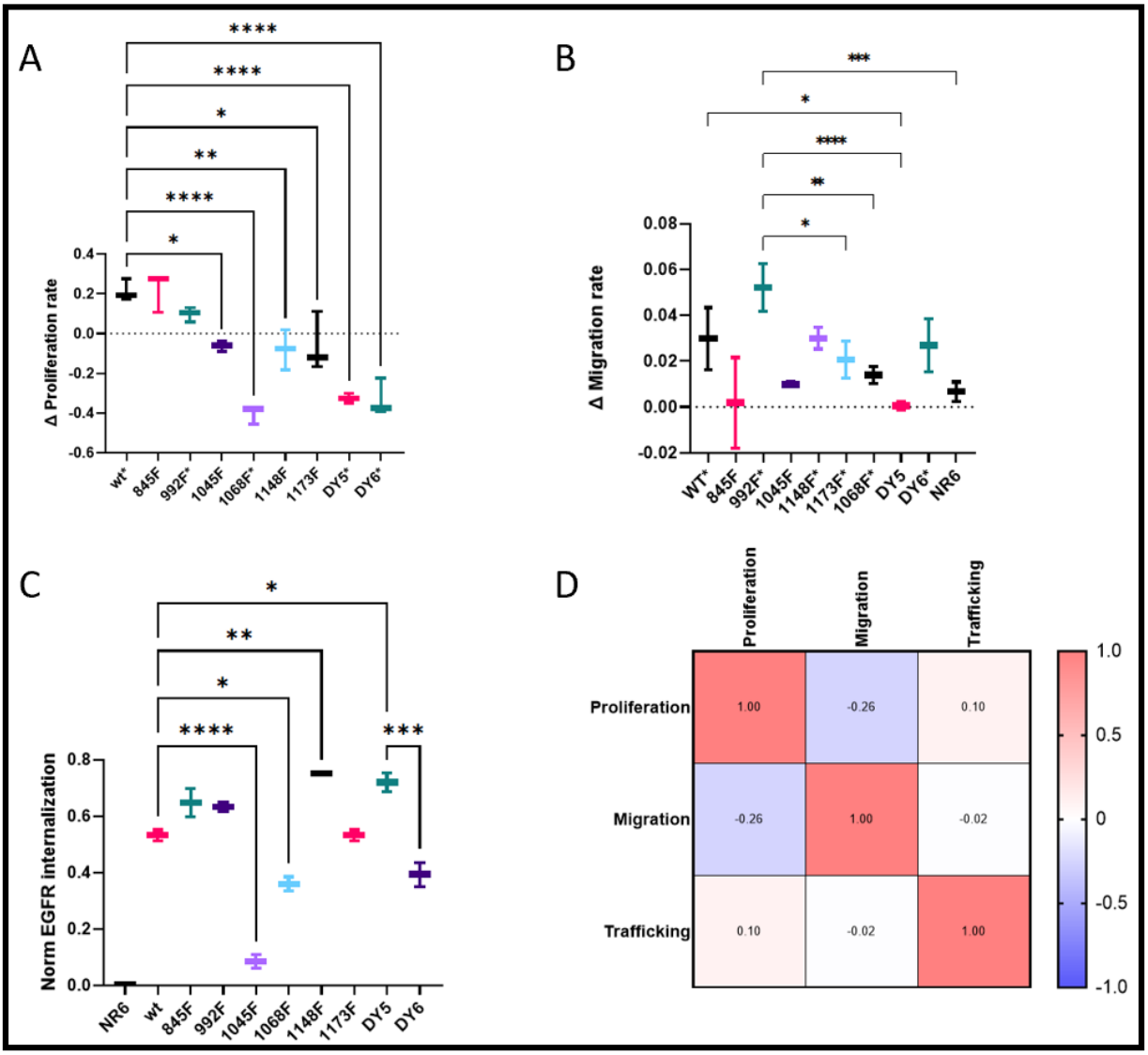


Figure 14: ***EGFR mutants demonstrate different phenotypic response to EGF***
*A. Proliferation rates measured by confluency % over 24 hrs, Data are presented as mean +- SD. *P≤0.05 (One-way ANOVA), n=3. * indicates significant difference compared to wt. * at label indicates significant difference between EGF and no EGF condition. B. Migration rates determined by measuring scratch wound assay closing rate over 24 hrs. Data are presented as mean +- SD. *P≤0.05 (One-way ANOVA), n=3. * indicates significant difference. * at label indicates significant difference between EGF and no EGF condition. C. Receptor trafficking as determined by flow cytometry and pHrodo-EGF conjugate. Untransfected NR6 were used as negative control and normalizing channel. Data are presented as mean +- SD. *P≤0.05 (One-way ANOVA), n=3. D. Correlation analysis between phenotypes, data presented are Pearson correlation coefficients. n=3.*

To evaluate EGF-stimulated cell migration rates, temporal dynamics of wound healing were quantified in confluent plates for each cell line, in the presence or absence of EGF. EGF

stimulated cell migration at a similar rate for most of the mutant isoform expressing lines as compared to wtEGFR, with only DY5 being significantly different (Fig 4B). In comparing the mutant lines to each other, Y992F-expressing cells were significantly more responsive to EGF-stimulated migration as compared to most other lines, despite decreased phosphorylation on many signaling nodes.

In addition to proliferation and migration as measures of downstream phenotypes, receptor trafficking to endosomes was evaluated as a more proximal phenotype. The pHrodo-EGF conjugate demonstrates increased fluorescence in acidic environments. We used this conjugate to quantify internalization and localization of the EGF-EGFR-bound complex to low pH endosomes. The results of this assay highlight significant differences in trafficking between wtEGFR and 4 mutant isoforms (Fig 4C). Perhaps not surprisingly, minimal endosomal trafficking was detected for Y1045F, in accordance with its reported role in CBL binding, ubiquitination, and receptor trafficking. Y1068F also showed decreased endosomal trafficking compared to wildtype. This effect may be explained by Y1068's function as a Grb2-binding site, as Grb2 is reported to be important in early stages of endocytosis where EGF-activated receptors are recruited into clathrin-coated pits [29]. Interestingly, DY6 also showed decreased trafficking, however when Y845 function was recovered (DY5), trafficking was at least as high as wildtype levels, although both isoforms still contained the Y1068F mutation (amongst others), which by itself caused decreased trafficking.

Correlation analysis of the different phenotypes revealed a slight negative correlation between migration and proliferation, which supports the principle of phenotype-switching as cells decide which phenotype to pursue (Fig 4D) [30]. Curiously, there appears to be minimal association

between network-level phosphorylation and phenotypic outcome in response to EGFR stimulation, as most mutant isoforms displayed relatively minor signaling changes at the network level relative to wtEGFR, yet several had significantly altered phenotypic response. On the other side, cells containing the Y992F mutant isoform were the most different from wtEGFR in signaling, yet this same mutation appeared to have minimal phenotypic impact.

Partial Least Squares Regression model reveals Epsin 8- and PLCγ-1-mediated mechanisms associated with cellular phenotype

To gain greater insight into the associations between protein site phosphorylation dynamics and phenotypic response to EGF stimulation, we used partial least square regression (PLSR) to integrate these quantitative data sets and determine drivers of the observed phenotypic differences. After selecting the optimal number of latent variables, the model was generated based on leave-one-out cross validation (LOOCV). The model was predictive of phenotype with a fit to data average R^2 of 0.97, while cross-validation provided highly predictive Q^2 values >0.75 for all phenotypes and signaling data timepoints (Supplemental Fig 7), as well as using the Area Under the Curve (AUC) as a metric of total phosphorylation for each site over the time course. From this model, variable importance in projection (VIP) scores were calculated to identify the most critical contributors to the PLSR model.

To assess the accuracy of the predicted model associations, we checked the phosphorylation sites predicted to be positively associated with cell proliferation (Supplemental Table 2). In agreement with previous reports, phosphorylation of the activation loop of the ERK MAPKs (MAPK1 and MAPK3) at 2 and 5 minutes following stimulation was found to be positively

associated with proliferation, as was phosphorylation of SCL38A2 (Y41) [31–33]. Additionally, CDK1/2 Y15 is positively associated with cell proliferation in our model; while this site is known to inhibit progression through the cell cycle, cells that proliferate more frequently have increased CDK Y15 phosphorylation, agreeing with the model prediction. On the other side, tyrosine phosphorylation of Afadin (Y1230, Y1285), Girdin (Ccdc88a) (Y1801), and Erbin (Y1097) are negatively associated with proliferation. Tyrosine phosphorylation of these proteins has been shown to promote cell migration, further suggesting a cell decision process downstream of EGFR, where phosphorylation of migration associated proteins is negatively correlated with cell proliferation [34,35]. For the PLSR model integrating protein phosphorylation with cell migration, the above sites on Afadin, Girdin, and Erbin were all positively associated with cell migration, as were sites on Delta-Catenin (Y228, Y865, Y904) and PLC ɣ-1 (Y771, Y1253), also supporting the predictive power of this model. The PLSR model integrating protein phosphorylation and receptor endocytosis has similar predictive power, with a Q2 = 0.87, yet the signaling networks associated with receptor internalization and trafficking are less characterized, so interpretation of the model predictions is more challenging. For instance, phosphorylation of Grb2-associated binding protein 1 (Gab1) on Y628 and Y660 and phosphorylation of the Erk 1/2 MAPK activation loops (T183/Y185; T203/Y205) are positively associated with receptor endocytosis, in agreement with a previous publication suggesting that Gab1 is recruited to early endosomes and is the primary mediator of Erk MAPK signaling following receptor endocytosis [36].

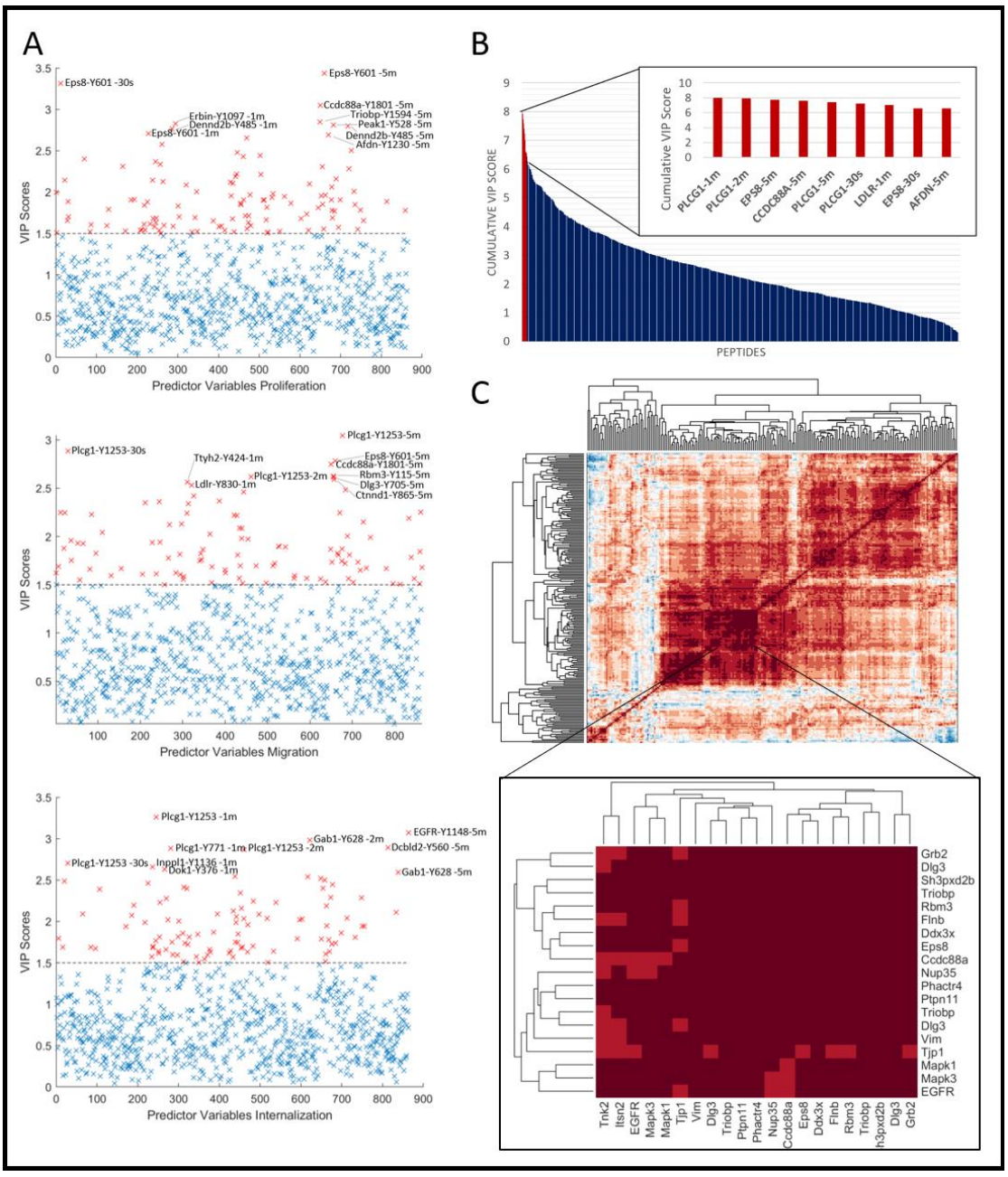


Figure 5: ***PLSR reveals PLCy1 and Eps8- mediated mechanisms predictive of cell phenotypes.*** *A. Variable Importance in the Projection (VIP) statistics depicting most predictive nodes for each phenotype based on signaling data (AUC). Data presented based on 3 component PLSR, cross validated model. B. Cumulative VIP scores across models, rank-ordered. C. Co-correlation map of basal-normalized pTyr data (mean log2 fold change) highlighting subcluster containing Eps8 peptide (Y602). PLSR model built using LOOCV and 3 components, R2 and prediction accuracy (Q2) values can be found in Supplemental Figure 7.*

However, the roles of other sites in this phenotype, including a strong negative association between EGFR internalization and PLCɣ-1 phosphorylation (Y771, Y783, Y1253) are less clear. It may be that EGFR phosphorylation of PLCɣ-1 occurs on the cell membrane, and that therefore increased internalization of EGFR leads to decreased phosphorylation of PLCɣ-1, but this hypothesis would need to be shown experimentally.

One of the strongest contributors to the proliferation and migration models is pY601 of Epidermal growth factor receptor pathway substrate 8 (Eps8) (Fig 5A and Fig 5B). Eps8 is a substrate of EGFR and has been shown to be involved in regulating cancer progression [37–39]. In our PLSR models, Eps8 was found to be a strong negative predictor for proliferation, whereas it was a robust positive predictor for migration. In agreement with our model predictions, phosphorylation of Eps8 Y601 has been reported to block proliferation and to promote cell migration [40]. In cell lines bearing the Y1068F mutation, Eps8 Y601 is significantly decreased in phosphorylation, suggesting that Eps8 and 1068F are directly or indirectly functionally correlated. To gain further insight into other protein phosphorylation sites that may be co-regulated with Eps8, we performed co-correlation clustering. This analysis showed that Eps8 was most closely correlated with Y104 of the DEAD box helicase DDX3X, a protein whose phosphorylation has not been previously linked to the EGFR signaling pathway, yet DDX3X has been recently suggested to play a role in EGFR Tyrosine Kinase Inhibitor (TKI) resistance [41]. Also highly co-regulated with Eps8 are Girdin, Filamin b (Flnb), RNA-binding motif protein 3 (Rbm3), TRIO and F-actin binding protein (Triobp), and Grb2, all of which are predicted to be positively associated with cell migration and most of which are negatively associated with cell proliferation (Fig 5C). Intriguingly, several of these proteins, including DDX3X, Rbm3, and

Triobp, have not been previously linked to EGFR signaling, yet through our analysis of signaling networks associated with different mutant isoforms of EGFR, we were able to not only connect these protein phosphorylation sites to EGFR activity, but also suggest associations with phenotypic outcomes downstream of EGFR activation.

Discussion

In Given the well-established role of EGFR as a driver oncogenic kinase in many cancer types, coupled with the need to improve EGFR-targeting strategies in the clinic and the dependence on phosphorylation of C-terminal tyrosines for signal propagation, we took a mechanistic, phosphoproteomics-based approach to investigate the contribution of individual EGFR C-terminal tail tyrosines to the EGFR signaling network. In this work, we utilized NR6 murine fibroblasts that lack endogenous ErbB expression as a model system. NR6 cells have previously been used to assess the EGFR signaling network response to ligand stimulation, and have been shown to respond through increased proliferation and migration, suggesting intact downstream signaling networks and compatibility of murine cells to expression of human EGFR [22–24]. Here, we generated individual and combined Y>F mutant isoforms and stably expressed each of these isoforms in NR6 cells to evaluate loss of function effects for selected C-terminal tyrosine phosphorylation sites.

Moderate to high levels of EGFR expression can lead to receptor activation in the absence of ligand. Basal activity of the stably expressed EGFR isoforms in our NR6 cells resulted in a slightly altered baseline signaling state for each isoform-expressing line. Although we were able to correct for this baseline signaling by normalizing the temporal data to the non-stimulated

condition (e.g., the 'zero' time point), other model system choices may provide for a more straightforward comparison. For instance, stable expression allows the cells to adapt/rewire to each mutant EGFR isoform. One could opt for a transient expression system to limit the amount of rewiring in the cells before stimulation, although the dynamic nature of transient expression, coupled with the difficulty of obtaining consistency in EGFR expression levels across replicates and lines, would be highly challenging. CRISPR-based gene editing to mutate selected tyrosine phosphorylation sites in endogenously expressed EGFR would also enable interrogation of loss-of-function of each site without exogenous stable expression. However, our attempts to achieve gene editing of EGFR sites in several epithelial lines were not successful, potentially due to the inefficiency of current gene editing approaches, making it highly challenging to achieve editing in both genomic copies of EGFR. Moreover, many epithelial cells are dependent on EGFR expression, and thus mutation of one or both copies may have been negatively selected.

Given some of the established connections between adaptor proteins and selected EGFR C-terminal tail phosphorylation sites, our initial hypothesis at the start of the project was that loss of a given phosphorylation site might lead to loss of signal for one or more adaptor proteins and associated signaling pathways. However, to our surprise, the EGFR signaling network was highly resilient, and signaling network response was highly conserved for each loss-of-function mutant isoform, even those with combined mutation of 5 or 6 tyrosines. The strong similarity in network response to EGF stimulation regardless of loss of tyrosine function suggests dynamic cellular rewiring in the context of each mutant isoform. This dynamic adaptation may be a result of a built-in mechanism, for example by using redundancy in adaptor protein binding sites to help compensate for loss of tyrosine residues. Unfortunately, the ability of the signaling

network to compensate for loss of one or more tyrosines made it challenging to isolate the functional interactions of a given tyrosine site.

Previous work has connected signaling network nodes in the ErbB network with consequent phenotypic response to EGFR or Her3 stimulation [31]. We used a similar modeling framework to generate integrative computational models focused on three phenotypes that are known to be regulated by EGFR and cover various degrees of cellular response. Proliferation and migration occur on the hours-to-days timescale, whereas receptor internalization is an earlier phenotypic response that could be measured on the minutes-timescale. Using quantitative data for hundreds of tyrosine phosphorylation sites across multiple time points per cell line and 9 different cell lines, we were able to generate robust and highly predictive models for each phenotype. Intriguingly, although cellular responses such as migration and proliferation take hours-to-days to occur, phosphorylation changes in the immediate-early timepoints, as early as 30 seconds or 1 minute, were found to be highly important in predictive models for these longer-term responses. These model predictions highlight the importance of measuring early phosphorylation events in the signaling network, as these events effectively set the course for the future cellular response to stimulation.

Phosphorylation of PLCɣ-1 and Eps8 emerged as some of the most important sites in the models, with Eps8 as a major feature in predicting cell proliferation (negative association) and cell migration (positive association). Among the EGFR phosphorylation sites, Eps8 was strong correlated with the 1068 residue. We hypothesize that the association between Eps8 and Y1068 may relate to complex formation of Eps8 with Sos1 and Abl1 when Grb2 is unable to bind pY1068, a known binding site of Grb2, in the Y1068F mutant isoform. Under this condition,

increased Eps8-Sos1 complex formation may lead to decreased ERK activity and consequent decreased cell proliferation [42]. Eps8 is known to promote cell migration and also emerged as a strong positive factor in this model, suggesting that phosphorylation of this protein may be part of the switch mechanism utilized by the cell for phenotype decision making [40]. Small molecule inhibitors targeting the EGFR/Eps8 complex in NSCLC and breast cancer models have shown promising results, supporting Eps8 as a potential therapeutic target in cancer [37]. PLCɣ-1 also emerged one of the strongest contributors to the models, with a positive association in the cell migration model and a negative association in receptor endocytosis. Although the role of PLCɣ-1 in promoting cell migration downstream of receptor activation is well established, the connection between receptor endocytosis and PLCɣ-1 requires further investigation. Regression based models provide associations, but often fail to provide mechanism. In this case, it is not obvious whether increased receptor endocytosis leads to decreased PLCɣ-1 phosphorylation, or whether decreased PLCɣ-1 phosphorylation leads to increased receptor trafficking, potentially through altered cAMP and Ca^{2+} signaling. In addition to Eps8 and PLCɣ-1, our regression models also highlight the role of other protein phosphorylation sites in mediating the cell response decision between migration and proliferation. For instance, Girdin (CCdc88a, also known as GIV) emerged as a highly important protein in the proliferation and migration models. Several manuscripts over the past few years have suggested that Girdin may serve to connect RTKs such as EGFR to G-protein coupled receptors to thereby modulate cAMP levels and control the proliferation/migration cell decision process [43–45]. Our data suggest that Y1801 on Girdin may play an important role in this process, as phosphorylation of this site was negatively associated with cell proliferation and positively associated with cell migration. Similarly,

phosphorylation of RBM3 Y115 also emerged as a strong predictor in the proliferation and migration models. While this protein has been previously linked to cell migration, in agreement with our model prediction, RBM3 has not been associated with EGFR signaling previously and Y115 appears to be a novel phosphorylation site. Additional studies are needed to characterize the role of this phosphorylation site on RBM3 function and to determine the mechanism underlying association with cell migration.

This mechanistic approach to uncover novel insights into the EGFR signaling network is easily translatable to other RTKs. It would be interesting to see if all RTKs have similar resilience in signaling response to loss-of-function mutation to one or more tyrosines, or whether there is a hierarchy of sites on selected RTKs. Using regression models to couple quantitative phosphotyrosine phosphoproteomics with quantitative measurements of cell biological response led to predictions as to a potential association for poorly characterized phosphorylation sites with particular cell responses. Application of this same approach to less studied RTKs should provide a plethora of novel associations that should lead to much better understanding of these signaling networks.

Materials and Methods

Plasmid preparation and extraction

Seven out of nine plasmid constructs were kindly donated by F. Furnari Laboratory (UCSD, CA). Wild-type and 1045F mutant plasmids were generated from 845F plasmid using Genscript site-directed mutagenesis service and sequences verified with Sanger sequencing (EtonBio). Plasmids were pLNCX2 retroviral vectors with a neomycin and Ampicillin resistance gene, as

well as the EGFR coding sequence containing a variety of Y-to-F mutations (Y845F, Y992F, Y1045F, Y1068F, Y1148F, Y1173F, DY6 (6 Y-to-F mutations (Y845F,Y992F, Y1068F,Y1086F,Y1148F,Y1173F)), DY5 (DY6 except F845Y) and wild-type.

Plasmid DNA was transformed into DH5α E. coli (NEB 5-alpha competent E. Coli (High Efficiency)) according to NEB protocol. Cells from -80^{o}C freezer were thawed on ice for 5 minutes, and 5µl of plasmid DNA was added to the cell mixture. The mixture was incubated on ice for 30 minutes and then heat shocked in a water bath at 42^{o}C for 45 seconds. The mixture was placed on ice for 5 minutes. 950 µL of Super Optimal broth with Catabolite expression (SOC) medium was added to the mixture, which was then shaken at 37^{o}C for 60 minutes. Subsequently, 100 µL was spread onto a Luria Broth (LB) agar with ampicillin plate and incubated overnight at 37^{o}C. Individual colonies were picked and grown for ~12 hours at 37^{o}C in 10mL of LB media with ampicillin (100mg/mL, 1000x). Culture was spun down at 4^{o}C and 10,000 rpm for 10 minutes. Plasmid DNA was isolated using the Qiagen Spin Miniprep Kit. Nanodrop (Thermo) was used to determine DNA concentration.

Retroviral transfection

On day 0, 2e5 293T cells/well were plated in 6 well-plates. The next day, 4µl fuGENE HD Transfection Reagent (Promega, E2311) was added to 100µl of Opti-MEM (Gibco) and vortexed to mix and incubate for 5 minutes at room temperature (RT). Envelope protein VSV-G and packaging genes Gag/Pol were combined with plasmid DNA; 333ng, 666ng and 500ng, respectively. The DNA mix was added dropwise to the solution containing fuGENE and HD transfection reagent, mixed, and left to incubate at RT for 30 minutes. The resulting mixture

was then added dropwise to the 293T cells. On day 2, the media was replaced with regular DMEM containing 1% Pen/Strep and 10% FBS using a P1000 to minimally disturb the cells. NR6 cells were plated in 6 well plates at 5e4 cells/well. The following day, virus was harvested from the 293T cells by collecting media and passing it through a .45µm filter. Polybrene (Sigma) was added to the solution at 6µg/mL, and the media for NR6 cells was replaced with the filtered DMEM containing the virus/polybrene mixture. 8 hours later, the media was removed and DMEM supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptavidin (P/S) added onto the cells. On day 6, Geneticin (G418, Gibco) was added at 45µL/well. One day later, the cells were split and kept on G418 selection pressure.

Flow cytometry for EGFR expression and endocytosis

To select successfully transfected cells as well as to ensure comparable expression levels between cell lines and replicates, EGFR expression was determined by flow cytometry. Biotinylated EGF, complexed to Alexa Fluor 488 (Invitrogen, E13345) was added to each NR6 cell line at a concentration of 0.5 µg/mL and incubated at 37^{o}C for 5 minutes. Cells were put on ice, spun down at 4^{o}C at 1,200 *g* for 5 min and washed twice with cold PBS before analysis on the BD FACScanto Clinical Flow Cytometry System. When selecting successfully transfected cells, the BD FACSAria Sorter was used and gating for EGFR-expressing cells was determined with respect to a negative control.

To evaluate EGFR internalization and trafficking, pHrodo EGF 488 conjugate (Thermo P35375) was added to each NR6 cell line at a concentration of 0.5 µg/mL and incubated at 37^{o}C for 5 minutes. Cells were put on ice, collected from plate using Accutase, spun down at 4^{o}C at 1,200

g for 5 min and washed twice with cold PBS before analysis. Data was acquired using the BD FACSDIVA Software and further analyzed using FlowJo.

Cell culture and EGF stimulation

NR6 mouse fibroblasts (gifted by F. Furnari Laboratory, UCSD, CA) were maintained in DMEM (Corning) medium supplemented with 1% Penicillin-Streptomycin (5,000 U/mL, Gibco), 2% G418 (Gibco) and 10% FBS (Gibco). Approximately 2e5 cells were seeded and incubated for 2 days to reach ~80% confluency in 10cm dishes. For each line and each biological replicate, one plate of cells was used to determine EGFR expression level using Alexa Fluor 488-conjugated EGF and flow cytometry, as described above. Other plates of cells (one 10 cm plate/time point) were stimulated with 2 nM EGF (Peprotech) in serum-free media for 30 seconds, 1 minute, 2 minutes or 5 minutes or left untreated as a control (0minute timepoint).

Sample preparation for MS analysis

After EGF stimulation, media was promptly discarded and cells were snap frozen on liquid N2 for instant arrest of all signaling events. Cells were lysed on ice using 500µL 8M Urea (Sigma) per 10cm plate. A bicinchoninic acid (BCA) protein concentration assay (Pierce) was performed according to the manufacturer's protocol to estimate the protein concentration in each lysate. Cell lysates were reduced with 10mM DTT for 1hr at 56^{o}C, alkylated with 55 mM iodoacetamide for 1hr at RT shielded from light, and diluted 5-fold with 100 mM ammonium acetate, pH 8.9, before trypsin (Promega) was added (20:1 protein:enzyme ratio) for overnight digestion at RT. The resulting solutions were acidified with 1 mL of acetic acid (HOAc) and loaded onto C18 Sep-Pak Plus Cartridges (Waters), rinsed with 10mL of 0.1% HOAc, and eluted with 10mL of 40% Acetonitrile (MeCN)/ 0.1% HOAc. Peptides were divided into XYZ microgram aliquots, and

sample volume was reduced using a vacuum centrifuge (Thermo) and then lyophilized to dryness for storage at -80^{o}C.

TMT labeling for multiplexed analysis was performed according to manufacturer's protocol. Samples, each containing ~200ug peptides, were resuspended in 35µL HEPES (pH 8.5), vortexed, and spun down at 13,400rpm for 1 minute. 400µg of a given channel of TMT10plex (Thermo) in anhydrous MeCN, was added per sample. Samples were shaken at 400rpm for 1hr, after which the labeling reaction was quenched using 5% Hydroxylamine (50%, Thermo). After another 15 minutes on the shaker, all samples were combined using the same pipette tip to reduce sample loss, and sample aliquots were washed twice with 40 µL 25% MeCN/0.1% HOAc which was added to the collection tube to improve yield. Sample volume was reduced using a vacuum centrifuge and then lyophilized to dryness for storage at -80^{o}C.

Phosphopeptide enrichment

Immunoprecipitation (IP) and IMAC were used sequentially to enrich samples for phosphotyrosine containing peptides. TMT-labeled samples were incubated in IP buffer consisting of 1% Nonidet P-40 with protein G agarose beads conjugated to 24 µg of 4G10 V312 IgG and 6 µg of PT-66 (P3300, Sigma) overnight at 4^{o}C. Peptides were eluted with 25 µl of 0.2% trifluoroacetic acid for 10 minutes at room temperature; this elution was performed twice to improve yield. Eluted peptides were subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)-based Fe-NTA spin column to reduce non-specific, non-phosphorylated peptide background. High-Select Fe-NTA enrichment kit (Pierce) was used according to manufacturer's instructions with following modifications. Eluted peptides from IP were incubated with Fe-NTA beads containing 25µL binding washing buffer for

30 minutes. Peptides were eluted twice with 20mL of elution buffer into a 1.7 mL microcentrifuge tube. Eluates were concentrated in speed-vac until ~1 µL of sample remained, and then resuspended in 10µL of 5% acetonitrile in 0.1% formic acid. Samples were loaded directly onto an in-house constructed fused silica capillary column [50 micron inner diameter (ID) x 10 cm] packed with 5µm C18 beads (YMC gel, ODS-AQ, AQ12S05) and with an integrated electrospray ionization tip (~2 micron tip ID).

LC-MS/MS analysis

LC-MS/MS of pTyr peptides were carried out on an Agilent 1260 LC coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptides were separated using a 140-minute gradient with 70% acetonitrile in 0.2 mol/L acetic acid at flow rate of 0.2 mL/minute with approximate split flow of 20 nL/minute. The mass spectrometer was operated in data-dependent acquisition with following settings for MS1 scans: m/z range: 350 to 2,000; resolution: 60,000; AGC target: 3×10^6; maximum injection time (maxIT): 50 ms. The top 15 abundant ions were isolated and fragmented by higher energy collision dissociation with following settings: resolution: 60,000; AGC target: $1x10^5$; maxIT: 350 ms; isolation width: 0.4 m/z, collisional energy (CE): 33%, dynamic exclusion: 20 seconds. Crude peptide analysis was performed on a Q Exactive Plus mass spectrometer to correct for small variation in peptide loadings for each of the TMT channels. Approximately 30 ng of the supernatant from pTyr IP was loaded onto an in-house packed precolumn (100µm ID x 10 cm) packed with 10mm C18 beads (YMC gel, ODS-A, AA12S11) and analyzed with a 70-minute LC gradient. MS1 scans were per-formed at following settings: m/z range: 350 to 2,000; resolution:70,000; AGC target: $3x10^6$;

maxIT: 50 ms. The top 10 abundant ions were isolated and fragmented with CE of 33% at a resolution of 35,000.

Peptide identification/quantification

Mass spectra were processed with Proteome Discoverer version 2.5 (Thermo Fisher Scientific) and searched against the mouse and human (for EGFR peptides) SwissProt database using Mascot version 2.4 (MatrixScience, RRID:SCR_014322). MS/MS spectra were searched with mass tolerance of 10 ppm for precursor ions and 20 mmu for fragment ions. Cysteine carbamidomethylation, TMT-labeled lysine, and TMT-labeled peptide N-termini were set as fixed modifications. Oxidation of methionine and phosphorylation of serine, threonine and tyrosine were searched as dynamic modifications. TMT reporter quantification was extracted and isotope corrected in Proteome Discoverer. Peptide spectrum matches (PSM) were filtered according to following parameters: rank=1, mascot ion score>15, isolation interference<40%, average TMT signal>1,000. Peptides with missing values across any channel were filtered out.

Phenotypic measurements

Proliferation assay EGFR mutant and wildtype expressing cells were seeded in 96-well tissue culture plastic plates (VWR) at 10,000 cells/well in culture media and allowed to adhere for 24 h. Media was replaced by media containing 1% FBS and 2nM EGF or 1% FBS and no EGF. Cells were analyzed by a live-cell imaging platform (Incucyte) every 3 hrs for 48 hrs. Proliferation rates were calculated using confluence measurements and correlation values between phenotypic measurements were calculated using delta rates with/without EGF stimulation.

Migration Wound scratch assay EGFR mutant and wildtype expressing cells were seeded at 1e5 cells/well in 96-well tissue culture plastic plates (Essen Bioscience) at 100,000 cells/well in

culture media and allowed to adhere for 24 h and reach ~90% confluency. The Wound Maker (Essen Bioscience) was used to create a scratch across the well surface. Wells were washed with PBS to remove floating cells and media was added containing 1% FBS and either 2nM EGF or no EGF. Wound closure was monitored by a live-cell imaging platform (Incucyte with Scratch wound cell migration software module) every 3 hrs for 48 hrs. Migration rates were calculated using wound width and wound confluency measurements.

Internalization/Receptor trafficking measurements Biotinylated EGF complexed with pH-sensitive fluorophore pHrodo (Thermo) was used to measure receptor internalization for each line. This fluorophore is weakly fluorescent outside the cells at neutral pH but become brightly fluorescent in acidic endosomes after EGFR internalization. Cells were plated in 6-well plates and cultured to 70-80% confluence. Plates were put on ice for 10 minutes to inhibit any ongoing endocytosis. Cells were then washed with Live Cell Imaging Solution (LCIS, Thermo) supplemented with 20mM Glucose. Cells were then incubated in 250uL LCIS/glucose with 0.5 ug/mL EGF-conjugate for 5 minutes at 37C. Then, cells were washed in cold LCIS, and kept on ice while being removed from the plate using Accutase (Thermo). Cells were spun down at 1000rpm for 5 minutes and washed twice in cold LCIS before resuspending in Glucose supplemented LCIS for analysis by Flow Cytometry.

Data analysis

Data analyses were performed in MATLAB R2020A, Microsoft Excel 2016 and Graphpad Prism 9. TMT reporter ion intensities from PSMs were summed for each unique phosphopeptide. For protein level quantification, TMT reporter intensities were summed for all unique peptides. Peptide or protein quantification were normalized with relative median values obtained from

crude lysate analysis to adjust for sample loading in TMT channels. A combination of Student t-test and One-way Anova was used to perform statistical analysis between conditions. Statistical significance was assigned for $p<0.05$. Unsupervised hierarchical clustering was performed on the basis of Person correlation distance metric, unless otherwise specified. Protein networks were obtained from STRING (version 11.0) database.

A self-organizing map (SOM) was used to cluster proteins that exhibited similar signaling dynamics following EGF treatment. Clustering analysis was performed using the Self Organizing Map Toolbox MATLAB package (http://www.cis.hut.fi/projects/somtoolbox). A 5-by-5 neural network was initiated with hexagonal lattice structure. The input was the log-2 fold-change in phosphorylation following EGF treatment relative to unstimulated condition for each cell line. The network was randomly initiated and used Euclidean distance as the metric for classifying proteins to specific neurons. The SOM algorithm was repeated 1,000 times, and a co-clustering map was generated indicating the frequency with which any two proteins clustered in the same neuron. This co-clustering map was then subjected to hierarchical clustering using Euclidean distance as the metric for clustering proteins. Biorender was used for schematics. PLSR model was evaluated for goodness of fit (R^2) and goodness of prediction (Q^2) using leave-one-out-cross validation (LOOCV).

References

1 Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. FEBS J Internet. 2010 cited 2022 Feb 4;277:301–308. Available from: https://pubmed.ncbi.nlm.nih.gov/19922469/.

2 Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol Internet. 2001 cited 2022 Feb 4;2:127–137. Available from: https://pubmed.ncbi.nlm.nih.gov/11252954/.

3 Reddy RJ, Gajadhar AS, Swenson EJ, et al. Early signaling dynamics of the epidermal growth factor receptor. Proc Natl Acad Sci U S A Internet. 2016 cited 2017 Jan 5;113:3114–3119. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26929352.

4 Zhang Y, Wolf-Yadlin A, Ross PL, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol Cell Proteomics Internet. 2005 cited 2022 Mar 16;4:1240–1250. Available from: https://pubmed.ncbi.nlm.nih.gov/15951569/.

5 Dengjel J, Akimov V, Olsen J V., et al. Quantitative proteomic assessment of very early cellular signaling events. Nat Biotechnol Internet. 2007 cited 2022 Jan 21;25:566–568. Available from: https://pubmed.ncbi.nlm.nih.gov/17450129/.

6 Jones RB, Gordus A, Krall JA, et al. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nat 2005 4397073 Internet. 2005 cited 2022 Feb 4;439:168–174. Available from: https://www.nature.com/articles/nature04177.

7 Tong J, Taylor P, Moran MF. Proteomic Analysis of the Epidermal Growth Factor Receptor (EGFR) Interactome and Post-translational Modifications Associated with Receptor Endocytosis in Response to EGF and Stress. Mol Cell Proteomics. 2014;13:1644–1658.

8 Rotin D, Margolis B, Mohammadi M, et al. SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: identification of Tyr992 as the high-affinity binding site for SH2 domains of phospholipase C gamma. EMBO J Internet. 1992 cited 2022 Jan 17;11:559. Available from: /pmc/articles/PMC556487/?report=abstract.

9 Emlet DR, Moscatello DK, Ludlow LB, et al. Subsets of epidermal growth factor receptors during activation and endocytosis. J Biol Chem Internet. 1997 cited 2022 Jan 17;272:4079–4086. Available from: https://pubmed.ncbi.nlm.nih.gov/9020117/.

10 Levkowitz G, Waterman H, Ettenberg SA, et al. Ubiquitin Ligase Activity and Tyrosine Phosphorylation Underlie Suppression of Growth Factor Signaling by c-Cbl/Sli-1. Mol Cell. 1999;4:1029–1040.

11 Yamauchi T, Ueki K, Tobe K, et al. Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. Nature Internet. 1997 cited 2022 Jan 17;390:91–96. Available from: https://pubmed.ncbi.nlm.nih.gov/9363897/.

12 Okutani T, Okabayashi Y, Kido Y, et al. Grb2/Ash binds directly to tyrosines 1068 and 1086 and indirectly to tyrosine 1148 of activated human epidermal growth factor receptors in intact cells. J Biol Chem. 1994;269:31310–31314.

13 Okabayashi Y, Kido Y, Okutani T, et al. Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of Shc in intact cells. J Biol Chem. 1994;269:18674–18678.

14 Sato KI, Nagao T, Iwasaki T, et al. Src-dependent phosphorylation of the EGF receptor Tyr-845 mediates Stat-p21waf1 pathway in A431 cells. Genes Cells Internet. 2003 cited 2022 Jan 17;8:995–1003. Available from: https://pubmed.ncbi.nlm.nih.gov/14750954/.

15 Tanos B, Pendergast AM. Abl tyrosine kinase regulates endocytosis of the epidermal growth factor receptor. J Biol Chem Internet. 2006 cited 2022 Feb 4;281:32714–32723. Available from: https://pubmed.ncbi.nlm.nih.gov/16943190/.

16 Östman A, Hellberg C, Böhmer FD. Protein-tyrosine phosphatases and cancer. Nat Rev Cancer Internet. 2006 cited 2022 Mar 2;6:307–320. Available from: https://pubmed.ncbi.nlm.nih.gov/16557282/.

17 Vouri M, Croucher DR, Kennedy SP, et al. Axl-EGFR receptor tyrosine kinase hetero-interaction provides EGFR with access to pro-invasive signalling in cancer cells. Oncogenesis Internet. 2016 cited 2022 Mar 2;5. Available from: https://pubmed.ncbi.nlm.nih.gov/27775700/.

18 Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med Internet. 2004 cited 2022 Mar 2;350:2129–2139. Available from: https://pubmed.ncbi.nlm.nih.gov/15118073/.

19 Gill K, Macdonald-Obermann JL, Pike LJ. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. J Biol Chem Internet. 2017 cited 2021 Jul 22;292:20744. Available from: /pmc/articles/PMC5733609/.

20 Kovacs E, Das R, Wang Q, et al. Analysis of the Role of the C-Terminal Tail in the Regulation of the Epidermal Growth Factor Receptor. Mol Cell Biol Internet. 2015 cited 2022 Jan 17;35:3083. Available from: /pmc/articles/PMC4525312/.

21 Pruss RM, Herschman HR. Variants of 3T3 cells lacking mitogenic response to epidermal growth factor. Proc Natl Acad Sci Internet. 1977 cited 2022 Feb 4;74:3918–3921. Available from: https://www.pnas.org/content/74/9/3918.

22 Jamison J, Lauffenburger D, Wang JCH, et al. PKCδ localization at the membrane increases matrix traction force dependent on PLCγ1/EGFR signaling. PLoS One Internet. 2013 cited 2022 Feb 23;8. Available from: https://pubmed.ncbi.nlm.nih.gov/24155954/.

23 Rosenfeld MG, Rosenfeld MG, Gill GN. Enhanced tumorigenesis of NR6 cells which express non-down-regulating epidermal growth factor receptors. Cancer Res Internet. 1991 cited 2022 Feb 23;51:6170–6175. Available from: https://pubmed.ncbi.nlm.nih.gov/1933876/.

24 Glading A, Chang P, Lauffenburger DA, et al. Epidermal growth factor receptor activation of calpain is required for fibroblast motility and occurs via an ERK/MAP kinase signaling pathway. J Biol Chem Internet. 2000 cited 2022 Feb 23;275:2390–2398. Available from: https://pubmed.ncbi.nlm.nih.gov/10644690/.

25 Schneider CA, Lim RW, Terwilliger E, et al. Epidermal growth factor-nonresponsive 3T3 variants do not contain epidermal growth factor receptor-related antigens or mRNA. Proc Natl Acad Sci Internet. 1986 cited 2022 Feb 4;83:333–336. Available from: https://www.pnas.org/content/83/2/333.

26 Curran TG, Zhang Y, Ma DJ, et al. MARQUIS: A multiplex method for absolute quantification of peptides and posttranslational modifications. Nat Commun Internet. 2015 cited 2021 Jun 6;6:1–11. Available from: https://www.nature.com/articles/ncomms6924.

27 Lemmon MA, Schlessinger J, Ferguson KM. The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases. Cold Spring Harb Perspect Biol Internet. 2014 cited 2022 Jan 14;6. Available from: /pmc/articles/PMC3970421/.

28 Thirukkumaran OM, Kluba M, Hofkens J, et al. Autophosphorylation of EGFR at Y954 Facilitated Homodimerization and Enhanced Downstream Signals. Biophys J Internet. 2020 cited 2022 Jan 14;119:2127–2137. Available from: https://pubmed.ncbi.nlm.nih.gov/33096081/.

29 Jiang X, Huang F, Marusyk A, et al. Grb2 Regulates Internalization of EGF Receptors through Clathrin-coated Pits. Mol Biol Cell Internet. 2003 cited 2022 Feb 7;14:858. Available from: /pmc/articles/PMC151565/.

30 Li GW, Xie XS. Central dogma at the single-molecule level in living cells. Nature Internet. 2011 cited 2022 Feb 6;475:308. Available from: /pmc/articles/PMC3600414/.

31 Wolf-Yadlin A, Kumar N, Zhang Y, et al. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Mol Syst Biol Internet. 2006 cited 2022 Feb 23;2. Available from: https://pubmed.ncbi.nlm.nih.gov/17016520/.

32 Eswarakumar VP, Lax I, Schlessinger J. Cellular signaling by fibroblast growth factor receptors. Cytokine Growth Factor Rev Internet. 2005 cited 2022 Feb 23;16:139–149. Available from: https://pubmed.ncbi.nlm.nih.gov/15863030/.

33 Wei Z, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res Internet. 2002 cited 2022 Feb 23;12:9–18. Available from: https://pubmed.ncbi.nlm.nih.gov/11942415/.

34 De Donatis A, Comito G, Buricchi F, et al. Proliferation versus migration in platelet-derived growth factor signaling: the key role of endocytosis. J Biol Chem Internet. 2008 cited 2022 Feb 23;283:19948–19956. Available from: https://pubmed.ncbi.nlm.nih.gov/18499659/.

35 Stallaert W, Brüggemann Y, Sabet O, et al. Contact inhibitory Eph signaling suppresses EGF-promoted cell migration by decoupling EGFR activity from vesicular recycling. Sci Signal Internet. 2018 cited 2022 Feb 23;11. Available from: https://www.science.org/doi/abs/10.1126/scisignal.aat0114.

36 Kostenko O, Tsacoumangos A, Crooks D, et al. Gab1 signaling is regulated by EGF receptor sorting in early endosomes. Oncogene Internet. 2006 cited 2022 Feb 23;25:6604–6617. Available from: https://pubmed.ncbi.nlm.nih.gov/16715136/.

37 Li M, Yang J, Zhang L, et al. A low-molecular-weight compound exerts anticancer activity against breast and lung cancers by disrupting EGFR/Eps8 complex formation. J Exp Clin Cancer Res Internet. 2019 cited 2022 Feb 3;38. Available from: https://pubmed.ncbi.nlm.nih.gov/31118055/.

38 Fazioli F, Minichiello L, Matoska V, et al. Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals. EMBO J Internet. 1993 cited 2022 Feb 7;12:3799. Available from: /pmc/articles/PMC413663/?report=abstract.

39 Thiel KW, Carpenter G. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. Proc Natl Acad Sci U S A Internet. 2007 cited 2022 Feb 4;104:19238–19243. Available from: https://pubmed.ncbi.nlm.nih.gov/18042729/.

40 Cunningham DL, Creese AJ, Auciello G, et al. Novel Binding Partners and Differentially Regulated Phosphorylation Sites Clarify Eps8 as a Multi-Functional Adaptor. PLoS One Internet. 2013 cited 2022 Jan 14;8. Available from: /pmc/articles/PMC3634024/.

41 Nozaki K, Kagamu H, Shoji S, et al. DDX3X Induces Primary EGFR-TKI Resistance Based on Intratumor Heterogeneity in Lung Cancer Cells Harboring EGFR-Activating Mutations. PLoS One Internet. 2014 cited 2022 Feb 23;9:111019. Available from: /pmc/articles/PMC4208809/.

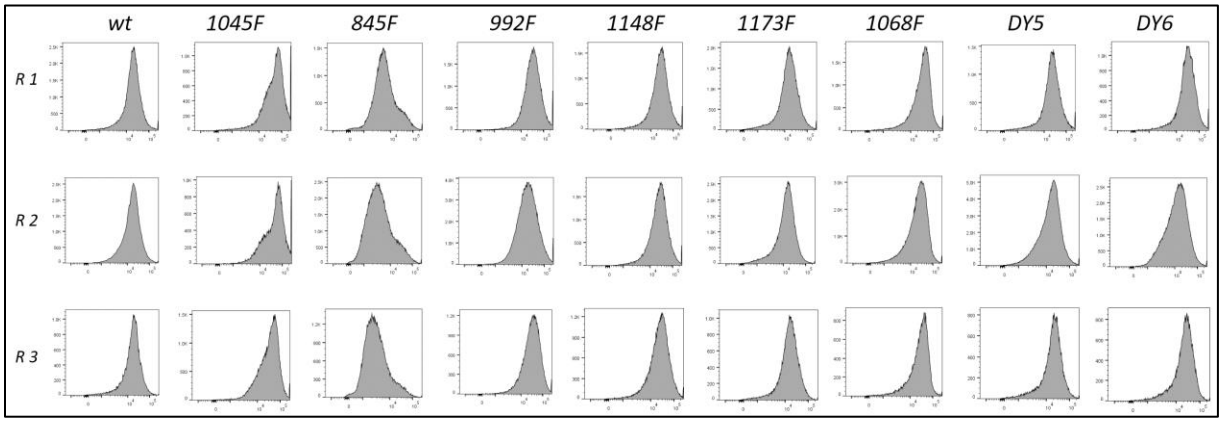
42 Castagnino P, Biesova Z, Wong WT, et al. Direct binding of eps8 to the juxtamembrane domain of EGFR is phosphotyrosine- and SH2-independent. Oncogene. 1995;10:723–729.

43 Getz M, Swanson L, Sahoo D, et al. A predictive computational model reveals that GIV/girdin serves as a tunable valve for EGFR-stimulated cyclic AMP signals. Mol Biol Cell Internet. 2019 cited 2022 Mar 2;30:1621–1633. Available from: https://pubmed.ncbi.nlm.nih.gov/31017840/.

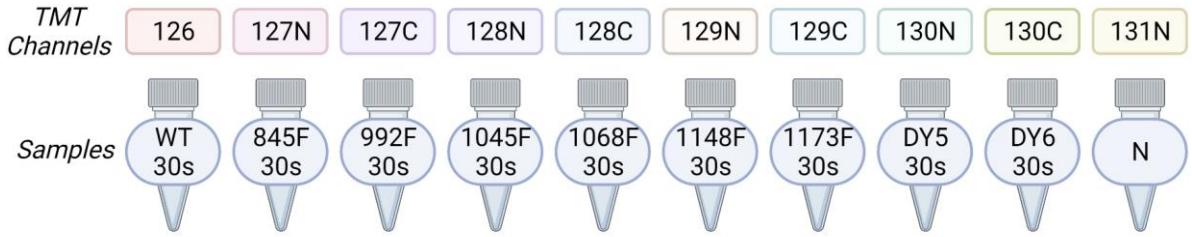
44 Lin C, Ear J, Midde K, et al. Structural basis for activation of trimeric Gi proteins by multiple growth factor receptors via GIV/Girdin. Mol Biol Cell Internet. 2014 cited 2022 Mar 2;25:3654–3671. Available from: https://pubmed.ncbi.nlm.nih.gov/25187647/.

45 Bhandari D, Lopez-Sanchez I, To A, et al. Cyclin-dependent kinase 5 activates guanine nucleotide exchange factor GIV/Girdin to orchestrate migration-proliferation dichotomy. Proc Natl Acad Sci U S A Internet. 2015 cited 2022 Mar 2;112:E4874–E4883. Available from: https://pubmed.ncbi.nlm.nih.gov/26286990/.

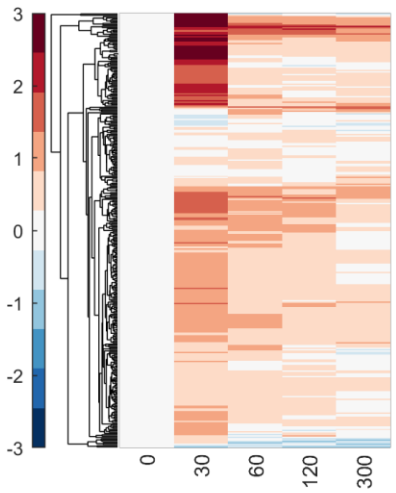
Supplemental Figures



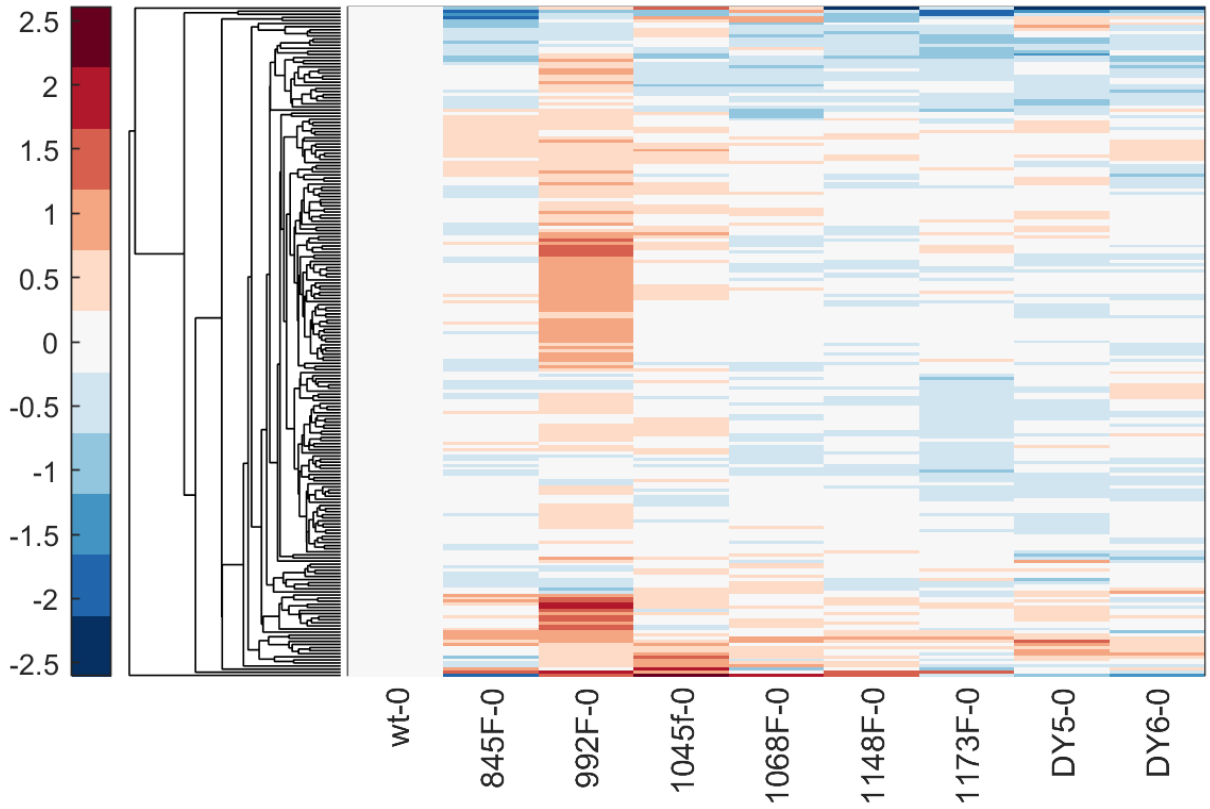
Supplementary figure 1: ***EGFR expression levels as determined by flow cytometry.*** *For each cell line (columns) and replicate (rows), EGFR surface-expression levels were determined using fluorophore-labeled EGF. Protocol as described in methods.*



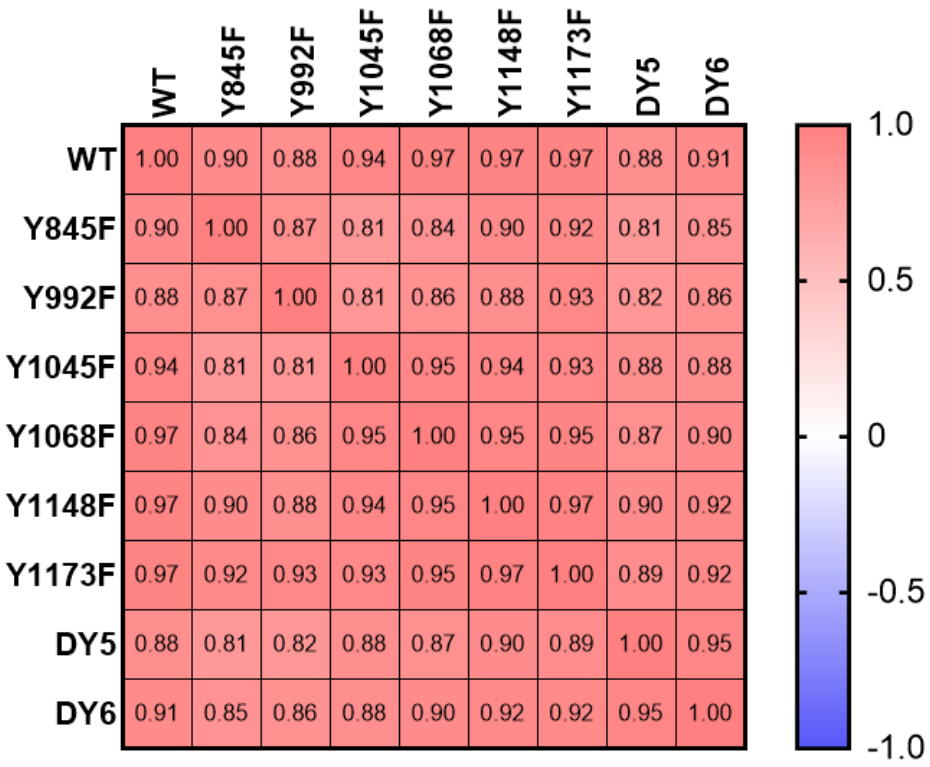
Supplementary figure 2: ***Sample labeling scheme.*** *Samples were multiplexed by timepoint to optimize quantitative comparisons between cell lines. One channel was utilized for a batch sample (WT stimulated with 2min EGF, depicted by N) to allow normalization and comparison between runs and replicates. Labeling scheme was scrambled between replicates to avoid channel bias.*



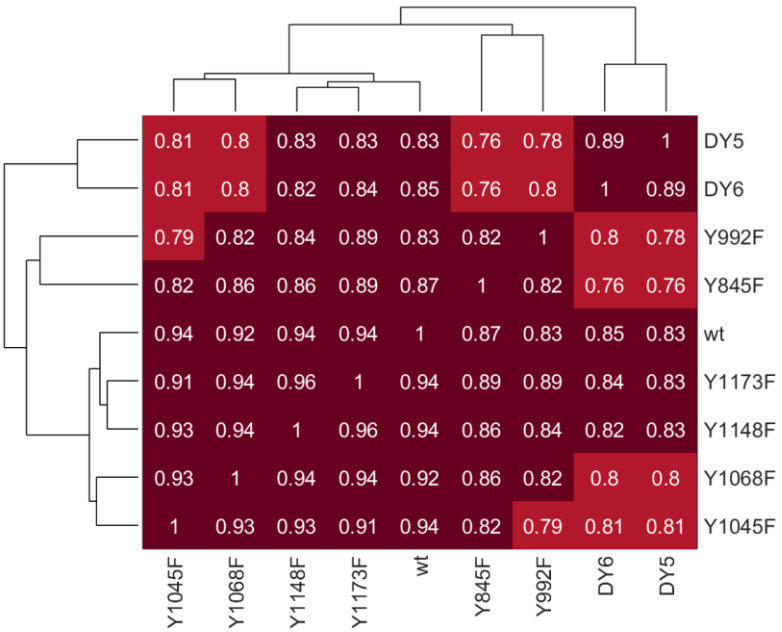
Supplementary figure 3: ***Parental NR6 cells in response to EGF.*** *Un-transfected parental NR6 cells were treated with 2nM EGF to confirm lack of EGFR phosphorylation and downstream response. Columns indicate timepoint in seconds. Data presented are log2 fold change compared to unstimulated condition.*



Supplementary figure 4: ***Basal signaling compared to wt.*** *Data presented are mean log2 fold change in unstimulated cells with respect to wildtype cells, n=3.*



Supplementary figure 5: ***Correlation of cell lines based average log2 fold change of all overlapping peptides, demonstrating high degree of similarity between cell lines.*** *Data presented are calculated Pearson r correlation coefficients.*



Supplementary figure 6: ***Correlation of EGFR peptide data shows high similarity between isoforms.*** *Data presented are calculated Pearson r correlation coefficients.*

	Proliferation		Migration		Endocytosis	
Timepoint	**R2**	**Q2**	**R2**	**Q2**	**R2**	**Q2**
30/30s	0.9427	0.7481	0.9741	0.8204	0.9769	0.8615
60/1m	0.899	0.835	0.9607	0.813	0.978	0.8415
120/2m	0.9443	0.8306	0.986	0.8239	0.9871	0.842
300/5m	0.9606	0.7947	0.9899	0.8607	0.9761	0.7896
Combined	0.9432	0.8027	0.981	0.8395	0.9897	0.8757
AUC	0.9828	0.8558	0.9882	0.736	0.9861	0.8159

Supplementary figure 7: ***PLSR prediction values.*** *Partial least squares regression using LOOCV was used to build predictive model connecting EGFR signaling and cellular behavior. Table indicates goodness of fit (R2) and prediction accuracy (Q2) for each model (3 phenotypes, separate or combined timepoints or Area Under the Curve (AUC)).*

CHAPTER 5

Using Green Fluorescent Protein in Signaling Studies

Jacqueline S. Gerritsen

Forest M. White

Manuscript in Preparation

Abstract

Biological research has benefited tremendously from the use of protein labels for the detection and quantification of their protein of interest (POI). Green Fluorescent Protein (GFP) in particular has been widely used due to its inherent luminescent properties. However, the applicability of these labels has been scrutinized when applied in assays with highly dynamic measurements, such as phosphorylation. The size of a label like GFP could potentially disrupt substrate interactions or hinder proper protein folding and trafficking. In this work, we evaluated the EGFR signaling network in NR6 cells expressing human wild type or mutant isoforms. One set of cell lines additionally contained a GFP at the C-terminal end, facilitating a variety of measurements but potentially interfering with normal EGFR function. The resulting signaling events as determined by LC/MS-MS analysis and phenotypic measurements (i.e. migration and proliferation) were compared. The results indicate that similar general patterns can be deduced from both systems, although slight differences in signaling between GFP and non-GFP systems are present. These differences are mostly reflected by the abundance of signal in response to ligand stimulation, which is generally higher in the GFP-lacking cells. Therefore, we propose that large labeling proteins like GFP can be used for hypothesis generation but that care should be taken when prioritizing quantification of these measurements, in particular when conducting signaling network studies focused on highly dynamic post-translational modifications such as tyrosine phosphorylation.

Introduction

Green Fluorescent protein (GFP) has long been a powerful tool for investigating a wide variety of biological events in live cells, including protein tracking, structure labeling and analyzing protein activity.[1] The major advantage of using GFPs is their ability to exhibit intrinsic fluorescence, as for most other fluorescent proteins, a cofactor is required. [2] This allows for monitoring of the dynamics of GFP-tagged proteins in living cells, addressing both the spatial and temporal aspects, and avoids potential artefacts due to cell fixation and permeabilization. However, as GFP is a large 27 kDa protein, investigators need to be to be wary of deleterious effects of such a large protein tag on the function and localization of the tagged protein. Especially when monitoring highly dynamic processes such as protein phosphorylation, concerns surrounding the effects on protein folding, internalization and steric hindrance exist. Here, we compared EGF-induced phosphoproteomic and phenotypic data from NR6 cells expressing human EGFR isoforms with or without a GFP-tag. In this model, the GFP tag allows for easy monitoring of EGFR expression levels across cell lines and replicates. Our data shows that although general trends are comparable across models, GFP seems to decrease the dynamic range of change observed in signaling as well as phenotypic data between mutant isoforms. Although GFP has clear advantages in biological research, caution must be exerted when interpreting this data.

Results

Development of EGFR-expression system

To evaluate and compare signaling effects in the EGFR network response, mouse fibroblasts (NR6) lacking ErbB expression were retrovirally transfected to stably express human wild type EGFR (wtEGFR) or mutant EGFR isoforms in which one or more tyrosines in the C-terminal tail were converted to a non-phosphorylatable phenylalanine residue (e.g., Y-to-F mutation) to evaluate loss of function. These lines were generated two-fold, one set additionally containing a GFP tag at the C-terminal end of the EGFR sequence (wild type, Y845F, Y992F, Y1045F, Y1068F, Y1148F and Y1173F, with or without GFP) (Figure 1).

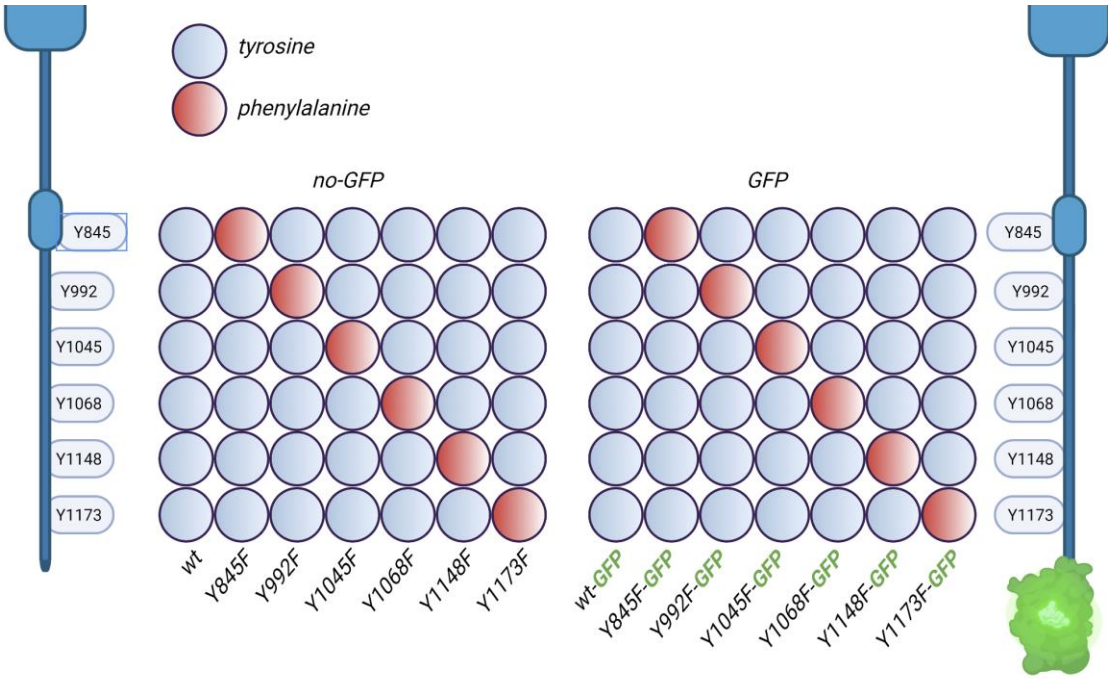


Figure 15: A panel of 14 cell lines were generated, 7 of which contained a GFP tag at the C-terminal end. The 7 isoforms included wild type and 6 single Y-to-F mutants. Red and Blue dots indicate presence of phenylalanine or tyrosine at residue location, respectively.

Following successful transfection, expression levels were determined for all cell liens and replicates by flow cytometry (Figure 2). Expression of each mutant isoform led to relatively slight changes in surface3 level receptor expression for most of the non-GFP cell lines, with Y845F and Y1045F being the major outliers in terms of surface expression. In the GFP-lines, EGFR expression was less variable, with the exception of the Y1173F mutant (Figure 2).

Phosphorylation of the Y1045 site has been associated with binding of the Cbl E3 ubiquitin ligase and subsequent ubiquitination of EGFR. Blocking phosphorylation at this site through Y-to-F mutation should therefore result in altered receptor trafficking, including increased recycling and decreased degradation, providing a potential explanation for the significantly greater cell surface expression relative to wtEGFR. These data indicate that GFP is functionally affecting surface level expression of the protein, potentially by slowing down protein trafficking.

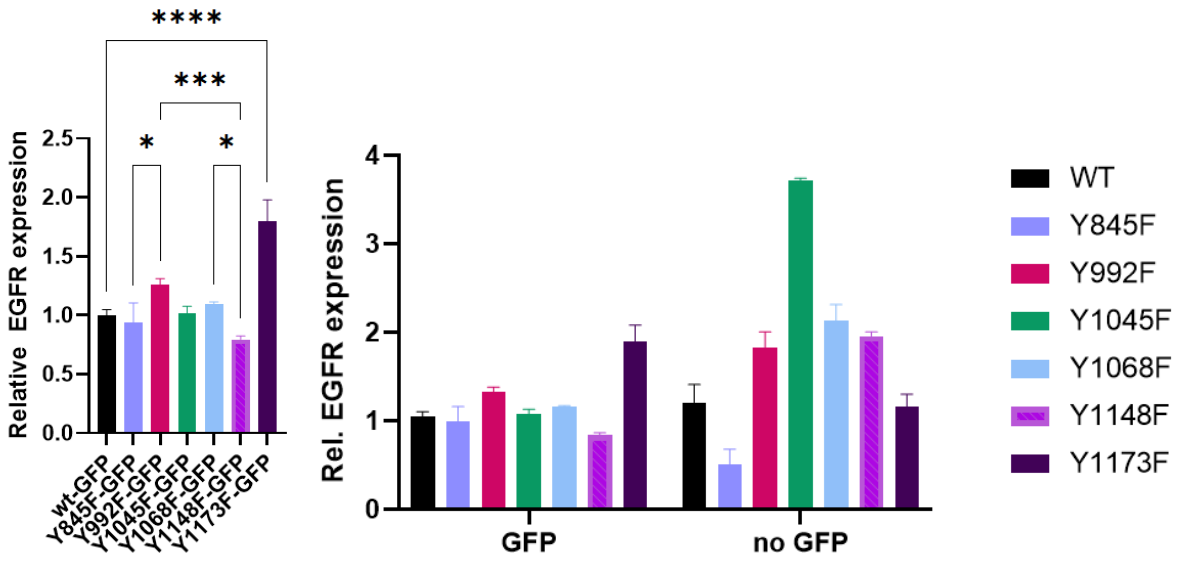


*Figure 16:EGFR surface expression levels for GFP-lines (left) and all lines (right). *P≤0.05 (one-way ANOVA), n=3.*

Comparing these cell lines to cell lines with known copy numbers, reveals that copy number is comparable to A549 levels [3] (Figure 3).

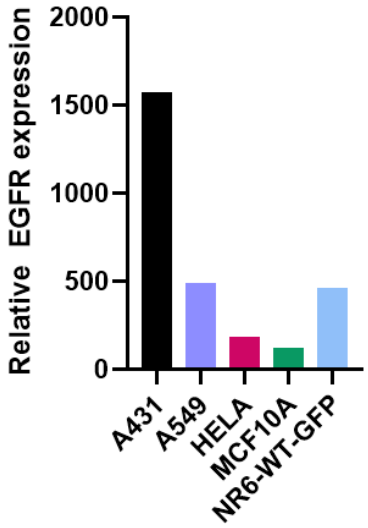


Figure 17: EGFR expression levels compared. Comparison of signal between NR6-EGFR expressing cell lines and other cell lines with known copy numbers of EGFR, showing comparable expression levels with A549, a high EGFR-expressing cell line.

After confirming that receptor surface expression was roughly similar across most of the mutant isoform expressing cell lines, to assess the effect of each mutation on the cellular signaling network, cells were stimulated with 2nM EGF for 0 seconds, 30 seconds, 1 minute, 2 minutes or 5 minutes. At the appropriate time following stimulation, cells were snap-frozen in liquid nitrogen, lysed in cold 8M Urea to preserve physiological signaling, and proteolytically digested to peptides. To accommodate the large number of samples (5 time points per cell line, 9 total cell lines, and 3 biological replicates per condition), peptides from each cell lysate were labeled with isobaric tandem mass tags and analyzed as 11-plex experiments with pooled normalization controls. Labeled samples were subjected to 2-step phosphotyrosine enrichment and subsequent liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis (Figure

4). These analyses yielded quantitative data for interrogate and compare pTyr temporal dynamics between GFP and non-GFP lines.

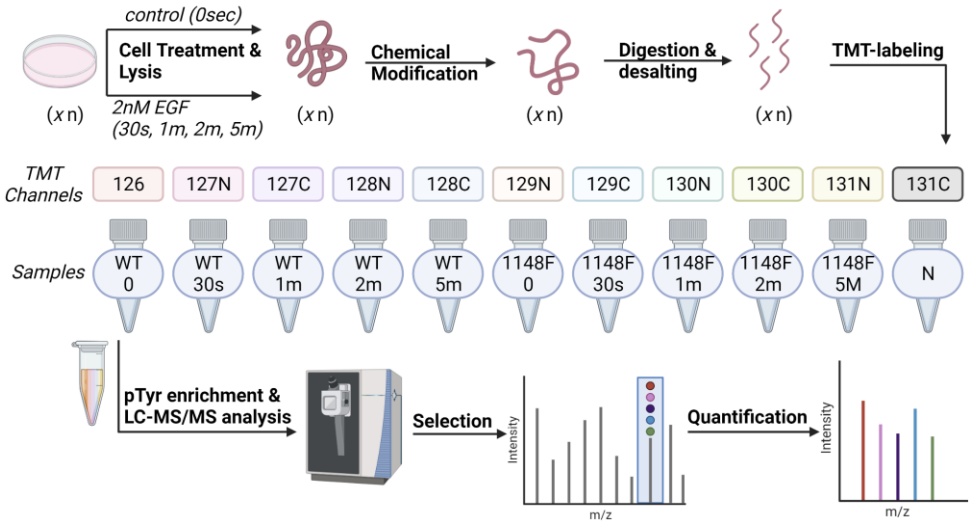


Figure 18: Workflow schematic of NR6 retroviral transfection, cell stimulation and sample processing, followed by TMT labeling and phosphotyrosine enrichments steps for LC-MS/MS analysis.

NR6 cells expressing EGFR-GFP and EGFR lacking GFP exhibit similar signaling network response to EGF stimulation

We interrogated this large data set by initially comparing EGFR network dynamics in response to EGF between wtEGFR-GFP and wtEGFR. To evaluate the EGF response, data was normalized to each cellline's corresponding unstimulated condition to correct for any differences in basal signaling. Overall, both lines respond to EGF with a strong increase in phosphorylation of the network(Figure 5). This response includes many of the canonical nodes of the EGFR network, including PLCy-1, SHIP2, GAB1, and ERK1/2.

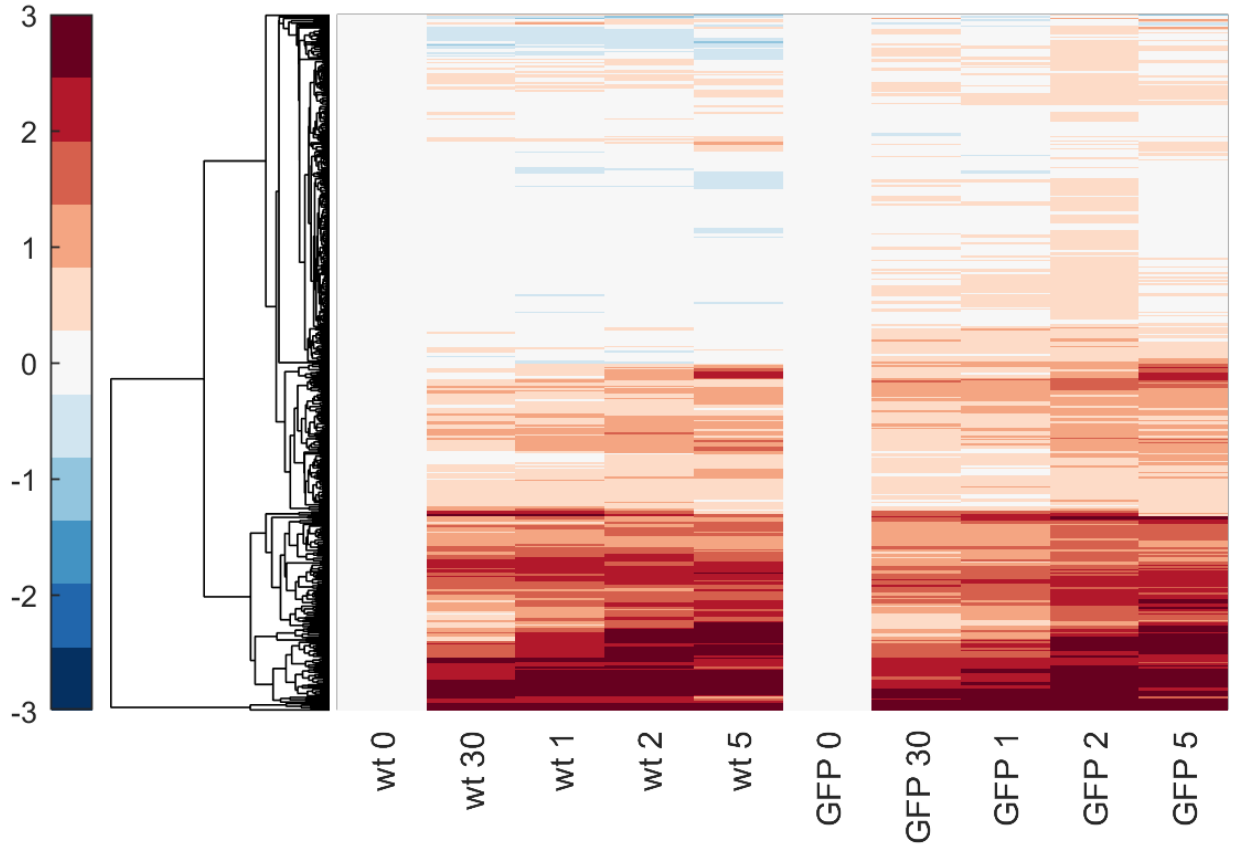


Figure 19: Comparison wtEGFR and wtEGFR-GFP dynamic network response to EGF. Data presented are log2fold change values over unstimulated(basal) condition.

Mutant EGFR isoforms exhibit stronger signaling network in absence of GFP

Interestingly, the response to EGF in GFP-containing mutant isoform lines was less affected as a result of the mutations compared to the non-GFP cell lines. Figure 6 demonstrates a lack of response in GFP-lines, whereas non-GFP lines display a variety of responses ranging from downregulation to upregulation of the network in response to EGF.

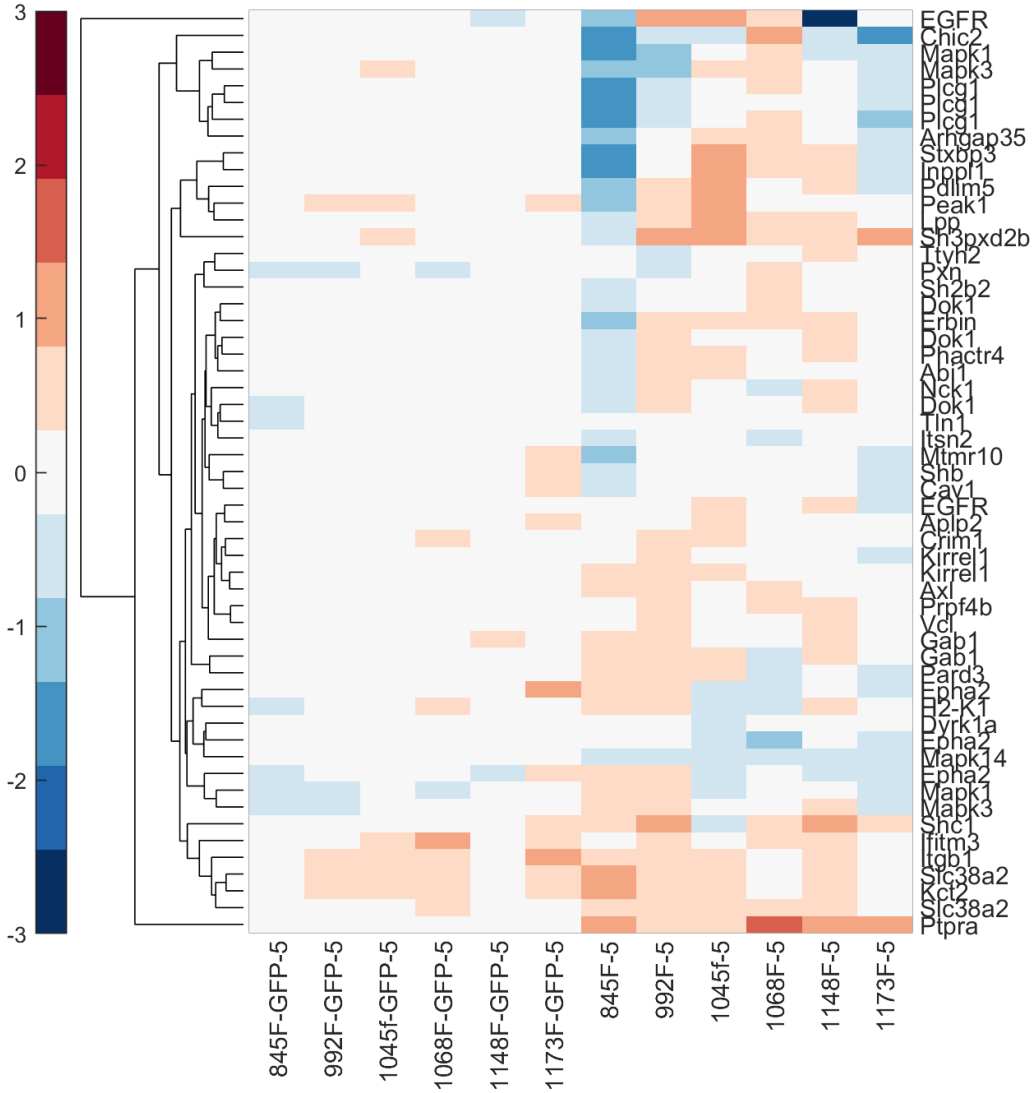


Figure 20: Hierarchical clustering of tyrosine phosphorylation data of all EGFR isoforms at t=5m. Log2 fold change over basal levels, data are presented as mean, n=3.

Phenotypes are differentially altered between GFP/non-GFP lines.

As the phosphoproteomic data reflect overall signaling network similarity as well as some mutant-specific differences in non-GFP lines, and little differences between mutant isoforms in GFP-lines, our findings raised the question as to whether and how this variation translates to

cellular response. To evaluate the phenotypic consequences of tyrosine loss-of-function, migration and proliferation rates were measured (Fig 7).

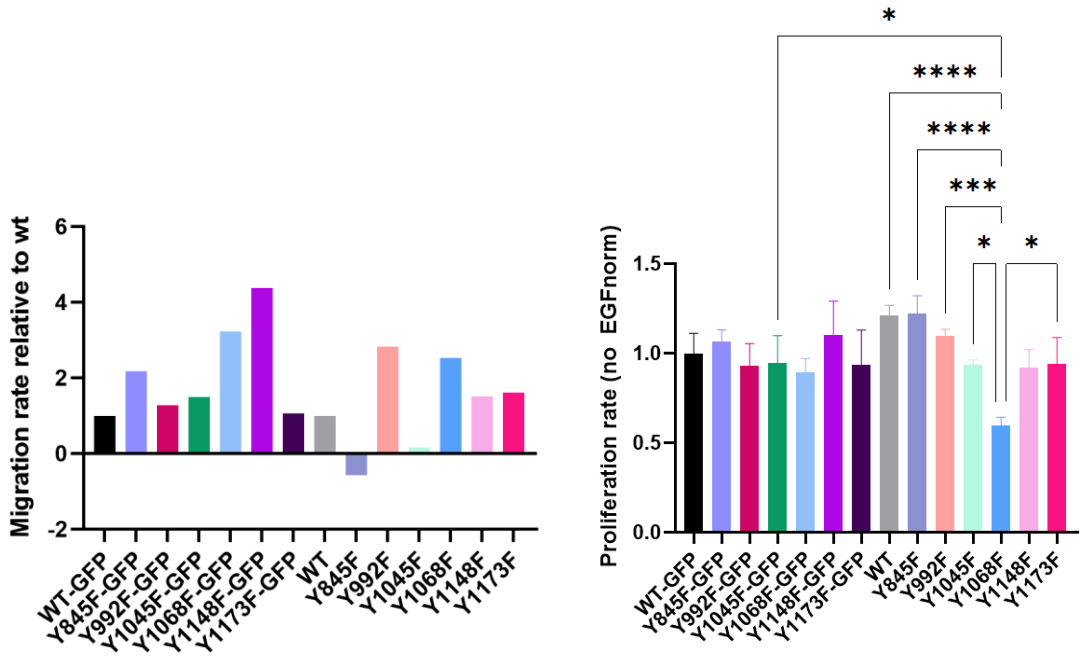


*Figure 21: Migration and Proliferation data for NR6 EGFR expressing cell lines. Left: Proliferation rates measured by confluency % over 24 hrs. Right: Migration rates determined by measuring scratch wound assay closing rate over 24 hrs. Data are presented as mean +- SD. *P≤0.05 (One-way ANOVA), n=3. * indicates significant difference.*

To evaluate EGF-stimulated cell migration rates, temporal dynamics of wound healing were quantified in confluent plates for each cell line, in the presence or absence of EGF. EGF stimulated cell migration at a similar rate for most of the mutant isoform expressing lines as compared to wtEGFR, however Y992F cells were significantly more responsive to EGF-stimulated migration as compared to most other lines. These results were not recapitulated in the GFP lines, where an increase in migration was observed for the Y1068F an Y1148F line, although they did not have a strong signaling response to EGF.

Confluency measurements taken over time for each line in the absence or presence of EGF were compared to quantify an EGF-response proliferation rate. Where migration rates were

clearly different between GFP-isoforms, proliferation rates were not significantly different between wtEGFR-GFP and mEGFR-GFP lines. In the non-GFP lines, significant differences between cell lines can be observed, most notably for the Y1068F line, nominating this site as a potential driver of proliferation.

To validate fair comparison between GFP and non-GFP lines, a direct comparison phenotypic experiment of wtEGFR and wtEGFR-GFP cell lines demonstrated similar trends upon EGF treatment (Fig 8). No significant differences were observed between non-stimulated conditions or between stimulated conditions. However, slightly greater EGF-induced migration and proliferation was observed in the non-GFP lines.

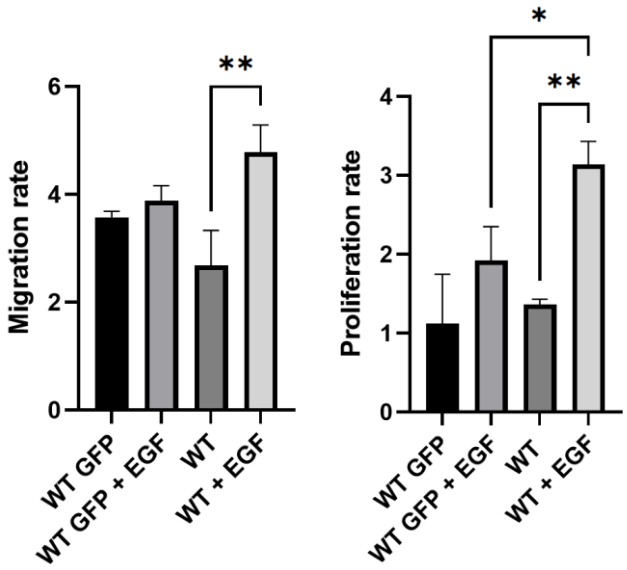


Figure 8: Migration and Proliferation rates for WT EGFR and WT EGFR-GFP expressing cell lines in absence or presence of EGF, directly compared. Left: Proliferation rates measured by confluency % over 24 hrs. Right: Migration rates determined by measuring scratch wound assay closing rate over 24 hrs. Data are presented as mean +- SD. *P≤0.05 (One-way ANOVA), n=3. * indicates significant difference

Discussion

In this study, we evaluated the effect of GFP on signaling dynamics of the EGFR network. We utilized NR6 cells tranfected with either wtEGFR or single mutant EGFR, in the presence or absence of a GFP for detection purposes. Although our wildtype lines both elicited a strong and similar response to EGF, we found that the size of the fold change varied. Furthermore, we found that our mutations caused greater change in phosphorylation profile in the non-GFP lines, potentially due to stabilizing effects of GFP. Furthermore, phenotypic data showed that although wildtype lines were comparable in terms of response direction, the size of response was significantly different. Furthermore, we found that phenotypic results were not easily comparable between GFP and non-GFP lines for the mutant isoforms. These experimental comparisons are important as they highlight the need to be cautious when using fluorophores and other labels in highly dynamic experimental setups.

Although EGFR in our model system remained responsive to EGF stimulation in wild type cells, it is clear that GFP has affected EGFR function, particularly at the phenotypic level. Changes in receptor surface level expression could very well be the result of decreased internalization or recycling of the receptor due to the addition of this 27 kDa protein. Furthermore, GFP may, to an extent, have sterically hindered dimerization and/or phosphorylation of the receptor as well as blocked substrate accessibility vital for signal transduction.

For future signaling studies, we propose limiting the use of these fluorophores in signaling studies to hypothesis generation purposes, and evaluate surface level expression using fluorophore-labeled ligands when available. Alternatively, fluorophores could be positioned

further away from interaction sites, such as the C-terminal tail, although caution must be exerted to prevent interference with the ligand binding domain.[4]

Materials and Methods

Plasmid preparation and extraction

Seven out of nine plasmid constructs were kindly donated by F. Furnari Laboratory (UCSD, CA). Wild-type and 1045F mutant plasmids were generated from 845F plasmid using Genscript site-directed mutagenesis service and sequences verified with Sanger sequencing (EtonBio). Plasmids were pLNCX2 retroviral vectors with a neomycin and Ampicillin resistance gene, as well as the EGFR coding sequence containing a variety of Y-to-F mutations (Y845F, Y992F, Y1045F, Y1068F, Y1148F, Y1173F, and wild-type.

Plasmid DNA was transformed into DH5α E. coli (NEB 5-alpha competent E. Coli (High Efficiency)) according to NEB protocol. Cells from -80^{o}C freezer were thawed on ice for 5 minutes, and 5µl of plasmid DNA was added to the cell mixture. The mixture was incubated on ice for 30 minutes and then heat shocked in a water bath at 42^{o}C for 45 seconds. The mixture was placed on ice for 5 minutes. 950 µL of Super Optimal broth with Catabolite expression (SOC) medium was added to the mixture, which was then shaken at 37^{o}C for 60 minutes. Subsequently, 100 µL was spread onto a Luria Broth (LB) agar with ampicillin plate and incubated overnight at 37^{o}C. Individual colonies were picked and grown for ~12 hours at 37^{o}C in 10mL of LB media with ampicillin (100mg/mL, 1000x). Culture was spun down at 4^{o}C and 10,000 rpm for 10 minutes. Plasmid DNA was isolated using the Qiagen Spin Miniprep Kit. Nanodrop (Thermo) was used to determine DNA concentration.

Retroviral transfection

On day 0, 2e5 293T cells/well were plated in 6 well-plates. The next day, 4µl fuGENE HD Transfection Reagent (Promega, E2311) was added to 100µl of Opti-MEM (Gibco) and vortexed to mix and incubate for 5 minutes at room temperature (RT). Envelope protein VSV-G and packaging genes Gag/Pol were combined with plasmid DNA; 333ng, 666ng and 500ng, respectively. The DNA mix was added dropwise to the solution containing fuGENE and HD transfection reagent, mixed, and left to incubate at RT for 30 minutes. The resulting mixture was then added dropwise to the 293T cells. On day 2, the media was replaced with regular DMEM containing 1% Pen/Strep and 10% FBS using a P1000 to minimally disturb the cells. NR6 cells were plated in 6 well plates at 5e4 cells/well. The following day, virus was harvested from the 293T cells by collecting media and passing it through a .45µm filter. Polybrene (Sigma) was added to the solution at 6µg/mL, and the media for NR6 cells was replaced with the filtered DMEM containing the virus/polybrene mixture. 8 hours later, the media was removed and DMEM supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptavidin (P/S) added onto the cells. On day 6, Geneticin (G418, Gibco) was added at 45µL/well. One day later, the cells were split and kept on G418 selection pressure.

Flow cytometry for EGFR expression and endocytosis

To select successfully transfected cells as well as to ensure comparable expression levels between cell lines and replicates, EGFR expression was determined by flow cytometry. Biotinylated EGF, complexed to Alexa Fluor 488 (Invitrogen, E13345) was added to each NR6 cell line at a concentration of 0.5 µg/mL and incubated at 37^{o}C for 5 minutes. Cells were put on

ice, spun down at 4°C at 1,200 *g* for 5 min and washed twice with cold PBS before analysis on the BD FACScanto Clinical Flow Cytometry System. When selecting successfully transfected cells, the BD FACSAria Sorter was used and gating for EGFR-expressing cells was determined with respect to a negative control.

To evaluate EGFR internalization and trafficking, pHrodo EGF 488 conjugate (Thermo P35375) was added to each NR6 cell line at a concentration of 0.5 µg/mL and incubated at 37°C for 5 minutes. Cells were put on ice, collected from plate using Accutase, spun down at 4°C at 1,200 *g* for 5 min and washed twice with cold PBS before analysis. Data was acquired using the BD FACSDIVA Software and further analyzed using FlowJo.

Cell culture and EGF stimulation

NR6 mouse fibroblasts (gifted by F. Furnari Laboratory, UCSD, CA) were maintained in DMEM (Corning) medium supplemented with 1% Penicillin-Streptomycin (5,000 U/mL, Gibco), 2% G418 (Gibco) and 10% FBS (Gibco). Approximately 2e5 cells were seeded and incubated for 2 days to reach ~80% confluency in 10cm dishes. For each line and each biological replicate, one plate of cells was used to determine EGFR expression level using Alexa Fluor 488-conjugated EGF and flow cytometry, as described above. Other plates of cells (one 10 cm plate/time point) were stimulated with 2 nM EGF (Peprotech) in serum-free media for 30 seconds, 1 minute, 2 minutes or 5 minutes or left untreated as a control (0minute timepoint).

Sample preparation for MS analysis

After EGF stimulation, media was promptly discarded and cells were snap frozen on liquid N2 for instant arrest of all signaling events. Cells were lysed on ice using 500µL 8M Urea (Sigma) per 10cm plate. A bicinchoninic acid (BCA) protein concentration assay (Pierce) was performed according to the manufacturer's protocol to estimate the protein concentration in each lysate. Cell lysates were reduced with 10mM DTT for 1hr at 56^{o}C, alkylated with 55 mM iodoacetamide for 1hr at RT shielded from light, and diluted 5-fold with 100 mM ammonium acetate, pH 8.9, before trypsin (Promega) was added (20:1 protein:enzyme ratio) for overnight digestion at RT. The resulting solutions were acidified with 1 mL of acetic acid (HOAc) and loaded onto C18 Sep-Pak Plus Cartridges (Waters), rinsed with 10mL of 0.1% HOAc, and eluted with 10mL of 40% Acetonitrile (MeCN)/ 0.1% HOAc. Peptides were divided into XYZ microgram aliquots, and sample volume was reduced using a vacuum centrifuge (Thermo) and then lyophilized to dryness for storage at -80^{o}C.

TMT labeling for multiplexed analysis was performed according to manufacturer's protocol. Samples, each containing ~200ug peptides, were resuspended in 35µL HEPES (pH 8.5), vortexed, and spun down at 13,400rpm for 1 minute. 400µg of a given channel of TMT10plex (Thermo) in anhydrous MeCN, was added per sample. Samples were shaken at 400rpm for 1hr, after which the labeling reaction was quenched using 5% Hydroxylamine (50%, Thermo). After another 15 minutes on the shaker, all samples were combined using the same pipette tip to reduce sample loss, and sample aliquots were washed twice with 40 µL 25% MeCN/0.1% HOAc which was added to the collection tube to improve yield. Sample volume was reduced using a vacuum centrifuge and then lyophilized to dryness for storage at -80^{o}C.

Phosphopeptide enrichment

Immunoprecipitation (IP) and IMAC were used sequentially to enrich samples for phosphotyrosine containing peptides. TMT-labeled samples were incubated in IP buffer consisting of 1% Nonidet P-40 with protein G agarose beads conjugated to 24 µg of 4G10 V312 IgG and 6 µg of PT-66 (P3300, Sigma) overnight at 4^{o}C. Peptides were eluted with 25 µl of 0.2% trifluoroacetic acid for 10 minutes at room temperature; this elution was performed twice to improve yield. Eluted peptides were subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)-based Fe-NTA spin column to reduce non-specific, non-phosphorylated peptide background. High-Select Fe-NTA enrichment kit (Pierce) was used according to manufacturer's instructions with following modifications. Eluted peptides from IP were incubated with Fe-NTA beads containing 25µL binding washing buffer for 30 minutes. Peptides were eluted twice with 20mL of elution buffer into a 1.7 mL microcentrifuge tube. Eluates were concentrated in speed-vac until ~1 µL of sample remained, and then resuspended in 10µL of 5% acetonitrile in 0.1% formic acid. Samples were loaded directly onto an in-house constructed fused silica capillary column [50 micron inner diameter (ID) x 10 cm] packed with 5µm C18 beads (YMC gel, ODS-AQ, AQ12S05) and with an integrated electrospray ionization tip (~2 micron tip ID).

LC-MS/MS analysis

LC-MS/MS of pTyr peptides were carried out on an Agilent 1260 LC coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptides were separated using a 140-minute gradient with 70% acetonitrile in 0.2 mol/L acetic acid at flow rate of 0.2 mL/minute with

approximate split flow of 20 nL/minute. The mass spectrometer was operated in data-dependent acquisition with following settings for MS1 scans: m/z range: 350 to 2,000; resolution: 60,000; AGC target: 3 x 10^6; maximum injection time (maxIT): 50 ms. The top 15 abundant ions were isolated and fragmented by higher energy collision dissociation with following settings: resolution: 60,000; AGC target: 1x10^5; maxIT: 350 ms; isolation width: 0.4 m/z, collisional energy (CE): 33%, dynamic exclusion: 20 seconds. Crude peptide analysis was performed on a Q Exactive Plus mass spectrometer to correct for small variation in peptide loadings for each of the TMT channels. Approximately 30 ng of the supernatant from pTyr IP was loaded onto an in-house packed precolumn (100µm ID x 10 cm) packed with 10mm C18 beads (YMC gel, ODS-A, AA12S11) and analyzed with a 70-minute LC gradient. MS1 scans were per-formed at following settings: m/z range: 350 to 2,000; resolution:70,000; AGC target: 3x10^6; maxIT: 50 ms. The top 10 abundant ions were isolated and fragmented with CE of 33% at a resolution of 35,000.

Peptide identification/quantification

Mass spectra were processed with Proteome Discoverer version 2.5 (Thermo Fisher Scientific) and searched against the mouse and human (for EGFR peptides) SwissProt database using Mascot version 2.4 (MatrixScience, RRID:SCR_014322). MS/MS spectra were searched with mass tolerance of 10 ppm for precursor ions and 20 mmu for fragment ions. Cysteine carbamidomethylation, TMT-labeled lysine, and TMT-labeled peptide N-termini were set as fixed modifications. Oxidation of methionine and phosphorylation of serine, threonine and tyrosine were searched as dynamic modifications. TMT reporter quantification was extracted

and isotope corrected in Proteome Discoverer. Peptide spectrum matches (PSM) were filtered according to following parameters: rank=1, mascot ion score>15, isolation interference<40%, average TMT signal>1,000. Peptides with missing values across any channel were filtered out.

Phenotypic measurements

Proliferation assay EGFR mutant and wildtype expressing cells were seeded in 96-well tissue culture plastic plates (VWR) at 10,000 cells/well in culture media and allowed to adhere for 24 h. Media was replaced by media containing 1% FBS and 2nM EGF or 1% FBS and no EGF. Cells were analyzed by a live-cell imaging platform (Incucyte) every 3 hrs for 48 hrs. Proliferation rates were calculated using confluence measurements and correlation values between phenotypic measurements were calculated using delta rates with/without EGF stimulation.

Migration Wound scratch assay EGFR mutant and wildtype expressing cells were seeded at 1e5 cells/well in 96-well tissue culture plastic plates (Essen Bioscience) at 100,000 cells/well in culture media and allowed to adhere for 24 h and reach ~90% confluency. The Wound Maker (Essen Bioscience) was used to create a scratch across the well surface. Wells were washed with PBS to remove floating cells and media was added containing 1% FBS and either 2nM EGF or no EGF. Wound closure was monitored by a live-cell imaging platform (Incucyte with Scratch wound cell migration software module) every 3 hrs for 48 hrs. Migration rates were calculated using wound width and wound confluency measurements.

Data analysis

Data analyses were performed in MATLAB R2020A, Microsoft Excel 2016 and Graphpad Prism 9. TMT reporter ion intensities from PSMs were summed for each unique phosphopeptide. For protein level quantification, TMT reporter intensities were summed for all unique peptides. Peptide or protein quantification were normalized with relative median values obtained from crude lysate analysis to adjust for sample loading in TMT channels. A combination of Student t-test and One-way Anova was used to perform statistical analysis between conditions. Statistical significance was assigned for $p<0.05$. Unsupervised hierarchical clustering was performed on the basis of Pearson correlation distance metric, unless otherwise specified. Protein networks were obtained from STRING (version 11.0) database.

References

1 Chiesa A, Rapizzi E, Tosello V, et al. Recombinant aequorin and green fluorescent protein as valuable tools in the study of cell signalling. Biochem J Internet. 2001 cited 2022 Feb 12;355:1–12. Available from: https://pubmed.ncbi.nlm.nih.gov/11256942/.

2 Tavaré JM, Fletcher LM, Welsh GI. Using green fluorescent protein to study intracellular signalling. J Endocrinol Internet. 2001 cited 2022 Feb 12;170:297–306. Available from: https://pubmed.ncbi.nlm.nih.gov/11479127/.

3 Zhang F, Wang S, Yin L, et al. Quantification of Epidermal Growth Factor Receptor Expression Level and Binding Kinetics on Cell Surfaces by Surface Plasmon Resonance Imaging. Anal Chem Internet. 2015 cited 2022 Feb 15;87:9960. Available from: /pmc/articles/PMC4836855/.

4 Sánchez-Rico C, Voith von Voithenberg L, Warner L, et al. Effects of fluorophore attachment on protein conformation and dynamics studied by spFRET and NMR. Chemistry [Internet]. 2017 [cited 2022 Mar 23];23:14267. Available from: /pmc/articles/PMC5862035/.

CHAPTER 6

Mapping of AXL signaling networks driving resistance to EGFR inhibition in lung cancer

Marc Creixell

Jacqueline S. Gerritsen

Anurima Majumder

Eric B. Haura

Forest M. White

Aaron S. Meyer

Manuscript in Preparation

Abstract

The receptor tyrosine kinase AXL has been a target of interest due to its suggested role in resistance to inhibitors by EGFR in lung cancer. It is proposed that AXL is transactivated by EGFR in a ligand-independent manner, mediating downstream signaling and augmenting EGFR signaling. To improve our understanding of these mechanisms and the AXL signaling network, we have used mass spectrometry-based phosphoproteomics to interrogate signaling networks in the PC9 lung cancer cell line under a variety of conditions. In this study, PC9 cells were treated with erlotinib (EGFR inhibitor) and AF154 (AXL activating antibody in a variety of AXL Y-to-F mutants. Integrating these signaling data with cell phenotype data has helped define the network nodes driving cell proliferation, apoptosis, migration and island effect under these different conditions. Various modeling approaches identified the YAP pathway and several other nodes as potentially critical components in AXL signaling and may provide future targeting strategies to overcome AXL-mediated resistance to EGFR TKIs.

Introduction

Combination therapy holds considerable promise for overcoming intrinsic and acquired resistance to targeted therapies but relies on our ability to precisely identify the best drug combination for tumors. While immense focus exists on using genomic information to direct therapeutic approach, many resistance mechanisms can also arise from entirely tumor-extrinsic factors within the microenvironment. The receptor tyrosine kinase (RTK) AXL is widely implicated in resistance to targeted therapies such as those directed against EGFR. Regulation of AXL by phosphatidylserine (PS), as opposed to mutation, amplification, or autocrine ligand, make identifying the tumors that will respond to AXL-targeted therapy especially challenging.[1] In lung cancer, resistance to EGFR tyrosine kinase inhibitors such as erlotinib have prompted researchers to evaluate potential mechanisms underlying this phenomenon in an attempt to prevent it. AXL was found to be highly predictive of lack of response to ErbB targeted inhibitors. AXL is transactivated by EGFR, thereby diversifying the downstream signaling pathways employed by EGFR.[2] With the goal of better understanding the underlying signaling mechanisms of AXL-mediated resistance to EGFR TKIs, we have used mass spectrometry-based phosphoproteomics to interrogate signaling networks in the PC9 lung cancer cell line under a variety of conditions. We previously treated PC9 and AXL knockout cells with erlotinib (EGFR inhibitor) and R428 (AXL inhibitor), or the combination of erlotinib+R428. PC9 cells were also treated with a combination of erlotinib and various growth factors, either HGF, FGF or IGF to provide insight into bypass resistance mechanisms. Data from three biological replicates motivated collection of phenotypic data as well as another

phosphoproteomic dataset, which includes signaling data for a number of cell lines containing Y-to-F mutations in AXL, in response to erlotinib or a combination of erlotinib and AF154 (AXL activating antibody) to determine functions of individual tyrosine residues on AXL. In addition, this dataset included phosphotyrosine data for PC9 cells, AXL knock-out, AXL knock-in, and AXL kinase-dead cells as controls. Phenotypic data was collected for these lines as well, as cell viability, cell death, cell migration and island formation was measured. Integrating these measurements with quantitative modeling will identify the connectivity between receptors, interacting adapters, and downstream signaling events, thereby defining the essential set of signaling network changes required for tumor cell survival in response to targeted therapeutics. Indeed, a combination of clustering and motif analysis revealed proteins of the YAP pathway as well as CK2 and NEKs to be highly predictive of the malignant phenotypes that were evaluated. These results were further validated using RNAseq and western blot, highlighting potential signaling nodes that could link EGFR and AXL together. In summary, this work has helped define the network nodes driving various phenotypes under different conditions as potential targets in AXL-mediated resistance to EGFR TKIs.

Results

Development and phenotypic characterization of AXL-mutant cell lines

To evaluate the individual contributions of the C-terminal tyrosines in the AXL signaling network, PC9 cells were transfected to express AXL wildtype (WT), knock-out(KO), knock-in(KI), kinase-dead or containing 1 of 6 C-terminal Y-to-F mutations (Fig 1A and S1). Cells were treated with EGFR inhibitor Erlotinib (Erl), or a combination of Erlotinib and AXL activating antibody AF154, and subjected to phosphoproteomic analysis. Furthermore, 4 different phenotypes were measured: cell death (Erlotinib only), cell viability (Erl+AF154), cell migration and island effect (i.e. clustering of cells) (Fig 1B and S1). Fig 1C shows decreased cell viability for all cell lines when treated with Erlotinib, even in the presence of AF154, suggesting AXL activity alone is not sufficient to maintain cell viability (Fig S1 - S3). Apoptosis results demonstrated less cell death in the AXL isoform cell lines compared to WT in the Erlotinib treated condition, suggesting an AXL-mediated mechanism at play. Interestingly, the migration data demonstrated an increase in migration in WT under treated conditions, while all AXL mutants showed the opposite effect in the Erl condition. This effect was negated for several but not all mutants in the presence of AF154. The island effect response patterns that were unique to certain Y-to-F mutants, highlighted the importance of evaluating individual function and mechanism of these tyrosine sites. When plotting our cell lines and phenotype data points in Principle Component Analysis (PCA) space, a clear distinction based on phenotype can be observed (Fig 1D). As

consistent with prior knowledge, apoptosis and viability, as well as island effect and migration are opposing phenotypes.

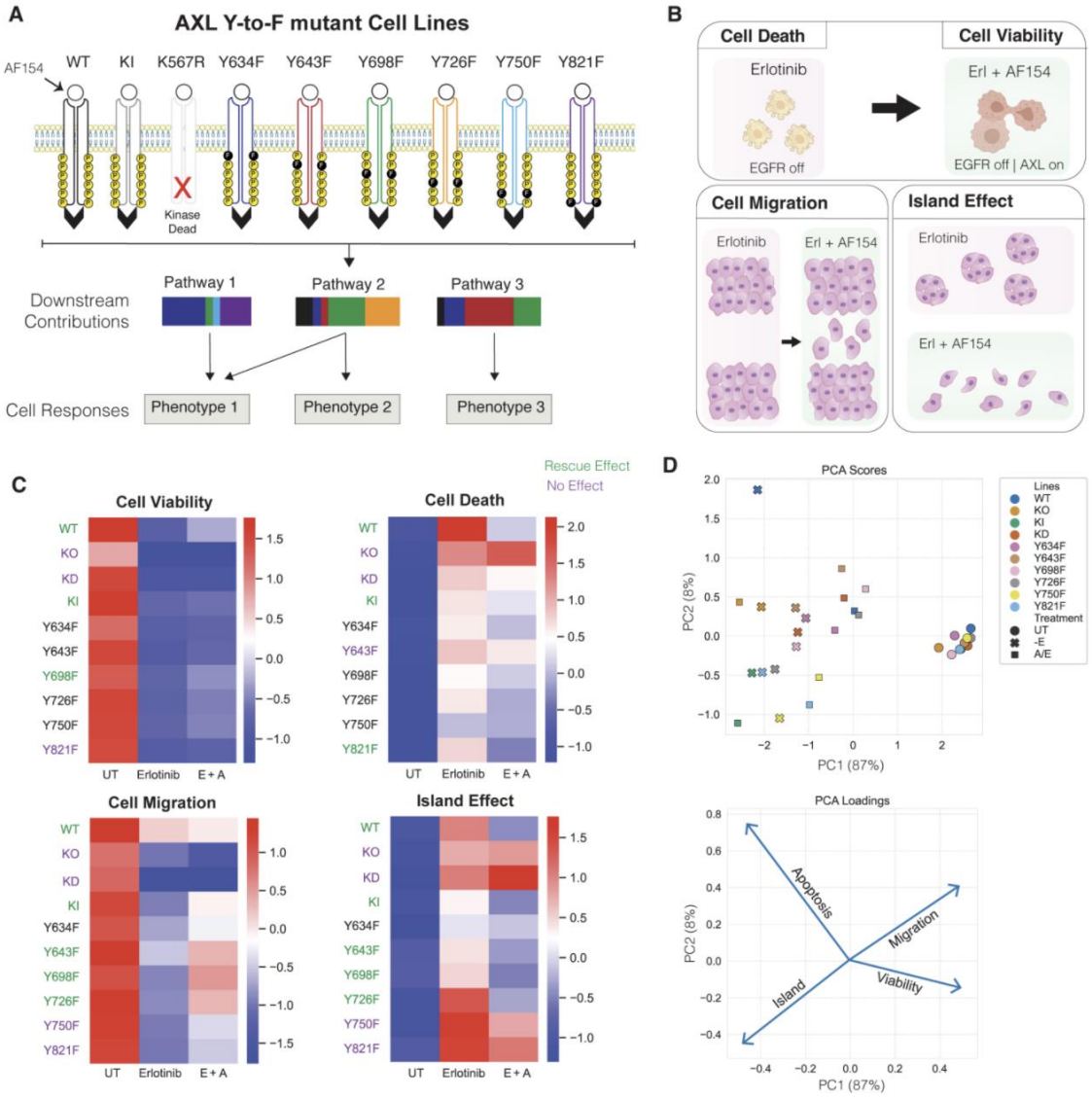


Figure 22: AXL Y-to-F mutations distinctly affect cell viability and migration phenotypes. *(**A**) A panel of PC9 AXL WT, KO, KD, KI, and Y-to-F mutants have different contributions AXL-specific downstream signaling pathways which in turn mediate various phenotypes. (**B**) Cartoon of the observed cell phenotypes in this study using the AXL mutant cell lines in the presence of 1uM erlotinib and 300ng/mL AF154. Cell proliferation was monitored via live imaging during 4 days alongside the number of YOYO+ apoptotic cells. In parallel, the ability of cells to migrate and was measured through a scratch wound assay (bottom left). Erlotinib-only treated cells form cluster together and form cell "islands" which subsequently spread after the activation of AXL with AF154. (**C**) Heatmap of phenotypic responses of PC9 untreated (UT), erlotinib-only, and erlotinib plus AF154. (**D**) PCA analysis of cell phenotypes.*

Modeling phosphoproteomic clusters driving AXL-mediated malignant phenotypes.

Using the Meyer lab's Dual data-Motif Clustering method is a clustering and kinase prediction strategy that identifies signaling nodes by grouping phosphosites according to their phosphorylation signal and amino acid sequence (Fig 2A) [3]. Fig 2B shows the log2 fold change in abundance of phosphorylated peptides (columns) for each of the AXL mutant isoforms (rows) (Fig S4). AXL phosphorylation is increased upon stimulation with the activating antibody, except in the KO and KD cell lines, as expected (Fig 2C). Y481 may be utilized by AXL to compensate for loss of other tyrosines, as this tyrosine shows no increase in phosphorylation in WT, but a strong increase in several of the single mutant cell lines. Furthermore, decreased phosphorylation of Y759 in the Y750F mutant suggests a potential co-dependence of these sites for signaling. DDMC identified 5 clusters of peptides as shown in Fig 2D and 2E. Combination of the signaling data with the phenotypic data through partial least squares regression (PLSR) identified several critical nodes as drivers of these phenotypes, most notably Src family kinases (SFKs) (Figs 2F, 2G, 2H).

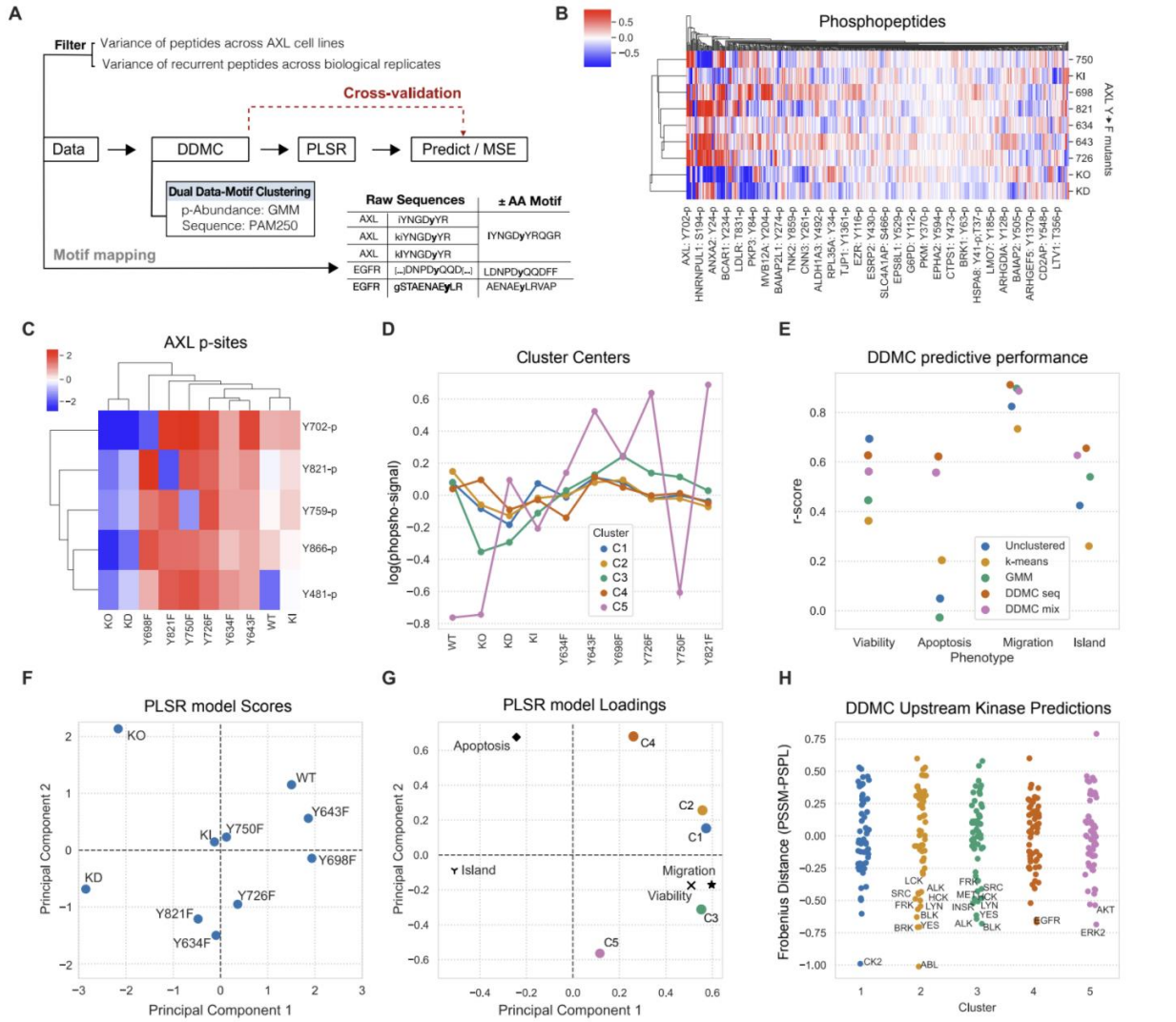


Figure 23: Linking AXL downstream signaling and phenotypic responses. *(**A**) Diagram of modeling strategy. Phosphoproteomic data was preprocessed and then clustered using DDMC based on the p-sites phosphorylation abundance and sequence motifs. Then, the cluster centers were fit into a PLSR model to predict the AXL-driven cell phenotypes. (**B**) Hierarchical clustering of phosphoproteomic data. (**C**) AXL phosphosites across mutants. (**D**) Cluster centers. (**E**) PLSR prediction performance calculated by measured vs predicted correlation scores fitting different clustering strategies. (**F-G**) PLSR scores and loadings. (**H**) DDMC upstream kinase predictions.*

AXL activates key components of the YAP pathway

Having identified SFKs as major drivers of malignant phenotypes in these data, this mechanism was further investigated by evaluating cell viability in the presence of SFK inhibitor Dasatinib (Fig 3A). Analysis showed that cell viability was affected in otherwise untreated cells when exposed to dasatinib. No effect was observed in cells already treated with Erl. Since Yes-associated protein (YAP) has been reportedly involved in AXL-mediated resistance to EGFR TKIs[4], and the proposed role of SFKs in predicting phenotype in our data, phosphorylated YAP was measured under various conditions (Fig 3B). This data showed decreased YAP in the presence of Erlotinib, particularly in a low cell density environment. This effect was negated with increased concentration of dasatinib (Fig 3C). GSEA analysis showed enrichment of YAP in our data(Fig 3D), and YAP phosphorylation was down in treated conditions (Fig 3E). Peptides consistently downregulated in response to dasatinib in WT and KO lines included classic SFK substrates, including ABL, FRK, LCK, LYN and YES which are key components of the YAP pathway (Fig 3F and S5). DDMC clusters 2 and 3 were strongly enriched for SFK and ABL motifs in dasatinib treated conditions (Fig 3G).

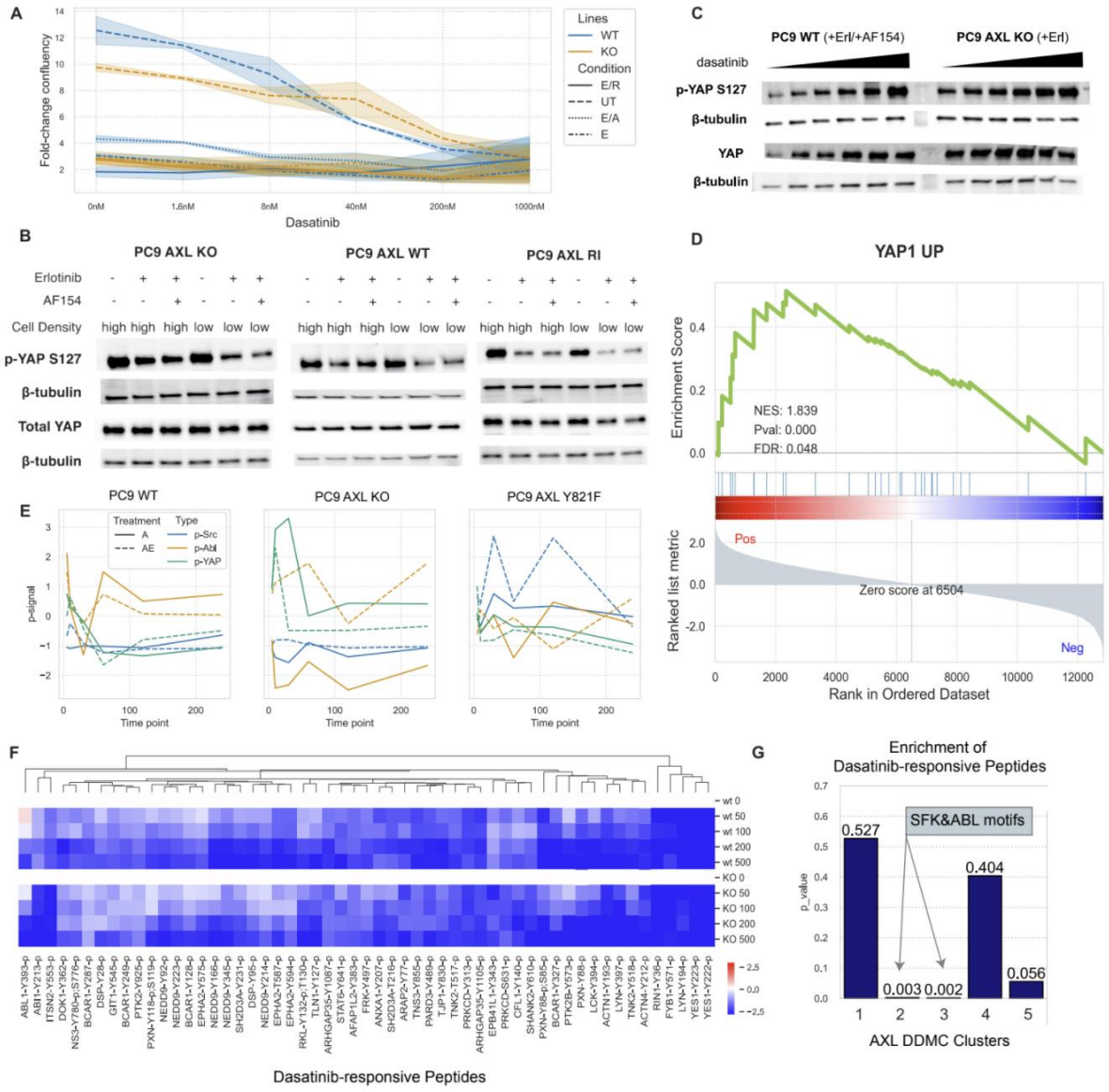


Figure 24: Key YAP signaling components drive AXL-driven phenotypes. *(**A**) Dasatinib dose response of PC9 WT cells after 72h treated with erlotinib only, erlotinib and AF154, or erlotinib and R428. (**B**) Western blots of total YAP and the inhibitory p-YAP S127 in the presence and/or absence of AF154 at high or low cell density in AXL KO, WT, or AXL re-introduced cells. (**C**) YAP activation status in response to dasatinib. (**D**) YAP transcriptional signature using RNAseq data of the AXL mutants. (**E**) Phospho-YAP, -Src, and -Abl signal over time. (**F**) Dasatinib-response phosphopeptides in PC9 AXL WT and KO cells. (**G**) Enrichment of dasatinib-response peptides in AXL DDMC phosphoproteomic clusters.*

CK2/NEKs involvement in DNA damage response

Besides YAP arising from the analysis as a strong driver of mechanism, CK2 was also identified as a potential linking node between AXL and EGFR. Given CK2s function in cell cycle regulation, GSEA analysis showed enrichment in DNA Damage Response related pathways (Fig 4A). Viability measurements of WT and KO cells subjected to CK2 inhibitor CX-4945 treatment showed strong decreased viability, which was also the case when treated with PLK1 inhibitor Volasertib (Fig 4B). PLK1 reportedly works in concert with CK2 to control DNA damage repair pathways[5]. Interestingly, western blot showed an increase in PLK1 activation with increased concentration of Volasertib when co-treated with CX-4945 (Fig 4C), suggesting a range of potential mechanisms responsible for these effects.

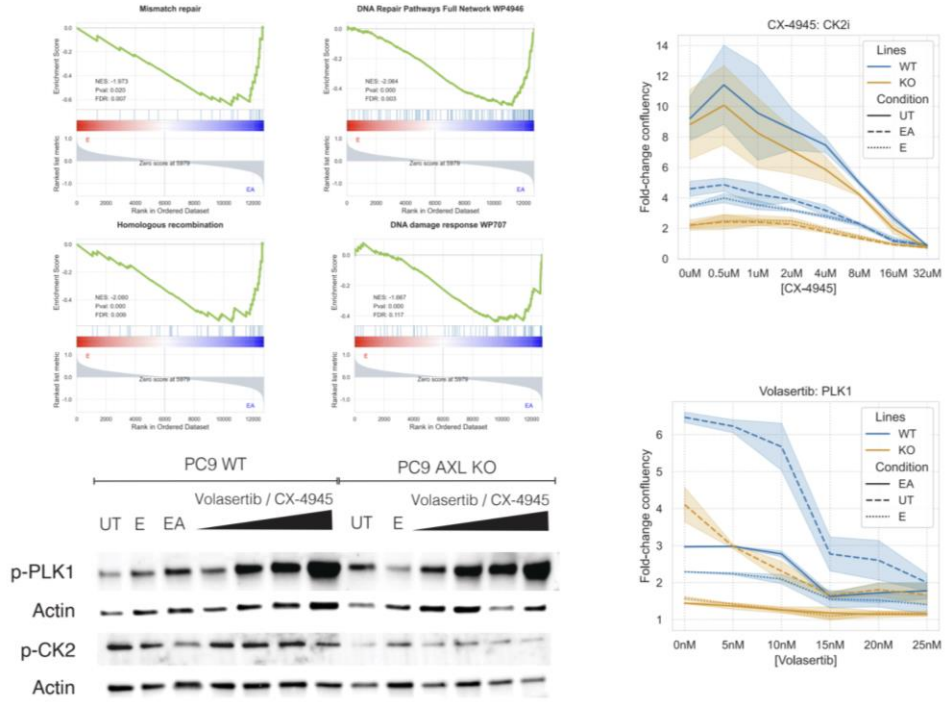


Figure 25: CK2/NEKs involvement in DNA damage response (DDR).

Discussion

The goal of this work was to obtain an improved understanding of the AXL signaling network, particularly as it pertains to AXL-mediated resistance to EGFR TKIs. A set of AXL mutant isoforms were expressed in PC9 cells to evaluate the individual function of its C-terminal tyrosines. Cells were treated with AXL activating antibody AF154 in the presence or absence of EGFR TKI Erlotinib, and signaling network evaluated using a phosphoproteomics approach. To obtain a holistic view of cell response, cell viability, cell death, cell migration and island formation were taken. These phenotypic and signaling data sets were subjected to DDMC and PLSR analysis that identified the YAP pathway and SFKs as a potential signaling axis mediating resistance to erlotinib. Furthermore, CK2 and NEKs were identified as drivers of malignant phenotypes and linked to the DNA damage response pathway. Further experiments are necessary to validate the link between CK2 and AXL. Other assays that do not require the use of chemo agents could be utilized to measure DDR as a phenotype.

Since RNAseq data has also shown an enrichment in trafficking in AXL activated samples, future trafficking assays may reveal more potential nodes and mechanisms or further elucidate the ones explored thusfar. Finally, given the DDA nature of the phosphoproteomic approach, Luminex or PLA measurements of particular nodes of interest or a targeted mass spec approach may allow for evaluation of particular proteins of interest in the network, and linking this data to the RNAseq data for added biological insight. Fig S6 shows potential drugs that could be used to experimentally validate the YAP pathway as a viable way to prevent resistance.

In all, this approach could potentially lead to the identification and pursuit of future targeting strategies in our combat against EGFR TKI resistance in lung cancer.

Personal contributions to this work:

- Processed samples for LC-MS/MS analysis
- Performed LC-MS/MS analysis
- Writing LC-MS/MS methods

AXL Genotype	Known Intracellular interactors	ScanSite	NetworKIN	NetPhorest	PDB	Known Phenotype
Wild-type	PI3K, AKT, GRB2, p85, PLCγ, SRC, LCK	N/A	N/A	N/A	N/A	Cell survival, proliferation, migration, and immunosuppression
Knock-Out	N/A	N/A	N/A	N/A	N/A	Inverse phenotyes of WT
Transduced Wild-type	N/A	N/A	N/A	N/A	N/A	N/A
K567R	N/A	N/A	N/A	N/A	N/A	Kinase dead, typically recapitulates KO phenotypes
Y634F	-	SHC	-	CRKL, SRC, PTP R3	-	-
Y643F	-	p85	-	-	-	-
Y698F	-	-	INPPL1, BCAR3	FLT3, CSF1R, KIT, PDGFR, PTP R3	-	-
Y726F	-	-	CBL, SYK	PTP R3	-	-
Y750F	-	-	-	SH2D	-	-
Y821F	PI3K, AKT, GRB2, p85, PLCγ, SRC, LCK	p85, FGR, GRB2	GRB2, PI3K, GRAP, MIST SHP2	PIK3R, GRB2, SRC, PTP R3	-	-

Search Parameters:

Scansite: Stringency = “High” for Y821, and “medium” for the rest. NetworKIN: Minimum score = 1.11 for Y812 ; 0.89 for the rest. NetPhorest: Minimum score = 0.20 for Y634, Y643, and Y821 ; = 0.10 for Y726 and Y750

Reagents and Cell Culture

Erlotinib was obtained from LC Laboratories. The AXL activating antibody AF154 was obtained from R&D Systems. PC9 (Sigma Aldrich) cells were grown in RPMI-1640 media supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin. HEK293T cells were grown in DMEM supplemented with 10% FBS and 1% GlutaMAX (Thermo Fisher Scientific).

Generation of Mutant Cell Lines

The PC9 AXL KO cell line was generated by transfecting cells with a CRISPR/Cas9 and GFP vector containing a gRNA targeting the AXL kinase domain. gRNA sequence, as well as cell culturing and sorting methods, have been previously described (*13*). Plasmids containing the AXL phosphosite mutations were generated from an AXL-IRES-Puro vector (Addgene #65627) using a site directed mutagenesis kit (New England Biosciences).

Each mutant was copied from its respective plasmid, digested, and inserted into a lentiviral vector containing a puromycin resistance gene (Addgene #17448). For viral packaging, HEK 293T cells were seeded 4.5 x 10^6 per 10 cm dish. After 24 hours, the lentiviral AXL expression vector, VSV-G envelope vector, and packaging vector (Addgene #12259 and #12260 respectively) were combined in a 10:1:10 mass ratio and diluted in Opti-MEM (Thermo Fisher Scientific). TransIT-LT1 (Mirus Bio) was added dropwise and the solution was mixed gently by swirling and incubated at room temperature for 20 minutes. The solution was then added dropwise to the cells. After 18 hours, transfection media was replaced by media supplemented with 1% BSA fraction V (Thermo Fisher Scientific). Cells were incubated for 24 hours, after which the virus-containing media was removed and stored at 4℃. The media was replaced and the cells incubated a further 24 hours to generate a second batch of viral

media. The harvested batches were then pooled, filtered through a .45µm PVDF membrane to remove packaging cells, and flash frozen followed by storage at -80ºC until use.

PC9 AXL KO cells were seeded with antibiotic-free media in a 6-well plate at a density of 1.5 x 105 cells per well and incubated for 24 hours. Cells were then infected with viral particles in antibiotic-free media supplemented with polybrene (MilliporeSigma). After 18 hours, the media was replaced with fresh antibiotic-free media. Cells were observed for a GFP positive population and passaged into a 10cm plate until confluent. The virally transduced cell lines were then sorted for GFP expressing cells using a BD FACSAria cell sorter. The mutant cell populations were then subcultured for later experiments.

Preparation of Cell Lysates

Cell lines were grown to confluence in 10 cm dishes over the course of 72-96 hours, washed, and treated by addition of media containing 1 µM erlotinib. Cells were incubated for 4 hours at 37ºC and then additionally treated with media containing 1 µM erlotinib and 300 ng/mL AXL activating antibody for 10 minutes. The cells were then placed immediately on ice, washed with ice-cold phosphate-buffered saline, and lysed with cold 8 M urea containing Phosphatase Inhibitor Cocktail I and Protease Inhibitor Cocktail I (Boston BioProducts). The lysates were then centrifuged at 20,000xg, 4℃ to pellet cell debris, and the supernatants removed and stored at -80℃.

Sample processing

After vehicle or Erlotinib/AF154 treatment, cells were lysed on ice using 500µL 8M Urea (Sigma) per 10cm plate. A bicinchoninic acid (BCA) protein concentration assay (Pierce) was performed according to the manufacturer's protocol to estimate the protein concentration in each lysate.

Cell lysates were reduced with 10mM DTT for 1hr at 56^{o}C, alkylated with 55 mM iodoacetamide for 1hr at RT shielded from light, and diluted 5-fold with 100 mM ammonium acetate, pH 8.9, before trypsin (Promega) was added (20:1 protein:enzyme ratio) for overnight digestion at RT. The resulting solutions were acidified with 1 mL of acetic acid (HOAc) and loaded onto C18 Sep-Pak Plus Cartridges (Waters), rinsed with 10mL of 0.1% HOAc, and eluted with 10mL of 40% Acetonitrile (MeCN)/ 0.1% HOAc. Peptides were divided into 200 microgram aliquots, and sample volume was reduced using a vacuum centrifuge (Thermo) and then lyophilized to dryness for storage at -80^{o}C.

TMT labeling for multiplexed analysis was performed according to manufacturer's protocol. Samples, each containing ~200ug peptides, were resuspended in 35µL HEPES (pH 8.5), vortexed, and spun down at 13,400rpm for 1 minute. 400µg of a given channel of TMT10plex (Thermo) in anhydrous MeCN, was added per sample. Samples were shaken at 400rpm for 1hr, after which the labeling reaction was quenched using 5% Hydroxylamine (50%, Thermo). After another 15 minutes on the shaker, all samples were combined using the same pipette tip to reduce sample loss, and sample aliquots were washed twice with 40 µL 25% MeCN/0.1% HOAc which was added to the collection tube to improve yield. Sample volume was reduced using a vacuum centifuge and then lyophilized to dryness for storage at -80^{o}C.

Phosphopeptide enrichment

Immunoprecipitation (IP) and IMAC were used sequentially to enrich samples for phosphotyrosine containing peptides. TMT-labeled samples were incubated in IP buffer consisting of 1% Nonidet P-40 with protein G agarose beads conjugated to 24 µg of 4G10 V312

IgG and 6 µg of PT-66 (P3300, Sigma) overnight at 4^{o}C. Peptides were eluted with 25 µl of 0.2% trifluoroacetic acid for 10 minutes at room temperature; this elution was performed twice to improve yield. Eluted peptides were subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)-based Fe-NTA spin column to reduce non-specific, non-phosphorylated peptide background. High-Select Fe-NTA enrichment kit (Pierce) was used according to manufacturer's instructions with following modifications. Eluted peptides from IP were incubated with Fe-NTA beads containing 25µL binding washing buffer for 30 minutes. Peptides were eluted twice with 20mL of elution buffer into a 1.7 mL microcentrifuge tube. Eluates were concentrated in speed-vac until ~1 µL of sample remained, and then resuspended in 10µL of 5% acetonitrile in 0.1% formic acid. Samples were loaded directly onto an in-house constructed fused silica capillary column [50 micron inner diameter (ID) x 10 cm] packed with 5µm C18 beads (YMC gel, ODS-AQ, AQ12S05) and with an integrated electrospray ionization tip (~2 micron tip ID).

LC-MS/MS analysis

LC-MS/MS of pTyr peptides were carried out on an Agilent 1260 LC coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptides were separated using a 140-minute gradient with 70% acetonitrile in 0.2 mol/L acetic acid at flow rate of 0.2 mL/minute with approximate split flow of 20 nL/minute. The mass spectrometer was operated in data-dependent acquisition with following settings for MS1 scans: m/z range: 350 to 2,000; resolution: 60,000; AGC target: 3 x 10^{6}; maximum injection time (maxIT): 50 ms. The top 15 abundant ions were isolated and fragmented by higher energy collision dissociation with

following settings: resolution: 60,000; AGC target: $1x10^5$; maxIT: 350 ms; isolation width: 0.4 m/z, collisional energy (CE): 33%, dynamic exclusion: 20 seconds. Crude peptide analysis was performed on a Q Exactive Plus mass spectrometer to correct for small variation in peptide loadings for each of the TMT channels. Approximately 30 ng of the supernatant from pTyr IP was loaded onto an in-house packed precolumn (100µm ID x 10 cm) packed with 10mm C18 beads (YMC gel, ODS-A, AA12S11) and analyzed with a 70-minute LC gradient. MS1 scans were per-formed at following settings: m/z range: 350 to 2,000; resolution:70,000; AGC target: $3x10^6$; maxIT: 50 ms. The top 10 abundant ions were isolated and fragmented with CE of 33% at a resolution of 35,000.

Peptide identification/quantification

Mass spectra were processed with Proteome Discoverer version 2.5 (Thermo Fisher Scientific) and searched against the human SwissProt database using Mascot version 2.4 (MatrixScience, RRID:SCR_014322). MS/MS spectra were searched with mass tolerance of 10 ppm for precursor ions and 20 mmu for fragment ions. Cysteine carbamidomethylation, TMT-labeled lysine, and TMT-labeled peptide N-termini were set as fixed modifications. Oxidation of methionine and phosphorylation of serine, threonine and tyrosine were searched as dynamic modifications. TMT reporter quantification was extracted and isotope corrected in Proteome Discoverer. Peptide spectrum matches (PSM) were filtered according to following para-meters: rank=1, mascot ion score>15, isolation interference<40%, average TMT signal>1,000. Peptides with missing values across any channel were filtered out.

Cell Viability Assay

Cells were seeded in a 96-well plate at a density of 1.05 x 10^3 cells per well. After 24 hours, treatments were added in media containing 300nm YOYO-3 (Thermo Fisher Scientific). Cells were cultured and imaged every 3 hours using an IncuCyte S3 (Essen Bioscience) at 20x magnification with 9 images per well. The phase, green, and red channels were manually thresholded and then analyzed by IncuCyte S3 software (Essen Bioscience) to determine cell counts and fraction of area covered.

Cell Migration Assay

96-well IncuCyte ImageLock plates (Essen Bioscience) were coated with a Collagen-I solution (Thermo Fisher Scientific), washed twice, and then seeded with 4 x 10^4 cells per well. After a 4 hour incubation, cells were wounded using the IncuCyte WoundMaker, washed twice to remove detached cells, and then treated with respective conditions. Images of the center of the wound were taken every 2 hours at a magnification of 10x, one image per well. The phase and green channels were thresholded and analyzed as above to determine migration measurements.

Cell Island Effect

Phase contrast images used for the cell island measurements were taken from image sets gathered in the cell viability assay. For endpoint readings, images at the 48 hour post-treatment

time point were used. Representative images were chosen across experimental replicates. Images were opened in ImageJ and the center of each cell was manually marked. Dead cells, identified using YOYO-3 based fluorescence, were not marked. The 2D coordinates of all cell centers in an image were then exported for analysis. The amount of clustering present in a particular image was then measured by applying Ripley's K function to the set of coordinates. The implementation of Ripley's K function used was taken from the astropy Python package.

All analysis was implemented in Python, and can be found at **https://github.com/meyer-lab/resistance-MS**.

References

1 Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. FEBS J Internet. 2010 cited 2022 Feb 4;277:301–308. Available from: https://pubmed.ncbi.nlm.nih.gov/19922469/.

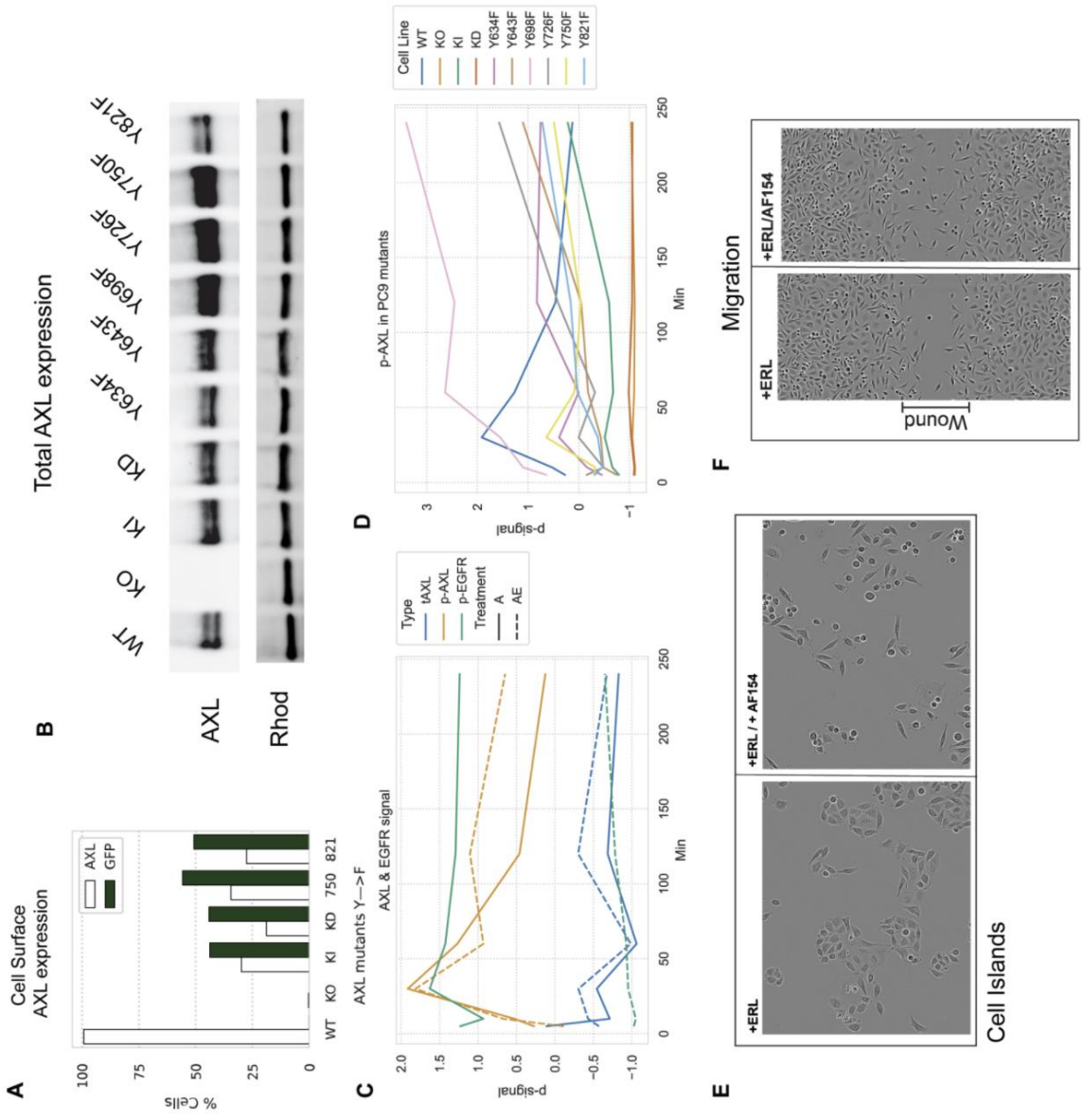
2 Meyer AS, Miller MA, Gertler FB, et al. The receptor AXL diversifies EGFR signaling and limits the response to EGFR-targeted inhibitors in triple-negative breast cancer cells. Sci Signal Internet. 2013 cited 2015 Feb 8;6:ra66. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3947921&tool=pmcentrez&rendertype=abstract.

3 Creixell M, Meyer AS. Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data. Cell Reports Methods. 2022;2:100167.

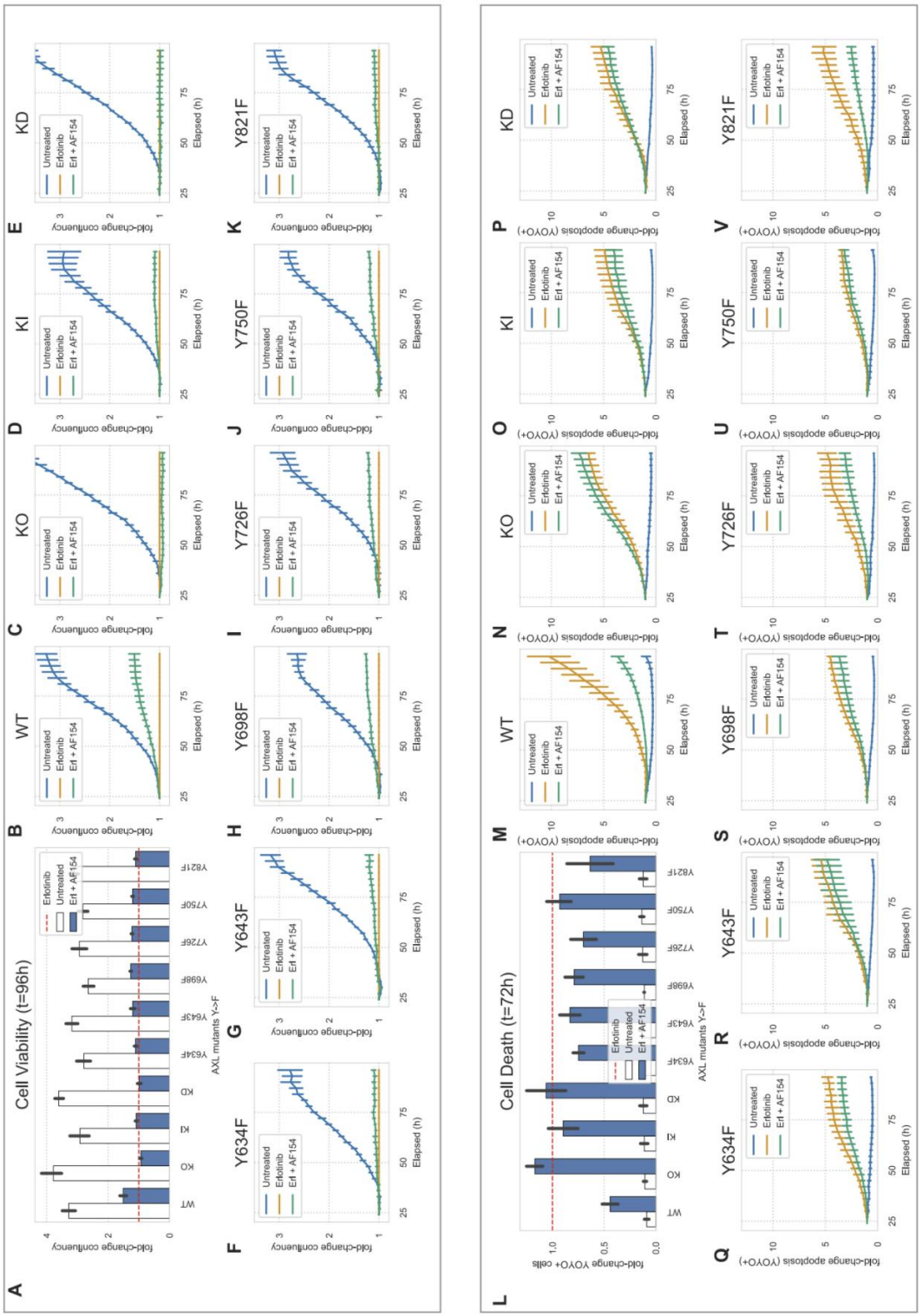
4 Saab S, Chang OS-S, Nagaoka K, et al. The potential role of YAP in Axl-mediated resistance to EGFR tyrosine kinase inhibitors. Am J Cancer Res Internet. 2019 cited 2022 Mar 21;9:2719. Available from: /pmc/articles/PMC6943358/.

5 Yata K, Lloyd J, Maslen S, et al. Plk1 and CK2 Act in Concert to Regulate Rad51 during DNA Double Strand Break Repair. Mol Cell Internet. 2012 cited 2022 Mar 21;45:371. Available from: /pmc/articles/PMC3280358/.

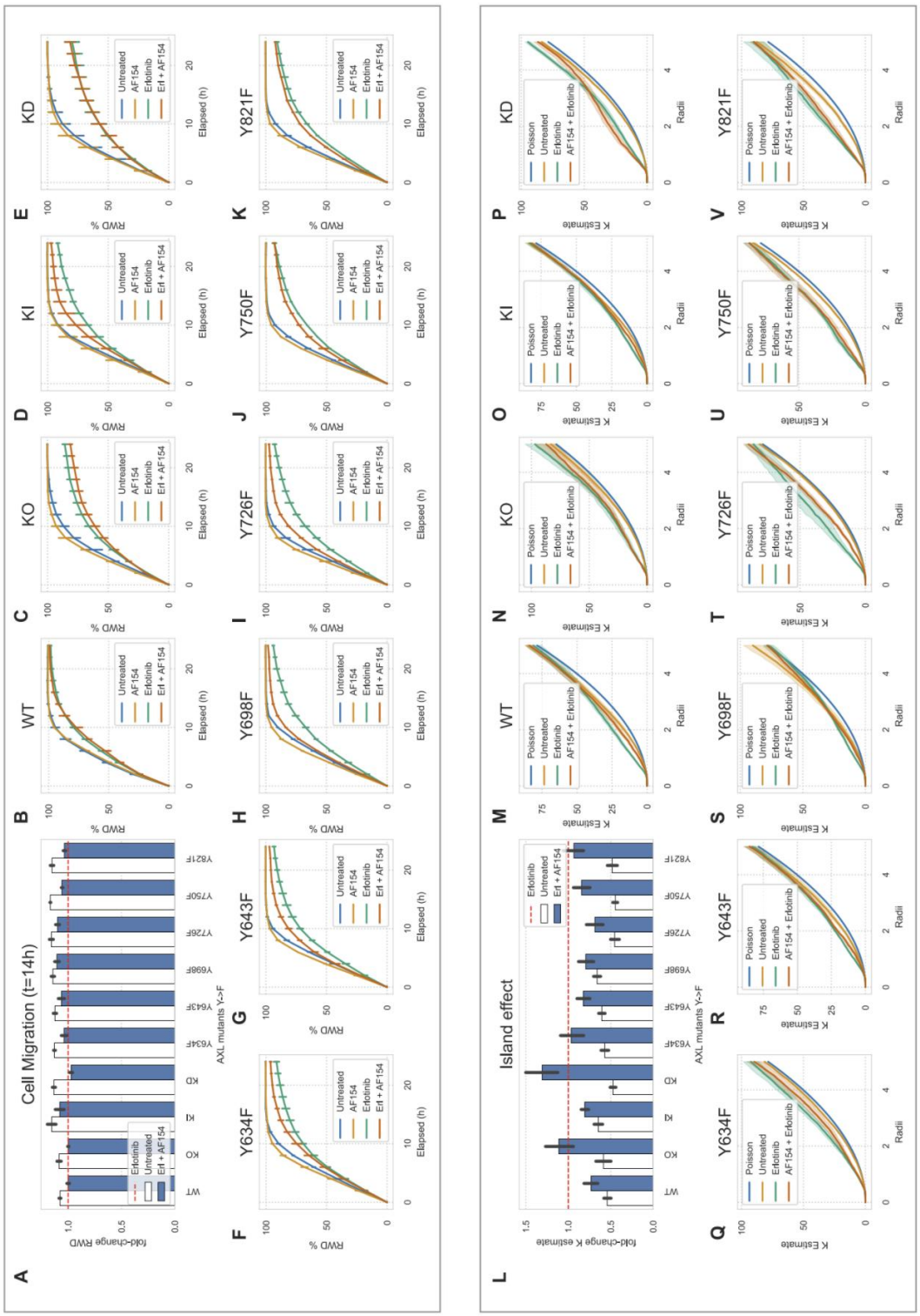
Supplemental Figures



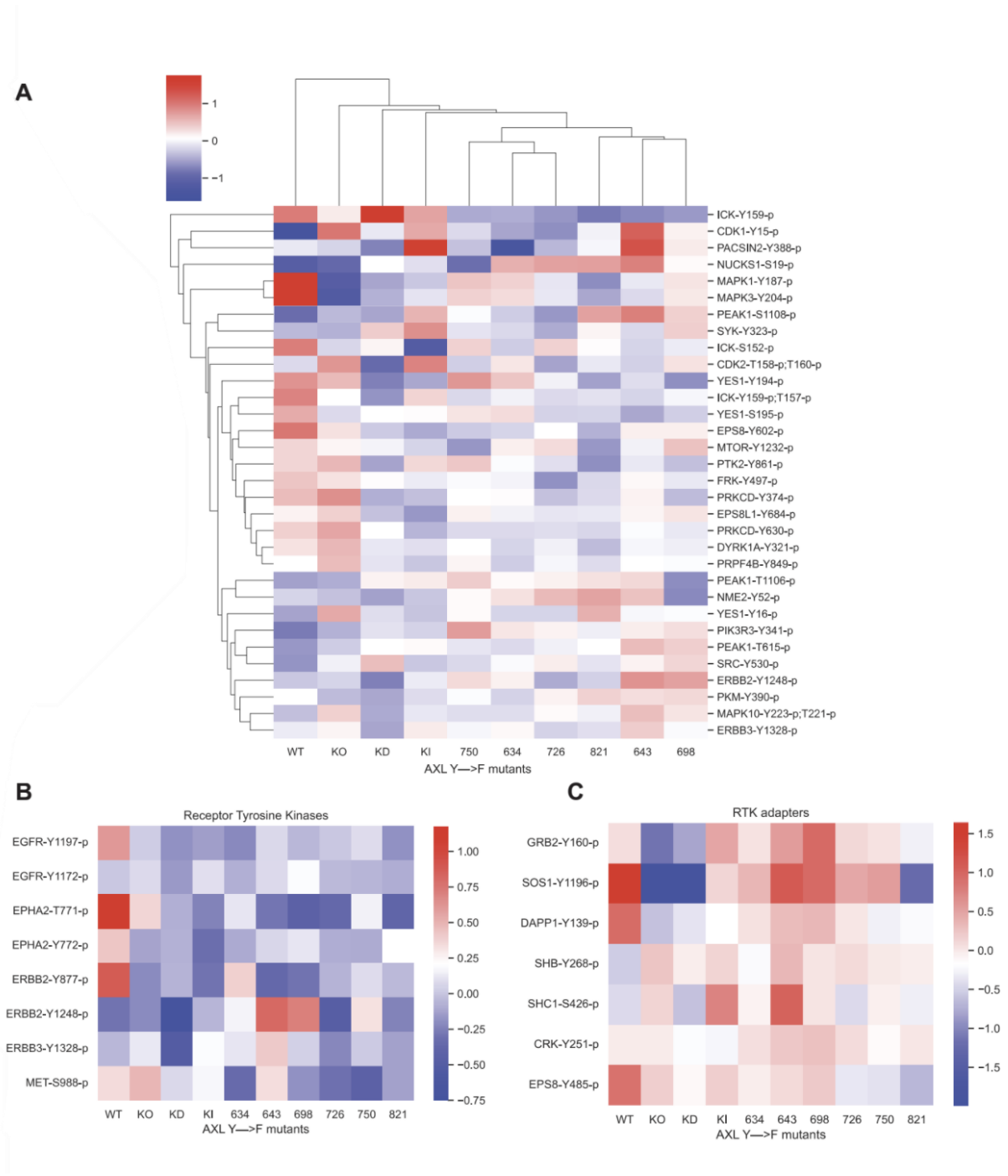
Supplementary figure 8: Ectopic AXL expression, activation and migration images.



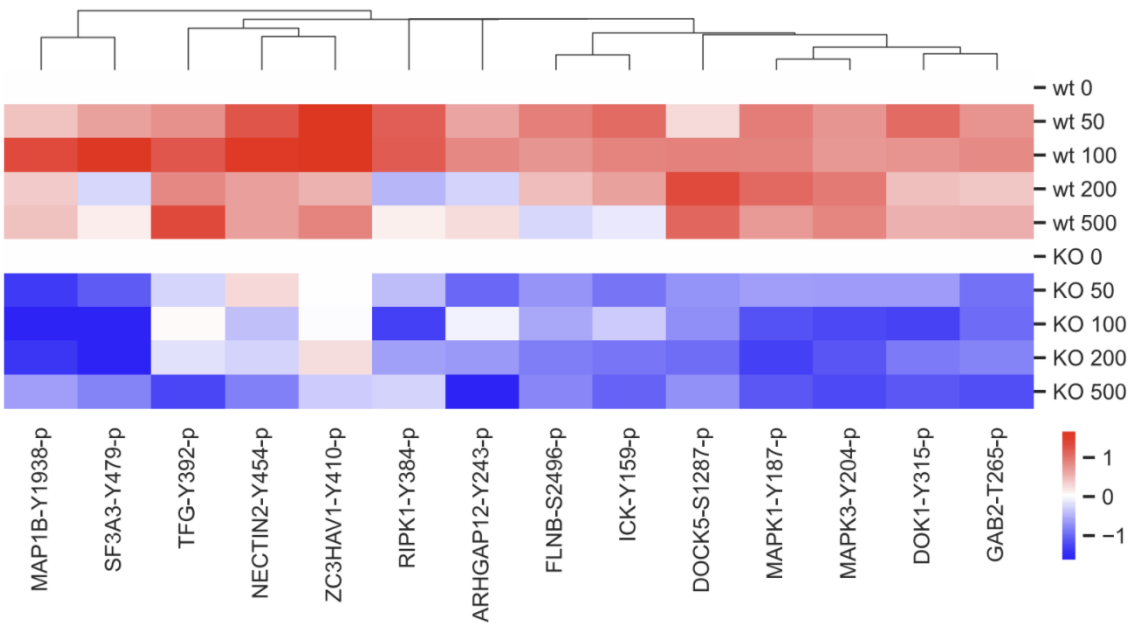
Supplementary figure 9: Cell viability and death time lapse.



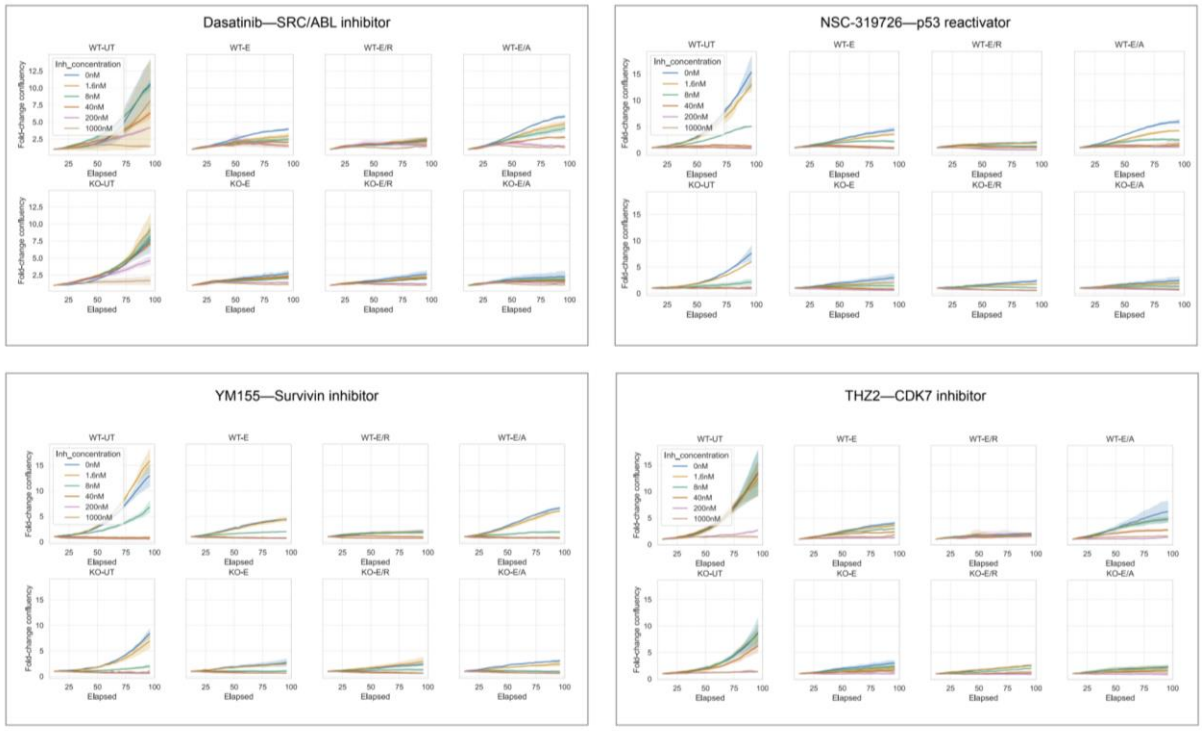
Supplementary figure 10: Cell migration and island effect time lapse.



Supplementary figure 11: Signal of kinases, RTKs and adapters across AXL mutants.



Supplementary figure 12: AXL-induced feedback.



Supplementary figure 13: Most responsive drugs to YAP-dependent cell lines on DepMap.

CHAPTER 7

Regulation of EGFR signaling by disrupting PTPRJ dimerization – exploring potential mechanisms

Jacqueline S. Gerritsen

Sophie Rizzo

Eden Sikorski

Will Hart

Matthew J. Lazarra

Forest M. White

Damien Thevenin

Abstract

PTPRJ is a receptor protein tyrosine phosphatase (RPTP) (also known as DEP1) that plays a critical role in regulating signaling cascades. Underlying mechanisms are poorly understood, in part aggravated by lack of targetability of PTPRJs due to lack of endogenous ligands. Interestingly, RPTP activity is suppressed upon dimerization, hypothetically because of prevention of RTK substrate interaction as a dimer. Certain transmembrane mutations in PTPRJ, including the G983L mutation, were found to destabilize homodimerization, thereby increasing phosphatase activity, and downregulating substrate activity, including phosphorylation of EGFR.

Here, we aim to elucidate the mechanisms underlying the PTPRJ-EGFR signaling axis, paving the way for usage of peptide-based agents that disrupt RPTP dimerization that enter the cell under acidic conditions, as is the case in the tumor microenvironment. These agents could be used as tools to manage the increased activation of RTK signaling networks often observed in cancers.

Introduction

The activity of Receptor Tyrosine Kinases (RTKs) is counterbalanced by phosphatases that dephosphorylate their substrates, thereby decreasing their signal transduction.[1–4] The receptor -like phosphatases (RPTPs) are a subgroup of phosphatases that span the cell membrane, and are active in monomer formation.[5,6] As opposed to RTKs, dimerization of RPTPs causes an inactive dimer due to lack of binding ability to its substrate.

One RPTP of particular interest is PTPRJ (also known as DEP1), due to its associated regulatory function of the C-terminal tail of EGFR, which is well-known to regulate many critical cellular functions.[7] As EGFR is a frequent driver of cancers, PTPRJ and other phosphatases have gained interest as a potential therapeutic target to indirectly regulate EGFR activity.[8] To increase the activity of PTPRJ and further decrease EGFR activity, previous work has identified several transmembrane domain mutations preventing dimerization and hence disfavoring the inactive conformation of PTPRJ.[9] The G983L mutation showed particularly effective at shifting the equilibrium towards active monomer.[10] The ultimate goal is to deliver a synthetic peptide that fulfills the same function as the mutation by preventing homodimerization, and to engineer this peptide to specifically enter cells in an acidic environment like the tumor microenvironment, thereby achieving a specific and effective treatment strategy for EGFR-driven tumors.[11]

To better understand the underlying mechanisms and employed signaling axes connecting PTPRJ and EGFR, here we analyzed and compared the signaling network in PTPRJ WT and G983L cell lines using a phosphoproteomic approach. Migration was also measured as a phenotypic outcome, enabling regression analysis to find predictive nodes of this phenotype. These insights may further inform development of a novel therapeutic strategy that employs PTPRJ to regulate EGFR activity in human cancers.

Results

To evaluate the potential of PTPRJ modulation to regulate EGFR activity, SCC2 cells were transfected with empty vector, wild type PTPRJ or PTPRJ G983L (blocking dimerization of PTPRJ). As discussed above, the G983L transmembrane mutation was found to inhibit formation of the inactive PTPRJ homodimer, thereby increasing phosphatase activity and downregulation of the EGFR network (Fig 1).

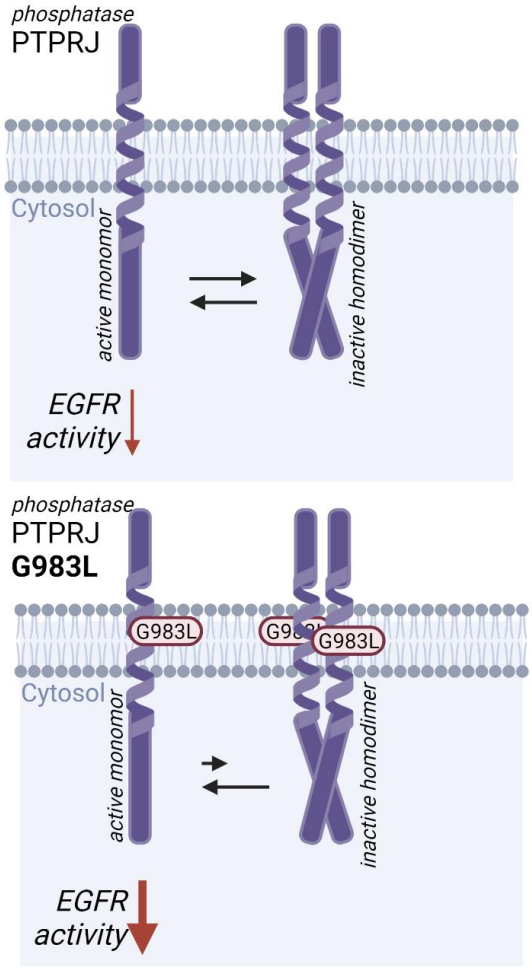


Figure 26: G983L mutation in PTPRJ increases PTPRJ phosphatase activity

Following successful transfection, SCC2 cells were stimulation with vehicle or 10mM EGF for 10 minutes, lysed in cold 8M Urea to preserve physiological signaling, and proteolytically digested to peptides. To reduce inter-sample variability, samples were labeled with isobaric tandem mass tags and analyzed as 6-plex experiments. Labeled samples were subjected to 2-step phosphotyrosine enrichment and subsequent liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, yielding quantitative phosphotyrosine data (Fig 2).

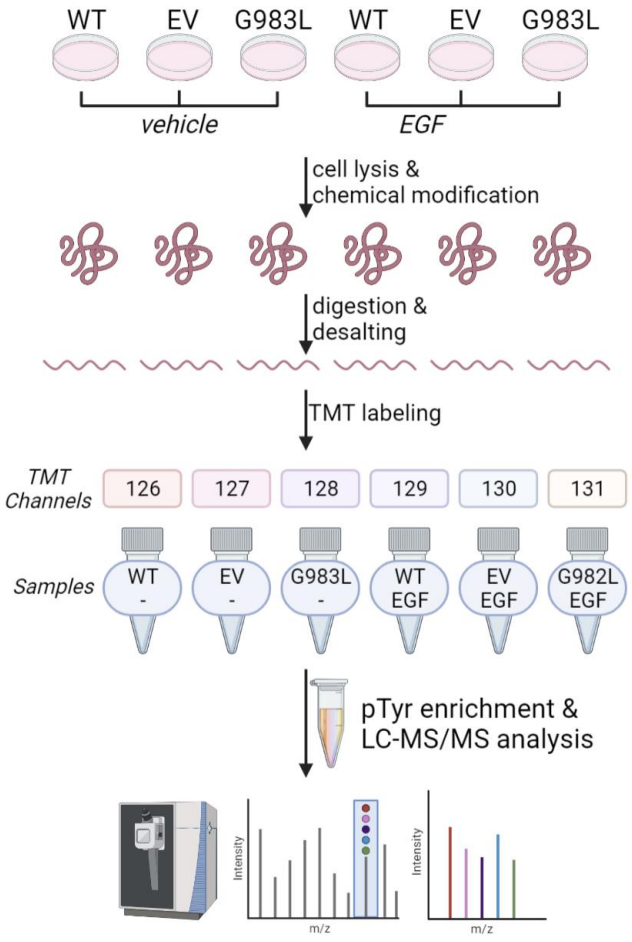


Figure 27: Workflow schematic of SCC2 cell stimulation and sample processing, TMT labeling and phosphotyrosine enrichment steps for LC-MS/MS analysis.

Initial evaluation of these data showed downregulation of canonical members of the EGFR network as compared in the mutant cell line compared to WT, including for MAPK1/3 and

several EGFR peptides (Fig 3, left), which could not be explained by differences in basal signaling levels (Fig 3, right). However, an increase in signaling was also observed when comparing empty vector to wild type, which could not immediately be explained.

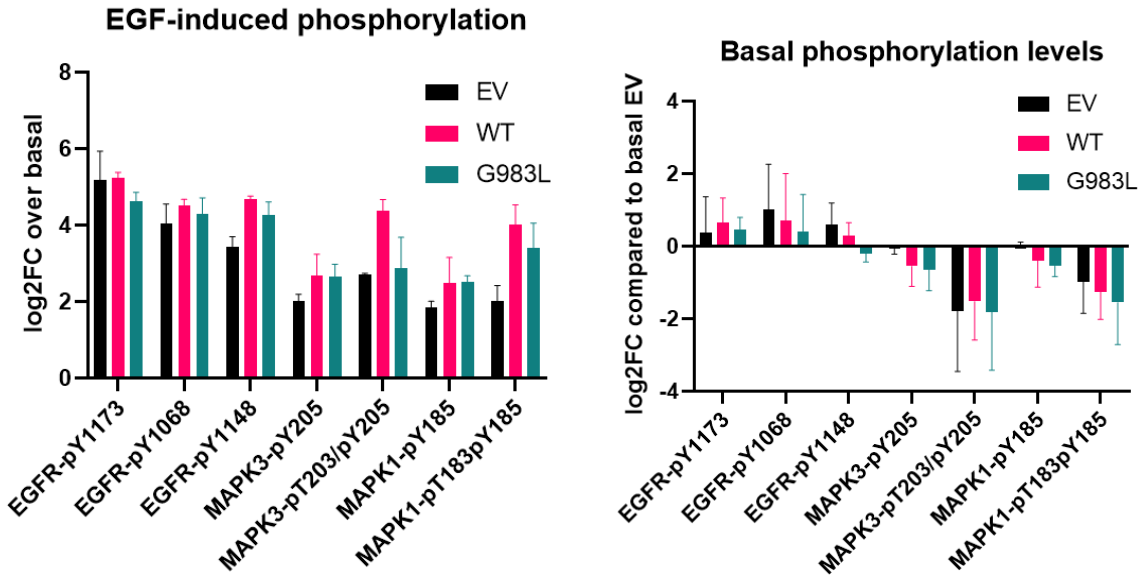


Figure 28: EGFR and MAPK phosphorylation in response to EGF.

When evaluating the overall network response, many EGFR network nodes were upregulated, including ERK1/2, SHC and PTK2.

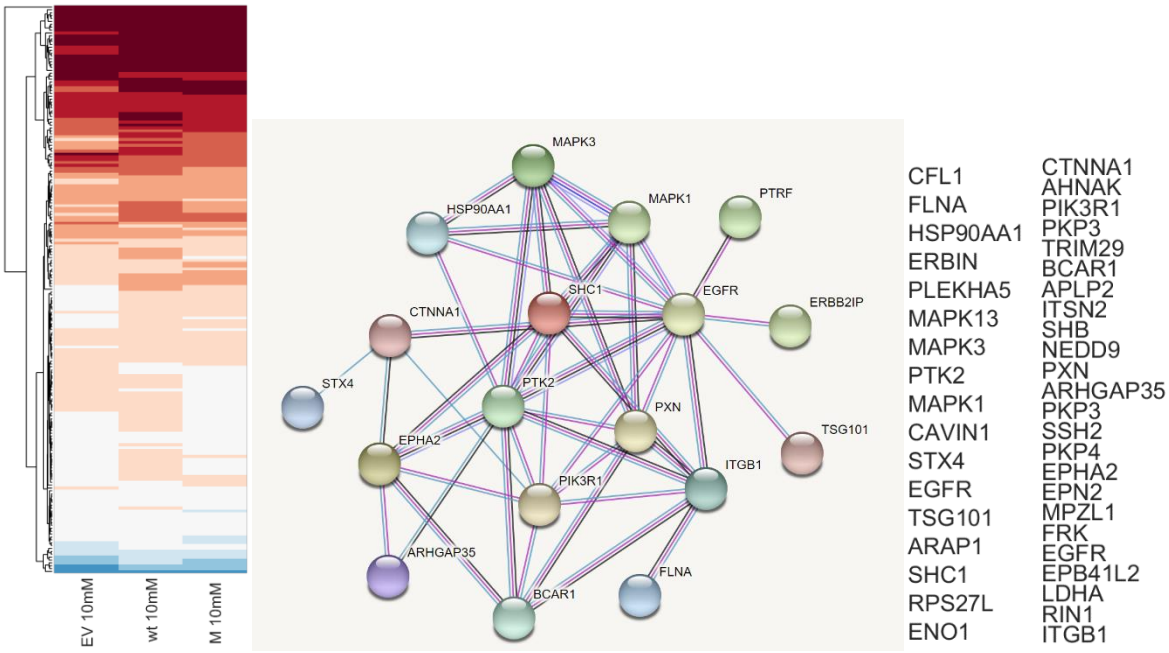


Figure 29: Upregulated network in response to EGF. Left: Hierarchical Clustering presenting data as log2foldchange compared to unstimulated condition, n=3. Middle: STRING analysis of upregulated pathway nodes. Right: List of input nodes for STRING analysis from highly phosphorylated top cluster from heatmap.

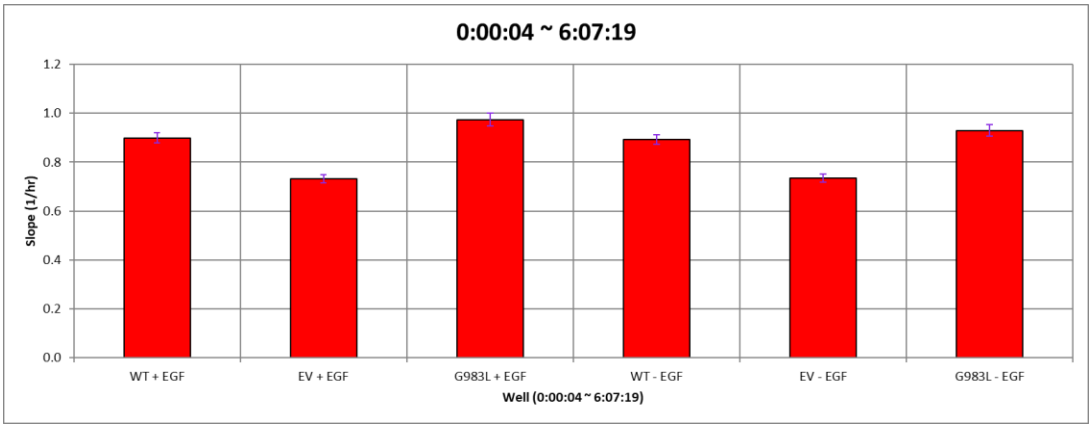
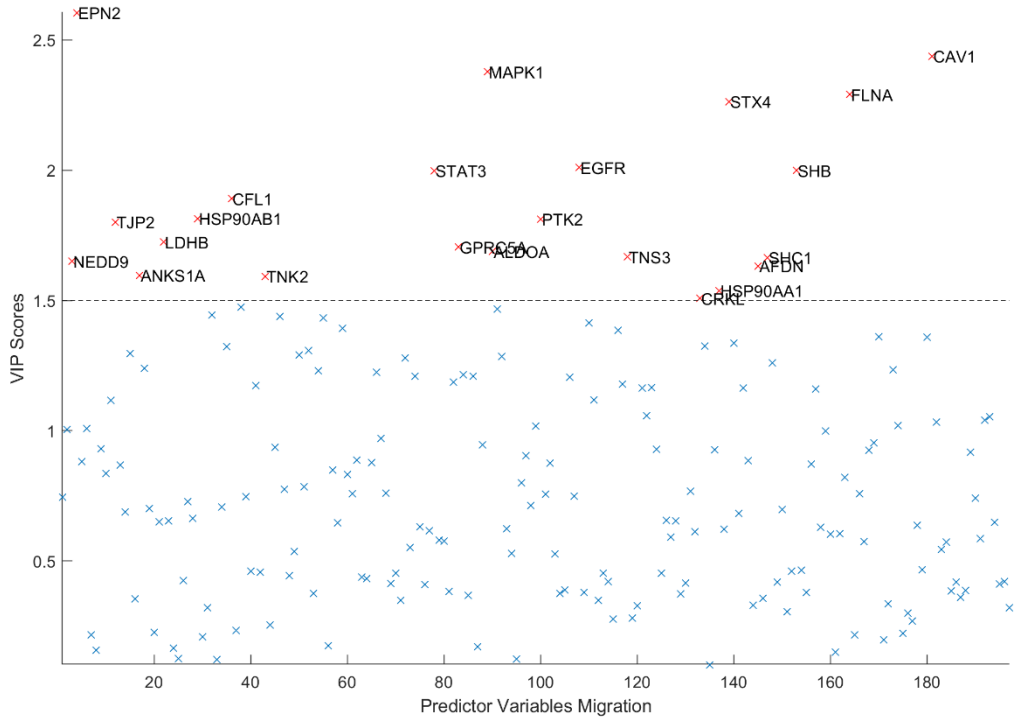


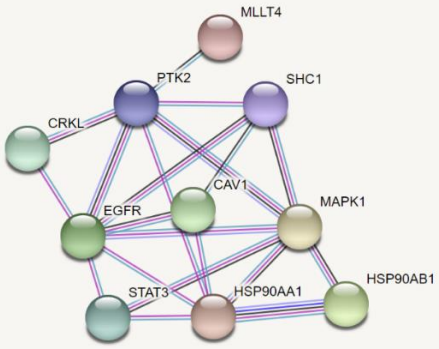
Figure 30: Migration data as collected by XCelligence assay. Data presented are migration rates calculated from wound closure rates.

Having observed decreased EGFR network signaling in the PTPRJ mutant cell line, we aimed to evaluate whether this effect would translate phenotypically as well. Using the XCelligence to measure wound closure by measuring changes in resistance due to increased cell density, we

did not see a significant decrease of migration in the mutant as compared to wt. If anything, the migration rate had slightly increased compared to the wildtype line, and these experiments will be repeated for validation.

Despite the counterintuitive results from our migration assay, we performed Partial Least Squares Regression (PLSR) analysis in an attempt to build a predictive model linking our signaling and phenotypic data together. As can be observed from our VIP score calculations in Figure 6 (top), several nodes not often associated with EGFR signaling were found to be predictive of migration. When analyzed using STRING, CAV1, STAT3 and MAPK1 were the linking nodes that connected these to EGFR activation.





Molecular Function (Gene Ontology)	
description	*false discovery rate*
Epidermal growth factor binding	0.0040
Nitric-oxide synthase regulator activity	0.0070
TPR domain binding	0.0106
MHC class II protein complex binding	0.0293
DNA polymerase binding	0.0348
Phosphotyrosine residue binding	0.00012
Protein tyrosine kinase binding	0.0015
Cadherin binding	2.97e-07
Phosphatase binding	0.0131

Figure 31: PLSR analysis of migration and signaling data. Top: VIP score plot with cutoff of 1.5 highlights signaling nodes predictive of migration response in our dataset. Bottom: STRING analysis highlights possible pathway connections these nodes have in common for hypothesis generation.

Discussion

The main goal of this work was to improve our mechanistic understanding of PTPRJ-mediated modulation of the EGFR network. We showed that inhibition of the dimer formation by instituting the G983L mutation led to decreased activation of the EGFR network. Furthermore, our analysis highlighted certain nodes less connected to the EGFR network that may provide hints to possible signaling axes that PTPRJ engages to affect EGFR downstream signaling.

Despite the richness of this dataset, more experimental validation and follow-up needs to be done to make meaningful claims regarding potential mechanisms of action. Finally, our data cannot provide any confirmation that targeting PTPRJ using a synthetic peptide will render sufficient or sustained downregulation of EGFR to justify it as a viable therapeutic strategy in the future. However, increased mechanistic insight may help provide alternative targeting strategies as well to modulate the signaling network in EGFR-overexpressing cancers.

Personal contributions to this work:

- Processed samples for LC-MS/MS analysis
- Performed LC-MS/MS analysis
- Data Analysis
- Writing

Materials and Methods for phosphoproteomic analysis

Sample processing

After vehicle or 10mM EGF treatment, cells were lysed on ice using 500µL 8M Urea (Sigma) per 10cm plate. A bicinchoninic acid (BCA) protein concentration assay (Pierce) was performed according to the manufacturer's protocol to estimate the protein concentration in each lysate. Cell lysates were reduced with 10mM DTT for 1hr at 56^{o}C, alkylated with 55 mM iodoacetamide for 1hr at RT shielded from light, and diluted 5-fold with 100 mM ammonium acetate, pH 8.9, before trypsin (Promega) was added (20:1 protein:enzyme ratio) for overnight digestion at RT. The resulting solutions were acidified with 1 mL of acetic acid (HOAc) and loaded onto C18 Sep-Pak Plus Cartridges (Waters), rinsed with 10mL of 0.1% HOAc, and eluted with 10mL of 40% Acetonitrile (MeCN)/ 0.1% HOAc. Peptides were divided into 200 microgram aliquots, and sample volume was reduced using a vacuum centrifuge (Thermo) and then lyophilized to dryness for storage at -80^{o}C.

TMT labeling for multiplexed analysis was performed according to manufacturer's protocol. Samples, each containing ~200ug peptides, were resuspended in 35µL HEPES (pH 8.5), vortexed, and spun down at 13,400rpm for 1 minute. 400µg of a given channel of TMT6plex (Thermo) in anhydrous MeCN, was added per sample. Samples were shaken at 400rpm for 1hr, after which the labeling reaction was quenched using 5% Hydroxylamine (50%, Thermo). After another 15 minutes on the shaker, all samples were combined using the same pipette tip to reduce sample loss, and sample aliquots were washed twice with 40 µL 25% MeCN/0.1% HOAc which was added to the collection tube to improve yield. Sample volume was reduced using a vacuum centifuge and then lyophilized to dryness for storage at -80^{o}C.

Phosphopeptide enrichment

Immunoprecipitation (IP) and IMAC were used sequentially to enrich samples for phosphotyrosine containing peptides. TMT-labeled samples were incubated in IP buffer consisting of 1% Nonidet P-40 with protein G agarose beads conjugated to 24 µg of 4G10 V312 IgG and 6 µg of PT-66 (P3300, Sigma) overnight at 4^{o}C. Peptides were eluted with 25 µl of 0.2% trifluoroacetic acid for 10 minutes at room temperature; this elution was performed twice to improve yield. Eluted peptides were subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)-based Fe-NTA spin column to reduce non-specific, non-phosphorylated peptide background. High-Select Fe-NTA enrichment kit (Pierce) was used according to manufacturer's instructions with following modifications. Eluted peptides from IP were incubated with Fe-NTA beads containing 25µL binding washing buffer for 30 minutes. Peptides were eluted twice with 20mL of elution buffer into a 1.7 mL microcentrifuge tube. Eluates were concentrated in speed-vac until ~1 µL of sample remained, and then resuspended in 10µL of 5% acetonitrile in 0.1% formic acid. Samples were loaded directly onto an in-house constructed fused silica capillary column [50 micron inner diameter (ID) x 10 cm] packed with 5µm C18 beads (YMC gel, ODS-AQ, AQ12S05) and with an integrated electrospray ionization tip (~2 micron tip ID).

LC-MS/MS analysis

LC-MS/MS of pTyr peptides were carried out on an Agilent 1260 LC coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptides were separated using a 140-minute gradient with 70% acetonitrile in 0.2 mol/L acetic acid at flow rate of 0.2 mL/minute with

approximate split flow of 20 nL/minute. The mass spectrometer was operated in data-dependent acquisition with following settings for MS1 scans: m/z range: 350 to 2,000; resolution: 60,000; AGC target: 3 x 10^6; maximum injection time (maxIT): 50 ms. The top 15 abundant ions were isolated and fragmented by higher energy collision dissociation with following settings: resolution: 60,000; AGC target: 1x10^5; maxIT: 350 ms; isolation width: 0.4 m/z, collisional energy (CE): 33%, dynamic exclusion: 20 seconds. Crude peptide analysis was performed on a Q Exactive Plus mass spectrometer to correct for small variation in peptide loadings for each of the TMT channels. Approximately 30 ng of the supernatant from pTyr IP was loaded onto an in-house packed precolumn (100µm ID x 10 cm) packed with 10mm C18 beads (YMC gel, ODS-A, AA12S11) and analyzed with a 70-minute LC gradient. MS1 scans were per-formed at following settings: m/z range: 350 to 2,000; resolution:70,000; AGC target: 3x10^6; maxIT: 50 ms. The top 10 abundant ions were isolated and fragmented with CE of 33% at a resolution of 35,000.

Peptide identification/quantification

Mass spectra were processed with Proteome Discoverer version 2.5 (Thermo Fisher Scientific) and searched against the human SwissProt database using Mascot version 2.4 (MatrixScience, RRID:SCR_014322). MS/MS spectra were searched with mass tolerance of 10 ppm for precursor ions and 20 mmu for fragment ions. Cysteine carbamidomethylation, TMT-labeled lysine, and TMT-labeled peptide N-termini were set as fixed modifications. Oxidation of methionine and phosphorylation of serine, threonine and tyrosine were searched as dynamic modifications. TMT reporter quantification was extracted and isotope corrected in Proteome Discoverer.

Peptide spectrum matches (PSM) were filtered according to following para-meters: rank=1, mascot ion score>15, isolation interference<40%, average TMT signal>1,000. Peptides with missing values across any channel were filtered out.

Phenotypic measurements

Migration Wound scratch assay Migration rates of PTPRJ mutant and wildtype expressing cells were determined using the XCelligence assay in the presence or absence of EGF.

Data analysis

Data analyses were performed in MATLAB R2020A, Microsoft Excel 2016 and Graphpad Prism 9. TMT reporter ion intensities from PSMs were summed for each unique phosphopeptide. For protein level quantification, TMT reporter intensities were summed for all unique peptides. Peptide or protein quantification were normalized with relative median values obtained from crude lysate analysis to adjust for sample loading in TMT channels. A combination of Student t-test and One-way Anova was used to perform statistical analysis between conditions. Statistical significance was assigned for $p<0.05$. Unsupervised hierarchical clustering was performed on the basis of Person correlation distance metric, unless otherwise specified. Protein networks were obtained from STRING (version 11.0) database.

References

1 Bixby JL. Ligands and signaling through receptor-type tyrosine phosphatases. IUBMB Life. 2001;51:157–163.

2 Barr AJ, Ugochukwu E, Lee WH, et al. Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. Cell. 2009;136:352–363.

3 Tonks NK. Protein tyrosine phosphatases - From housekeeping enzymes to master regulators of signal transduction. FEBS J. 2013;280:346–378.

4 Tonks NK. Protein tyrosine phosphatases: From genes, to function, to disease. Nat Rev Mol Cell Biol. 2006;7:833–846.

5 Nunes-Xavier CE, Martín-Pérez J, Elson A, et al. Protein tyrosine phosphatases as novel targets in breast cancer therapy. Biochim Biophys Acta - Rev Cancer. 2013;1836:211–226.

6 Östman A, Hellberg C, Böhmer FD. Protein-tyrosine phosphatases and cancer. Nat Rev Cancer Internet. 2006 cited 2022 Mar 2;6:307–320. Available from: https://pubmed.ncbi.nlm.nih.gov/16557282/.

7 Tarcic G, Boguslavsky SK, Wakim J, et al. An Unbiased Screen Identifies DEP-1 Tumor Suppressor as a Phosphatase Controlling EGFR Endocytosis. Curr Biol. 2009;19:1788–1798.

8 Xu Y, Tan LJ, Grachtchouk V, et al. Receptor-type protein-tyrosine phosphatase-κ regulates epidermal growth factor receptor function. J Biol Chem. 2005;280:42694–42700.

9 Bilwes AM, Den Hertog J, Hunter T, et al. Structural basis for inhibition of receptor protein-tyrosine phosphatase-α by dimerization. Nature. 1996;382:555–559.

10 Bloch E, Sikorski EL, Pontoriero D, et al. Disrupting the transmembrane domain-mediated oligomerization of protein tyrosine phosphatase receptor J inhibits EGFR-driven cancer cell phenotypes. J Biol Chem Internet. 2019 cited 2022 Mar 23;294:18796–18806. Available from: https://pubmed.ncbi.nlm.nih.gov/31676686/.

11 Bennasroune A, Fickova M, Gardin A, et al. Transmembrane peptides as inhibitors of ErbB receptor signaling. Mol Biol Cell. 2004;15:3464–3474.

CHAPTER 8

RNA-binding protein ELAVL4/HuD ameliorates Alzheimer's disease-related molecular changes in human iPSC-derived neurons

Robert J. van der Linden, M.Sc.

Jacqueline Gerritsen, M.Sc.

Meichen Liao, Ph.D.

Joanna Widomska, M.Sc.

Richard Pearse, Ph.D.

Forest M. White, Ph.D.

Barbara Franke, Ph.D.

Tracy L. Young-Pearse, Ph.D.

Geert Poelmans, M.D, Ph.D.

Submitted to Molecular Neurodegeneration, October 2021

Abstract

Background: The RNA-binding protein ELAVL4/HuD has been implicated in dementia since it was identified as a target of autoantibodies in paraneoplastic encephalomyelitis. Knock-out (KO) and overexpression studies in mice have revealed that ELALV4 regulates learning and memory, and ELALV4 expression was found to be decreased in the hippocampus of patients with Alzheimer's disease (AD), the most common form of dementia. Furthermore, ELAVL4 regulates the translation and splicing of multiple AD candidate genes. Therefore, ELAVL4may be a promising target for AD drug development.

Methods: We generated ELAVL4 KO human induced pluripotent stem cell-derived neurons in both familial AD (fAD) and fAD-corrected human neurons to study the effect that ELAVL4has on AD-related cellular phenotypes. To gain further insight into the molecular cascades involved in ELAVL4 signaling in neurons, we also conducted pathway and upstream regulator analyses of transcriptomic and proteomic data from these neurons.

Results: ELAVL4KO significantly increased the levels of specific APP isoforms and intracellular phosphorylated tau, molecular changes that are related to the pathological hallmarks of AD. In contrast, overexpression of ELAVL4in wild-type neurons and rescue experiments in ELAVL4KO cells showed opposite effects—decreased levels of specific APP isoforms and intracellular phosphorylated tau—and also led to a reduction of the extracellular amyloid-beta (Aβ)42/40 ratio. All of these observations were made in both familial AD (fAD) and fAD-corrected neurons. Further, analyses of the generated transcriptomic and proteomic data revealed that ELAVL4 affects multiple biological pathways linked to AD, including those involved in synaptic function, as well as gene expression downstream of APP and tau signaling. These analyses also suggest that ELAVL4 expression is regulated by insulin receptor-FOXO1 signaling in neurons.

Conclusions: ELAVL4expression ameliorates AD-related molecular changes in neurons and affects multiple synaptic pathways, making it a promising target for novel drug development.

Background

ELAVL4 (other name: HuD) is a member of the Hu/ELAV-like family of RNA-binding proteins (RBPs)that is predominantly expressed in neurons and, at lower levels, in the pancreas and testis (1). In addition to ELAVL4, the Hu family consists of two RBPs that are enriched for expression in neurons—i.e., HuB (ELAVL2) and HuC (ELAVL3)—and the ubiquitously expressed HuR (ELAVL1) (2). By post-transcriptionally processing (pre-)mRNAs, RBPs add an additional layer to gene regulation and produce a more diverse assortment of mRNAs and proteins (3). Hu proteins contain three RNA recognition motifs (RRMs) that directly interact with AU-rich elements (RRM1andRRM2) and poly-A tails (RRM3) of specific target mRNAs (4, 5). Through protein-protein interactions mediated by RRM3 and the hinge region that separates RRM2 and RRM3, Hu proteins can recruit additional proteins to mRNA-protein complexes(2, 6, 7). By directly binding mRNAs, Hu proteins are involved in multiple aspects of posttranscriptional gene regulation, including mRNA polyadenylation, alternative splicing, trafficking, turnover, and translation (8).

As for human diseases associated with ELAVL4/HuD, links have been established with both dementia and type 2 diabetes mellitus (T2DM). As for dementia, it was already reported in 1986 that high titers of ELAVL4 autoantibodies had been found in the sera from a subset of patients with small-cell lung tumors who also developed a paraneoplastic syndrome characterized by dementia and other neurological disabilities (9, 10). In this respect, it was demonstrated that the autoantibodies produced against ELAVL4 ectopically expressed by lung tumor cells induced neuronal loss and inflammatory infiltrates in the brain, leading to the observed neurological symptoms (3, 9, 10).In addition, mice with a prenatal ELAVL4 knock-out

showed deficits in hippocampus-dependent learning and memory (11), while it was also shown that upon learning and memory tasks, ELAVL4 expression is upregulated in the hippocampus of mice and rats (12, 13).However, mice that constitutively overexpress ELAVL4 in their forebrain also have impaired hippocampus-dependent learning and memory (14).At the molecular level,ELAVL4 has been shown to strengthen synapses of individual dendritic branches by stabilizing the mRNA and increasing the expression of proteins with important synaptic functions, such as CAMK2Aand BDNF(15, 16). Additionally, ELAVL4 stabilizes and upregulates the expression of GAP43, another protein with a key role in hippocampal neuronal plasticity(17). Based on all these findings, it was concluded that ELAVL4 functions as an essential regulator of synaptic plasticity during learning and memory tasks in mature neurons, likely having roles in both neurodevelopment and in the adult brain (14, 18, 19).

The most common form of dementia is the progressive neurodegenerative disorder Alzheimer's disease (AD) (20). The two hallmark neuropathological lesions that are found in AD patients' brains are extracellular neuritic plaques consisting of amyloid-β (Aβ)—which is generated through the processing of Amyloid Precursor Protein (APP)—and intracellular neurofibrillary tangles containing excessively phosphorylated tau protein(21). As for Aβ, the longer form consisting of 42 amino acids (Aβ42) is more prone to aggregate into plaques than the most common form with40 amino acids (Aβ40), and familial AD mutations either raise the Aβ42 levels or lower the Aβ40 levels, each of which leads to a higher Aβ42/40 ratio. Therefore, an increased Aβ42/40 ratio is considered to be indicative of more neurotoxic Aβ (22-25). In addition to being involved in regulating hippocampal memory-related processes in general, decreased expression of neuronal Hu proteins (including ELAVL4) has been reported in the

hippocampus of AD patients (26, 27). Furthermore, ELAVL4 has been shown to bind and stabilize many mRNAs encoding proteins that have been specifically implicated in AD. As such, ELAVL4 binds and stabilizes the mRNAs—leading to increased protein expression—of both APP and BACE1, a β-secretase that is involved in amyloidogenic APP to Aβ processing (28). However, ELAVL4can also promote the non-amyloidogenic processing of APP by stabilizing the mRNA and upregulating the expression of the α-secretase ADAM10 (29). Neuronally expressed Hu proteins—including ELAVL4—have also been reported to regulate the alternative splicing of APP mRNA(30). Moreover, ELAVL4 has been shown to bind and stabilize the mRNAs of neprilysin, an Aβ degrading enzyme (31), and neuroserpin, an extracellular enzyme that can both negatively and positively regulate Aβ clearance (32, 33). In addition to Aβ-related targets, ELAVL4 can directly bind, stabilize and upregulate the expression of tau (mRNA) in neuronal cells (34). Further, ELAVL4 stabilizes the mRNA and hence upregulates the expression of acetylcholinesterase (ACHE) (35), which provides another link with AD as ACHE inhibitors are currently the most often used drugs for symptomatic treatment of the disease (36). Lastly, a recent co-expression network analysis of RNAseq data generated from human postmortem brain samples identifiedELAVL4 as one of ten 'hub genes' in AD-related synaptic pathways (37). Information on how ELAVL4expression and function is regulated is scarce. Some evidence comes from work on its contribution to T2DM. Firstly, silencing of ELAVL4 expression increases apoptosis of pancreatic beta cells—that are functionally impaired in diabetes—implying that ELAVL4 is an important regulator of beta cell function and survival (38). Together with the finding that ELAVL4 expression is markedly reduced in beta cells of diabetic patients (39), this indicates that ELAVL4 deficiency contributes to diabetes etiology. Furthermore, it was

demonstrated in beta cells that the expression of ELAVL4 is regulated by insulin receptor-FOXO1 signaling (40), which is the only molecular mechanism for ELAVL4 expression regulation described thus far.

In this study, we determined the effects that ELAVL4knock-out and overexpression have on AD-related phenotypes—i.e., APP (isoform) levels, Aβ42/40 ratio, and (phosphorylated)tau levels—in human induced pluripotent stem cell-derived glutamatergic cortical neurons. Subsequently, to gain insight into the molecular cascades underlyingELAVL4 signaling in human neurons, we analyzed RNA sequencing and proteomic data from neurons derived from the different cell lines generated in this study.

Results

Generation of ELAVL4/HuD knock out, overexpression and rescue induced neurons

We used CRISPR-Cas9 to generate two monoclonal *ELAVL4* KO induced pluripotent stem cell (iPSC) lines. Both lines were derived from an early onset AD patient carrying the *APP* London mutation (41). One line harbored the APPV717I mutation (to be referred to as 'fAD'), while in the second, isogenic line, the mutation was corrected to wild type (WT) *APP* (referred to as fAD-corrected or 'fADcorr') (42). We sequenced the genomic region around the binding site of the sgRNA and selected the fADcorr *ELAVL4* KO clone B1c2, which has 27 nucleotide and 46 nucleotide deletions, and the fAD *ELAVL4* KO B1c20 containing a 123 nucleotide deletion with a 23 nucleotide insertion and a one nucleotide deletion, for “induced neuron” (iN) differentiation (Figure 1B). Monoclonal unedited CRISPR control lines ('CRC') that had been subjected to the same process were also obtained. Common CNV analysis in the monoclonal lines using the Nanostring nCounter Human CNV codeset indicated that the editing process did not result in large chromosomal abnormalities (Supplementary Figure 1). In iPSCs, ELAVL4 expression is low and not readily detectable by either qPCR or Western blotting. Hence, to confirm loss of ELAVL4 protein levels in the knock out, we transduced the iPSCs with lentivirus encoding NGN2 to produce iNs. The used iN protocol yields a homogeneous population of cells with neuronal morphologies and transcriptional profiles consistent with layer 2/3 excitatory projection neurons within two weeks (43, 50). On iN differentiation DIV21, RNA and protein were collected and subjected to qPCR and Western blotting using ELAVL4-specific primers and antibodies.

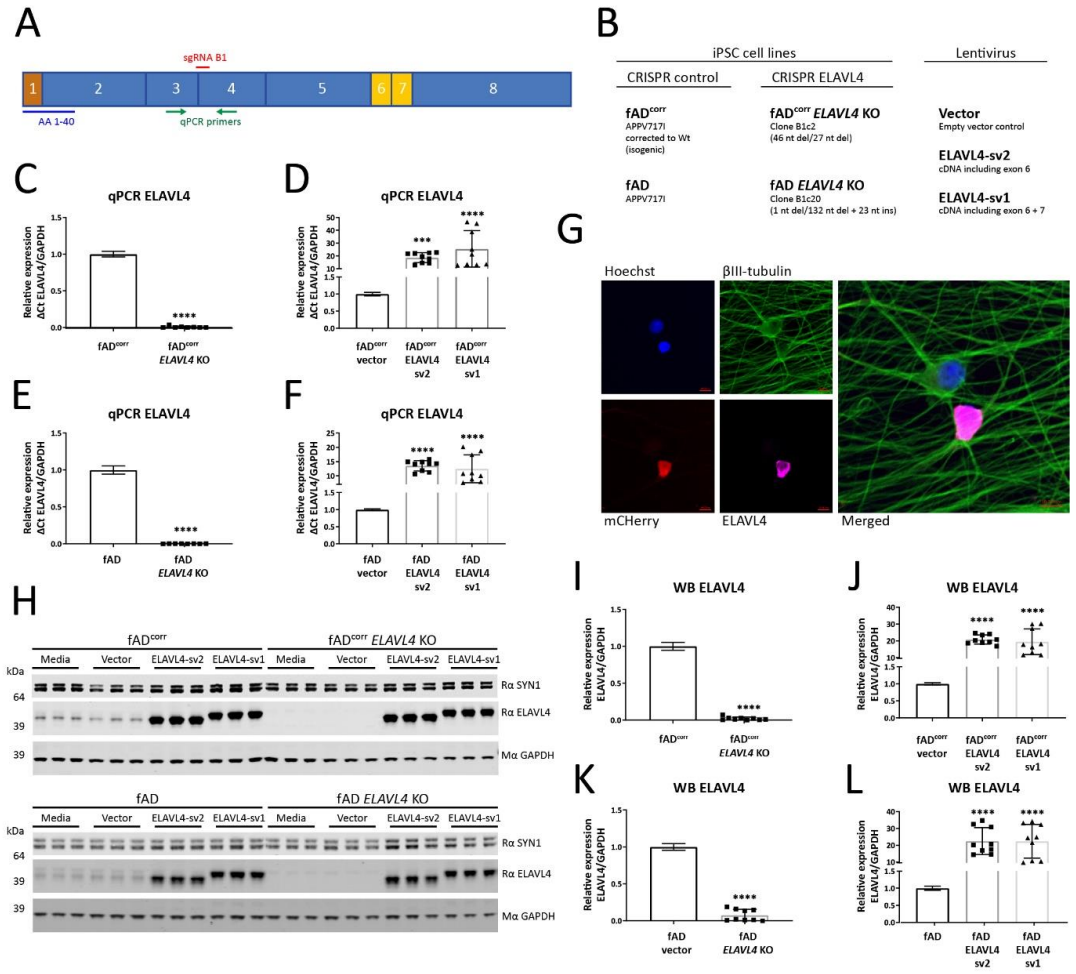


Figure 32: Validation of *ELAVL4* knock out (KO), over expression (OE), and rescue in iPSCs and iNs. (A) Interaction sites for the primary antibody, qPCR primers, and sgRNA for ELAVL4 used in this study projected on a schematic representation of ELAVL4. (B) Overview of the cell lines and lentiviral conditions generated/used in this study. (C) Full KO of ELAVL4 mRNA in fADcorr iNs ($P < 0.0001$, ttest). (D) 18.6- and 25.5-fold increase in ELAVL4 mRNA upon ELAVL4 sv2 and ELAVL4 sv1 OE expression in fADcorr iNs respectively ($P < 0.001$ and $P < 0.0001$, Dunnett's following a significant ANOVA). (E) Full KO of ELAVL4 mRNA in fAD iNs ($P<0.0001$, t-test). (F) 13.5- and 12.4-fold increase in ELAVL4 mRNA upon ELAVL4 sv2 and ELAVL4 sv1 OE expression in fADcorr iNs respectively ($P <$ 0.0001, Dunnett's following a significant ANOVA). (G) Representative image of immunocytochemistry of ELAVL4 rescue with ELAVL4-sv1 in fADcorr *ELAVL4* KO iNs, stained with the nuclear DNA stain Hoechst, the neuronal marker βIII-tubulin and ELAVL4, mCHerry signal from the overexpression plasmid, scale bars: 10 µm. (H) Example western blots are shown for fADcorr iNs and fAD iNs, the pre-synaptic marker SYN1 is not significantly affected by manipulating the expression levels of ELAVL4. (I) Full KO of ELAVL4 protein in fADcorr iNs ($P < 0.0001$, t-test). (J) 20.8- and 19.5-fold increase in ELAVL4 protein upon ELAVL4 sv2 and ELAVL4 sv1 OE expression in fADcorr iNs respectively ($P < 0.0001$, Dunnett's following a significant ANOVA). (K) Full KO of ELAVL4 protein in fAD iNs ($P<0.0001$, t-test). (L) 22.4- and 22.3-fold increase in ELAVL4 mRNA upon ELAVL4 sv2 and ELAVL4 sv1 OE expression in fADcorr iNs respectively ($P < 0.0001$, Dunnett's following a significant *ANOVA).*

Full KO of both mRNA and protein was observed in fADcorr *ELAVL4* KO (clone B1c2) and fAD *ELAVL4* KO (clone B1c20) (Figures 1C, 1H, and 1I and Figures 1E, 1H, and 1K, respectively). Overexpression (OE) and rescue of ELAVL4 expression in the *ELAVL4* KO lines was performed by lentiviral transduction of ELAVL4 splice variant (sv) 1 (ENST00000371823.8) or sv2 (ENST00000371824.7) cDNA on DIV6 and resulted in a sustained 10- to 20-fold increased mRNA levels on DIV21 (Figures 1D,1F). Western blotting showed that sv2 was the most predominantly expressed splice variant in DIV21 iNs (Figure 1H). Protein expression of the presynaptic neuronal marker SYN1 was not affected by ELAVL4 KO, OE, or rescue in iNs (Figure 1H).

ELAVL4 expression affects APP phenotypes in induced neurons

After we successfully manipulated ELAVL4 expression in iNs, we examined the effect of these manipulations on APP mRNA, protein, and processing. Alternative splicing of APP is known to result in three major isoforms: APP751 and APP770, which are ubiquitously expressed, and the predominantly neuronally expressed APP695 isoform (51).

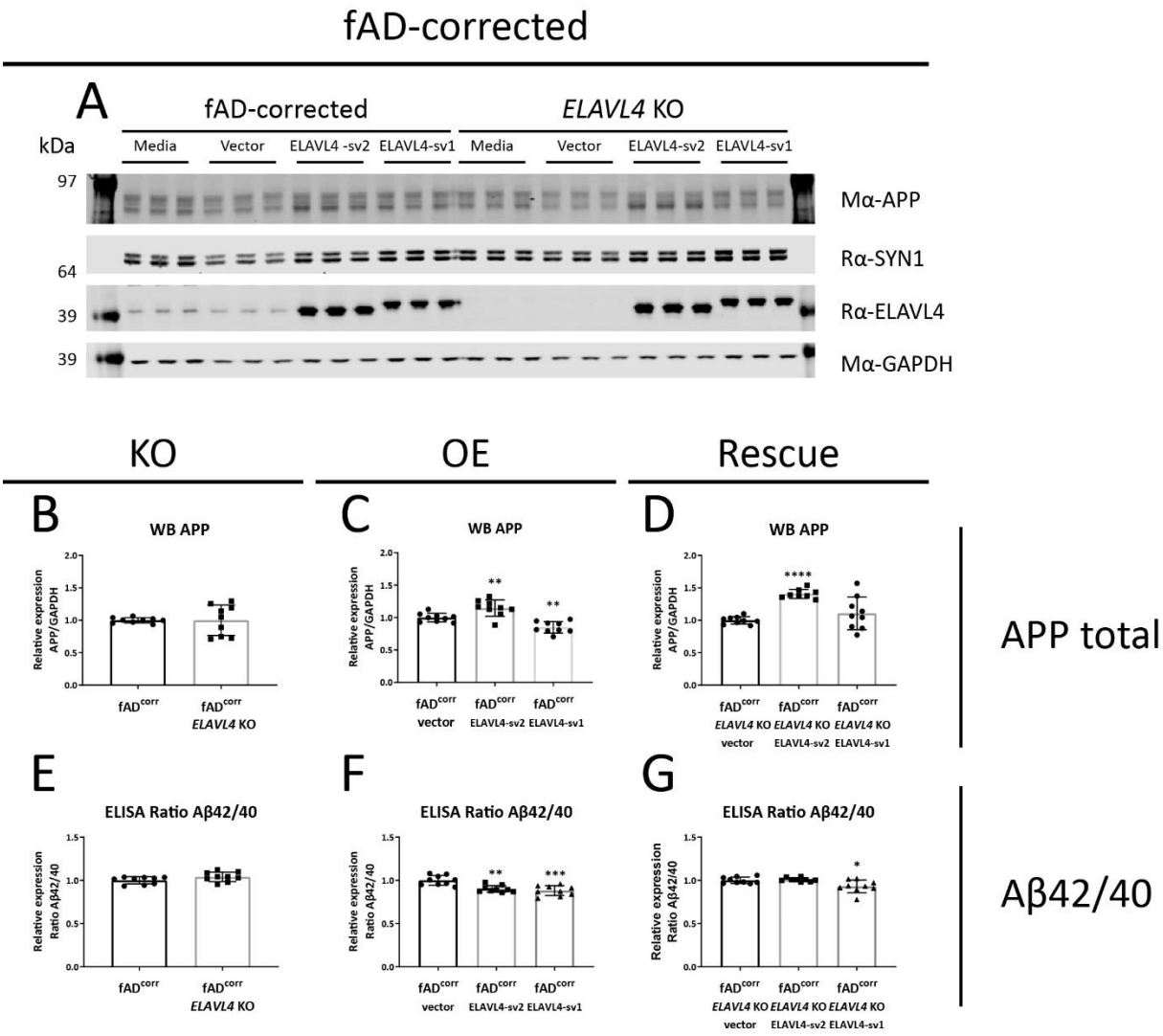


Figure 33: ELAVL4 affects APP protein levels and Aβ42/40 ratios in fADcorr iNs.
(A) Example western blot and (B-D) quantifications are shown for *ELAVL4* knock out (KO) overexpression (OE) and rescue of ELAVL4 in fAD-corrected iNs. In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where *ELAVL4* was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, 48h condition media was collected and cells were lysed. (E-G) Aβ40 and Aβ42 levels were measured via multiplexed ELISA (MSD), and normalized to the fAD-corrected CRISPR control for each differentiation. Quantifications (B-G) are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, P-values were calculated with Dunnett's multiple comparison that followed after a significant ANOVA test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

In both the fADcorr and fAD cell lines, loss-of-function of ELAVL4 resulted in the increased formation of longer APP splice variants (increased APP751 and APP770 isoforms) (Table 1, Supplementary Figures 2 and 3). OE and rescue of ELAVL4 expression in both *ELAVL4* KO lines had the opposite effect: increased APP695 mRNA and reduced APP751 and APP770 mRNA (Table 1, Supplementary Figures 2 and 3). In the OE experiments, we observed an increase in total APP mRNA while in the KO and rescue experiment, total APP mRNA levels were not significantly affected (Table 1, Supplemental Figures 2 and 3). At the protein level, we observed lower APP expression after KO of *EALVL4* in fAD neurons and increased APP protein expression in neurons with rescued or overexpressed ELAVL4 (Table 1, Figures 2A-D and 3A-D). We previously had found, and confirmed here, that fAD neurons express a higher Aβ42/40 ratio (41). When we measured Aβ species that result from APP processing, we found that *ELAVL4* KO lowered Aβ38 and Aβ40 levels in the culture media from fADcorr iNs and Aβ38, Aβ40, and Aβ42 levels in media of fAD iNs (Supplementary Figures 4 and 5). On the other hand, while OE of ELAVL4 sv2 increased the extracellular levels of Aβ38 and Aβ40, all other OE and rescue conditions did not affect the levels of the Aβ species measured (Supplementary Figures 4 and 5). However, the ratio of Aβ42/40 was consistently reduced in the ELAVL4 OE and rescue experiments (Table 1, Figures 2E-G and 3E-G).

ELAVL4 expression affects tau phenotypes in induced neurons

We further evaluated the effect of ELAVL4 manipulation on AD-relevant phenotypes by examining tau, a protein that aggregates in the other pathological hallmark of AD: neurofibrillary tangles. The levels of tau mRNA increase and decrease in *ELAVL4* KO iNs in fADcorr and fAD backgrounds, respectively (Table 1, Figures 4A and 5A). Overexpression of

ELAVL4 sv1 increased tau mRNA in both backgrounds, and rescue with ELAVL4 sv1 increased tau mRNA in the fAD *ELAVL4* KO iNs (Table 1, Figures 4B, C and 5B, C). The elevated phosphorylation of tau protein (pTau) facilitates intracellular tangle formation.

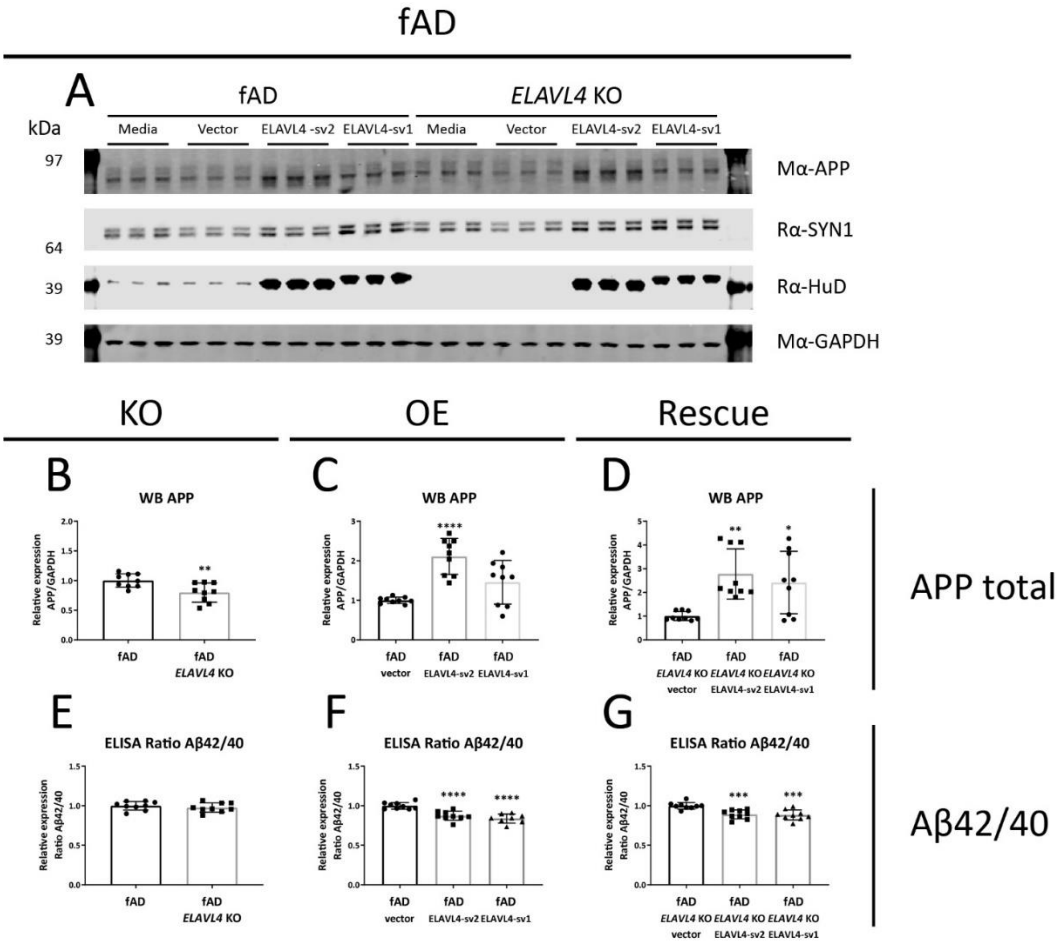


Figure 34: ELAVL4 affects APP protein levels and Aβ42/40 ratios in fAD iNs.
(A) Example western blot and (B-D) quantifications are shown for *ELAVL4* knock out (KO) overexpression (OE) and rescue of ELAVL4 in fAD iNs. In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where *ELAVL4* was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, 48h condition media was collected and cells were lysed. (E-G) Aβ40 and Aβ42 levels were measured via multiplexed ELISA (MSD), and normalized to the fAD-corrected CRISPR control for each differentiation. Quantifications (B-G) are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, P-values were calculated with Dunnett's multiple comparison that followed after a significant ANOVA test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

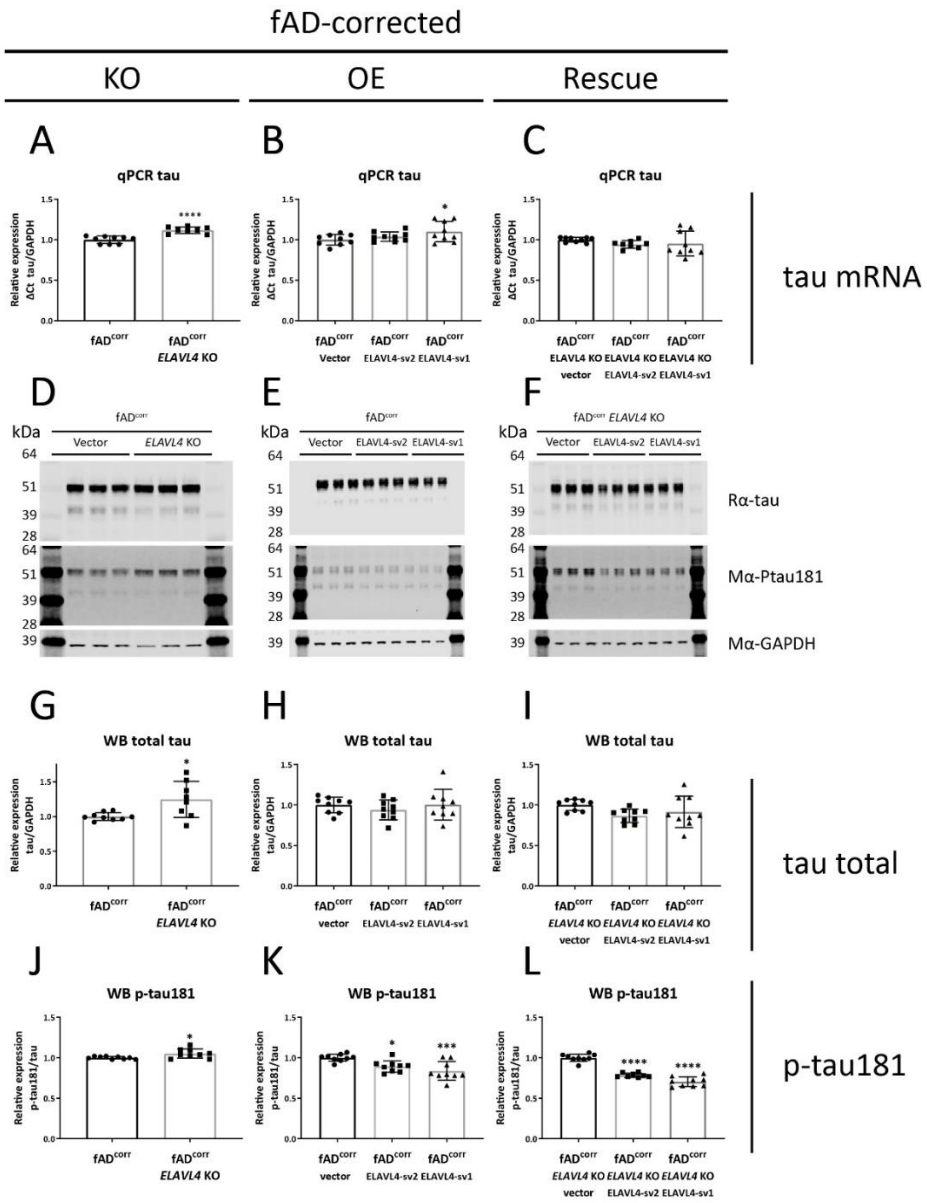


Figure 35: Increased ELAVL4 expression reduces tau phosphorylation in fAD-corrected iNs.
(A-C) Quantification of tau mRNA in iNs from *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD-corrected induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where *ELAVL4* was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or ELAVL4 KO iNs respectively. At day 21 of differentiation, cells were lysed and protein and RNA samples were harvested. (D-F) example blots and (G-L) quantifications tau and p-tau181 protein in ELAVL4 KO, OE, and rescue fAD-corrected iNs. Quantifications (A-C,G-L) are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, P-values were calculated with Dunnett's multiple comparison that followed after a significant ANOVA test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

We previously had found, and confirmed here, that fAD neurons express higher pTau protein levels (41). *ELAVL4* KO, OE and rescue of ELAVL4 did not significantly alter total tau protein levels in most cells, with the exception of increased tau protein in the fADcorr *ELAVL4* KO iNs and ELAVL4 sv2 OE in fAD iNs (Table 1, Figures 4G-I and 5G-I). However, *ELAVL4* KO did consistently increase tau phosphorylation, specifically phosphorylation on residue threonine 181 (Thr181) in both the fADcorr and fAD cell lines (Table 1, Figures 4J and 5J). On the other hand, increasing ELAVL4 levels through OE and rescue experiments had the opposite effect in both fADcorr and fAD iNs, reducing phosphorylated Thr181 tau compared to total tau (Table 1, Figures 4K,L and 5K,L).

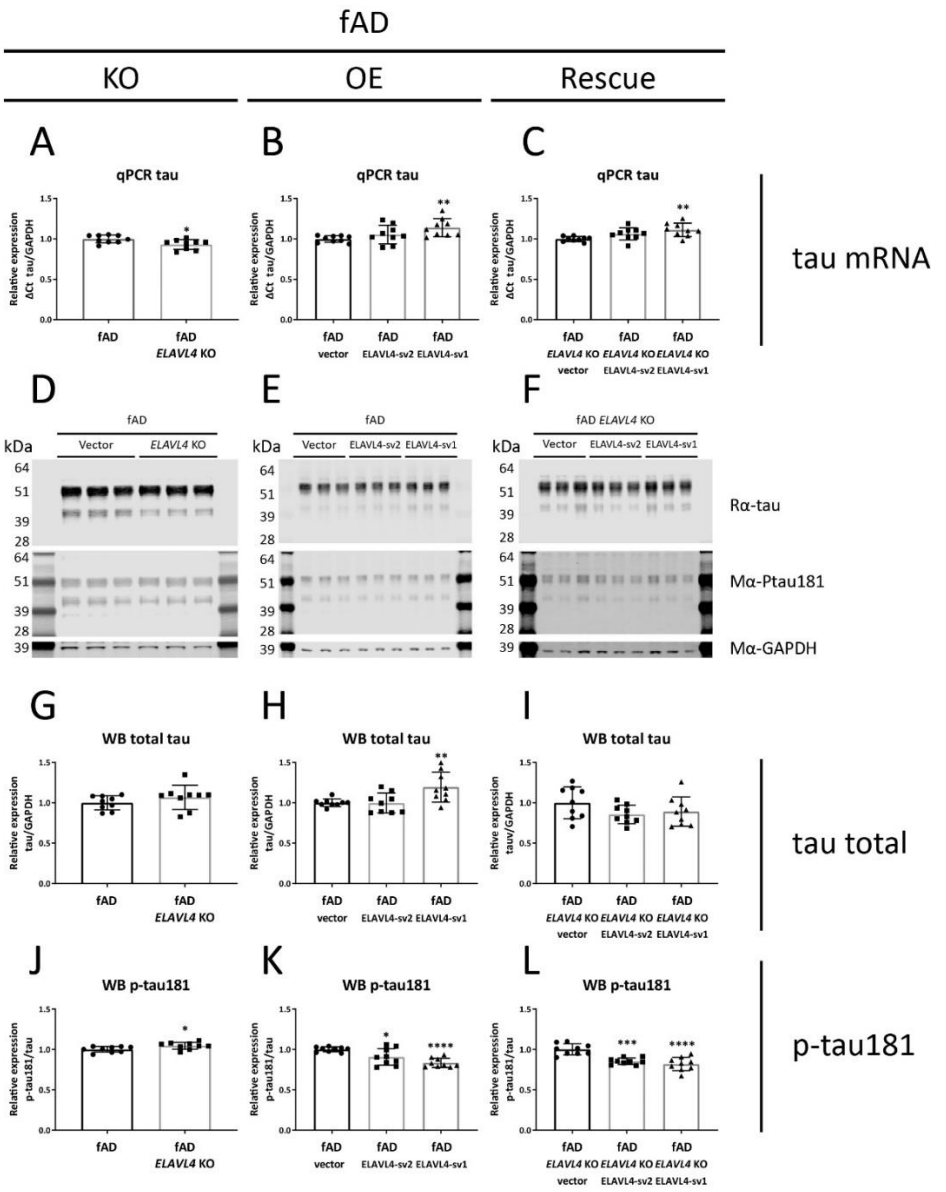


Figure 36: Increased ELAVL4 expression reduces tau phosphorylation in fAD iNs.
(A-C) Quantification of tau mRNA in iNs from *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where *ELAVL4* was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, cells were lysed and protein and RNA samples were harvested. (D-F) example blots and (G-L) quantifications tau and p-tau181 protein in ELAVL4 KO, OE, and rescue fAD iNs. Quantifications (A-C,G-L) are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, Pvalues were calculated with Dunnett's multiple comparison that followed after a significant ANOVA *test, * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.*

Canonical pathway and upstream regulator analysis results

After confirming that decreasing or increasing ELAVL4 expression affects AD-related molecular changes in iNs, we aimed to gain further insight into the molecular cascades involved in ELAVL4 signaling through conducting analyses of transcriptomic and proteomic data from iNs derived from the different cell lines that we had generated. For further analysis, we defined 10 sets of differentially expressed genes (DEGs) generated through comparing total RNA sequencing data from 48 samples of iNs that included 3 biological replicates for each of 16 conditions (Figure 6A, Table 2). First, we analyzed the DEGs in fADcorr *ELAVL4* KO iNs compared to fADcorr iNs with WT *ELAVL4* (Table 2.1, Figure 6B), the DEGs in fADcorr iNs in which ELAVL4 was overexpressed (combined analysis of ELAVL4 sv1 and sv2 OE) compared to fADcorr iNs with WT *ELAVL4* (Table 2.2, Figure 6C), and the genes that were differentially expressed in the opposite direction (DEGs-OD) in the fADcorr ELAVL4 sv1+sv2 OE iNs and the fADcorr *ELAVL4* KO (Table 2.3). Then, we analyzed the DEGs from the rescue experiments: the DEGs in fADcorr *ELAVL4* KO iNs with ELAVL4 sv1+sv2 rescue compared to fADcorr *ELAVL4* KO iNs (Table 2.4, Figure 6D), as well as the DEGs-OD in the fADcorr ELAVL4 sv1+sv2 rescue iNs and the fADcorr *ELAVL4* KO iNs (Table 2.5). Subsequently, we performed the same five KO, OE, and rescue analyses based on RNAseq data from fAD iNs (Table 2.6-2.10, Supplementary Figure 8A-C). In addition to RNAseq data, we analyzed four sets of differentially expressed proteins (DEPs) and differentially expressed phosphoproteins (DPPs): DEPs and DPPs in fAD *ELAVL4* KO iNs compared to fAD iNs (Table 2.11-12, Figure 7A,B), and DEPs and DPPs in fAD *ELAVL4* KO iNs with ELAVL4 sv2 rescue compared to fAD *ELAVL4* KO iNs (Table 2.13-14, Figure 7C,D).

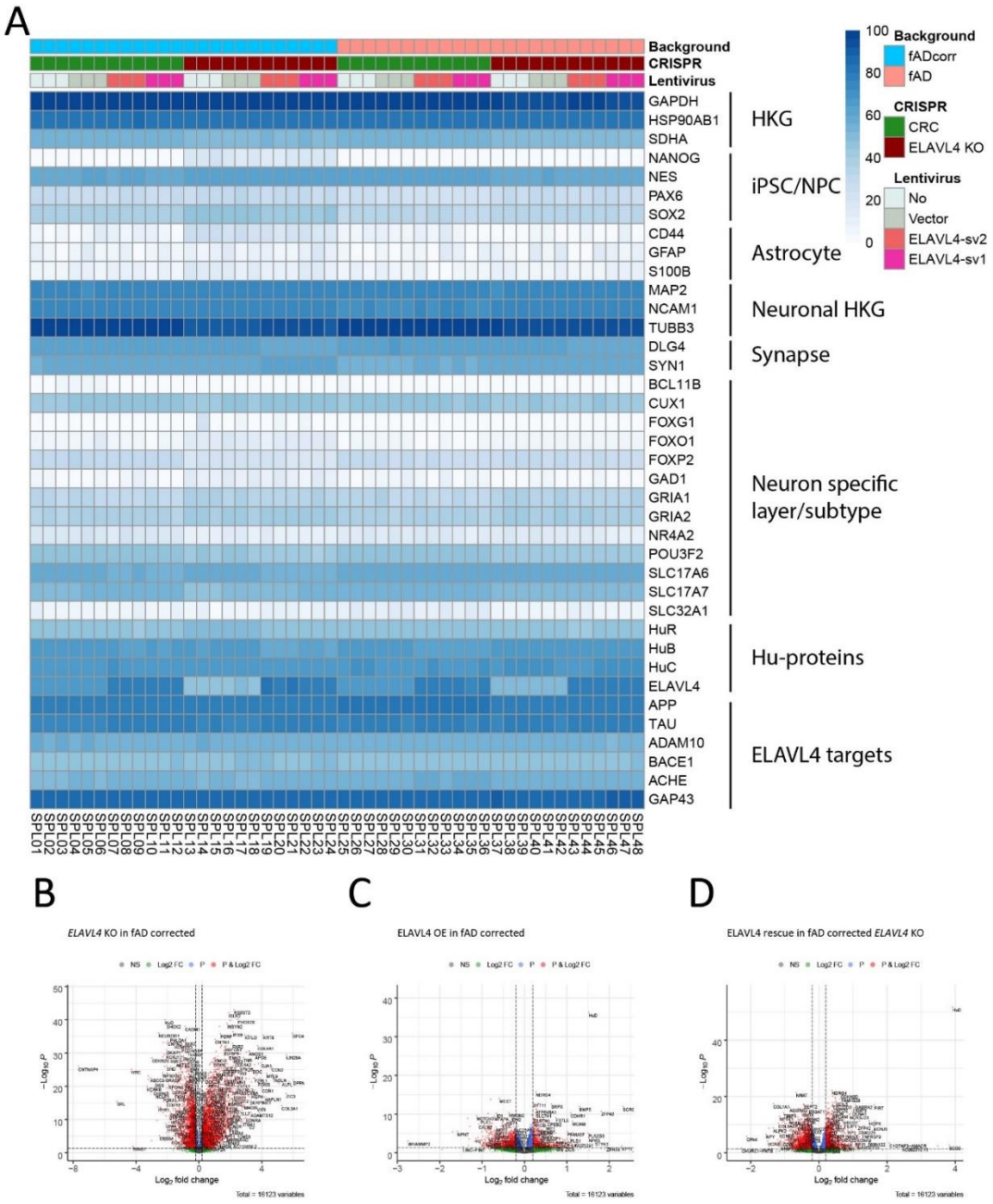


Figure 37: RNAseq characterization of iPSC lines generated in this study.
(A) Purified RNA of cell lysates collected from iNs DIV21 were analyzed by RNAseq. Genes were selected that mark subsets of neuronal and glial cells and a heat map was created using the pheatmap package in R. HKG, housekeeping gene iPSC/NPC, induced pluripotent stem cell/neural progenitor cell. (B) Differentially expressed genes (DEGs) in fADcorr *ELAVL4* KO iNs compared to fADcorr iNs (C) DEGs in fADcorr iNs in which ELAVL4 was overexpressed (combined analysis of ELAVL4 sv1 and sv2 OE) compared to fADcorr iNs. (D) DEGs from the rescue experiments: the DEGs in fADcorr *ELAVL4* KO iNs with ELAVL4 sv1+sv2 rescue compared to fADcorr *ELAVL4* KO iNs. DEGs with FDR corrected P-values < 0.05 and fold changes ≥ |1.20| are indicated in red. Volcano DEGs plots were generated using the EnhancedVolcano package in R.

The data sets generated provide a rich resource for interrogating ELAVL4/HuD biology. Supplementary Table 1 includes the results of each of the above-described analyses, which can be readily probed for any gene or protein of interest. For example, as mentioned in the introduction, several genes have been described in the literature to be targets of ELAVL4, and in Table 2, we have listed these genes, with their expression changes in human neurons with ELAVL4 knock out or overexpression.

In order to obtain a global view of the pathways that are altered following ELAVL4 modulation, we used IPA to conduct canonical pathway analyses of the above-described lists of DEGs, DEPs, and DPPs. When considering the top five most significantly enriched pathways for each comparison, 40 different pathways were enriched in the DEGs, DEPs, and DPPs (Table 3; full canonical pathway analysis results in Supplementary Table 2). Canonical pathways that were found to be among the top five enriched pathways in more than one comparison are ("X"=number of comparisons showing enrichment): "Axonal guidance signaling" (6X), "Cell Cycle: G2/M DNA Damage Checkpoint Regulation (4X)", "Cyclins and Cell Cycle Regulation" (4X), "Kinetochore Metaphase Signaling Pathway" (4X), "Hepatic Fibrosis / Hepatic Stellate Cell Activation" (3X), "Hepatic Fibrosis Signaling Pathway" (3X), "Human Embryonic Stem Cell Pluripotency" (3X), "Molecular Mechanisms of Cancer" (3X), "14-3-3-mediated Signaling" (2X), "Calcium Signaling" (2X), "Mitotic Roles of Polo-Like Kinase" (2X), "Sirtuin Signaling Pathway" (2X), and "Synaptogenesis Signaling Pathway" (2X) (Table 3).

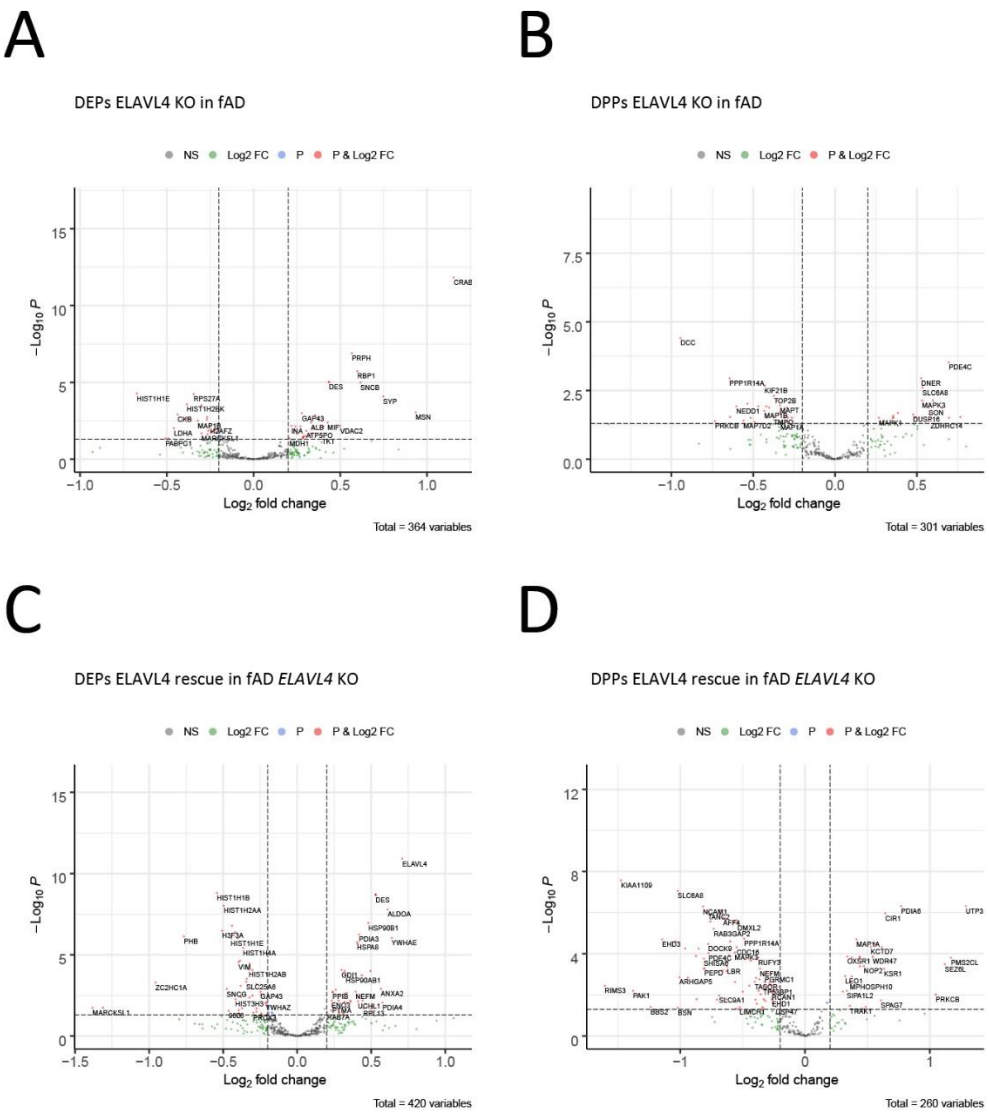


Figure 38: Proteomic characterization of iPSC lines generated in this study.
(A,C) Protein and (B,D) phosphoprotein quantification of DIV21 iNs from (A,B) *ELAVL4* KO and (C,D) rescue in fAD induced neurons. Differentially expresses peptides (DEPs) and differentially phosporylated peptides (DPPs) with FDR corrected P-values < 0.05 and fold changes ≥ |1.20| are indicated in red. Volcano DEPs and DPPs plots were generated using the EnhancedVolcano package in R.

We then performed focused 'upstream regulator' analyses to identify regulators that influence the expression of multiple downstream targets, and we have used similar analyses previously to find upstream regulators for movement disorders (52-54). We focused on assessing if altering ELAVL4 expression in neurons affects genes and proteins that are downstream transcriptional

targets of four regulators. These regulators were APP and tau, the key components of the AD brain lesions, as well as the insulin receptor (INSR) and FOXO1, the only known transcriptional regulators of ELAVL4 expression (a mechanism that was demonstrated in pancreatic beta cells (40)). The analyses revealed that, with the exception of analysis 8 (DEGs-OD in the fAD ELAVl4 sv1+sv2 OE iNs versus the fAD *ELAVL4* KO), all analyzed DEGs, DEPs and DPPs are significantly enriched for transcriptional targets downstream of APP signaling (Table 4). In analysis 1 (DEGs in the fADcorr *ELAVL4* KO iNs and the fADcorr iNs), APP was also predicted to be activated (Z-score 2,36) (Table 4). In addition, half of the analyses (7/14) revealed a significant enrichment of transcriptional targets downstream of tau signaling (Table 4). Further, the upstream regulator analysis of the DEGs in both the fADcorr *ELAVL4* KO and fAD *ELAVL4* KO iNs (analyses 1 and 6) revealed that the INSR downstream targets are enriched and the INSR is predicted to be inhibited (Z-scores -3,14 and -3,15). In contrast, an enrichment of INSR targets and predicted activation of the INSR (Z-score 2,21) was identified in analysis 11 (DEPs in the *ELAVL4* KO in fAD iNs) (Table 4). In 7 of the remaining 11 analyses, a significant enrichment INSR downstream targets was found. As for FOXO1, its downstream transcriptional targets were enriched and it was predicted to be inhibited in 3 OE analyses (analyses 2,3, and 7) and 3 rescue analyses (4,5 and 9), while FOXO1 target enrichment was identified in 3 of the remaining 8 analyses (Table 4).

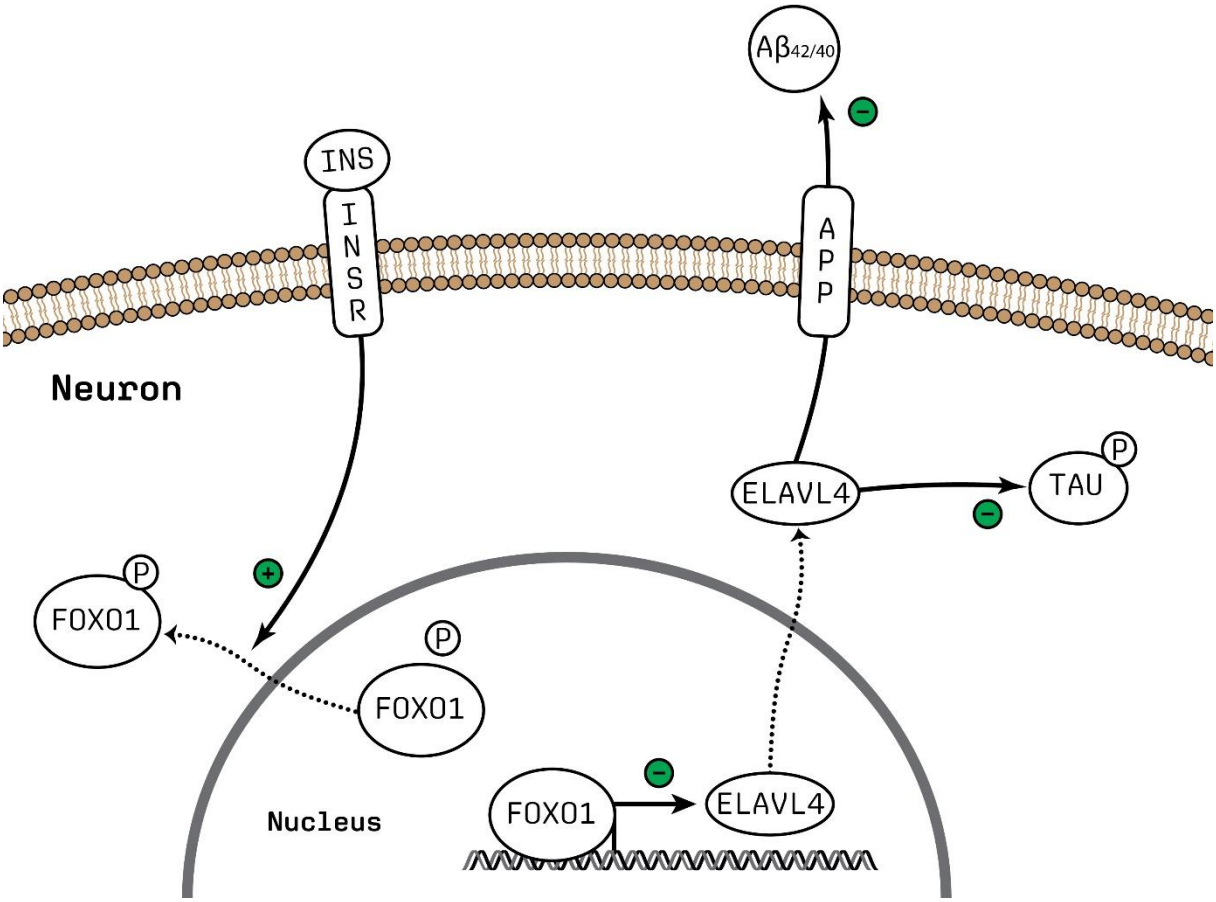


Figure 39: Schematic representation of INSR-FOXO1 signaling putatively regulating ELAVL4 expression and the downstream effects of ELAVL4 on APP processing and tau phosphorylation. The figure was made using Adobe Illustrator. Abbreviations: Aβ; amyloid beta, P; phosphate group.

Discussion

In this study, we showed that manipulating ELAVL4/HuD expression in iPSC-derived induced neurons (iNs) affects AD-relevant phenotypes in human neuronal cells. We found that increasing ELAVL4 expression in WT and *ELAVL4* KO iNs ameliorated AD-associated molecular changes—i.e., reduced levels of specific APP isoforms, decreased Aβ42/40 ratio, and reduced levels of phosphorylated tau— from both familial AD (fAD) and fAD-corrected iPSCs. For APP splicing isoforms and phosphorylated tau levels, the opposite effect was observed in *ELAVL4* KO cells.

Three main alternatively spliced isoforms of APP exist, i.e., APP751 and APP770 that are ubiquitously expressed and APP695, a predominantly neuronally expressed isoform (51). In both fAD-corrected and fAD iNs, knocking out *ELAVL4* resulted in altered splicing of APP mRNA (increased APP751 and APP770), while overexpression and rescue of ELAVL4 expression in *ELAVL4* KO cells had the opposite effect (decreased APP751 and APP770) and also led to an increase of APP695 levels. These findings are in line with the literature, where ELAVL4 was reported to regulate APP splicing, and a strong positive correlation between Hu protein expression and APP695 levels was found in the brain (30). Specifically, ELAVL4 was found to modulate APP splicing by favoring exclusion of exons 7 and 8 that are coding for the longer APP isoforms (APP751 and APP770) (30). In AD, a decrease of APP695 and an increase of the APP751/APP770 isoforms have been reported (55, 56), which suggests that ELALV4 function is impaired in AD. Further, except for most of the OE experiments—in which we observed an increase in total APP mRNA—we found that total APP mRNA levels were not significantly

affected by manipulating ELAVL4 expression. At the protein level, increased APP expression was seen in most experiments with ELAVL4 OE or rescued iNs, but lower APP expression was observed following *ELAVL4* KO in fAD iNs. This being said, when we examined APP processing, we found that the Aβ42/40 ratio was reduced in almost all (7/8) of the ELAVL4 OE and rescue experiments, with higher Aβ42/40 ratios having been correlated to more severe AD pathology (22, 57). Again, these findings are in line with ELAVL4 function being reduced in AD.

According to the Aβ cascade hypothesis, pathogenic Aβ oligomers trigger a signaling cascade that results in an increase in phosphorylated tau, synaptic dysfunction, and neurodegeneration in the brain of AD patients (58). In this respect, increased CSF and plasma levels of tau phosphorylated at threonine 181 (Thr181) are used to predict elevated amounts of cerebral Aβ in the clinic (59, 60). In keeping with this, our OE and rescue experiments indeed showed that upon upregulating ELAVL4 expression, not only the Aβ42/40 ratio but also the levels of intracellular tau phosphorylated at Thr181 were markedly decreased. Conversely, *ELAVL4* KO also resulted in increased phosphorylated tau (at Thr181) levels. Interestingly, it was recently also found in iPSC-derived cerebral organoids carrying frontotemporal dementia-associated mutations in *MAPT*, ELAVL4 is increased in expression but mislocalized together with tau in stress granules, which results in impaired ELAVL4 function and altered synaptic signaling pathways (61).

Taken together, our experimental findings provide new, additional evidence *in human neurons* for an important role of ELAVL4 in modulating AD pathogenesis. Specifically, increased expression and hence the activity of ELAVL4 in iNs seems to counteract the two main

pathological hallmarks of AD, i.e., the formation of extracellular Aβ plaques (resulting from an increased Aβ42/40 ratio) and intracellular neurofibrillary tangles (resulting from an increase in Thr181-phosphorylated tau levels). Therefore, we submit that *ELAVL4* KO iNs represent a novel cellular model of dysregulation of tau and Aβ that can be probed to further elucidate mechanisms of AD. In particular, our findings in rescued iNs—i.e., upon upregulation of ELAVL4 levels in *ELAVL4* KO cells—indicate that novel approaches aimed at upregulating ELAVL4 expression levels may be beneficial therapeutically.

The database of RNA profiles (RNAseq) and protein levels ((phospho)proteomics) resulting from ELAVL4 modulation that we created provides a rich resource for us and others studying ELAVL4/HuD function to gain insights into the molecular events that result from manipulating ELAVL4 expression (KO, OE, or rescue) in human neurons. Analyzing these data, our results for three already established ELAVL4 targets *NEP*, *Tau*, and *ACHE* follow a pattern that is consistent with previous findings, i.e., that ELAVL4 stabilizes the mRNA of these genes. For the other known targets listed in Table 2, this pattern is less consistent, which may be due to the fact that cell models different from human iNs or animal models and/or different experimental conditions were used to demonstrate that these genes were ELAVL4 targets.

Our unbiased canonical pathway analysis identified multiple pathways that are significantly enriched among the DEGs, DEPs, and DPPs linked to differential ELAVL4 expression. These included pathways that have been linked to AD previously. Specifically, we identified several enriched synaptic pathways, i.e., "axonal guidance signaling", "semaphorin signaling in neurons", and "synaptogenesis signaling". Axonal guidance is generally associated with neurodevelopment. Indeed, Hu proteins—that include HuD/ELAVL4—are thought to be one of

the earliest markers of neuronal differentiation, and in dentate granule cells from ELAVL4 OE mice, increased expression of axonogenesis-related genes has been observed (62). However, axonal guidance molecules can also stimulate or inhibit inflammatory responses and play a pivotal role in the inflammation of the nervous system associated with AD (63, 64). A second set of enriched pathways consisted of DNA damage pathways, including "UVB/C-induced MAPK signaling", "ATM signaling", and "p53 signaling". Recently, we have shown that several synaptic pathways are impaired in the hippocampus of AD patients by analyzing transcriptomic data and have also linked these findings to somatic DNA damage (65). In this respect, an impaired p53-mediated DNA damage response has been observed in AD patient brains (66). Furthermore, aberrant re-entry into the cell cycle has been implicated in the pathogenesis of AD (67, 68). This fits with our finding of an enrichment of pathways related to cell cycle regulation ("Cyclins and cell cycle regulation", "Cell cycle: G2/M damage checkpoint regulation", "Cell cycle: G1/S checkpoint regulation", and "CDK5 signaling"). DNA damage and other stressors may also activate a senescence-like phenotype in neurons leading to inflammation and neuronal dysfunction (69), and therefore, it is interesting that oxidative stress has been shown to induce posttranscriptional gene regulation by RBPs—including Hu protein family member HuR—of mRNAs involved in regulating cellular senescence (70). Other significantly enriched pathways relate to NAD and sirtuin signaling, and intriguingly, these pathways have been linked to both AD and aging/longevity (71, 72) through regulating the activity of FOXO1 (73, 74), one of the upstream regulators for which we performed enrichment analyses. Lastly, some pathways that we identified are directly linked to phosphorylation of tau, i.e., BAG2 signaling (75), 14-3-3-mediated signaling (76, 77), and ERK/MAPK signaling (78).

Our upstream regulator analyses of the transcriptomic and proteomic data from most experiments that we performed—knocking out and/or overexpressing ELAVL4 in iNs—revealed that the differentially expressed genes/proteins are strongly enriched for transcriptional targets of both APP and tau. These findings suggest that ELAVL4 not only affects APP processing and tau phosphorylation but also signaling downstream from these two crucial AD-linked proteins. Furthermore, the results from our upstream regulator analyses of the transcriptomic data indicate that ELAVL4 expression in neurons may be regulated by insulin receptor (INSR)-FOXO1 signaling: in the *ELAVL4* KO iNs, INSR is predicted to be inhibited, whereas FOXO1 is predicted to be inhibited in ELAVL4 OE/rescue cells. This is the same mechanism that regulates ELAVL4 expression in pancreatic beta cells (40). More specifically, upon binding insulin, the INSR is activated, and this leads to a downstream signaling cascade that results in phosphorylation of FOXO1, a transcription factor that represses ELAVL4 transcription. Phosphorylated FOXO1 translocates from the nucleus and loses its ability to regulate transcription, resulting in increased transcription and expression of ELAVL4 (40). Subsequently, as we demonstrated, increased ELAVL4 expression results in a decreased Aβ42/40 ratio and phosphorylated tau levels, thereby ameliorating the key AD-related molecular changes. In Figure 8, we have schematically represented the directionality of INSR-FOXO1 signaling putatively regulating ELAVL4 expression, with its subsequent effects on APP processing and tau phosphorylation.

As mentioned above, ELAVL4 is (also) an important regulator of pancreatic beta-cell function (38), and ELAVL4 deficiency contributes to the etiology of diabetes type 2 (39). In this respect, it is interesting that insulin resistance (in the periphery and brain) has been implicated as a core mechanism underlying both diabetes and AD (79). In the brain, insulin resistance was found to

have a negative effect on synaptic plasticity, which in turn could contribute to AD progression (80). Diabetes and AD are often comorbid, and Aβ peptides have been found to compete directly with insulin for INSR binding, leading to decreased INSR functioning and ultimately insulin resistance (81). Therefore, we could speculate that in AD, aggregated Aβ causes decreased INSR activity and subsequently decreased FOXO1 phosphorylation and ELAVL4 expression, which further exacerbates amyloidogenic APP processing, and hence AD pathology.

This study has some strengths and limitations. A particular strength is that we were able to generate and validate ELAVL4 KO cells as a human cell model of AD that could be used in further studies aimed at further elucidating the exact role of ELAVL4 in the brain in general, and in AD in particular. A limitation is that the iN system we used does not model the full complexity of the human brain, as neurons in the brain are interconnected with other brain regions and reside among other cell types. Therefore, more complex (3D) model systems including not only neurons but also e.g., astrocytes and microglia may be generated to study ELAVL4 function in an experimental setting that more closely resembles a functioning brain. Further, it is unlikely that ELAVL4 function in the AD brain is completely reduced to zero in the way that we have modeled with full *ELAVL4* KO iNs. This being said, we also conducted OE and rescue experiments that showed effects opposite to those of the KO experiments, which strengthens the evidence that altered ELAVL4 expression leads to the observed phenotypic changes. Lastly, the putative ELAVL4 expression-regulating cascade in neurons that we identified through analyzing the RNAseq and proteomics data needs to be further studied and corroborated in follow-up experiments.

Conclusions

In conclusion, we show that increased ELAVL4/HuD expression ameliorates APP and tau pathology in neurons. In addition, our transcriptomic data analyses suggest that ELAVL4 affects signaling downstream of APP and tau and that INSR-FOXO1 signaling may regulate ELAVL4 expression in neurons. Taken together with the results from published studies, increasing ELAVL4 expression and hence functional protein levels in the brain may be beneficial in AD—especially through the positive effects that ELAVL4 has on synaptic function—making it a promising target for novel drug development.

Personal contributions to this work:

- Processed samples for LC-MS/MS analysis
- Performed LC-MS/MS analysis
- Partial LC-MS/MS data analysis
- Writing LC-MS/MS methods

Materials and Methods

Cell culture

Induced pluripotent stem cells (iPSCs) harboring the *APP* London mutation (V717I) (fAD) and isogenic CRISPR corrected lines (fADcorr) were reported previously (41, 42). iPSCs were maintained in Stemflex medium (ThermoFisher) on plates coated with growth factor reduced Matrigel (Corning). Generation of induced neurons (iNs) was performed as described previously (43). Briefly, iPSCs were plated at 95K/cm2 in mTeSR1 media one-day prior to viral transduction. Ultrahigh titer lentiviruses were obtained from Alstem for plasmids FUdeltaGW-rtTA (Addgene #19780), pTet-O-Ngn2-puro (Addgene #52047), and DRH-307 (Addgene #112670), and were transduced at MOIs 5, 2, and 2, respectively. On day in vitro (DIV) 1, differentiation was induced by Doxycycline (2µg/ml). On DIV 2, puromycin (5 µg/ml) was added to select for cells expressing Ngn2. On DIV 4, iNs were plated on Poly-L-Ornithine laminin coated plates and cultured in Neurobasal medium containing Glutamax, dextrose, NEAA, B27, Doxycycline, puromycin, BDNF, CNTF, and GDNF. Culturing media was conditioned starting at DIV 18, and media, protein, and RNA samples were harvested at DIV 21.

CRISPR-Cas9 knock out

single guide RNAs targeting exons common to all human *ELAVL4* splicing isoforms were designed using the Broad CRISPR design tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrnadesign) (figure 1A). The top three sgRNAs were cloned into the pXPR-003 (Addgene #52963) backbone and transfected with SpCas9 plasmid (Addgene #78166) into M17D cells using FuGENE HD (Promega) to assess

sgRNA knockdown (KD) efficiency by qPCR. The most efficient M17D KD of ELAVL4 was observed using sgRNA B1, which was subsequently co-transfected with SpCas9 into fAD and fADcorr iPSCs using Lipofectamine2000. Two days post-transfection, iPSCs were selected with puromycin (5 µg/ml) and blasticidin (4 µg/ml) for enrichment. Editing efficiency in the resulting polyclonal lines was assessed with the GeneArt Genomic Cleavage Detection Kit (ThermoFisher). Monoclonal cell lines were obtained by limiting dilution and examined by sequencing over the sgRNA target region. After monoclonal selection, the cellular karyotype was assessed using the NanoString nCounter CNV codeset and visualized using copy number package in R (44).

Lentiviral overexpression of ELAVL4

The EIF1α promoter of pLVX-EF1α-IRES-mCherry (Takara #631987) was switched to a shorter CAG promoter (Genewiz) by ClaI and NotI digestion for higher viral packaging efficiency and elevated expression in neurons (pLVX-CAG-IRES-mCherry). The two most common isoforms of ELAVL4 (splicing variants 1 or 2 (sv1 or sv2) in the postnatal brain arise from alternative splicing of ELAVL4 pre-mRNA in the region that codes for the hinge between RRM2 and RRM3 (2). cDNA for ELAVL4 sv1 (Origene #RC218612) and sv2 (Genscript #OHu28682D) was amplified by PCR using primers 5'-CTCTAGAGCCACCATGGTTATGA-3' (FW) and 5'-ATATGGATCCTCAGGACTTGTGGG-3' (REV). Amplified cDNA and pLVX-CAG-IRES mCherry vector were digested by XbaI and BamHI-HF and ligated using T4-ligase. The ligated product was transformed in stable competent E. coli (NEB #C3040H) and plated on ampicillin plates. For lentivirus production, 90% confluent HEK293T cells in 15 cm culture plates were

transfected with 24 µg pLVX-CAG-ELAVL4-IRES-mCherry, 12 µg envelope plasmid (Addgene #12259), and 18 µg packaging plasmid (Addgene #12260) using Lipofectamine2000. Media collected at 24H and 36H was pooled and concentrated with Lenti-X-Concentrator (Takara), virus particles were resuspended in PBS buffer.

Quantitative reverse transcription PCR

Samples for qPCR were prepared using a Fast SYBR Green Cells-to-CT Kit (ThermoFisher) according to manufacturer's guidelines, measured using three technical replicates on a ViiA 7 System (Applied Biosystems), and normalized to GAPDH values during comparative Ct analysis (45). Samples (n=3) from three replicate experiments were normalized to control conditions in the same experiment.

Western blotting

iN protein lysates were extracted using RIPA buffer (Abcam) containing complete protease inhibitor (Sigma) and phosphatase stop (Roche). Protein concentrations were determined by BCA (ThermoFisher), and samples were prepared with 4X sample buffer (LI-COR). Sample separation was performed using 4-12% Bis-Tris gels (ThermoFisher) in MOPS running buffer. Subsequently, samples were transferred onto nitrocellulose membranes. Membranes were blocked in blocking buffer (LI-COR) and probed overnight with primary antibodies at 4°C. Secondary anti-mouse and/or antirabbit antibodies (LI-COR), where appropriate, were incubated for 1 hour at room temperature (RT), and blots were scanned using the Odyssey CLx Imaging System (LI-COR). The primary antibodies used were: RαTau (DAKO, A0024),

MαPhospho tau Thr181 (Thermo Fisher, MN1050), RαSYN1 (EMDMillipore, 574777), MαAPP (Sigma-Aldrich, MAB348), RαELAVL4 (Proteintech, 24992-1-AP), and MαGAPDH (Proteintech, 60004-1-Ig). Samples (n=3) from three replicate experiments were normalized to control conditions in the same experiment.

Aβ-ELISA

A triplex electrochemiluminescence assay kit (K15200E-2, V-PLEX Aβ Peptide Panel 1 kit, (Meso Scale Diagnostics (MSD)) was used to simultaneously measure levels of Aβ38, Aβ40, and Aβ42 in conditioned media following manufacturer's instructions. Samples (n=3) from three replicate experiments were normalized to control conditions in the same experiment.

RNA sequencing

Total RNA of DIV21 iNs was purified using PureLink RNA mini kits (ThermoFisher). RNA-Seq Library preparation and Illumina HiSeq 2x150bp sequencing was performed by Genewiz, Inc. Trimmomatic was used for trimming the beginning and ending bases from each read and identifying and trimming adapter sequences from the reads (46). For quantifying gene level transcript abundances, we (pseudo)aligned the trimmed reads to a human reference genome (GRCh38) using Kallisto (47). Detection of differentially expressed genes (DEGs) was performed with Limma-voom using the count data (transcripts per million) obtained from Kallisto. *P*-values were corrected for multiple testing using the false discovery rate (FDR) method (48).

Immunostaining

DIV 21 iNs were fixed with 4% Paraformaldehyde solution at 4°C for 20 min at RT, blocked in blocking buffer (2% donkey serum, 0.1% Triton in PBS) for 1 hour at RT, and incubated with primary antibodies at 4°C overnight. The next day, secondary antibodies were incubated for 1 hour at RT followed by Hoechst DNA stain for 10 minutes at RT. Three 10 minutes PBS wash steps were performed between each step.

Phosphoproteomics and protein expression profiling

On DIV21, iN protein samples were collected in 4°C 1% SDS in H2O. Following cell lysis, samples were centrifuged at maximum speed (~21,000 x g) for 10 minutes at 4°C to pellet any cell debris. Protein concentrations were determined using a BCA assay (PierceTM). Proteins were reduced with 10 mM DTT for 1 hour at 56°C, then alkylated with 55 mM Iodoacetamide (IAA, Sigma) for 1 hour at RT on a rotator protected from light. SP3 beads were washed three times with milliQ water and 500ug beads were added per sample. Acetonitrile was added in 1:1 ratio, and the sample was incubated for 8 minutes. After incubating an additional 2 minutes on the magnetic rack, the supernatant was removed, and the beads were washed twice with 70% ethanol and once with 100% acetonitrile. After letting the beads air dry for 30 seconds to remove residual acetonitrile, 10ug of trypsin in 200ul of 50mM HEPES buffer was added for on bead digestion. Beads were sonicated for 1 minute and incubated o/n on the rotator. The next day, the beads were incubated on the magnetic rack for 2 minutes, and the supernatant containing digested peptides was transferred into a new tube. Peptides were dried down using a speedvac and then lyophilized. To enable multiplexing and peptide quantification, samples were labeled using isobaric Tandem Mass Tags (TMT, ThermoFisher). Lyophilized samples were

resuspended in 50 mM HEPES buffer (pH 7.4), and TMT was resuspended in anhydrous acetonitrile. 400 µg of TMT was added per sample to a final volume of 50 µl. After allowing the labeling reaction to proceed for 1 hour at RT on a platform shaker at 400 rpm, the reaction was quenched by adding 3.2 µl of 5% Hydroxylamine in HPLC-grade water. After an additional 15 minutes on the shaker, samples were pooled and dried by vacuum centrifugation. For phosphoserine and phosphothreonine mass spectrometry (MS) analysis, samples were then resuspended in 400 µl of buffer (100 mM Tris-HCl, 0.3% NP-40, pH 7.4), and phosphopeptides were

enriched using High-Select™ Fe-NTA phosphopeptide enrichment kits (Thermo). The phosphopeptide eluent was reduced in volume to 1-5 µl using vacuum centrifugation and resuspended by adding 10 µl of 5% acetonitrile in 0.1% formic acid. Subsequently, the sample was loaded onto a 50 µm i.d. capillary (Polymicro technologies) liquid chromatography (LC) column containing 5 µm C18 beads and a tip fritted with a mixture of Lithisil (PQ corp), Tetramethylammonium (TFA) Silicate (Sigma), and Formamide (Sigma). Phosphopeptide enrichment flow-through was preserved for use in a global proteomics run, where 2 µl of 1:500 diluted sample in 0.1% acetic acid was loaded onto a similarly

constructed 50 µm i.d. capillary LC column. The column was placed in line with an HPLC, and the following gradient was run at a flow rate of 2 µl/min (A = 0.1% acetic acid, B = 80% MeCN in 0.1% acetic acid): 0-4 min, 0-14% B; 4-50 min, 14-42% B; 50-57 min, 42-60% B; 57-60 min, 60-100% B, 60-68 min, 100% B, 68-69min, 100-0%B; 69-75 min, 0% B. The samples were analyzed on a ThermoFisher QExactive Plus mass spectrometer. The instrument was operated in data-dependent acquisition mode, with the top 15 most abundant precursors with charge +2 or

greater selected for fragmentation and dynamic exclusion set to 15 seconds. Precursors were isolated with a 0.4 m/z window and fragmented at 33 NCE via HCD. Precursor scan settings were set to AGC = 3e6, maximum IT = 50 ms, and resolution of 70,000. MS2 scan settings were set to AGC = 1e5, maximum IT = 300 ms, and resolution of 35,000. The total acquisition time was 75 minutes per sample. MS datafiles were searched on MASCOT version 2.4 with fixed modifications for NEM alkylation on cysteines (+125.047 Da), addition of TMT 6-plex to N-termini and lysine residues (+229.163 Da), oxidation of methionine residues (+15.995 Da) and phosphorylation on tyrosine, threonine, and serine residues (+79.966 Da). Precursor tolerance was ppm, fragment tolerance was 15 mmu, two missed cleavages were allowed. Peptides were considered to be positively identified if they had an ion score of at least 25. Peptides were discarded if they had TMT reporter ion intensities less than 1000 in any one channel, if they did not contain a phospho- or TMT-modification for the phosphoproteomics runs, or if they had isolation interference of more than 40%. To control for technical variation between channels, values were normalized to the median of each channel of the global proteomics runs. Differential expression analysis for peptides and phophopeptides were performed using limma-voom in R. P-values were corrected for multiple testing using the FDR method (48).

Enrichment analyses

By performing enrichment analyses, we aimed to gain insight into the molecular mechanisms that underlie the functional outcome of manipulating ELAVL4 expression in neurons. First, we used Ingenuity Pathway Analysis (IPA) (http://www.ingenuity.com; QIAGEN Bioinformatics, Redwood City, CA, USA) to assess the global enrichment of canonical pathways within the

generated transcriptomic and proteomic data. IPA uses the Ingenuity Knowledge Base, a combined repository of data from publicly accessible databases and data that are manually curated through systematically reviewing published literature, and currently, 712 canonical pathways have been annotated by IPA. For pathway enrichment, we analyzed all differentially expressed genes (DEGs), differentially expressed proteins (DEPs), and differentially phosphorylated proteins (DPPs) in the different neuronal lines with a FDRcorrected *P*-value (FDR P) < 5.0 X 10-2 and a fold change (FC) ≥ |1.20|. Peptides and phosphorylated peptides were mapped to one master protein, and the lowest *P*-value and the corresponding FC per protein was used for analysis. *P*-values for enriched canonical pathways were corrected using the FDR method (48). Subsequently, we used IPA to perform 'upstream regulator' analyses (49), using the same lists of DEGs, DEPs and DPPs (i.e., with FDR *P* < 5.0 X 10-2 and FC ≥ |1.20|). Upstream regulators can be any type of molecule—e.g., proteins, hormones, drugs and chemical compounds, and transcription factors—that can affect the expression of 'target' genes/proteins. For these analyses, we specifically focused on four upstream regulators, i.e., APP and tau, the key components of the main pathological lesions in AD (21), and the insulin receptor (INSR) and FOXO1, the only known transcriptional regulators of ELAVL4 expression (in pancreatic beta cells) (40). IPA calculates *P*-values that reflect the enrichment of targets of the upstream regulator in an input list (in this study our RNAseq and phosphoproteomics data), i.e., the amount of overlap between all known targets of the upstream regulator and the number of targets that are found in the input list. Furthermore, based on the direction of the expression changes of the input genes/proteins (up- or downregulated), IPA calculates an activation score (Z-score) that reflects the activation state of the upstream regulator. Z-scores ≥ 2 or ≤ -2 –

reflecting predicted activation or inhibition of the upstream regulator-dependent, downstream effects on target gene expression, respectively – are considered significant.

References

1. Hinman MN, Lou H. Diverse molecular functions of Hu proteins. Cell Mol Life Sci. 2008;65(20):3168-81.
2. Bronicki LM, Jasmin BJ. Emerging complexity of the HuD/ELAVl4 gene; implications for neuronal development, function, and dysfunction. RNA. 2013;19(8):1019-37.
3. Nussbacher JK, Tabet R, Yeo GW, Lagier-Tourenne C. Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. Neuron. 2019;102(2):294-320.
4. Park S, Myszka DG, Yu M, Littler SJ, Laird-Offringa IA. HuD RNA recognition motifs play distinct roles in the formation of a stable complex with AU-rich RNA. Mol Cell Biol. 2000;20(13):4765-72.
5. Bolognani F, Contente-Cuomo T, Perrone-Bizzozero NI. Novel recognition motifs and biological functions of the RNA-binding protein HuD revealed by genome-wide identification of its targets. Nucleic Acids Res. 2010;38(1):117-30.
6. Fukao A, Sasano Y, Imataka H, Inoue K, Sakamoto H, Sonenberg N, et al. The ELAV protein HuD stimulates cap-dependent translation in a Poly(A)- and eIF4A-dependent manner. Mol Cell. 2009;36(6):1007-17.
7. Fujiwara T, Fukao A, Sasano Y, Matsuzaki H, Kikkawa U, Imataka H, et al. Functional and direct interaction between the RNA binding protein HuD and active Akt1. Nucleic Acids Res. 2012;40(5):1944-53.
8. Pascale A, Amadio M, Quattrone A. Defining a neuron: neuronal ELAV proteins. Cell Mol Life Sci. 2008;65(1):128-40.
9. Graus F, Elkon KB, Cordon-Cardo C, Posner JB. Sensory neuronopathy and small cell lung cancer. Antineuronal antibody that also reacts with the tumor. Am J Med. 1986;80(1):45-52.
10. Szabo A, Dalmau J, Manley G, Rosenfeld M, Wong E, Henson J, et al. HuD, a paraneoplastic encephalomyelitis antigen, contains RNA-binding domains and is homologous to Elav and -lethal. Cell. 1991;67(2):325-33.
11. DeBoer EM, Azevedo R, Vega TA, Brodkin J, Akamatsu W, Okano H, et al. Prenatal deletion of the RNA-binding protein HuD disrupts postnatal cortical circuit maturation and behavior. J Neurosci. 2014;34(10):3674-86.
12. Quattrone A, Pascale A, Nogues X, Zhao W, Gusev P, Pacini A, et al. Posttranscriptional regulation of gene expression in learning by the neuronal ELAV-like mRNA-stabilizing proteins. Proc Natl Acad Sci U S A. 2001;98(20):11668-73.
13. Pascale A, Gusev PA, Amadio M, Dottorini T, Govoni S, Alkon DL, et al. Increase of the RNAbinding protein HuD and posttranscriptional up-regulation of the GAP-43 gene during spatial memory. Proc Natl Acad Sci U S A. 2004;101(5):1217-22.
14. Bolognani F, Qiu S, Tanner DC, Paik J, Perrone-Bizzozero NI, Weeber EJ. Associative and spatial learning and memory deficits in transgenic mice overexpressing the RNA-binding protein HuD. Neurobiol Learn Mem. 2007;87(4):635-43.
15. Sosanya NM, Cacheaux LP, Workman ER, Niere F, Perrone-Bizzozero NI, Raab-Graham KF.

Mammalian Target of Rapamycin (mTOR) Tagging Promotes Dendritic Branch Variability through the Capture of Ca2+/Calmodulin-dependent Protein Kinase II alpha (CaMKIIalpha) mRNAs by the RNAbinding Protein HuD. J Biol Chem. 2015;290(26):16357-71.

16. Vanevski F, Xu B. HuD interacts with Bdnf mRNA and is essential for activity-induced BDNF synthesis in dendrites. PLoS One. 2015;10(2):e0117264.
17. Bolognani F, Tanner DC, Nixon S, Okano HJ, Okano H, Perrone-Bizzozero NI. Coordinated expression of HuD and GAP-43 in hippocampal dentate granule cells during developmental and adult plasticity. Neurochem Res. 2007;32(12):2142-51.
18. Deschenes-Furry J, Perrone-Bizzozero N, Jasmin BJ. The RNA-binding protein HuD: a regulator of neuronal differentiation, maintenance and plasticity. Bioessays. 2006;28(8):822-33.
19. Lee YS, Lee JA, Kaang BK. Regulation of mRNA stability by ARE-binding proteins in synaptic plasticity and memory. Neurobiol Learn Mem. 2015;124:28-33.
20. Patterson C. World Alzheimer Report 2018: the state of the art of dementia research: new frontiers. Alzheimer's Disease International (ADI): London, UK. 2018.
21. Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14(4):535-62.
22. Hansson O, Lehmann S, Otto M, Zetterberg H, Lewczuk P. Advantages and disadvantages of the use of the CSF Amyloid beta (Abeta) 42/40 ratio in the diagnosis of Alzheimer's Disease. Alzheimers Res Ther. 2019;11(1):34.
23. Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, et al. Secreted amyloid betaprotein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nat Med. 1996;2(8):864-70.
24. Bentahir M, Nyabi O, Verhamme J, Tolia A, Horre K, Wiltfang J, et al. Presenilin clinical mutations can affect gamma-secretase activity by different mechanisms. J Neurochem. 2006;96(3):732-42.
25. Liu L, Lauro BM, Wolfe MS, Selkoe DJ. Hydrophilic loop 1 of Presenilin-1 and the APP GxxxG transmembrane motif regulate gamma-secretase function in generating Alzheimer-causing Abeta peptides. J Biol Chem. 2021;296:100393.
26. Amadio M, Pascale A, Wang J, Ho L, Quattrone A, Gandy S, et al. nELAV proteins alteration in Alzheimer's disease brain: a novel putative target for amyloid-beta reverberating on AbetaPP processing. J Alzheimers Dis. 2009;16(2):409-19.
27. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Sci Data. 2018;5:180185.
28. Kang MJ, Abdelmohsen K, Hutchison ER, Mitchell SJ, Grammatikakis I, Guo R, et al. HuD regulates coding and noncoding RNA to induce APP-->Abeta processing. Cell Rep. 2014;7(5):1401-9.
29. Marchesi N, Amadio M, Colombrita C, Govoni S, Ratti A, Pascale A. PKC Activation Counteracts ADAM10 Deficit in HuD-Silenced Neuroblastoma Cells. J Alzheimers Dis. 2016;54(2):535-47.

30. Fragkouli A, Koukouraki P, Vlachos IS, Paraskevopoulou MD, Hatzigeorgiou AG, Doxakis E. Neuronal ELAVL proteins utilize AUF-1 as a co-partner to induce neuron-specific alternative splicing of APP. Sci Rep. 2017;7:44507.
31. Lim CS, Alkon DL. PKCepsilon promotes HuD-mediated neprilysin mRNA stability and enhances neprilysin-induced Abeta degradation in brain neurons. PLoS One. 2014;9(5):e97756.
32. Subhadra B, Schaller K, Seeds NW. Neuroserpin up-regulation in the Alzheimer's disease brain is associated with elevated thyroid hormone receptor-beta1 and HuD expression. Neurochem Int. 2013;63(5):476-81.
33. Kinghorn KJ, Crowther DC, Sharp LK, Nerelius C, Davis RL, Chang HT, et al. Neuroserpin binds Abeta and is a neuroprotective component of amyloid plaques in Alzheimer disease. J Biol Chem. 2006;281(39):29268-77.
34. Aranda-Abreu GE, Behar L, Chung S, Furneaux H, Ginzburg I. Embryonic lethal abnormal vision-like RNA-binding proteins regulate neurite outgrowth and tau expression in PC12 cells. J Neurosci. 1999;19(16):6907-17.
35. Deschenes-Furry J, Mousavi K, Bolognani F, Neve RL, Parks RJ, Perrone-Bizzozero NI, et al. The RNA-binding protein HuD binds acetylcholinesterase mRNA in neurons and regulates its expression after axotomy. J Neurosci. 2007;27(3):665-75.
36. Hogan DB. Long-term efficacy and toxicity of cholinesterase inhibitors in the treatment of Alzheimer disease. Can J Psychiatry. 2014;59(12):618-23.
37. Hu RT, Yu Q, Zhou SD, Yin YX, Hu RG, Lu HP, et al. Co-expression Network Analysis Reveals Novel Genes Underlying Alzheimer's Disease Pathogenesis. Front Aging Neurosci. 2020;12:605961.
38. Juan-Mateu J, Rech TH, Villate O, Lizarraga-Mollinedo E, Wendt A, Turatsinze JV, et al. Neuron-enriched RNA-binding Proteins Regulate Pancreatic Beta Cell Function and Survival. J Biol Chem. 2017;292(8):3466-80.
39. Hong Y, Tak H, Kim C, Kang H, Ji E, Ahn S, et al. RNA binding protein HuD contributes to betacell dysfunction by impairing mitochondria dynamics. Cell Death Differ. 2020;27(5):1633-43.
40. Lee EK, Kim W, Tominaga K, Martindale JL, Yang X, Subaran SS, et al. RNA-binding protein HuD controls insulin translation. Mol Cell. 2012;45(6):826-35.
41. Muratore CR, Rice HC, Srikanth P, Callahan DG, Shin T, Benjamin LN, et al. The familial Alzheimer's disease APPV717I mutation alters APP processing and Tau expression in iPSC-derived neurons. Hum Mol Genet. 2014;23(13):3523-36.
42. Muratore CR, Zhou C, Liao M, Fernandez MA, Taylor WM, Lagomarsino VN, et al. Cell-type Dependent Alzheimer's Disease Phenotypes: Probing the Biology of Selective Neuronal Vulnerability. Stem Cell Reports. 2017;9(6):1868-84.
43. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron. 2013;78(5):785-98.
44. Nilsen G, Liestol K, Van Loo P, Moen Vollan HK, Eide MB, Rueda OM, et al. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics. 2012;13:591.
45. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

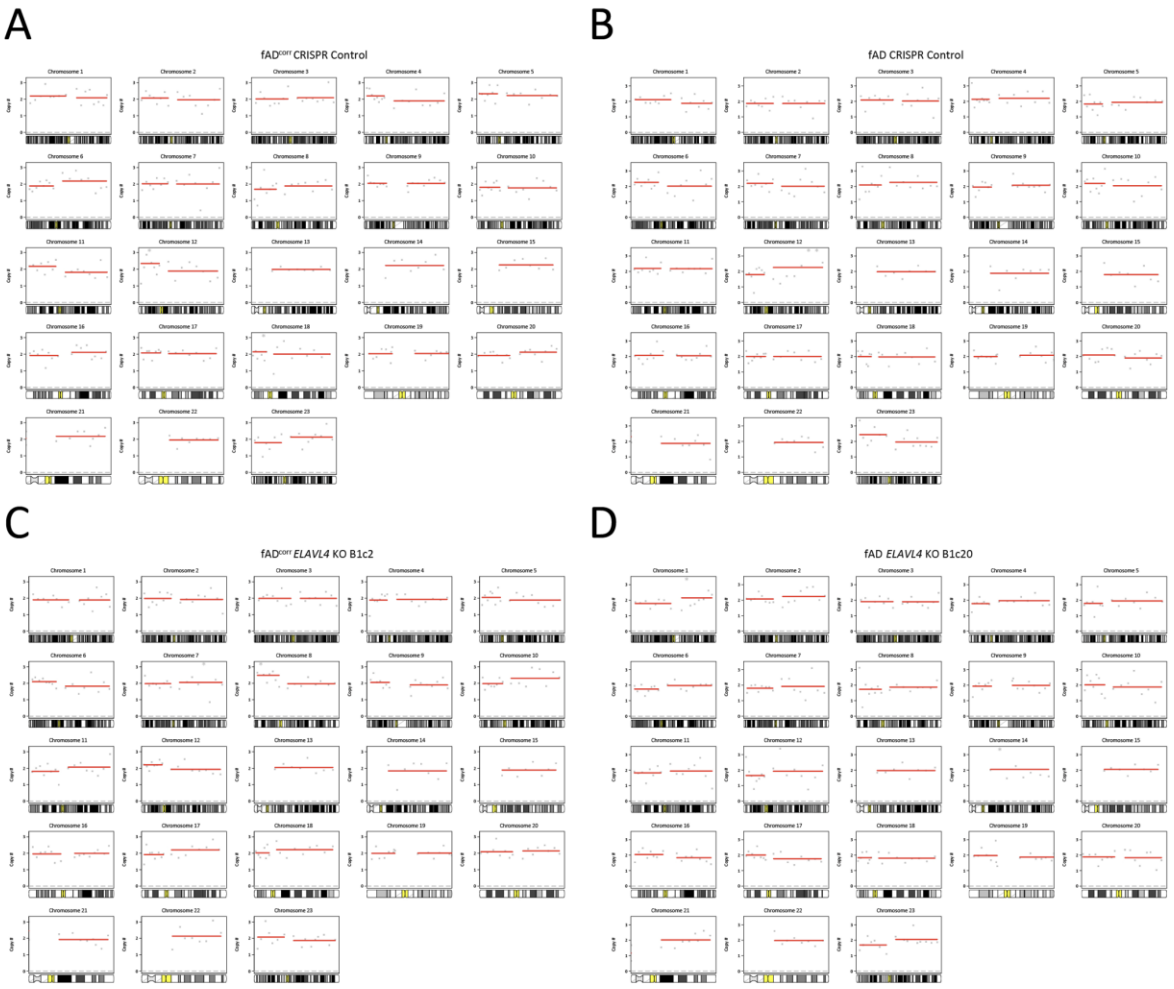
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
47. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525-7.
48. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
49. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30.
50. Srikanth P, Lagomarsino VN, Pearse RV, 2nd, Liao M, Ghosh S, Nehme R, et al. Convergence of independent DISC1 mutations on impaired neurite growth via decreased UNC5D expression. Transl Psychiatry. 2018;8(1):245.
51. Wang X, Zhou X, Li G, Zhang Y, Wu Y, Song W. Modifications and Trafficking of APP in the Pathogenesis of Alzheimer's Disease. Front Mol Neurosci. 2017;10:294.
52. Klemann C, Xicoy H, Poelmans G, Bloem BR, Martens GJM, Visser JE. Physical Exercise Modulates L-DOPA-Regulated Molecular Pathways in the MPTP Mouse Model of Parkinson's Disease. Mol Neurobiol. 2018;55(7):5639-57.
53. Macri S, Spinello C, Widomska J, Magliozzi R, Poelmans G, Invernizzi RW, et al. Neonatal corticosterone mitigates autoimmune neuropsychiatric disorders associated with streptococcus in mice. Sci Rep. 2018;8(1):10188.
54. Pagliaroli L, Widomska J, Nespoli E, Hildebrandt T, Barta C, Glennon J, et al. Riluzole Attenuates L-DOPA-Induced Abnormal Involuntary Movements Through Decreasing CREB1 Activity: Insights from a Rat Model. Mol Neurobiol. 2019;56(7):5111-21.
55. Johnson SA, Rogers J, Finch CE. APP-695 transcript prevalence is selectively reduced during Alzheimer's disease in cortex and hippocampus but not in cerebellum. Neurobiol Aging. 1989;10(6):755-60.
56. Moir RD, Lynch T, Bush AI, Whyte S, Henry A, Portbury S, et al. Relative increase in Alzheimer's disease of soluble forms of cerebral Abeta amyloid protein precursor containing the Kunitz protease inhibitory domain. J Biol Chem. 1998;273(9):5013-9.
57. Kwak SS, Washicosky KJ, Brand E, von Maydell D, Aronson J, Kim S, et al. Amyloid-beta42/40 ratio drives tau pathology in 3D human neural cell culture models of Alzheimer's disease. Nat Commun. 2020;11(1):1377.
58. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol Med. 2016;8(6):595-608.
59. Fagan AM, Mintun MA, Shah AR, Aldea P, Roe CM, Mach RH, et al. Cerebrospinal fluid tau and ptau(181) increase with cortical amyloid deposition in cognitively normal individuals: implications for future clinical trials of Alzheimer's disease. EMBO Mol Med. 2009;1(8-9):371-80.
60. Mielke MM, Hagen CE, Xu J, Chai X, Vemuri P, Lowe VJ, et al. Plasma phospho-tau181 increases with Alzheimer's disease clinical severity and is associated with tau- and amyloid-positron emission tomography. Alzheimers Dement. 2018;14(8):989-97.
61. Bowles KR, Silva MC, Whitney K, Bertucci T, Berlind JE, Lai JD, et al. ELAVL4, splicing, and glutamatergic dysfunction precede neuron loss in MAPT mutation cerebral organoids. Cell.

2021;184(17):4547-63 e17.
62. Perrone-Bizzozero NI, Tanner DC, Mounce J, Bolognani F. Increased expression of axogenesisrelated genes and mossy fibre length in dentate granule cells from adult HuD overexpressor mice. ASN Neuro. 2011;3(5):259-70.
63. Lee WS, Lee WH, Bae YC, Suk K. Axon Guidance Molecules Guiding Neuroinflammation. Exp Neurobiol. 2019;28(3):311-9.
64. Zhang L, Qi Z, Li J, Li M, Du X, Wang S, et al. Roles and Mechanisms of Axon-Guidance Molecules in Alzheimer's Disease. Mol Neurobiol. 2021.
65. Soheili-Nezhad S, van der Linden RJ, Olde Rikkert M, Sprooten E, Poelmans G. Long genes are more frequently affected by somatic mutations and show reduced expression in Alzheimer's disease: Implications for disease etiology. Alzheimers Dement. 2021;17(3):489-99.
66. Farmer KM, Ghag G, Puangmalai N, Montalbano M, Bhatt N, Kayed R. P53 aggregation, interactions with tau, and impaired DNA damage response in Alzheimer's disease. Acta Neuropathol Commun. 2020;8(1):132.
67. Lee HG, Casadesus G, Zhu X, Castellani RJ, McShea A, Perry G, et al. Cell cycle re-entry mediated neurodegeneration and its treatment role in the pathogenesis of Alzheimer's disease. Neurochem Int. 2009;54(2):84-8.
68. Moh C, Kubiak JZ, Bajic VP, Zhu X, Smith MA, Lee HG. Cell cycle deregulation in the neurons of Alzheimer's disease. Results Probl Cell Differ. 2011;53:565-76.
69. Fielder E, von Zglinicki T, Jurk D. The DNA Damage Response in Neurons: Die by Apoptosis or Survive in a Senescence-Like State? J Alzheimers Dis. 2017;60(s1):S107-S31.
70. Abdelmohsen K, Kuwano Y, Kim HH, Gorospe M. Posttranscriptional gene regulation by RNAbinding proteins during oxidative stress: implications for cellular senescence. Biol Chem. 2008;389(3):243-55.
71. Bonda DJ, Lee HG, Camins A, Pallas M, Casadesus G, Smith MA, et al. The sirtuin pathway in ageing and Alzheimer disease: mechanistic and therapeutic considerations. Lancet Neurol. 2011;10(3):275-9.
72. Mouchiroud L, Houtkooper RH, Moullan N, Katsyuba E, Ryu D, Canto C, et al. The NAD(+)/Sirtuin Pathway Modulates Longevity through Activation of Mitochondrial UPR and FOXO Signaling. Cell. 2013;154(2):430-41.
73. Giannakou ME, Partridge L. The interaction between FOXO and SIRT1: tipping the balance towards survival. Trends Cell Biol. 2004;14(8):408-12.
74. Zhang M, Zhang Q, Hu Y, Xu L, Jiang Y, Zhang C, et al. miR-181a increases FoxO1 acetylation and promotes granulosa cell apoptosis via SIRT1 downregulation. Cell Death Dis. 2017;8(10):e3088.
75. Qin L, Guo J, Zheng Q, Zhang H. BAG2 structure, function and involvement in disease. Cell Mol Biol Lett. 2016;21:18.
76. Sluchanko NN, Gusev NB. Probable participation of 14-3-3 in tau protein oligomerization and aggregation. J Alzheimers Dis. 2011;27(3):467-76.
77. Foote M, Zhou Y. 14-3-3 proteins in neurological disorders. Int J Biochem Mol Biol. 2012;3(2):152-64.
78. Cai T, Che H, Yao T, Chen Y, Huang C, Zhang W, et al. Manganese induces tau

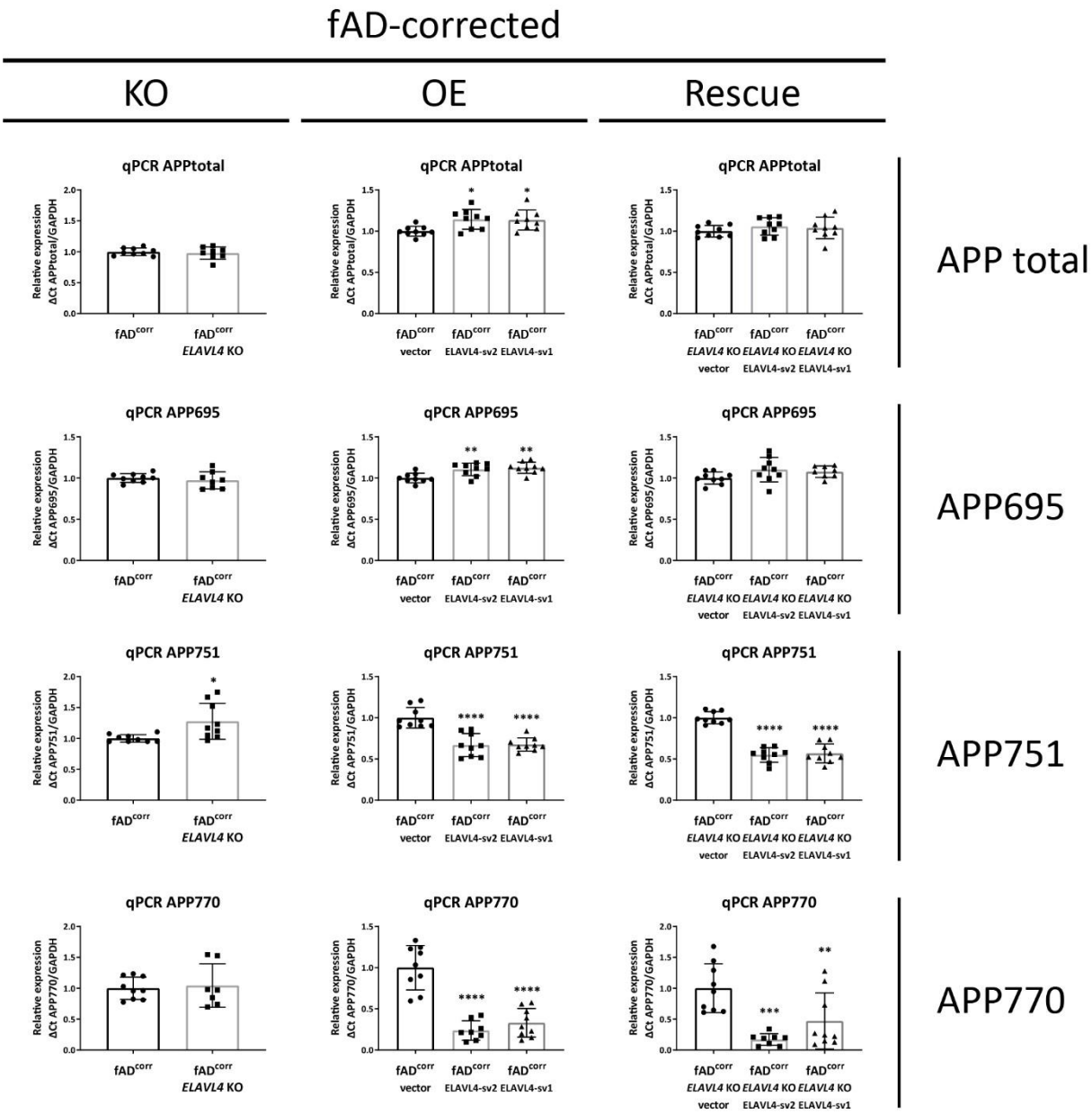
hyperphosphorylation through the activation of ERK MAPK pathway in PC12 cells. Toxicol Sci. 2011;119(1):169-77.

79. Arnold SE, Arvanitakis Z, Macauley-Rambach SL, Koenig AM, Wang HY, Ahima RS, et al. Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. Nat Rev Neurol. 2018;14(3):168-81.
80. Ferrario CR, Reagan LP. Insulin-mediated synaptic plasticity in the CNS: Anatomical, functional and temporal contexts. Neuropharmacology. 2018;136(Pt B):182-91.
81. Xie L, Helmerhorst E, Taddei K, Plewright B, Van Bronswijk W, Martins R. Alzheimer's betaamyloid peptides compete for insulin binding to the insulin receptor. J Neurosci. 2002;22(10):RC221.

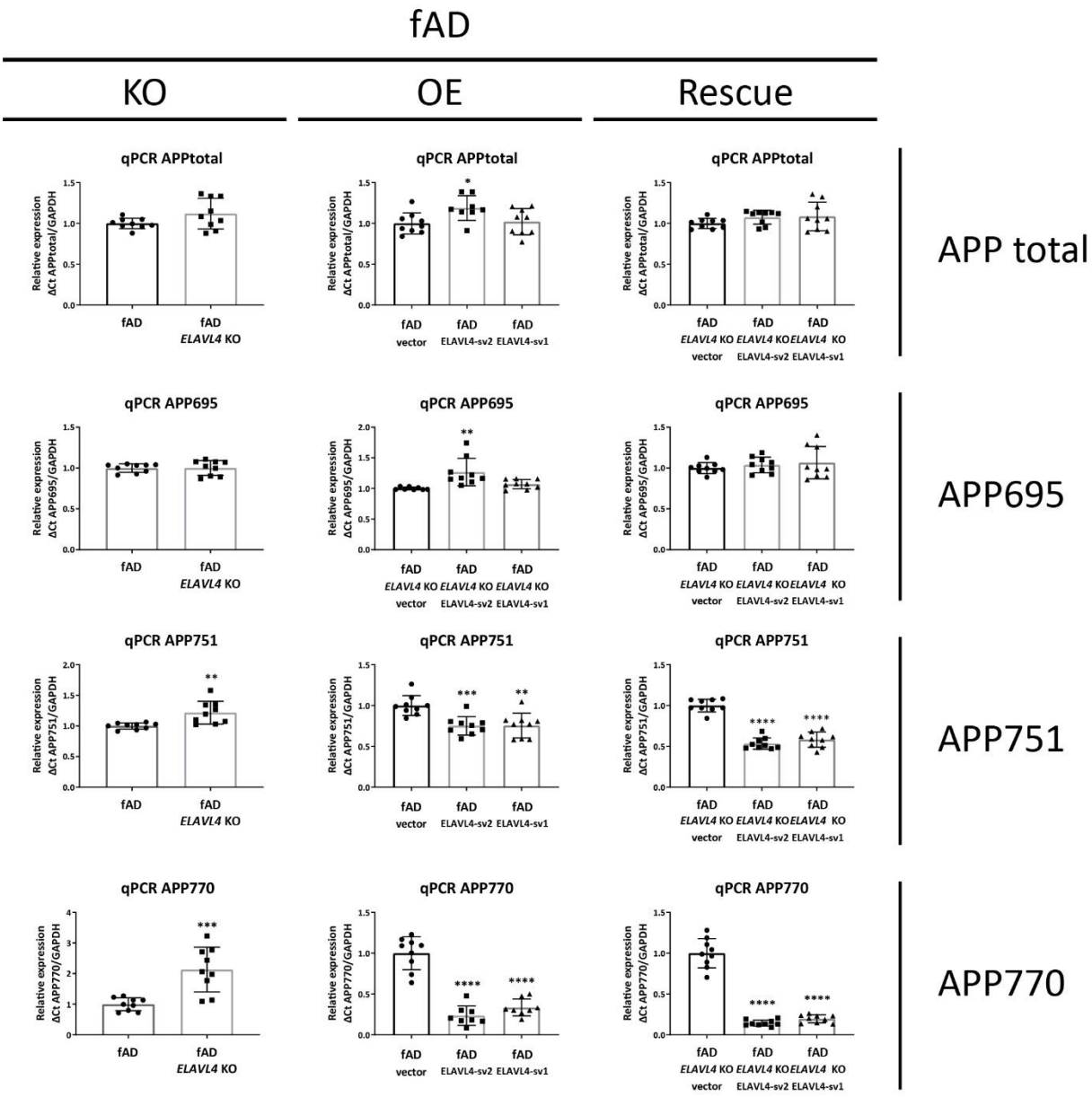
Supplemental Figures



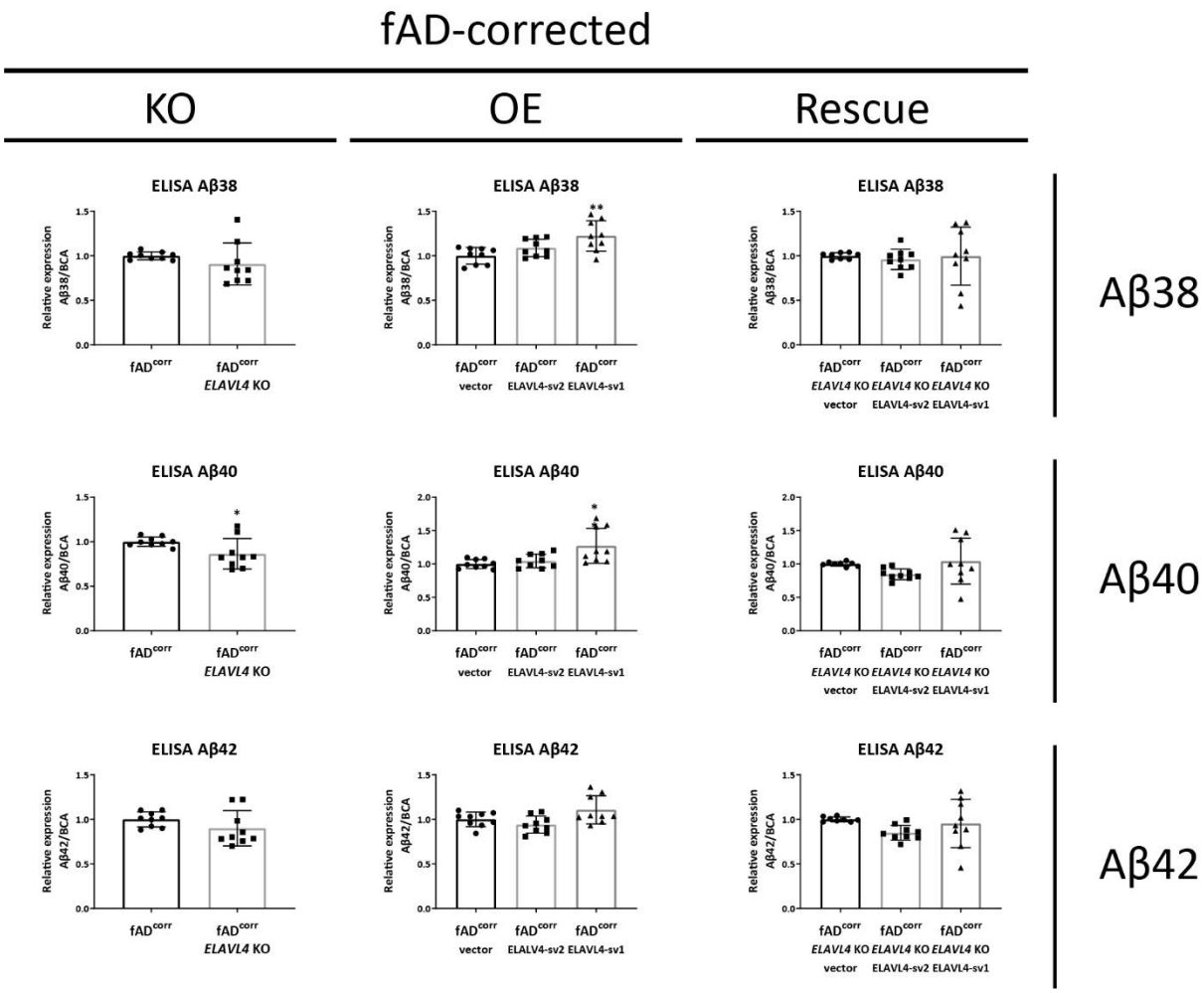
Supplementary figure 14: CRISPR/Cas9 editing did not result in large chromosomal abnormalities in iPSC lines generated in this study. Common CNV analysis in the monoclonal lines using the Nanostring nCounter Human CNV codeset for (A) fAD corrected CRISPR control (B) fAD CRISPR control (C) fAD corrected *ELAVL4* knock out B1c2 (D) fAD *ELAVL4* knock out B1c20



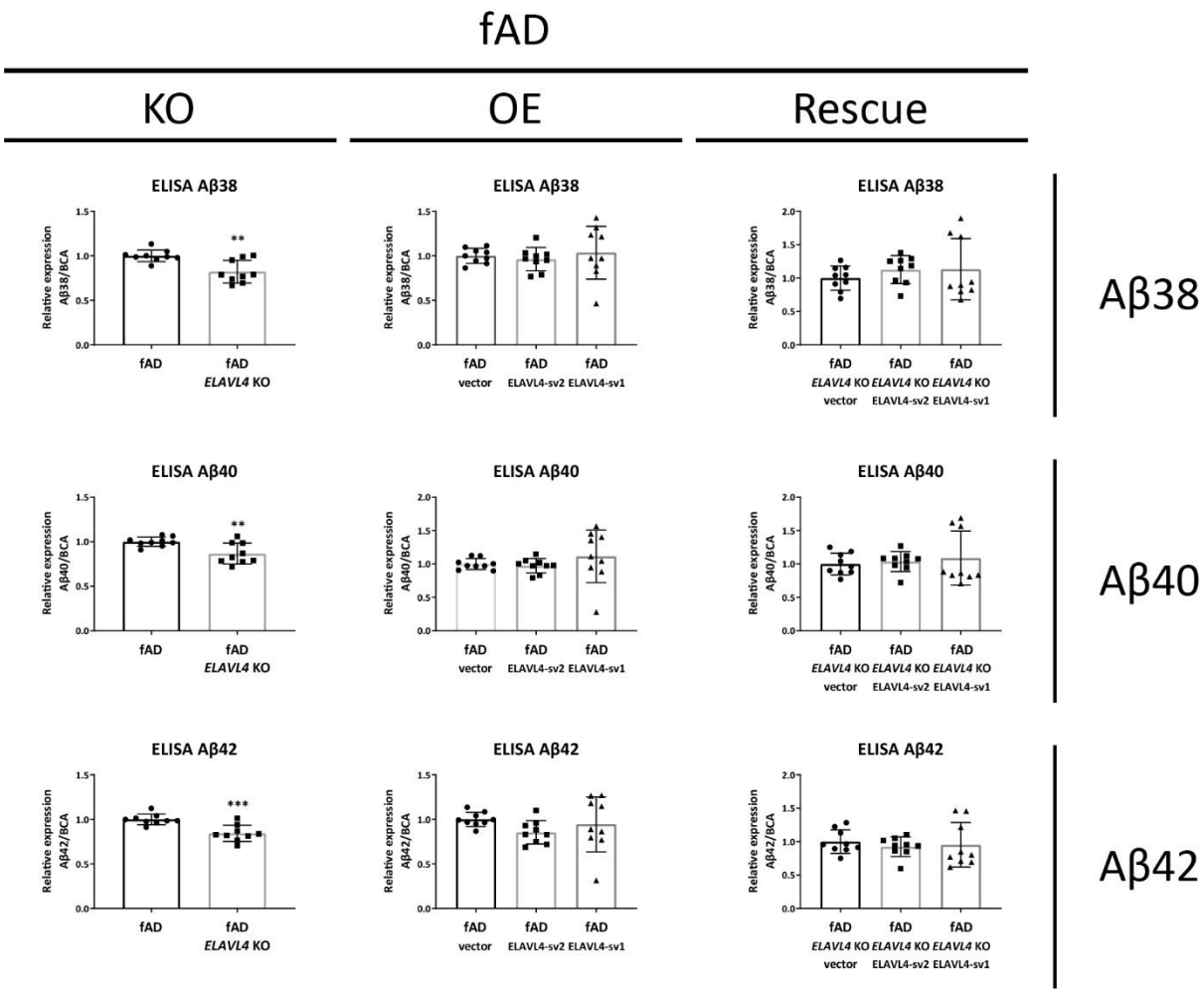
Supplementary figure 15: ELAVL4 affects APP mRNA splicing in fAD-corrected iNs.
Quantification of specific APP mRNA splice forms in iNs from *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD-corrected induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where ELAVL4 was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, cells were lysed and protein and RNA samples were harvested. Quantifications are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, Pvalues
were calculated with Dunnett's multiple comparison that followed after a significant ANOVA *test, * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.*



Supplementary figure 16: ELAVL4 affects APP mRNA splicing in fAD iNs.
Quantification of specific APP mRNA splice forms in iNs from *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where ELAVL4 was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, cells were lysed and protein and RNA samples were harvested. Quantifications are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, Pvalues were calculated with Dunnett's multiple comparison that followed after a significant ANOVA *test, * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.*



Supplementary figure 17: ELAVL4 affects extracellular Aβ species in fAD-corrected iNs
Quantification of extracellular Aβ species (Aβ38, Aβ40, and Aβ42) in culture media of *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD-corrected induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where ELAVL4 was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, 72 h conditioned media were collected and cells lysed. Quantifications are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, P-values were calculated with Dunnett's multiple comparison that followed after a significant ANOVA test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.



Supplementary figure 18: ELAVL4 affects extracellular Aβ species in fAD iNs.
Quantification of extracellular Aβ species (Aβ38, Aβ40, and Aβ42) in culture media of *ELAVL4* knock out (KO) overexpression (OE) and rescue in fAD induced neurons (iNs). In the KO, changes relative to CRISPR control line that underwent mock targeting with an empty sgRNA vector transfected alongside with Cas9, and monoclonal subclones were isolated and analyzed in parallel to subclones where *ELAVL4* was targeted. For OE and rescue conditions, changes relative to empty vector OE plasmid in the CRISPR control or *ELAVL4* KO iNs respectively. At day 21 of differentiation, 72 h conditioned media were collected and cells lysed. Quantifications are from 3 independent differentiations with n=3. In the KO experiments P-values were calculated with t tests, while in the OE and rescue experiments, P-values were calculated with Dunnett's multiple comparison that followed after a significant ANOVA test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Table 1. Effects of ELAVL4 expression manipulation on APP and tau phenotypes

	APP Phenotypes			Tau Phenotypes	
	RNA	**Protein**	**Processing**	**RNA**	**Protein**
Knock out					
fAD^{corr} *ELAVL4* KO	Total ↓ 97.9 % n.s. 695 ↓ 97.2 % n.s. **751 ↑ 127.6 % *** 770 ↑ 104.4 % n.s.	Total ↑100.1 % n.s.	Aβ42/40 ↑ 104.1 % n.s. Aβ42 ↓ 90.0 % n.s. **Aβ40 ↓ 86.4 % *** Aβ38 ↓ 90.8 % n.s.	**↑ 111.9 % ******	**Total tau ↑ 124.8 % *** **P-tau181 ↑ 105.2 % ***
fAD *ELAVL4* KO	Total ↑112.0 % n.s. 695 ↑100.2 % n.s. **751 ↑ 121.9 % **** **770 ↑ 213.5 % *****	**Total ↓ 79.7 % ****	Aβ42/40 ↓ 97.7 % n.s. **Aβ42 ↓ 84.3 % ***** **Aβ40 ↓ 86.6 % **** **Aβ38 ↓ 82.2 % ****	**↓ 93.2 % ***	Total tau ↑ 106.8 % n.s. **P-tau181 ↑ 104.4 % ***
Over Expression					
fAD^{corr} ELAVL4-sv2	**Total ↑ 114.3 % *** **695 ↑ 110.4 % **** **751 ↓ 66.8 % ****** **770 ↓ 23.7 % ******	**Total ↑ 114.7 % ****	**Aβ42/40 ↓ 90.1 % **** Aβ42 ↓ 94.2 % n.s. Aβ40 ↑ 104.6 % n.s Aβ38 ↑ 108.9 % n.s	↑ 103.9 % n.s.	Total tau ↓ 94.0 % n.s. **P-tau181 ↓ 89.3 % ***
fAD ELAVL4-sv2	**Total ↑ 118.9 % *** **695 ↑ 126.8 % **** **751 ↓ 75.2 % ***** **770 ↓ 23.5 % ******	**Total ↑ 210.9 % ******	**Aβ42/40 ↓ 87.4 % ****** Aβ42 ↓ 85.4 % n.s Aβ40 ↓ 97.4 % n.s Aβ38 ↓ 96.2 % n.s.	↑105.5 % n.s.	DAKO ↓ 99.7 % n.s. **P-tau181 ↓ 90.6 % ***
fAD^{corr} ELAVL4-sv1	**Total ↑ 113.6 % *** **695 ↑ 112.5 % **** **751 ↓ 67.5 % ****** **770 ↓ 33.1 % ******	**Total ↓ 85.1 % ****	**Aβ42/40 ↓ 88.1 % ***** Aβ42 ↑ 110.7 % n.s **Aβ40 ↑ 127.1 % **** **Aβ38 ↑ 122.4 % ****	**↑ 110.0 % ***	Total tau ↑ 100.4 % n.s. **P-Tau181 ↓ 83.7 % *****
fAD ELAVL4-sv1	Total ↑102.1 % n.s. 695 ↑ 107.0 % **751 ↓ 75.5 % **** **770 ↓ 33.6 % ******	Total ↑ 145.9 % n.s.	**Aβ42/40 ↓ 83.9 % ****** Aβ42 ↓ 94.4 % n.s Aβ40 ↑ 111.5 % n.s. Aβ38 ↑ 103.6 % n.s.	**↑ 114.0 % ***	**Total tau ↑ 119.4 % **** **P-tau181 ↓ 83.3 % ******
Rescue in *ELAVL4* KO					
fAD^{corr} *ELAVL4* KO ELAVL4-sv2	Total ↑ 105.7 % n.s. 695 ↑ 110.1 % n.s. **751 ↓ 54.7 % ****** **770 ↓ 17.0 % *****	**Total ↑ 140.8 % ******	Aβ42/40 ↑ 100.8 % n.s Aβ42 ↓ 85.0 % n.s. Aβ40 ↓ 84.6 % n.s. Aβ38 ↓ 96.0 % n.s.	↓ 94.2 % n.s	DAKO ↓ 86.7 % n.s. **P-tau181 ↓ 78.4 % ******
fAD *ELAVL4* KO ELAVL4-sv2	Total ↑ 107.6 % n.s. 695 ↑ 103.9 % n.s. **751 ↓ 53.5 % ****** **770 ↓ 14.5 % ******	Total ↑ 277.6 % **	**Aβ42/40 ↓ 89.1 % ***** Aβ42 ↓ 92.4 % n.s Aβ40 ↑ 103.6 % n.s. Aβ38 ↑ 112.6 % n.s.	↑106.3 % n.s.	DAKO ↓ 85.7 % n.s. **P-tau181 ↓ 85.2 % *****
fAD^{corr} *ELAVL4* KO ELAVL4-sv1	Total ↑ 103.9 % n.s. 695 ↑ 107.7 % n.s. **751 ↓56.7 % ****** **770 ↓ 47.0 % **	Total ↑ 110.7 % n.s.	**Aβ42/40 ↓ 93.0 % *** Aβ42 ↓ 95.4 % n.s. Aβ40 ↑ 104.4 % n.s Aβ38 ↓ 99.6 % n.s.	↓ 95.4 % n.s	DAKO ↓ 91.6 % n.s. **P-tau181 ↓ 70.4 % ******
fAD *ELAVL4* KO ELAVL4-sv1	Total ↑ 108.6 % n.s. 695 ↑ 106.6 % n.s. **751 ↓ 58.3 % ****** **770 ↓ 19.7 % ******	Total ↑ 241.5 % *	**Aβ42/40 ↓ 88.4 % ***** Aβ42 ↓ 95.3 % n.s Aβ40 ↑ 108.9 % n.s. Aβ38 ↑ 113.3 % n.s.	**↑ 111.3 % ****	Total tau ↓ 89.1 % n.s. **P-tau181 ↓ 81.8 % ******

NOTE. For all conditions: per condition, three replicates from three independent differentiations were used. For KO conditions: changes relative to CRISPR control line, *P*-value calculated with t tests. For OE and rescue conditions: changes relative to empty vector control, *P*-values calculated with Dunnett's multiple comparison that followed after ANOVA test. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001

Table 2. Expression of ELAVL4 and ELAVL4 targets after ELAVL4 manipulation

	***ELAVL4* knock out**						**ELAVL4 overexpression**				**ELAVL4 rescue in knock out**					
	RNAseq				**Protein**		**RNAseq**				**RNAseq**				**Protein**	
	fAD corrected		**fAD**		**fAD**		**fAD corrected**		**fAD**		**fAD corrected**		**fAD**		**fAD**	
Gene	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**	**FC**	**FDR-P**
ELAVL4/HuD	-4,47	2,86E-37	-3,24	8,13E-34	N.A.	N.D.	2,86	8,49E-33	2,62	2,25E-31	15,14	3,11E-48	9,40	1,13E-45	1,635	1,21E-11
CAMK2A	1,92	8,26E-12	1,10	2,42E-01	N.A.	N.D.	-1,26	5,71E-02	-1,29	3,13E-02	-1,96	1,30E-10	-1,53	1,43E-05	N.A.	N.D.
BDNF	1,17	4,83E-03	-1,19	1,55E-01	N.A.	N.D.	-1,13	1,89E-01	-1,04	7,80E-01	-1,18	1,13E-02	-1,15	3,58E-02	N.A.	N.D.
GAP43	-1,07	4,00E-03	1,11	1,19E-04	1,21	1,05E-03	1,09	1,17E-02	1,12	1,07E-03	1,14	9,46E-06	1,07	3,02E-02	-1,19	1,89E-03
APP	-1,16	7,03E-13	-1,06	4,32E-04	N.A.	N.D.	1,02	4,51E-01	-1,10	8,38E-02	-1,05	4,18E-03	-1,03	1,16E-01	N.A.	N.D.
BACE1	1,13	6,71E-04	1,03	4,78E-01	N.A.	N.D.	1,08	1,26E-01	-1,05	8,54E-01	1,01	8,06E-01	1,02	6,84E-01	N.A.	N.D.
ADAM10	-1,10	1,69E-03	-1,02	6,53E-01	N.A.	N.D.	-1,10	2,73E-02	-1,04	4,74E-01	-1,22	5,78E-07	-1,12	3,48E-03	N.A.	N.D.
NEP	-3,21	7,98E-08	1,23	3,39E-01	N.A.	N.D.	-1,35	2,77E-01	-1,14	7,80E-01	-1,07	8,49E-01	-1,26	4,01E-01	N.A.	N.D.
SERPINI1	1,29	3,55E-13	1,01	7,54E-01	N.A.	N.D.	-1,04	4,02E-01	-1,06	2,30E-01	-1,01	7,64E-01	-1,07	4,68E-02	N.A.	N.D.
TAU	-1,02	5,66E-01	-1,02	4,84E-01	1,09	6,02E-01	1,15	1,48E-04	1,38	6,24E-01	1,16	2,85E-06	1,11	4,07E-04	1,08	5,07E-01
ACHE	1,02	6,09E-01	-1,10	5,86E-03	N.A.	N.D.	1,18	2,15E-04	1,18	1,26E-04	1,30	1,63E-09	1,20	3,80E-06	N.A.	N.D.

Abbreviations: FC, Fold change; FDR-P, FDR corrected *P*-value

Table 3. Canonical pathway analysis of RNAseq and (phospho)proteomics HuD KO, overexpression, and rescue neuronal cell lines, using IPA.

RNAseq								phosphoproteomics			
fAD-corrected				fAD				fAD			
(1) 6260 DEGs HuD KO in fAD^{corr}				**(6) 3661 DEGs HuD KO in fAD**				**(11) 40 DEPs HuD KO in fAD**		**(12) 48 DPPs HuD KO in fAD**	
Canonical Pathway	***P*-value**			**Canonical Pathway**	***P*-value**			**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**
Axonal Guidance Signaling	1,00E-19			Synaptogenesis Signaling Pathway	3,16E-11			Sirtuin Signaling Pathway	6,76E-05	CDK5 Signaling	2,29E-04
Synaptogenesis Signaling Pathway	1,00E-13			Axonal Guidance Signaling	2,75E-08			Mitochondrial Dysfunction	5,25E-04	UVB-Induced MAPK Signaling	2,29E-04
Molecular Mechanisms of Cancer	2,51E-12			Sperm Motility	8,13E-07			NAD Signaling Pathway	4,47E-03	UVC-Induced MAPK Signaling	2,29E-04
p53 Signaling	5,75E-09			Calcium Signaling	9,77E-07			Protein Ubiquitination Pathway	2,24E-02	Apelin Cardiomyocyte Signaling Pathway	2,29E-04
Kinetochore Metaphase Signaling Pathway	5,75E-09			Neuropathic Pain Signaling In Dorsal Horn Neurons	1,07E-05			DNA Methylation and Transcriptional Repression Signaling	2,51E-02	14-3-3-mediated Signaling	2,34E-04
(2) 609 DEGs HuD sv1+2 OE in fAD^{corr}		**(3) 207 DEGs-OD HuD sv1+sv2 OE in fAD^{corr} vs. HuD KO in fAD^{corr}**		**(7) 360 DEGs HuD sv1+2 OE in fAD**		**(8) 74 DEGs-OD HuD sv1+sv2 OE in fAD vs. HuD KO in fAD**					
Canonical Pathway	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**				
Kinetochore Metaphase Signaling Pathway	6,61E-09	Kinetochore Metaphase Signaling Pathway	5,01E-14	Human Embryonic Stem Cell Pluripotency	2,57E-08	cAMP-mediated signaling	1,12E-02				
Cyclins and Cell Cycle Regulation	9,33E-08	Cell Cycle: G2/M DNA Damage Checkpoint Regulation	9,12E-10	Cyclins and Cell Cycle Regulation	2,57E-08						
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	1,20E-07	Mitotic Roles of Polo-Like Kinase	3,16E-07	Cell Cycle: G2/M DNA Damage Checkpoint Regulation	1,66E-07						
Cell Cycle: G1/S Checkpoint Regulation	3,80E-07	Cyclins and Cell Cycle Regulation	1,48E-06	Mouse Embryonic Stem Cell Pluripotency	2,09E-07						
Mitotic Roles of Polo-Like Kinase	2,57E-05	ATM Signaling	5,75E-05	Regulation of the Epithelial-Mesenchymal Transition Pathway	1,07E-06						
(4) 2032 DEGs HuD sv1+sv2 rescue in fAD^{corr} HuD KO		**(5) 920 DEGs-OD HuD sv1+ sv2 Rescue in fAD^{corr} KO vs. HuD KO in fAD^{corr}**		**(9) 1253 DEGs HuD sv1+sv2 rescue in fAD HuD KO**		**(10) 383 DEGs-OD HuD sv1+sv2 Rescue in fAD KO vs. HuD KO in fAD**		**(13) 67 DEPs HuD sv2 Rescue in fAD KO**		**(14) 103 DPPs HuD sv2 Rescue in fAD KO**	
Canonical Pathway	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**	**Canonical Pathway**	***P*-value**
Axonal Guidance Signaling	1,58E-11	Axonal Guidance Signaling	3,89E-09	Axonal Guidance Signaling	4,79E-09	Axonal Guidance Signaling	1,00E-03	Protein Kinase A Signaling	3,24E-06	Semaphorin Signaling in Neurons	5,62E-08
Molecular Mechanisms of Cancer	1,23E-09	Hepatic Fibrosis / Hepatic Stellate Cell Activation	5,62E-09	Factors Promoting Cardiogenesis in Vertebrates	3,39E-06	Calcium Signaling	1,00E-03	Role of PKR in Interferon Induction and Antiviral Response	4,37E-06	Renin-Angiotensin Signaling	5,13E-07
Hepatic Fibrosis / Hepatic Stellate Cell Activation	1,12E-08	Human Embryonic Stem Cell Pluripotency	1,70E-08	Molecular Mechanisms of Cancer	3,39E-06	Hepatic Fibrosis Signaling Pathway	2,40E-03	14-3-3-mediated Signaling	5,01E-06	Semaphorin Neuronal Repulsive Signaling Pathway	1,55E-06
Hepatic Fibrosis Signaling Pathway	1,55E-07	Hepatic Fibrosis Signaling Pathway	1,32E-07	Human Embryonic Stem Cell Pluripotency	3,39E-06	Hepatic Fibrosis / Hepatic Stellate Cell Activation	1,15E-02	BAG2 Signaling Pathway	9,12E-06	ERK/MAPK Signaling	1,91E-06
Breast Cancer Regulation by Stathmin1	2,09E-07	Kinetochore Metaphase Signaling Pathway	1,78E-07	Cyclins and Cell Cycle Regulation	3,39E-06	RhoGDI Signaling	1,15E-02	Sirtuin Signaling Pathway	1,12E-05	LPS-stimulated MAPK Signaling	3,98E-06

NOTE *P*-values are FDR-corrected

Table 4. Upstream regulator analyses of RNAseq and phosphoproteomics HuD KO, overexpression, and rescue neuronal cell lines, using IPA.

	RNAseq												phosphoproteomics					
	fAD-corrected						fAD						fAD					
KO	**(1) 6260 DEGs HuD KO in fAD^{corr}**						**(6) 3661 DEGs HuD KO in fAD**						**(11) 40 DEPs HuD KO in fAD**			**(12) 48 DPPs HuD KO in fAD**		
Upstream regulator	**Z-score**	***P*-value**	**# Targets**				**Z-score**	***P*-value**	**# Targets**				**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**
APP	**2,36**	3,83E-10	315				N.A.	2,00E-06	180				N.A.	3,90E-10	14	N.A.	1,34E-02	6
Tau	N.A.	4,53E-06	172				N.A.	7,05E-03	91				N.A.	2,80E-12	13	N.A.	6,92E-02	6
INSR	**-3,14**	6,75E-06	155				**-3,15**	9,29E-06	97				**2,21**	1,46E-05	7	N.A.	N.S	N.A.
FOXO1	N.A.	1,15E-04	145				N.A.	N.S	N.A.				N.A.	4,50E-02	3	N.A.	N.S	N.A.
OE	**(2) 609 DEGs HuD sv1+2 OE in fAD^{corr}**			**(3) 207 DEGs-OD HuD sv1+sv2 OE in fAD^{corr} vs. HuD KO in fAD^{corr}**			**(7) 360 DEGs HuD sv1+2 OE in fAD**			**(8) 74 DEGs-OD HuD sv1+sv2 OE in fAD vs. HuD KO in fAD**								
Upstream regulator	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**						
APP	N.A.	4,79E-03	37	**-2,40**	1,05E-02	16	N.A.	1,05E-02	24	N.A.	N.S	N.A.						
Tau	N.A.	3,65E-02	20	N.A.	N.S	N.A.	N.A.	N.S	N.A.	N.A.	N.S	N.A.						
INSR	N.A.	4,26E-05	27	N.A.	8,52E-03	10	N.A.	5,10E-03	15	N.A.	N.S	N.A.						
FOXO1	**-3,31**	1,14E-08	34	**-3,89**	6,68E-10	21	**-3,63**	2,06E-04	18	N.A.	N.S	N.A.						
Rescue	**(4) 2032 DEGs HuD sv1+sv2 rescue in fAD^{corr} HuD KO**			**(5) 920 DEGs-OD HuD sv1+ sv2 Rescue in fAD^{corr} KO vs. HuD KO in fAD^{corr}**			**(9) 1253 DEGs HuD sv1+sv2 rescue in fAD HuD KO**			**(10) 383 DEGs-OD HuD sv1+sv2 Rescue in fAD KO vs. HuD KO in fAD**			**(13) 67 DEPs HuD sv2 Rescue in fAD KO**			**(14) 103 DPPs HuD sv2 Rescue in fAD KO**		
Upstream regulator	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**	**Z-score**	***P*-value**	**# Targets**
APP	N.A.	7,98E-04	106	N.A.	3,32E-03	53	N.A.	5,85E-05	76	N.A.	3,11E-02	23	N.A.	2,16E-12	20	N.A.	3,62E-03	11
Tau	N.A.	4,94E-04	65	N.A.	N.S	N.A.	N.A.	N.S	N.A.	N.A.	N.S	N.A.	N.A.	2,56E-17	20	N.A.	8,60E-08	14
INSR	N.A.	3,26E-04	60	N.A.	6,11E-03	29	N.A.	7,77E-03	36	N.A.	N.S	N.A.	N.A.	1,21E-02	5	N.A.	N.S	N.A.
FOXO1	**-3,37**	9,67E-08	71	**-2,96**	1,26E-04	34	**-3,36**	3,30E-08	52	N.A.	2,97E-02	13	N.A.	N.S	N.A.	N.A.	N.S	N.A.

NOTE IPA upstream regulator analyses results for APP, TAU, insulin receptor (INSR), and FOXO1 in the differentially expressed genes (DEGs) are shown. Significant *P*-values ($P < 0.05$) and Z-scores ($Z \leq -2$ or $Z \geq 2$) indicated in bold. In green: pathways are predicted to inhibited based on direction of differential gene expression. In red: pathways are predicted to be activated based on direction of differential gene expression.

CHAPTER 9

Conclusions & Future Perspectives

In this thesis, the signaling mechanisms of receptor tyrosine kinases were extensively studied and discussed. Insights in mechanism and novel nodes were presented. This chapter concludes the findings of these studies, and discusses their impact on drug discovery and potential future research directions.

Conclusions & Future perspectives

EGFR and the RTK signaling network

The incredible resilience of EGFR and other members of the RTK family makes for an incredibly complex mechanism that is challenging to artificially modulate. Our work on the EGFR signaling network has highlighted this resilience as well as the ability to use our approach to identifiy novel associations that lead to a better understanding of these signaling networks. As our work on AXL and PTPRJ illustrated as well, there are as many potential ways to intervene and affect EGFR signaling as there are ways for EGFR to circumvent this. One important aspect we highlighted in these studies is to make a link between signaling response and phenotype. Too often, studies focus on one or the other, but to truly understand these networks, one should aim to create a holistic view to determine whether any of the signaling nodes are in fact responsible for a phenotypic outcome, or which mechanism is underlying that particular outcome.

Phosphoproteomics and signaling studies

Quantitative MS-based phosphoproteomics has emerged as a powerful technique to identify oncogenic signaling networks in a variety of biological systems. While global phosphoproteomics can provide an impressive overview of the phosphorylation state of thousands of proteins within a given sample, this analysis method may fail to identify and quantify low-abundance sites, e.g., pTyr and many oncogenic kinase substrates, that may be critical regulators of biological function, especially in cancer cell signaling. Although more technically challenging to detect due to sample losses, these data provide critical and unique

insight into activated tyrosine kinase signaling networks. As tools and methods continue to advance, e.g. higher affinity reagents for enrichment, targeted approaches for monitoring selected sites of interest, and instrumentation with increased sensitivity, these technique are becoming more widely accessible.

One potential application of phosphoproteomics would be in interrogating signaling networks in patient derived tumor models, combining signaling and genetic data to inform combination therapies to advance efforts in personalized medicine. As methods become larger-throughput thanks to development of auto-sampler platforms and commercial columns, implementing phosphoproteomics in an industry setting is more conceivable every day.

Final notes

In conclusion, phosphoproteomics in particular when applied for mechanistic purposes holds great potential to expand our understanding of signaling mechanisms, identify novel biomarkers, and informing therapeutic strategy in cancer and beyond. As the field continues to accelerate and is pushed to focus on "translatable" research in PDX samples, tumor tissues and other human samples, my hope is that we do not forget about the power of fundamental research, and the replicability of in vitro studies (human samples often show great heterogeneity in signaling response making it challenging to distill meaningful insight). Furthermore, we should not get lost in inefficient giant library screens to find biomarkers for our diseases with no regards for underlying mechanism. Rather, we should aim for a deep mechanistic understanding that informs biomarker identification and validation studies. This is something that the approaches outlined in this thesis can help to achieve, and hopefully will continue to be used to pave the way for many therapeutic discoveries to be made.

Summary

This thesis has provided novel insights in the signaling mechanisms of various members of the receptor tyrosine kinase family, with EGFR being our main node of interest. EGFR is part of the ErbB subfamily of RTKs and often dysregulated in many different diseases, in particular cancers. The many therapies that have been developed over the past few decades have had various successes, stressing the need for an improved understanding of the underlying signaling mechanisms. **Chapter 1** introduces the main subjects of this thesis, including RTK signaling, EGFR function, tyrosine phosphorylation and phosphoproteomics. The latter is extensively discussed in **Chapter 2**, which highlights the recent advances in tools and methods that have enabled valuable discoveries to be made using phosphoproteomic approaches as it relates to cancer. One particularly powerful approach that can provide mechanistic insight when combined with 'omics approaches, are Y-to-F mutational studies. As described in **Chapter 3**, these site-specific mutations allow for systematic interrogation of function of tyrosine sites, and have been used for a variety of applications. This review discusses valuable insight gained so far using such an approach, and the potential when combined with other advanced tools to link signaling and phenotype together. **Chapter 4** describes the main thesis project, where a Y-to-F mutational approach was combined with phosphoproteomics and phenotypic assays to obtain a holistic view of the function of the individual tyrosines on the C-terminal tail of EGFR. Using various clustering and regression approaches, novel network nodes were identified as well as phenotype predictors that could serve as biomarkers in future applications. One of our earlier model systems for this study contained a GFP tag at the end of the C-terminal tail, causing signaling data to deviate from non-GFP labeled constructs. As the use of these types of large

reporters should be carefully considered in highly dynamic assays such as signaling studies, these data are presented and discussed in **Chapter 5**. Our Y-to-F approach combined with phosphoproteomics is applicable beyond EGFR, as shown by our similar study for AXL, another RTK often dysregulated in cancer. **Chapter 6** discusses the signaling and phenotypic consequences of stimulation with AXL stimulants and EGFR inhibitors for various mutant isoforms. Beyond understanding direct RTK mechanism, others aim to regulated RTKs through neighboring proteins, e.g. phosphatases. One such phosphatase, PTPRJ, is known to be inactive as a dimer, prompting research to find ways to shift the equilibrium towards monomers in an attempt to downregulate EGFR in cancer. One mutation that prevents dimerization of PTPRJ is G983L, yet the mechanism through which PTPRJ downregulates EGFR is poorly understood. **Chapter 7** evaluates potential mechanisms through a phosphoproteomic approach. As highlighted in **Chapter 8**, phosphoproteomics can have a profound impact beyond cancer. Using a combination of genomics and phosphoproteomics, ELAVL4 was identified as a potential novel drug target for Alzheimer's disease. In summary, this thesis highlights the power of phosphoproteomics in identifying signaling mechanisms and biomarkers in RTKs and their interactors, in cancer and beyond. **Chapter 9** provides a thoughtful overview of conclusions from this work, as well as a discussion of potential for this work moving forward. In all, these results may guide future research and drug development efforts, ultimately enabling more effective modulation of signaling networks in disease.

List of publications

Gerritsen JS, White FM, *Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells*, Expert Review of Proteomics,18(8), 2021, 661-674.

Gerritsen JS, Faraguna JS, Bonavia R, Furnari FB, White FM, *Predictive data-driven modeling of C-terminal tyrosine function in the EGFR signaling network*. Manuscript submitted to Molecular Systems of Biology, March 2022.

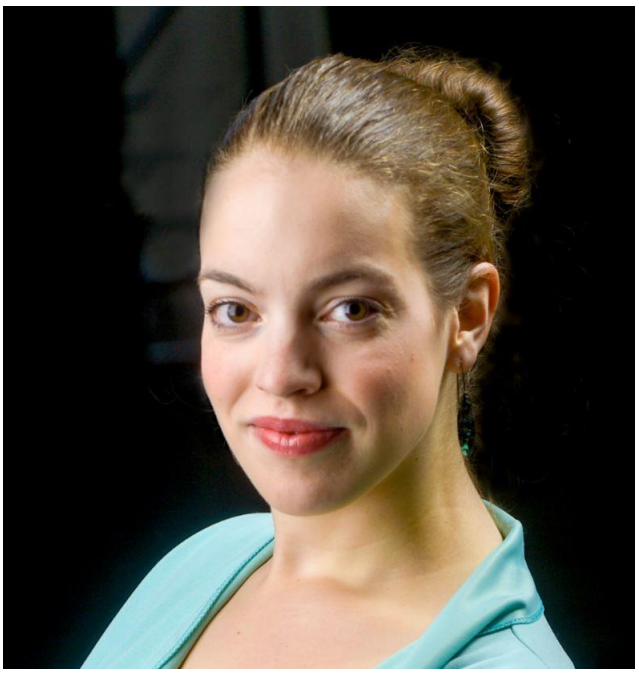
Gerritsen JS*, Creixell M*, White FM, Meyer AS, *Elucidating phosphorylation-mediated signaling networks in receptor tyrosine kinases through Y-to-F mutational studies*. Manuscript in preparation. *These authors contributed equally.

Gerritsen JS, White FM, *Using Green Fluorescent Protein in Signaling Studies*. Manuscript in preparation.

Van der Linden RJ, Gerritsen JS, Liao M, Widomska J, Pearse R, White FM, Franke B, Young-Pearse TL, Poelmans G, *RNA-binding protein ELAVL4/HuD ameliorates Alzheimer's disease-related molecular changes in human iPSC-derived neurons*. Manuscript submitted to Molecular Neurodegeneration, October 2021.

Creixell M, Gerritsen JS, Majumder A, Haura EB, White FM, Meyer AS, *Mapping of AXL signaling networks driving resistance to EGFR inhibition in lung cancer*. Manuscript in preparation.

Bibliography



Jacqueline started her scientific career at the Vrije Universiteit Amsterdam where she pursued her B.Sc. in Pharmaceutical Sciences. During her Bachelor studies, she participated in the honours program, and spent a semester abroad at Copenhagen University where she attended graduate level classes, including Advanced Manufacturing of Pharmaceuticals, Biopharmaceuticals: Formulation of peptides and proteins, Synthetic organic chemistry, and Pharmaceutics & Drug development. Her Bachelor thesis work focused on generating and AVI-tagging bivalent nanobodies targeting CXCR4 and CXCR7 receptors, two chemokine receptors dysregulated in a variety of diseases. Motivated by these experiences, she pursued her Master's degree in Biopharmaceutical Sciences in the department of Medicinal Chemistry at Leiden University, where she investigated genetic influences on GPCR drug responses using cell lines from the Netherlands Twin Registry and the XCelligence Assay platform. During her second year, she jumped at an opportunity to spend a semester at MIT, researching crosstalk between GPCRs and RTK in the lab of professor Douglas Lauffenburger, under direct supervision of Dr. Annelien Zweemer. A mutually productive and pleasant collaboration lead to her starting the PhD program in September 2016, followed by her joining Forest White's research group in the Koch Institute to pursue her degree. Besides research, Jacqueline has been involved in several other on campus initiatives, including the BE REFS (peer to peer coaching for conflict management), BE Communication lab (peer to peer coaching for science communication), the GSC Diversity Equity and Inclusion committee (As international student representative) and the BE Retreat committee.

Acknowledgements

To show my genuine appreciation for the people that have helped me achieve this milestone at MIT, I have to start by thanking the people that helped me get to MIT in the first place. Prof. Ad IJzerman who I worked with at the department of Medicinal Chemistry during my Master studies at Leiden University, connected me with one of his stellar former graduate students Dr. Annelien Zweemer, who was a postdoc at the time in Prof. Doug Lauffenburger's lab. They allowed me to get a taste of doing research in the unique environment that is MIT, and if it wasn't for Doug's encouragement and support, I would not have been able to start the BE PhD program one year later. Doug is one of those rare people that truly care about their students, and saw potential in me beyond my own dreams.

Starting the PhD program, I could not have been more fortunate to meet Forest, who has supported and taught me more than one could hope for. He has created such a welcoming environment in the lab which I was immediately drawn to, and ensured me that as long as students are good people and self-motivated, everything else can be taught. His never wavering enthusiasm, and extraordinary ability to make one feel good about even the most failed experiment, was a blessing and I could not be more grateful for having been able to spend my time under his guidance.

To my other committee member Prof. Sebastian Lourido, a huge thanks for the support on both the technical AND practical aspects of my research. You've given invaluable feedback at critical moments during my time here, and I respect your knowledge and appreciate your patience so much.

To my lab colleagues, current and former, thank you for all the collaboration, training, support and adventures we've been through together. From learning how to operate the mass spec and making analytical columns, to traveling to India with Kristina and Lauren to attend Vibhuti's wedding, it has been a pleasure getting to know each and every one of you. Daniel, Vibhuti, Antje, Kristina, Nader, Lauren, Ishwar, Jen, Jason, Tigist, Angela, Cameron, Sean, Elizabeth, Alicia, Owen and Yufei, my deepest thanks and all the best of luck to you in your future endeavors.

Prof Frank Furnari, Dr. Rudy Bonavia, Marc Creixell, Dr. Aaron Meyer, Prof. Damien Thevenin, Dr. Eden Sikorski, Will Hart, Prof. Matthew Lazarra, Robert van der Linden, Prof. Geert Poelmans and more: thank you for the opportunity to collaborate with and learn from you over the years. Thank you for putting your faith in me, allowing me to contribute in such meaningful ways to your projects and sharing your resources with me.

Other members of the MIT BE Community include the MIT BE 2016 cohort, which struggled together through the first year and continued to stay in touch through occasional dinners and check-ins. Prerna Bhargava, who lead the BE Communication lab, for believing in me and giving me that ultimate opportunity to be a part of the Comm lab. I appreciate your support more than you know. Sean Clarke, Dan Darling, Gerri Powers, thank you for always encouraging me

and your support throughout my journey here. I have loved every minute of working with you, whether through the BE Comm lab, BE Retreat committee or unrelated. The department would not be the same without your love and care.

Another big thank you to the amazing staff running the Koch Institute Core facilities, allowing me to get up and running quickly and always ready to lend a hand. Jamie Cheah, Jeffrey Kuhn, Glenn Paradis, Michelle, Mike, I appreciate all you do on the scenes and behind the scenes to make our experiments run smoothly and help drive innovative research.

To the people who have provided me with a home away from home by basically adopting me into the Boston Dance Community, Jenna and Luis, I am eternally grateful for finding you and Bachata as it has been my joy and air for the past 6+ years. Thanks to your love and guidance, I found myself a dance family and discovered a passion that I will always cherish. To many of my long-time team members and dancers who have become great friends over the years: Janna, Welina, Tigist, Yumi, Eleni, Jackie, Patricia, Alejandro, Jose, Kevin, Amanda, Edward and many others; thank you for always being there to lift me up and share a dance with me.

Love knows no boundaries, as proven by my dear friends from the Netherlands, many of which have visited over the years, or have offered a listening ear or word of encouragements when needed: Susan, Celine, Karlijn, Elaine, Laura, Sandra, Lisanne, Esther, Myrthe and many others, I miss you all and appreciate your continued support the most!

To my family, who has always made me feel like I can do anything I set my mind to and has supported this adventure even when it meant I needed to cross the Atlantic for such an extended amount of time, thank you from the bottom of my heart. Your selflessness and care have allowed me to experience something I never dreamt possible, and set me up for life. I will forever be in your debt. Ria and Georges, Kim and Klaas Jan, Chris and Giel, Manon and Julie, Anna and Joris, Tim and Maya, thank you for believing in me, being proud of me and being patient with me.

To my dad, even though this journey has made us live far apart, I feel like we've grown closer over the years and I couldn't be happier for you to have met Alexandra. I love how we've stayed in touch and how often you've made the effort to come out and visit me, you're the real MVP! This thesis is dedicated to you and mom, as both of you've always inspired me to be braver and better and empowered me to be the best version of myself. Thank you for always expressing how proud you are of me and for all the valuable lessons over the years. I love you.

And finally, a big thank you to the three beings that have supported me on the daily over the years: Amado and my two fur babies Maya and Troy. Thank you for always providing a listening ear or a laugh when I needed it the most.