



TECHNISCHE UNIVERSITÄT MÜNCHEN

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und
Umwelt

Lehrstuhl für Proteomik und Bioanalytik

**Profiling the phosphotyrosine interactome of receptor
tyrosine kinases by affinity enrichment-mass spectrometry**

Runsheng Zheng

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Dieter Langosch

Prüfer der Dissertation: 1. Prof. Dr. Bernhard Kuster

2. Prof. Dr. Pascal Falter-Braun

Die Dissertation wurde am 24.01.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 12.03.2019 angenommen.

Contents

| | |
|---|----|
| Abstract..... | 1 |
| Chapter 1 Introduction | 3 |
| 1.1 The human receptor tyrosine kinases..... | 3 |
| 1.1.1 Family members and cellular impact..... | 3 |
| 1.1.2 Activation of cell signaling pathways..... | 6 |
| 1.1.3 Mutation of receptor tyrosine kinases | 8 |
| 1.2 Phosphotyrosine-binding domains | 9 |
| 1.2.1 The Src homology 2 (SH2) domain..... | 9 |
| 1.2.2 The phosphotyrosine-binding (PTB) domain | 11 |
| 1.2.3 The protein tyrosine phosphatase (PTP) domain and PTP-like domain | 13 |
| 1.2.4 The Hakai-tyrosine binding (HYB) domain | 14 |
| 1.2.5 The C2 domain of PRKCD and PRKCO..... | 15 |
| 1.2.6 Other binding domains or proteins | 16 |
| 1.3 Approach for studying the interaction between pY-peptide and proteins | 18 |
| 1.3.1 Yeast two-hybrid system..... | 18 |
| 1.3.2 Co-immunoprecipitation..... | 20 |
| 1.3.3 Affinity enrichment-mass spectrometry assay | 21 |
| 1.4 Objective and outline | 23 |
| Chapter 2 Experimental procedures..... | 24 |
| 2.1 Cell culture and lysis..... | 24 |
| 2.2 Peptide sequence design | 25 |
| 2.3 Peptide synthesis | 26 |
| 2.3.1 Phosphotyrosine peptides for the AE-MS assay..... | 26 |
| 2.3.2 Peptides for the competition assay | 27 |
| 2.4 Generation of the affinity matrices..... | 29 |
| 2.5 Affinity enrichment (AE) assay | 30 |
| 2.6 Competition pull-down assays | 31 |
| 2.7 In-gel protein digestion | 32 |
| 2.7.1 pY-peptide affinity enrichment samples | 32 |
| 2.7.2 Cell line full proteome..... | 32 |
| 2.8 Peptide fractionation | 33 |
| 2.9 LC-MS/MS analysis | 34 |
| 2.9.1 Interactome screening and full proteome profiling | 34 |
| 2.9.2 Competition assay..... | 35 |
| 2.10 Protein identification and quantification | 36 |
| 2.11 Data analysis | 37 |
| 2.11.1 Data normalization and estimation of enrichment factors | 37 |
| 2.11.2 Database mapping | 40 |
| 2.11.3 Motif analysis | 41 |
| 2.11.4 Interactome network assembly | 42 |
| 2.11.5 Competition assay..... | 43 |
| Chapter 3 Results and discussion | 44 |

| | | |
|-----------|--|-----|
| 3.1 | Analysis of the synthetic peptides | 45 |
| 3.2 | Global profiling of the RTK interactome by AE-MS..... | 48 |
| 3.2.1 | Identification and quantification of significant interactors | 49 |
| 3.2.2 | Domain analysis of interactors | 51 |
| 3.3 | Analysis of the SH2 or PTB domain-containing interactors | 52 |
| 3.3.1 | Binding specificity and motif analysis..... | 52 |
| 3.3.2 | Comparison with public interactome databases..... | 59 |
| 3.3.3 | Interactome of wild-type tyrosine residues | 63 |
| 3.3.4 | Interactome of hot-spot mutations..... | 70 |
| 3.3.5 | Interactome of mutation-gain tyrosine residues..... | 72 |
| 3.4 | Non-SH2 or PTB domain-containing interactors..... | 76 |
| 3.4.1 | Interactors with HYB or C2 domain | 76 |
| 3.4.2 | Interactors with PTP or PTP-like domain | 78 |
| 3.4.3 | Other unclassified interactors | 81 |
| 3.5 | Validation of the interactome by competition assay..... | 83 |
| 3.5.1 | Interactor-centric selection of representative pY-peptides | 84 |
| 3.5.2 | Proof of concept | 85 |
| 3.5.3 | Re-elucidation of known peptide-protein interactions | 88 |
| 3.5.4 | Beyond the known interactors of phosphotyrosine residues | 93 |
| 3.5.5 | Identification of inconsistent interactors of phosphotyrosine residues | 95 |
| 3.5.6 | Exploration of unreported interactors of phosphotyrosine residues | 100 |
| 3.5.7 | The potential function of mutation-gain tyrosine residues | 103 |
| 3.5.8 | Beyond the SH2 or PTB domain-containing interactors..... | 107 |
| Chapter 4 | General discussion | 112 |
| 4.1 | Exploration of interactors | 114 |
| 4.2 | Exploration of pY-peptides..... | 117 |
| Chapter 5 | Outlook | 119 |
| | Abbreviations | 120 |
| | References | 122 |
| | Acknowledgment | 129 |
| | Curriculum Vitae | 130 |
| | Conference contribution and publication..... | 131 |
| | Appendix | 132 |

Abstract

The human proteome contains 58 receptor tyrosine kinases (RTKs). These are involved in multiple cell signaling events that influence cell responses such as cell proliferation and cell-cell communication. The activation of an RTK usually induces dimerization of the receptor, followed by phosphorylation of cytoplasmic tyrosine residues. These residues recruit signaling proteins such as Src-homology-2 (SH2) or phosphotyrosine-binding (PTB) domain-containing proteins and subsequently trigger downstream cell signaling pathways. Aberrant regulation of RTKs by either overexpression or genetic mutations that flank functional tyrosine residues are known to cause serious disorders such as cancer. Therefore, profiling the phosphotyrosine interactome of RTKs can aid in further understanding the background biology and subsequent development of disease. Ultimately, such information will then promote the investigation and development of new treatment regimens for the associated disorders.

In this study, a systematic quantification of the interactome of 1,144 unique wild-type and mutant phosphotyrosine residues was performed by affinity enrichment-mass spectrometry (AE-MS). Briefly, the synthetic phosphotyrosine peptides (pY-peptides) were incubated with cell lysates to enrich interacting proteins. This was followed by protein identification and quantification via a 'bottom-up' proteomic strategy. To cross-validate novel interactors, the resultant AE-MS data was mapped to multiple databases including PhosphoSitePlus, BioGRID, PeptideAtlas, and ProteomicsDB. Sixty-seven selected peptides were purified, and 300 competition assays performed to validate novel interactors (with or without known domains). The dissociation constant (K_d) and half maximal inhibitory concentration (IC_{50}) values were also determined.

From 6,862 proteins, this quantitative proteomic approach identified 2,184 potential interactors. Eighty-six of the proteins contained SH2 or PTB domains that occupied approximately 66% of the iBAQ (intensity-based absolute quantification) intensity. The significant enrichment of proteins with known domains confirmed the validity of the AE-MS approach, and also indicated the presence of novel interactors without known domains. The signal transducer and activator of transcription 5B (STAT5B) and 1-phosphatidylinositol 4, 5-bisphosphate phosphodiesterase gamma-1 (PLCG1) were the most predominant interactors; thus demonstrating the shared signaling capabilities of all RTKs. The AE-MS study also revealed that the E3 ubiquitin-protein ligase, CBL-B, can specifically bind to 6 phosphotyrosine residues with the motif [x][x][x][x][R][pY][x][x][x][P][x]; indicating the existence of a specific negative regulatory mechanism for RTKs during cell signaling. Such negative regulation machinery was also apparent through the novel-identified interaction with protein tyrosine phosphatases. In addition to altering cell signaling cascades in multiple RTKs, mutant pY-peptides can potentially enhance signaling intensity. For example, with the greatest number of mutation-gain tyrosine residues in the FGFR family, FGFR2 can exclusively interact with kinases such as SRC and ZAP70.

Mapping the AE-MS data set to multiple databases enabled the discovery of potential new signaling events, such as the role of INSR_pY1149 and INSR_pY1185 in the Jurkat cell line. Furthermore, motif analysis indicated that the AE-MS assay could determine the binding features to novel interactors. For example, the binding motif of STAT2 (containing an SH2 domain) was defined as [x][x][x][x][x][pY][x][x][Q/R][x][x]. Moreover, 5-formyltetrahydrofolate cyclo-ligase (MTHFS) that lacks a classical binding domain was significantly enriched by the motif [x][x][x][F/Y][x][pY][x][x][x][x][x]. Finally, the dose-dependent competition pull-down assays further validated the affinity of various novel interactors to pY-peptides. These included the interaction between IGF1R_pY1281 and MTHFS with the promising K_d value of 2 μ M.

Overall, this AE-MS study provides a comprehensive view of the interactome of all phosphotyrosine residues present in RTKs and how the interactome can alter when specific mutations occur. For the future, the landscape generated by this interactome has vast potential to generate numerous new biological hypotheses and drive treatment of disease.

Keywords: cell signaling, receptor tyrosine kinases, phosphorylation, phosphotyrosine peptide, proteomics, interactome, affinity enrichment-mass spectrometry, protein-protein interaction, peptide-protein interaction, protein domain, SH2 domain, PTB domain, PID domain, hot-spot mutation, mutation-gain tyrosine

Zusammenfassung

Das humane Proteom umfasst 58 Rezeptor-Tyrosinkinasen (RTKs). Diese sind an mannigfaltigen, zellulären Signalwegen beteiligt, die Zellantworten wie Proliferation und Zell-Zell-Kommunikation beeinflussen. Die Aktivierung einer RTK induziert normalerweise eine Dimerisierung des Rezeptors, gefolgt von der Phosphorylierung cytoplasmatischer Tyrosinreste. Diese Tyrosinmotive rekrutieren Signalproteine mit Src-Homology-2 (SH2) oder Phosphotyrosin-bindenden (PTB) Domänen und induzieren damit nachgelagerte Signalwege. Es ist bekannt, dass eine aberrante Regulation von RTKs, hervorgerufen durch Überexpression oder Genmutationen der flankierenden Bereiche der Tyrosinreste, schwerwiegende Krankheiten wie Krebs verursacht. Ein Profiling des Phosphotyrosin-Interaktoms von RTKs kann zu einem besseren Verständnis des biologischen Hintergrunds und der Entwicklung von Krankheiten beitragen. Letztendlich können solche Informationen die Erforschung und Entwicklung neuer Behandlungsregime für die entsprechenden Erkrankungen unterstützen.

In dieser Studie wurde eine systematische Quantifizierung des Interaktoms von 1.144 spezifischen Wildtyp- und mutierten Phosphotyrosinresten durch Affinitätsanreicherung kombiniert mit Massenspektrometrie (AE-MS) durchgeführt. Um interagierende Proteine anzureichern, wurden synthetische Phosphotyrosinpeptide (pY-Peptide) mit Zelllysaten inkubiert. Anschließend wurden diese Proteine über einen proteomischen „bottom-up“-Ansatz identifiziert und quantifiziert. Zur Validierung neuartiger Interaktoren wurden die resultierenden AE-MS-Daten mit mehrere Datenbanken einschließlich PhosphoSitePlus, BioGRID, PepspotDB und ProteomicsDB abgeglichen. Darüber hinaus wurden 67 gereinigte pY-Peptide ausgewählt und 300 Wettbewerbsassays durchgeführt, um neuartige Interaktoren (mit oder ohne bekannte Domänen) zu validieren. Zusätzlich wurden K_d- (Dissoziationskonstante) und IC₅₀-Werte (halbmaximale Hemmkonzentration) bestimmt.

Bei diesem quantitativen Proteomik-Ansatz wurden aus 6.862 Proteinen 2.184 potentielle Interaktoren identifiziert. Sechszwanzig dieser Proteine enthielten SH2- oder PTB-Domänen, die ungefähr 66% der Intensität der intensitätsbasierten absoluten Quantifizierung (iBAQ) einnahmen. Die signifikante Anreicherung von Proteinen mit bekannten Domänen bestätigte generell die Validität des AE-MS-Ansatzes und wies darüber hinaus auf neue Interaktoren ohne bekannte Domänenmotive. Der Signal transducer and activator of transcription 5B (STAT5B) und PLCG1 (1-Phosphatidylinosit 4, 5-bisphosphatphosphodiesterase-gamma-1) waren vorherrschende Interaktoren - ein Indiz für gemeinsame Strategien bei Signalwegen in der RTK-Familie. Die AE-MS-Studie zeigte auch, dass die E3-Ubiquitin-Protein-Ligase CBL-B spezifisch an sechs Phosphotyrosinreste mit dem Motiv [x][x][x][x][R][pY][x][x][x][P][x] binden kann, ein Hinweis auf einen spezifischen, negativen regulatorischen Mechanismus für RTKs in der Signaltransduktion. Ein solch negativer Regulationsmechanismus wurde auch für die unlängst identifizierte Wechselwirkung mit Proteintyrosinphosphatasen apparent. Zusätzlich zur Änderung von Signalkaskaden können mutierte pY-Peptide in RTKs die Signalintensität potenziell erhöhen. Beispielsweise kann innerhalb der FGFR-Familie ausschließlich FGFR2 durch die über Mutation generierten Tyrosinreste mit Kinasen wie SRC und ZAP70 interagieren.

Durch Abgleich der AE-MS-Daten zu mehreren Datenbanken konnten neue potenzielle Signalereignisse nachgewiesen werden, z. B. die Rolle von INSR_pY1149 und INSR_pY1185 in der Zelllinie Jurkat. Darüber hinaus zeigte die Analyse der Bindemotive, dass der AE-MS-Assay Eigenschaften neuartiger Interaktoren aufzeigen kann. Beispielsweise wurde das Bindemotiv von STAT2, das eine SH2-Domäne enthält, mit [x][x][x][x][x][pY][x][x][Q/R][x][x][x][x] definiert. Darüber hinaus wurde MTHFS (5-Formyltetrahydrofolat-Cyclo-Ligase), der über keine klassische Bindedomäne verfügt, signifikant über das Motiv [x][x][x][F/Y][x][pY][x][x][x][x][x].angereichert. Schließlich bestätigten die dosisabhängigen Wettbewerbs-Pulldown-Tests die Affinität verschiedener neuer Interaktoren von pY-Peptiden. Dazu zählt die Interaktion zwischen IGF1R_pY1281 und MTHFS mit einem bemerkenswerten K_d-Wert von 2 µM.

Insgesamt präsentiert diese AE-MS-Studie einen umfassenden Überblick über das Interaktom aller in RTKs vorkommenden Phosphotyrosinreste und wie sich das Interaktom bei bestimmten Mutationen verändern kann. Für die Zukunft bietet dieses erweiterte Interaktom enormes Potenzial, um zahlreiche neue biologische Hypothesen zu erzeugen und die Therapie von Krankheiten zu unterstützen.

Schlüsselwörter: Zelluläres Signalwege, Rezeptortyrosinkinasen, Phosphorylierung, Phosphotyrosin-Peptid, Proteomik, Interaktom, Affinitätsanreicherungs-Massenspektrometrie, Protein-Protein-Interaktion, Peptid-Protein-Interaktion, Proteindomäne, SH2-Domäne, PTB-Domäne, PID-Domäne, Hot-Spot-Mutation Mutationsgewinn-Tyrosin

Chapter 1 Introduction

1.1 The human receptor tyrosine kinases

1.1.1 Family members and cellular impact

Since the discovery of the first receptor tyrosine kinases (RTKs)¹⁻⁴ over a quarter of a century ago, the human proteome is currently known to contain 58 RTKs. These proteins are cell-surface receptors that can transmit extracellular signals into the cytoplasmic region of the cell after activation by various ligands.

Consisting of 20 sub-families⁵ (Table 1), all RTKs (except AATK, LMTK3, and STYK1) share a similar structure (Figure 1). Typically, RTKs are comprised of an extracellular ligand-binding region, a transmembrane helix, and a cytoplasmic region that consists of a juxtamembrane regulatory region, a protein kinase domain and a C-terminal tail¹⁻⁴. There are numerous tyrosine residues located in the intracellular region, and these can act as a central hub for signaling transduction when an RTK is activated⁶.

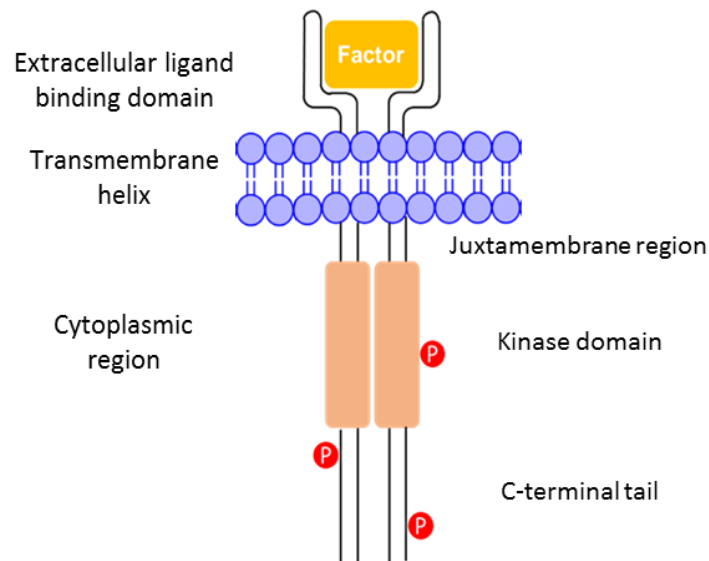


Figure 1 | Schematic representation of the general structure of receptor tyrosine kinases

Emerging studies have provided evidence that RTKs are critical regulators of normal cellular processes that include cell proliferation, migration, and differentiation^{7,8}. Additionally, it is also known that the RTKs are highly-correlated with cancer development and progression⁶. For example, the epidermal growth factor receptor (EGFR) displays a high level of expression in 62% of non-small-cell lung cancer (NSCLC) cases, and expression is usually associated with poor

prognosis^{9–11}. Therefore, resolving the cell signaling pathways triggered by aberrant RTK activation would significantly contribute to understanding health and disease.

Table 1 | General information of 58 receptor tyrosine kinases

| UniProt Entry | UniProt Entry Name | Gene Names | Cytoplasmic Positions | Subfamily |
|---------------|--------------------|---|-----------------------|-----------|
| Q6ZMQ8 | AATK | AATK, AATYK, KIAA0641, LMR1, LMTK1 | 0–1374 | LMR |
| Q9UM73 | ALK | ALK | 1060–1620 | ALK |
| P30530 | AXL | AXL, UFO | 473–894 | Axl |
| P07333 | CSF1R | CSF1R, FMS | 539–972 | PDGF |
| Q08345 | DDR1 | DDR1, CAK, EDDR1, NEP, NTRK4, PTK3A, RTK6, TRKE | 439–913 | DDR |
| Q16832 | DDR2 | DDR2, NTRKR3, TKT, TYRO10 | 422–855 | DDR |
| P00533 | EGFR | EGFR, ERBB, ERBB1, HER1 | 669–1210 | ErbB |
| P21709 | EPHA1 | EPHA1, EPH, EPHT, EPHT1 | 569–976 | Eph |
| Q5JZY3 | EPHA10 | EPHA10 | 587–1008 | Eph |
| P29317 | EPHA2 | EPHA2, ECK | 559–976 | Eph |
| P29320 | EPHA3 | EPHA3, ETK, ETK1, HEK, TYRO4 | 566–983 | Eph |
| P54764 | EPHA4 | EPHA4, HEK8, SEK, TYRO1 | 570–986 | Eph |
| P54756 | EPHA5 | EPHA5, BSK, EHK1, HEK7, TYRO4 | 595–1037 | Eph |
| Q9UF33 | EPHA6 | EPHA6, EHK2, HEK12 | 572–1036 | Eph |
| Q15375 | EPHA7 | EPHA7, EHK3, HEK11 | 577–998 | Eph |
| P29322 | EPHA8 | EPHA8, EEK, HEK3, KIAA1459 | 564–1005 | Eph |
| P54762 | EPHB1 | EPHB1, ELK, EPHT2, HEK6, NET | 564–984 | Eph |
| P29323 | EPHB2 | EPHB2, DRT, EPHT3, EPTH3, ERK, HEK5, TYRO5 | 565–1055 | Eph |
| P54753 | EPHB3 | EPHB3, ETK2, HEK2, TYRO6 | 581–998 | Eph |
| P54760 | EPHB4 | EPHB4, HTK, MYK1, TYRO11 | 561–987 | Eph |
| O15197 | EPHB6 | EPHB6 | 616–1021 | Eph |
| P04626 | ERBB2 | ERBB2, HER2, MLN19, NEU, NGL | 676–1255 | ErbB |
| P21860 | ERBB3 | ERBB3, HER3 | 665–1342 | ErbB |
| Q15303 | ERBB4 | ERBB4, HER4 | 676–1308 | ErbB |
| P11362 | FGFR1 | FGFR1, BFGFR, CEK, FGFBR, FLG, FLT2, HBGFR | 398–822 | FGF |
| P21802 | FGFR2 | FGFR2, BEK, KGFR, KSAM | 399–821 | FGF |
| P22607 | FGFR3 | FGFR3, JTK4 | 397–806 | FGF |
| P22455 | FGFR4 | FGFR4, JTK2, TKF | 391–802 | FGF |
| P17948 | FLT1 | FLT1, FLT, FRT, VEGFR1 | 781–1338 | VEGF |
| P36888 | FLT3 | FLT3, CD135, FLK2, STK1 | 564–993 | PDGF |
| P35916 | FLT4 | FLT4, VEGFR3 | 797–1363 | VEGF |
| P08069 | IGF1R | IGF1R | 960–1367 | Ins |

| | | | | |
|--------|--------|---|-----------|-------|
| P06213 | INSR | INSR | 980–1380 | Ins |
| P14616 | INSRR | INSRR, IRR | 944–1297 | Ins |
| P35968 | KDR | KDR, FLK1, VEGFR2 | 786–1356 | VEGF |
| P10721 | KIT | KIT, SCFR | 546–976 | PDGF |
| Q8IWU2 | LMTK2 | LMTK2, AATYK2, BREK, KIAA1079, KPI2, LMR2 | 64–1503 | LMR |
| Q96Q04 | LMTK3 | LMTK3, KIAA1883, TYKLM3 | 0–1460 | LMR |
| P29376 | LTK | LTK, TYK1 | 450–864 | ALK |
| Q12866 | MERTK | MERTK, MER | 527–999 | Axl |
| P08581 | MET | MET | 956–1390 | Met |
| Q04912 | MST1R | MST1R, PTK8, RON | 979–1400 | Met |
| O15146 | MUSK | MUSK | 517–869 | MuSK |
| P04629 | NTRK1 | NTRK1, MTC, TRK, TRKA | 440–796 | Trk |
| Q16620 | NTRK2 | NTRK2, TRKB | 455–822 | Trk |
| Q16288 | NTRK3 | NTRK3, TRKC | 454–839 | Trk |
| P16234 | PDGFRA | PDGFRA, PDGFR2, RHEPDGFRA | 550–1089 | PDGF |
| P09619 | PDGFRB | PDGFRB, PDGFR, PDGFR1 | 554–1106 | PDGF |
| Q13308 | PTK7 | PTK7, CCK4 | 726–1070 | PTK7 |
| P07949 | RET | RET, CDHF12, CDHR16, PTC, RET51 | 658–1114 | Ret |
| Q01973 | ROR1 | ROR1, NTRKR1 | 428–937 | Ror |
| Q01974 | ROR2 | ROR2, NTRKR2 | 425–943 | Ror |
| P08922 | ROS1 | ROS1, MCF3, ROS | 1883–2347 | Ros |
| P34925 | RYK | RYK, JTK5A | 253–604 | Ryk |
| Q6J9G0 | STYK1 | STYK1, NOK | 0–422 | STYK1 |
| Q02763 | TEK | TEK, TIE2, VMCM, VMCM1 | 770–1124 | Tie |
| P35590 | TIE1 | TIE1, TIE | 785–1138 | Tie |
| Q06418 | TYRO3 | TYRO3, BYK, DTK, RSE, SKY, TIF | 451–890 | Axl |

1.1.2 Activation of cell signaling pathways

In general, the activation of RTKs by ligands or growth factors induces homo- or hetero-dimerization of two proximal receptors. After dimerization, the RTKs promote a precisely orchestrated auto-phosphorylation of the cytosolic tyrosine residues in the dimers to enhance the catalytic activity of the kinases¹²⁻¹⁵. Only a subset of the RTKs can form oligomers without an activating ligand. For example, insulin-like growth factor 1 receptor (IGF1R) can exist on the cell surface as disulfide-linked dimers¹⁶, and EGFR may form pre-existing oligomers before accepting an interacting ligand^{17,18}. Subsequently, the phosphorylated tyrosine residues act as docking sites to direct the organized assembly of specific substrates that contain SH2 or PTB domains.

Via the Src homology 2 (SH2) or phosphotyrosine-binding (PTB) domain, protein interactors are usually directly recruited to RTKs. Alternatively, indirect interaction occurs by binding to phosphorylated tyrosine residues of docked proteins that were phosphorylated by RTKs after association with the receptor¹⁹. These substrates can further interact with downstream signaling proteins to propagate the extracellular signal to the nucleus. Here, the corresponding cell responses are promoted (Figure 2). For example, the phosphorylation of Y1063 in vascular endothelial growth factor receptor 3 (FLT4) is essential for initiating cell survival signaling by recruiting CRK and an isoform of CRK (CRK-II). Another example is where pY1230 and pY1231, and pY1337 coordinate by interacting with growth factor receptor-bound protein 2 (GRB2) or SHC-transforming protein 1 (SHC1), respectively, to influence endothelial cell proliferation, migration, and survival²⁰. Decades of research efforts using conventional biochemical approaches have shown the existence of redundant RTK interactors that have the ability to bind to different pY residues, *e.g.*, the adaptor protein GRB2 and the phosphatidylinositol 3-kinase regulatory subunits α , β , and γ (PIK3R1, 2, 3)²¹. The promiscuous binding of these proteins creates a conundrum to identify the full spectrum of RTK interactors. How can the same initiation point with signaling proteins lead to distinct outcomes, *e.g.*, cell proliferation versus migration? It is known that this could be a consequence of expression differences of RTKs and interactors, the phosphorylation status of the tyrosine residues, and the regulatory mechanisms (positive and negative feedback loops) in cells^{6,22}.

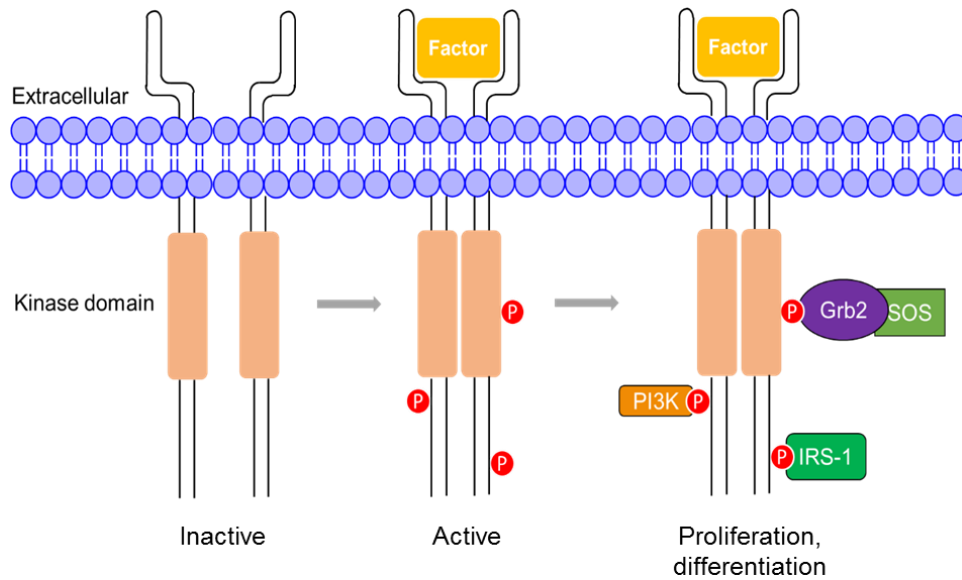


Figure 2 | The conventional activation scheme of receptor tyrosine kinases. After receiving a growth factor, two receptor tyrosine kinases dimerize, and one receptor phosphorylates the tyrosine residues on the counterpart. The phosphorylated tyrosine residues recruit different proteins containing an SH2 or PTB domain that influences the cellular response.

Is it also possible that specific interactors of pY-residues in RTKs are yet to be identified? This hypothesis is based on the following facts: (1) the current human proteome generated from different cell lines and tissue (<https://www.proteomicsdb.org/>)²³ suggests that there is a massive number of co-expressed RTKs and typical interactors thereof (containing SH2 or PTB domains); (2) several protein-protein interaction databases²¹ have curated a promising number and type of interactors to different RTKs; and (3) approximately 480 tyrosine residues from RTKs have been suggested as phosphorylated in cells and tissues²⁴ (<https://www.phosphosite.org/>); albeit evidence of functional annotation for many are still lacking. With the development of new technologies such as peptide/protein microarrays and proteomics, many more proteins have been confirmed to interact with phosphotyrosine residues^{25–28}. Despite this advancement, however, the entire interactome of RTK phosphotyrosine residues is yet to be compiled. Therefore, it is critically important to systematically reveal the affinity and specificity of all the phosphotyrosine residues in RTKs in the cell. This work would not only contribute to deconvoluting the role of all pY-residues in health, but would also enable the evaluation and improved understanding of the impact of mutations in diseases such as cancers.

1.1.3 Mutation of receptor tyrosine kinases

With the vast recent efforts in genome sequencing, many somatic mutations in human cancer have been discovered in different RTKs²⁹, including the PDGF, ErbB, and FGF receptor families. Similar to other proteins, missense and nonsense mutations can also occur in RTKs. Such mutations cause alterations in the RTK structure, *e.g.*, within the juxtamembrane region or kinase domain. When this occurs, multiple signaling pathways are initiated that affects cellular responses. For instance, mutations in the auto-inhibitory-juxtamembrane region of the KIT and PDGFR families can transform the receptors into a permanently active state that triggers signaling events that lead to the development of many cancers³⁰.

Several hotspot mutations flanking active phosphotyrosine residues in RTKs have also been linked clinically to cancer development. For example, ERBB2_pY772 is annotated as associated with various cancer types such as breast cancer³¹. A hotspot mutation of D769Y flanking this residue is known as a gain-of-function mutation, but treatment with the ERBB2-targeted inhibitor Neratinib has proven to be clinically effective³². The molecular mechanism as to how this mutation attenuates or enhances existing signaling cascades is yet to be determined; as is the effect of inhibitor treatment. In addition to the hypothesis that the hotspot mutation alters the interactors that are recruited by pY772, is it possible that the mutant residue D769Y directly recruits downstream proteins after phosphorylation? This kind of tyrosine (Y) residue replacing an amino acid by missense mutation is referred to as mutation-gain tyrosine (MGY) in this study. There are around 200 MGY residues in the RTK family³¹, and some of them are associated with drug resistance. For example, following treatment with either Quizartinib or Sorafenib, the mutation of FLT3_D835Y leads to constitutively-activated FLT3 in acute myelogenous leukemia patients^{29,33,34}. Therefore, dissecting the oncogenic mechanisms behind hotspot and MGY mutations from an interactome perspective could dramatically aid cancer treatment.

1.2 Phosphotyrosine-interacting domains

As mentioned previously, the phosphotyrosine residues of RTKs typically interact with proteins that contain an SH2 or PTB domain. Additionally, several proteins with atypical C2 or Hakai-tyrosine binding (HYB) domains have been proven to also interact with phosphotyrosine residues in RTKs. Moreover, protein tyrosine phosphatase (PTP) and PTP-like domains have been evidenced to naturally interact with phosphotyrosine residues (pYs). Last but not least, various studies including proteomics have led to the identification of novel interactors of pY-peptides or phosphorylated proteins^{35–39}. These discoveries emphasize the existence of new binding domains. The description of the binding features of these domains (or proteins) to phosphotyrosines is described below.

1.2.1 The Src homology 2 (SH2) domain

In 1986, the SH2 domain was first discovered from the insertional mutation study of v-fps/fes oncogene from the virus Fujinami sarcoma. The SH2 domain contains ~100 amino acids and is a non-catalytic module⁴⁰. Later, the domain was discovered in cytoplasmic protein phospholipase C gamma 1 (PLCG1) and Ras GTPase activating protein (RasGAP)^{41,42}. To date, ~121 SH2 domains in 111 proteins have been identified from the human genome⁴³; encompassing protein kinases, adaptor proteins, ubiquitin ligases, and protein phosphatases. It is well-known that SH2 domain-containing proteins are recruited by phosphotyrosine residues, but recruitment occurs in the context of a particular sequence motif. Additionally, it is known that the presence of SH2 domains can elicit different functions on the protein. For example, the SH2 domain in protein kinases can firstly act to retain the tyrosine phosphorylated protein, then to promote further phosphorylation of the interactors (or partner of interactors) through the kinase domain located near the SH2 domain^{44,45}.

Structurally, the representative SH2 domain of v-Src consists of a β sheets containing seven β -strands (β A- β G) that are sandwiched by another two α -helices (α A and α B) (Figure 3)^{46,47}. Located in the N-terminal half of the SH2 domain, arginine (R) 175 provides electrostatic contact with the phosphotyrosine residue (position 0), whilst the hydrophobic pocket formed between the EF and BG loop in the C-terminal half of the SH2 domain interacts with the isoleucine (I) residue in the phosphotyrosine +3 position⁴⁸. From structure and sequence

analyses, similar features have also been identified across ~70 SH2 domains. Various studies have suggested that the binding specificities of these SH2 domains to the C-terminal residues of pY-peptides are influenced by the distinct composition and configuration of the EF and BG loop

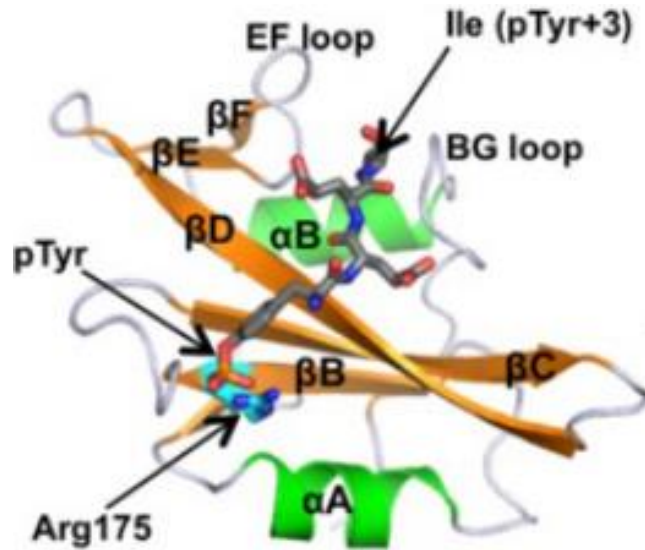


Figure 3 | The structure of the v-Src SH2 domain in a complex with the pYEEI peptide (PDB ID: 1SPS) (adapted from reference⁴⁸). Two α -helices (green) and seven β -strands (orange) are shown. Indicated by arrows on the peptide (grey sticks) are the phosphotyrosine and isoleucine residues.

Although the SH2 domains are recognized as the largest entity to bind to phosphotyrosine residues, there are still some cases where the interaction with the peptides occurs in a phosphotyrosine-independent manner. For example, SH2D1A can interact with both the phosphorylated and unphosphorylated version of the peptides similarly as the unphosphorylated peptide can insert into the same phosphotyrosine-binding pocket⁵⁵. Moreover, some SH2 domains contain secondary sites that interact with the peptides regardless of the phosphotyrosine status. This second pocket-binding mode was posited to exist between the interactions of the kinase domain of fibroblast growth factor receptor (FGFR) and the N-terminal SH2 domain of PLCG1⁵⁶. Therefore, it is known that the binding specificity of the SH2 domain can be quite variable. To fully deconvolute it, further investigation is required.

1.2.2 The phosphotyrosine-binding (PTB) domain

Initially identified in the adaptor protein SHC^{57,58}, the PTB domain is the second most crucial domain that interacts with phosphotyrosine peptides. Based on the 3D structure, the PTB domain can be further divided into three groups: phosphotyrosine-dependent IRS-like, phosphotyrosine-dependent SHC-like, and phosphotyrosine-independent DAB-like. The latter two groups are also referred to as the phosphotyrosine interaction domain (PID or PI domain) and are comprised of approximately 160 amino acids⁴⁸.

Currently, the human proteome is known to contain approximately 60 PTB domain-containing proteins. Most act as adaptors or scaffold proteins to organize signaling complexes⁵⁹. Structural studies of PTB domains have revealed that these share a folding pattern referred to as pleckstrin homology (PH) domain 'superfold'. The 'superfold' is composed of 2 orthogonal β sheets formed by seven antiparallel β strands. An α -helix stabilizes the sandwich structure at the C-terminus, whereas Shc-like and Dab1-like contain two additional α -helices at the N-terminus and between β 1 and β 2 strands⁶⁰ (Figure 4).

Unlike the SH2 domain, the PTB domain primarily binds to pY-peptides by recognizing N-terminal amino acids. The binding motif is usually NPXpY or NPXY⁶¹. In addition to phosphotyrosine-dependent binding to pY-peptides, the PTB Shc-like and IRS-like domains can also recruit phospholipids^{59,62}. It has been shown that the pY-peptide SCFTNQGpYFF in IL-2R (interleukin-2 receptor) can compete with a phospholipid to interact with the PTB domain in SHC⁶³. For convenience and unless otherwise specified, these two groups are called PTB domains in this study. Proteins containing SH2 or PTB domain are listed in Supplementary File 1, with their full names and gene names annotated.

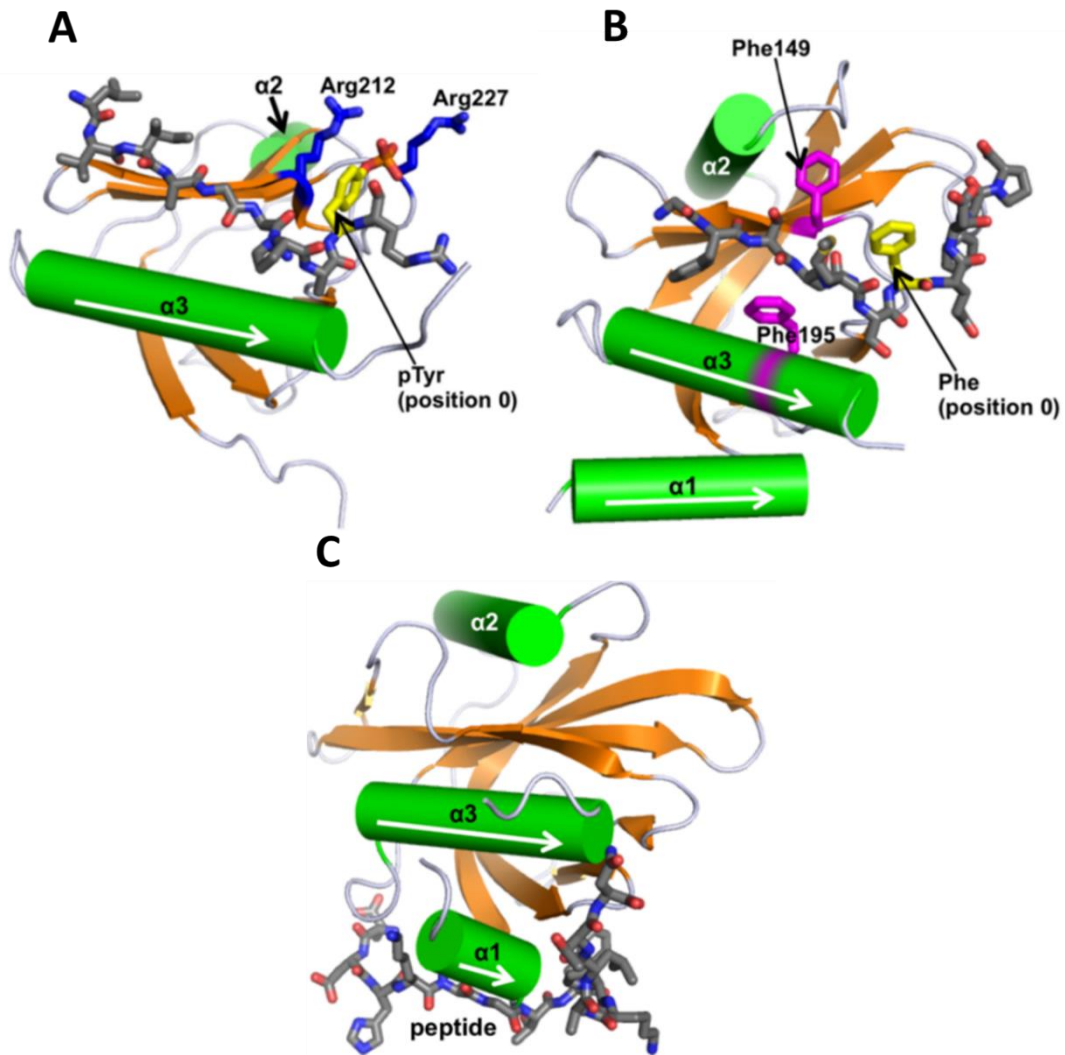


Figure 4 |The diversity of ligand recognition in the PTB domain family (adapted from reference⁴⁸). The α -helices (green) and β -strands (orange) are shown with bound peptides (grey sticks). A-C: PTB domain of IRS-1 (PDB ID: 1IRS), Numb (Dab1-like, PDB ID: 1DDM), TNS2 (Shc-like, PDB ID: 2LOZ). Numb and TNS2 have two additional α -helices.

1.2.3 The protein tyrosine phosphatase (PTP) domain and PTP-like domain

Contrary to protein tyrosine kinases that add phosphate groups to substrates, protein tyrosine phosphatases (PTP) reverse this biological process by removing the phosphate group and then positively- or negatively-regulating numerous cell signaling events triggered by tyrosine phosphorylation. This behavior indicates that together with protein tyrosine kinases, the PTPs equilibrate cellular tyrosine phosphorylation levels and monitor the rate and duration of phosphotyrosine signaling. Depending on selectivity and specificity of catalysis, PTPs are comprised of two groups: the classical phosphotyrosine-specific phosphatases and the dual-specificity phosphatases (DSPs)^{64,65}. The classical group consists of transmembrane receptor-like proteins (RPTPs) and cytoplasmic PTPs.

There are codes for approximately 100 PTP genes, and enzymatic activity is conventionally defined by the active site signature motif HCX5R. Structurally, it is well-known that the cysteine residue in the motif can interact with substrates via nucleophilicity and subsequent catalyzed dephosphorylation⁶⁶. Compared to classical PTPs, DSPs possess smaller catalytic domains that permit entrance of both phosphotyrosine residues and phosphoserine/phosphothreonine residues.

The PTP domain from the active tyrosine phosphatases PTPN1 and PTEN conservatively contains the signature motif, HCX5R, inside the catalytic domain. The cysteine residue is responsible for removing phosphate groups from the substrates⁶⁶. Interestingly, alignment of the sequences returned several PTP domain-containing proteins that lack the critical cysteine or other residues. The first protein discovered in this category was STYX. This protein has a similar sequence composition, although the cysteine residue has been replaced by glycine in the catalytic motif. Artificially mutating the glycine residue to cysteine reinstated the catalytic activity of STYX to dephosphorylate pY⁶⁷. After this discovery, many more PTP-like domains have been identified including Tensin-1 (TNS1) and Putative tyrosine-protein phosphatase auxilin (DNAJC6)⁶⁸. This PTP-like domain could be the phosphotyrosine-binding domain. However, the functions are still not yet fully characterized.

1.2.4 The Hakai-tyrosine binding (HYB) domain

In 2002, the HYB domain of the E3 ubiquitin ligase, CBLL1 (or Hakai), was identified as an atypical phosphotyrosine recognition domain⁶⁹. Structural analysis (Figure 5)⁷⁰ suggested that this domain is formed by the dimerization of two Hakai proteins. In total, these consist of 100 amino acid residues including the RING finger motifs. Six zinc ions coordinate 24 residues in the dimer, and four residues, H127, R189, H185, and Y176 from each monomer contribute to a positively-charged pocket that recruits a phosphotyrosine residue. Protein sequence alignment revealed that ZNF645 and LNX may also contain HYB domains with potentially different binding specificities.

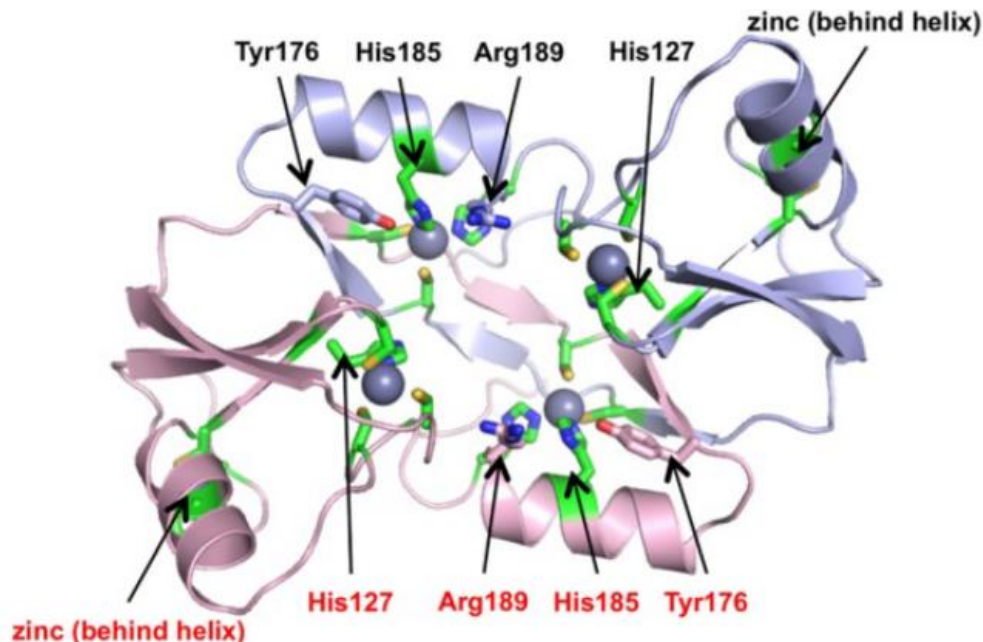


Figure 5 | The homodimeric structure of the HYB domain (PDB ID: 3VK6, adapted from reference⁴⁸). Shown are the two chains of CBLL1 (pink and purple) and the zinc ions (grey spheres). Coordination between the zinc ions and each monomer of CBLL1 is displayed as green sticks. Four residues (H127, R189, H185, and Y176) from each monomer provide a positively-charged pocket to interact with pY-peptides

1.2.5 The C2 domain of PRKCD and PRKCQ

Present in over 126 proteins, the protein C2 domain was initially believed to anchor the parent protein to the cell membrane. Together with this cell membrane targeting domain, such proteins can enzymatically function in the cell membrane region. For instance, after localizing to the cell membrane, PLCG1 promotes the production of the secondary messenger molecules diacylglycerol and inositol 1,4,5,-trisphosphate.

Currently, the C2 domain-containing proteins protein kinase C delta and theta (PRKCD and PRKCQ) are known to bind pY-peptides^{71,72}. The crystal structure shows that the C2 domain of PRKCD consists of 8 anti-parallel strands forming a β -sandwich that retains the phosphotyrosine residue of the ligand in a deep pocket. Inside the pocket, the positive charge in the side chains of the arginine (R) and lysine (K) residues is utilized to bind the negative charge of phosphotyrosine. An additional histidine (H) residue stabilizes the phenyl ring of tyrosine via a ring-stacking interaction (Figure 6).

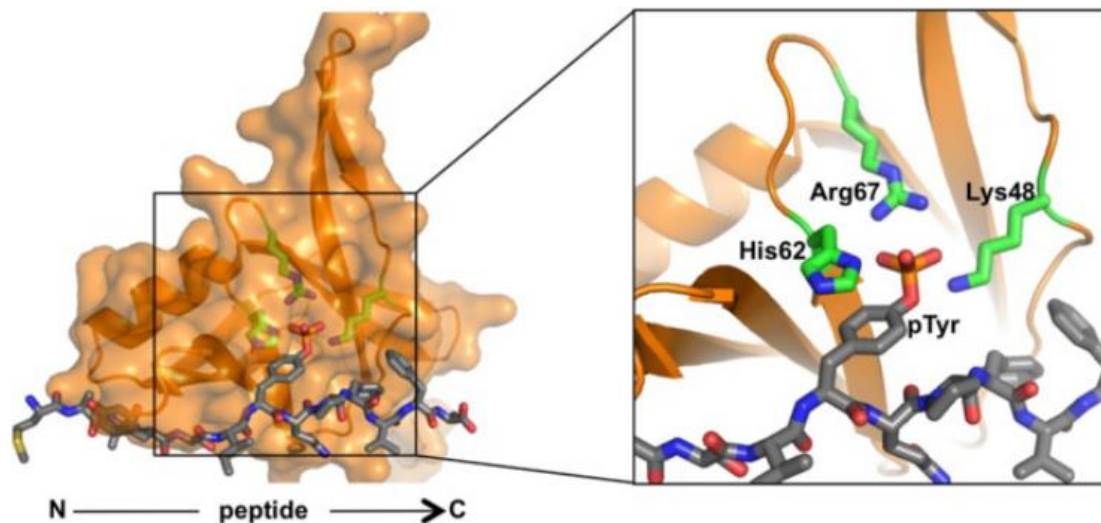


Figure 6 | The C2 domain of PRKCD interacts with a pY-peptide (grey sticks). In the right panel, the critical residues of the C2 domain (green sticks) form a positively-charged pocket to engage the phosphotyrosine residue (PDB ID: 1YRK, adapted from reference⁴⁸).

Although the similarity of the C2 domain between PRKCD and PRKCQ is only 30.9%, the essential residues of K, H, and R of PRKCD and PRKCQ are perfectly aligned. This feature infers that the proteins share a similar capacity to recruit pY-peptides but most probably with different specificities.

1.2.6 Other binding domains or proteins

Primarily through contributions from the development of peptide microarrays and proteomics, a few other domains/proteins that recruit pY-peptides have been uncovered. Such discoveries address the existence of unexplored protein regions that bind phosphotyrosine residues; however, whether these novel binding regions can interact with pY-peptides in RTKs is still unclear. The discoveries and functional annotation of these interactions are described below.

The M2 isoform of pyruvate kinase muscle (PKM) is known to selectively interact with tyrosine-phosphorylated peptides, and the K433 residue located near the binding pocket of fructose-1,6-bisphosphate (FBP) is vital for this interaction (Figure 7). The binding of pY-peptides in this pocket promotes the release of FBP resulting in the inhibition of PKM2 enzymatic activity³⁵. Although the optimal motif for binding to PKM2 has been defined as GGAVDDDpYAQFANGG, further experiments have revealed the existence of additional pY-peptides that also inhibit the activity of PKM2. Therefore, the unique motif binding that interacts with the pocket is still unclear.

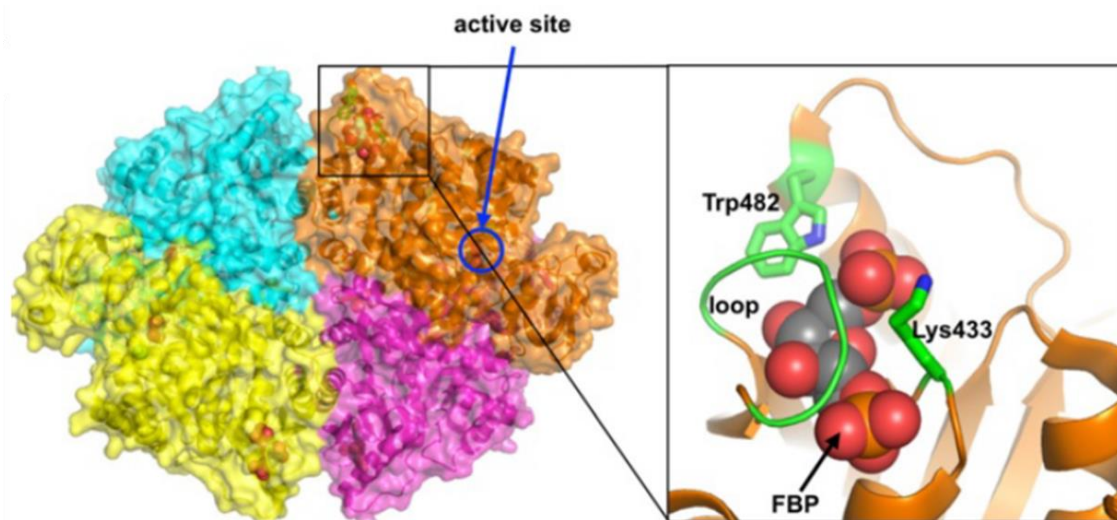


Figure 7 | The binding of FBP to the homo-tetrameric form of PKM2 (PDB ID: 3BJF, adapted from reference⁴⁸). The monomers of PKM2 are shown in different colors (left panel), and the binding site of FBP is magnified in the right panel. The interactor FBP is shown as a space-filling model, and the critical residues W482 and K433 are in green. The pY-peptides bind to this region to promote the release of the FBP and the subsequent inactivation of PKM2.

As another example, the PH domain of IQSEC1 (IQ motif and SEC7 domain-containing protein 1) has been shown to interact with pY1068 or/and pY1086³⁶ of EGFR. Due to a low affinity in

solution, the crystal structure of pY-peptides interacting with the PH domain is not yet available. A further example still is the Raf-1 kinase. This protein can interact with the inhibitory protein RKIP in the presence of self-phosphorylated S388 and Y341³⁷. An NMR titration experiment validated the interaction between the peptide (pS338pSYpY341) and RKIP to generate a dissociation constant (K_d) value of 45 μM ³⁸. Finally, proteomic studies using SILAC (stable isotope labeling by amino acids in cell culture) combined with a pY-peptide library has suggested several potential interactors such as WNK1 (serine/threonine-protein kinase WNK1), MTHFS (5-formyltetrahydrofolate cyclo-ligase), and GPD2 (glycerol-3-phosphate dehydrogenase)³⁹. Further validation concerning the interaction with RTKs is still required. Overall, these novel findings require more effort to identify the interaction between pY-peptides in RTKs and these atypical interactors without an SH2 or PTB domain.

1.3 Approach for studying the interaction between pY-peptide and proteins

There are various methods available to study the interaction between pY-peptides and target proteins. These methodologies include, but are not limited to, two-hybrid screening (Y2H), co-immunoprecipitation (co-IP), and affinity enrichment-mass spectrometry (AE-MS). For each method there are advantages and disadvantages for a pY-peptide interactome study and these are described below.

1.3.1 Yeast two-hybrid system

Introduced in 1989⁷³, the yeast two-hybrid system (Y2H) is a genetic method to investigate the physical interaction that occurs between two proteins, or a single protein and DNA. Central to the approach is that the transcriptional activator Gal4 from the yeast strain *Saccharomyces cerevisiae* is employed to detect a protein-protein interaction (Figure 8). The activator consists of two domains: a DNA-binding domain (DBD) at the N-terminus that specifically binds to DNA sequences (upstream activating sequences, UAS_G), and an activating domain (AD) at the C-terminus that can activate transcription. In an experimental procedure, these two domains are separately fused to bait and prey proteins. Once there is a physical interaction, the two components combine, and a transcriptional event is triggered. As activation of Gal4 leads to the utilization of galactose, a galactose selection experiment can be performed to examine the binding event.

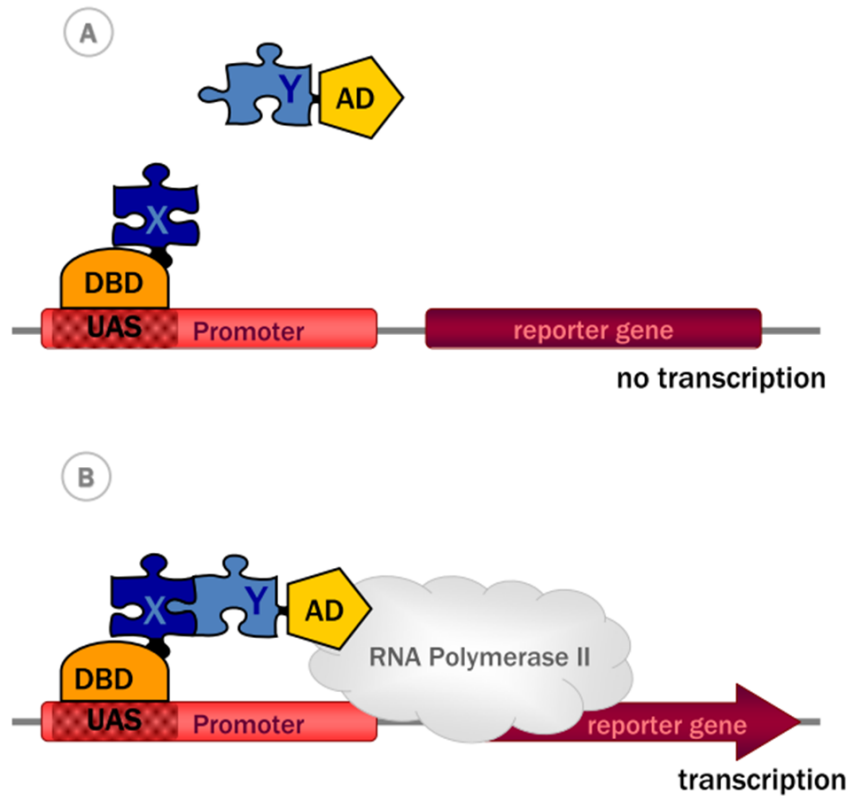


Figure 8 | The principle of the yeast two-hybrid system for protein-protein interactions (adapted from reference⁷⁴). The bait protein X is constructed with a DNA binding domain (DBD), and the potential interactor Y is fused to the activation domain (AD). Only when protein X and Y physically associate can the AD domain further recruit RNA polymerase II to promote the transcription of the reporter gene (panel B). The transcription can be observed by a galactose selection experiment.

Over the years this technique has been modified and applied to pY-dependent interactions with proteins containing SH2/PTB domains⁷⁵ or the interaction between RTKs and protein phosphatases⁷⁶.

As a low-technique approach, the Y2H-type experiment can be easily performed in many laboratories to initially screen interacting proteins. This technique, however, has disadvantages such as the different expression pattern of bait and prey proteins in yeast, protein modifications, localization, and misfolding *etc*^{77,78}. To generate high-confidence data, the result obtained from a Y2H experiment generally requires further validation such as co-immunoprecipitation.

1.3.2 Co-immunoprecipitation

In a typical co-immunoprecipitation (co-IP) experiment, the protein or peptide of interest (preys) is recognized by a specific antibody and subsequently isolated from a complex background by using protein A-Sepharose. Consequently, the interactors of the prey can be enriched and analyzed by immunoblotting (Figure 9)⁷⁹.

Coordinated with introduced point mutations, co-IP has been widely used for decades to study tyrosine phosphorylation-dependent protein-protein interactions. For example, to identify the phosphorylation site in FLT3, peptide-specific antibodies were used to pull-down selected FLT3 tryptic peptides. Radiosequencing subsequently revealed the four tyrosine residues that are auto-phosphorylated *in vivo*. Moreover, additional tyrosine mutation experiments suggested that the tyrosine-protein phosphatase non-receptor type 11 (PTPN11) binds directly to pY599 thereby contributing to FLT3 ligand-mediated mitogen-activated protein kinase (ERK) activation and proliferation⁸⁰.

Co-IP, however, is not without shortcomings as the approach is readily influenced by the antibody selected. Difficulties arise particularly when the protein-binding site overlaps with the binding site of the antibody⁷⁹. Furthermore, engineering of a mutant protein usually requires considerable effort. Therefore, throughput is too low and not suitable for large-scale interactome screening.

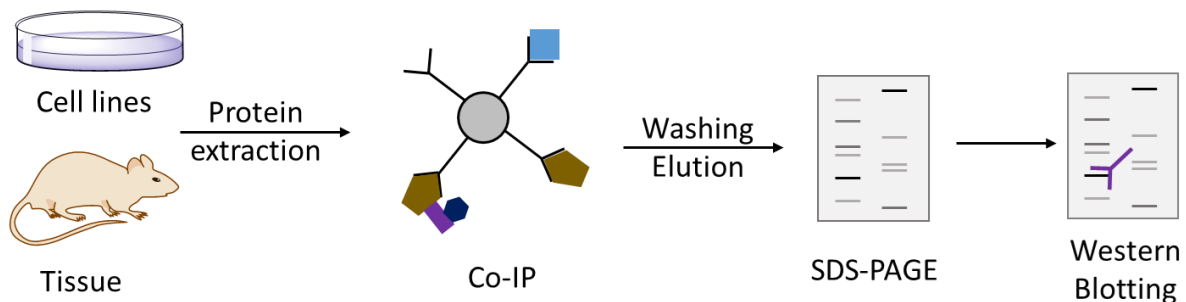


Figure 9 | The general workflow of a co-IP experiment

1.3.3 Affinity enrichment-mass spectrometry assay

In contrast to a co-IP experiment, affinity purification can directly measure the binding partners of a protein of interest. In a classical affinity-purification experiment, the bait is a GST-tagged recombinant protein of interest immobilized on Sepharose beads ⁷⁹. Alternatively, the bait can be a recombinant protein domain or a synthetic peptide. Following incubation with a biological matrix, the affinity-purified complex can be separated by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis), and immunoblotted to confirm presence/absence of specific proteins.

With the emerging of proteomics, a high-throughput approach termed affinity enrichment-mass spectrometry (AE-MS) was introduced to investigate protein-protein interactions^{27,28,81}. With this experimental strategy, the identification of interactors is decoupled from both separation by SDS and the utilization of antibodies for immunoblotting. Instead, the binding partners are digested by a protease and analyzed by nLC-MS/MS. To identify protein interactors, the standard workflow for AE-MS consists of 4 consecutive steps (Figure 10)⁸¹. Briefly, cell or tissue lysate was incubated together with the baits (proteins, domain or peptides) for the enrichment. Then bound proteins are digested in gel or solution with the protein protease (e.g., trypsin) to produce the featured peptide samples. After that, peptides are injected into an nLC system for the peptide separation, followed by the ionization step with the application of a voltage at the end of the LC column before the entrance of mass spectrometry. Subsequently, the m/z of the peptides are detected by the analyzer (e.g., Orbitrap, TOF), and the top N intensity of them are fragmented (e.g., Higher-energy collisional dissociation, HCD) for the peptide sequencing. Lastly, the peptide and protein identification and quantification are done by the database search with software (e.g., MaxQuant). Quantified proteins are further analyzed statistically to find out the critical binders.

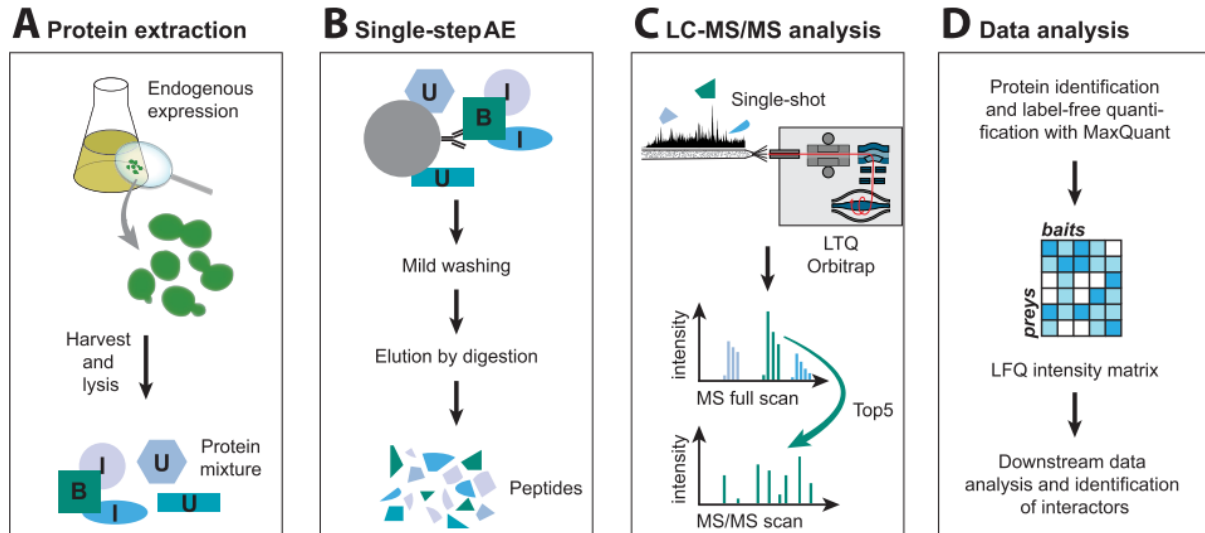


Figure 10 | The general AE-MS workflow (adapted from reference⁸¹). (A-B) Cell or tissue lysate is incubated with the bait protein (GFP-tagged protein, protein domain or peptide). Bound proteins are digested with a protease (*e.g.*, trypsin) to produce peptides. (C) Samples are injected onto an nLC system to separate the peptides followed by ionization and detection on an Orbitrap mass spectrometer. (D) Peptide and protein identification and quantification is achieved via a software-directed (*e.g.*, MaxQuant) database search. Quantified proteins are statistically analyzed to determine the critical protein binders.

This protein sequencing strategy is termed ‘bottom-up’ or ‘shotgun’ proteomics. As a high-throughput approach, it has the ability to simultaneously resolve all potential interactors regardless of cellular abundance and independent of any antibody.

Therefore, employing such an AE-MS strategy would address the conundrum of systematically profiling all interactors of the phosphotyrosine in RTKs. The phosphotyrosine interactome of several RTKs from human and mouse, including the EGFR, INSR, and DDR families have been profiled^{27,28,82}. Nevertheless, a complete picture of the phosphotyrosine interactome of human RTKs is still lacking.

1.4 Objective and outline

Receptor tyrosine kinases (RTKs) are critical cell surface receptors that can transmit extracellular signals into the intracellular regions via multiple signaling pathways. RTKs are essential regulators of different cellular processes such as cell proliferation, migration, and differentiation. Mutation of RTKs in cancer is known to induce the problem of drug resistance and ultimately poor patient prognosis.

It is well-recognized that the signaling intensity and direction of RTKs are controlled by phosphotyrosine residues, that recruit different functional proteins that contain SH2 or PTB domains. Systematically studying all the interactors of RTKs will provide a wealth of information that will lead to a deeper and more thorough understanding of cell signaling cascades. Consequently, the effectiveness of disease treatment will be improved. Additional efforts have led to the discovery that several atypical domains, such as PTP, C2, and HYB domains, possess the ability to interact with phosphotyrosine residues. The evidence of such interaction with RTKs is not yet sufficient.

With the outstanding development of proteomics, the AE-MS approach has a huge advantage in efficiently identifying protein interactomes. To date, some studies exist where the phosphotyrosine interactome of receptor tyrosine kinases have been profiled using this strategy; however, this is by no means comprehensive. Therefore, this thesis describes the global profiling of the phosphotyrosine interactome of RTKs by an AE-MS approach. Using synthetic pY-peptides as an affinity matrix, 1,144 wide-type and mutant residues were evaluated. The generated interactome was then integrated with data from four different databases to assess and increase the confidence of novel discovery. Furthermore, a subset of new interactors was validated by a competition assay, whereby both the dissociation constant (K_d) and IC_{50} values were calculated.

Ultimately, this systematic study will provide a deeper insight into the signaling cascades of RTKs and the specific interaction of RTKs to typical and atypical proteins. Moreover, this newly-resolved interactome of RTKs could become a comprehensive resource for further exploitation in the future.

Chapter 2 Experimental procedures

2.1 Cell culture and lysis

Four cell lines: Jurkat, MCF-7, MV4-11, and SK-N-BE (2) were purchased from the American Type Culture Collection (ATCC) and passaged with the recommended procedure at 37°C with 5% CO₂. Briefly, the adherent cell lines: MCF-7 and SK-N-BE (2) were cultured in Iscove basal medium and DMEM/Ham's F-12 (1:1) containing 10% FBS. The suspension cell lines: Jurkat and MV4-11 were both cultivated in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS). At approximately 80% confluency, cells were washed twice with PBS. For complete cell lysis, the cells were then incubated with 1× CP buffer (50 mM Tris/HCl pH 7.5, 5% glycerol, 1.5 mM MgCl₂, 150 mM NaCl, 1 mM Na₃VO₄, 25 mM NaF, 1 mM DTT), 0.8% NP-40 (IGEPAL CA-630, Sigma-Aldrich), phosphatase inhibitor I (25 μM Cantharidin (C7632, Sigma-Aldrich), 125 μM (-)-p-bromolevamisole oxalate (190047, Sigma-Aldrich), 25 nM microcystin-LR (ALX-350-012-C050, Enzo), phosphatase II (10 mM imidazole, 5 mM sodium fluoride, 5.75 mM sodium molybdate, 5 mM sodium orthovanadate, 20 mM sodium tartrate), phosphatase III (200 nM calyculin A (C-3987, LC Laboratories)) and 1× protease inhibitors (S8830, Sigma-Aldrich) at 4°C for 60 min. The cells were centrifuged at 21,100×g for 20 min to sediment debris. After determining the protein concentration using the Pierce™ Coomassie Plus (Bradford) assay kit (23236, ThermoFisher Scientific), supernatants were stored at -80°C.

2.2 Peptide sequence design

In total, there are 1,225 redundant tyrosine residues in the sequences of wild-type and mutant RTKs (Supplementary File 2). After redundancy removal, 1,144 residues were retained to generate synthetic peptides. Of these, 900 wild-types (WT), 24 isoforms (IF) peptides from 58 RTKs were extracted from UniProt⁵ (<http://www.uniprot.org/> and <https://www.uniprot.org/docs/humsavar>), and 220 mutant peptides including 16 hot-spot mutant (HSM) and 204 mutation-gain tyrosine (MGY) were extracted from MutationAligner³¹ (<http://mutationaligner.org/>).

All 1,144 pY-peptides for the AE-MS assay were designed as 11-mers with the phosphotyrosine residue located in the central position, plus an additional SGS linker on the N-terminus. In addition, 59 representative pY-peptides were also generated as 'free compounds' for the competition assay. The same design was chosen as for the assay peptides, but excluding the SGS linker. To further investigate significant differences between the interactomes of the wild-type and mutant peptides in the competition assay, one phosphoserine peptide (pS-peptide) and seven unphosphorylated peptides (wild-type peptides) were designed as 14-mers and 11-mers, respectively.

To clearly distinguish all peptides in the AE-MS assay, the wild-type and isoform peptides were named according to the following nomenclature: WT/IF + RTK name + central residue + position, *e.g.*, WT_DDR1_pY513 and IF_FGFR1_pY654. Similarly, mutation-gain tyrosine peptides were named: MGY + RTK name + central residue of wild-type + position + mutant residue, *e.g.*, MGY_ROR2_H743pY. Finally, the hot-spot mutation peptides were named as follows: HSM + RTK name + central residue + position + wild-type residue + position + mutant residue, *e.g.*, ERBB2_pY772_D769H. For the competition assays, the same nomenclature was adopted for the peptides, except the category (WT, IF, MGY, HSM) was removed.

2.3 Peptide synthesis

2.3.1 Phosphotyrosine peptides for the AE-MS assay

For the AE-MS assay, all 1,144 pY-peptides were synthesized with a free N-terminal amino group and a C-terminal amide group by solid phase peptide synthesis (SPPS) at a 2 μmol scale using Fmoc chemistry on an automated peptide synthesizer (Intavis Bioanalytical Instruments AG, Germany) (Figure 11). Each synthetic peptide was then cleaved from the resin (S30023, Rapp Polymere) with 500 μL cleavage cocktail (5% triisopropylsilane and 2.5% H_2O in anhydrous trifluoroacetic acid (TFA)), followed by 100 μL anhydrous TFA for 3 h. Approximately 6.6 pmol peptides were then loaded onto a MALDI (matrix-assisted laser desorption/ionization) target (pre-spotted AnchorChip PAC II 384 HCCA (α -cyano-4-hydroxycinnamic acid), Bruker) after diluting 1 μL elute with 500 μL H_2O . MALDI-TOF-MS (UltrafleXtreme, Bruker) analysis in positive-ion mode was then performed from m/z 600 to 3,000.

Additionally, 24 random peptides were synthesized and collected using the same procedure. The sequences were randomized from a kinetic amino acid mixture⁸³ without tyrosine (Y), phosphotyrosine (pY), or cysteine (C).

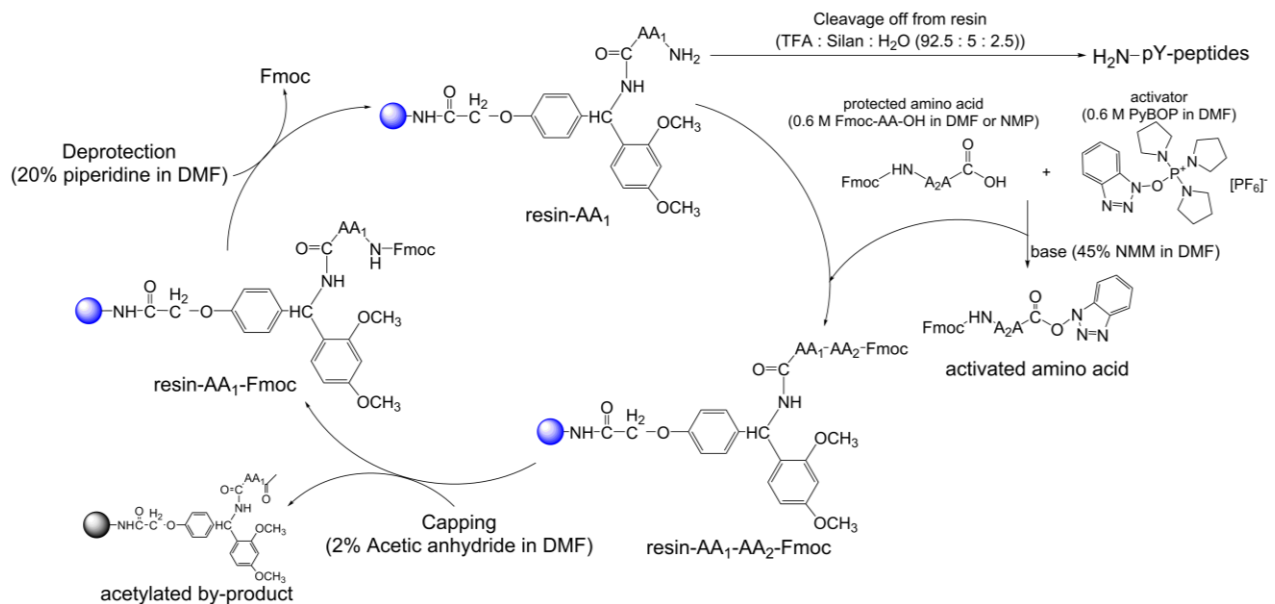


Figure 11 | Synthesis of pY-peptides using a solid-phase peptide synthesis strategy with Fmoc chemistry (modified from the instrument manual)

2.3.2 Peptides for the competition assay

All peptides were either synthesized internally or by Belyntic (Berlin, Germany) following a preliminary purification step with Peptide Easy Clean (PEC) technology. Internally-generated synthetic peptides were lyophilized by evaporating the TFA. After initially dissolving the peptides in H₂O at various pH conditions adjusted with formic acid and ammonium hydroxide, the peptides were prepared as stock solutions dissolved in DMSO. Two different pH conditions were utilized to dissolve the peptides and separate by HPLC.

Under the acidic condition, peptides were further diluted with different ratios of acetonitrile and H₂O with 0.1% TFA and injected onto an XBridge[®] Peptide BEH C18 column (3.5 μm, 2.1 mm × 150 mm, Waters, USA) connected to an Ultimate 3000 LC system (Dionex, ThermoFisher Scientific, USA). ACN-0.1% TFA and H₂O-0.1% TFA (pH 1.7) were used as mobile phases, and the UV wavelengths of 214, 267, and 280 nm were applied to monitor elution of the peaks. Due to the distinct hydrophobicities of these peptides, different gradients were applied to separate and collect the peptides from other impurities. Under the alkaline condition, the 0.1% TFA in the dissolving buffer and HPLC mobile phase was replaced with 0.1% NH₄OH (pH 10). The nature of the purified pY-peptides was further confirmed by LC-MS/MS or MALDI-TOF.

All pY-peptides were lyophilized, and after reconstitution in DMSO, quantified in a NanoDrop[™] 2000 spectrophotometer (Thermo Fisher Scientific, USA). The Lambert-Beer law shown below was then applied to calculate the concentration of the peptides.

$$A = \epsilon cl$$

In this equation, A is the UV absorption value at a specific wavelength, ϵ (L mol⁻¹ cm⁻¹) is the extinction coefficient of the peptide, c is the molar concentration (mol/L), and l is the path length in cm. With a path length of 0.1 cm in the device, the peptide concentration (nmol/μL) could be determined as:

$$c = 10^4 A / \epsilon$$

Due to different peptide compositions, three situations were considered to calculate the concentration. Firstly, at the UV wavelength of 267 nm together with the sum of the molar extinction coefficient of the pY-peptide, where 652 L mol⁻¹ cm⁻¹ for phosphotyrosine⁸⁴, 1,048 L mol⁻¹ cm⁻¹ for tyrosine, and 4,921 L mol⁻¹ cm⁻¹ for tryptophan

<https://omlc.org/spectra/PhotochemCAD/html>)^{85,86} were applied. Secondly, for the peptides with tyrosine instead of phosphotyrosine, 278 nm was utilized to quantify the concentration in DMSO. The extinction coefficient of 1,400 L mol⁻¹ cm⁻¹ and 127 L mol⁻¹ cm⁻¹ for tyrosine and cysteine, respectively, were applied⁸⁷. Lastly, a measurement was performed at 257 nm for peptides only containing phenylalanine but not the other amino acids above. After reconstitution in 30% ACN, an extinction coefficient of 194 L mol⁻¹ cm⁻¹ was used <https://omlc.org/spectra/PhotochemCAD/html>)^{85,86}.

2.4 Generation of the affinity matrices

The TFA was evaporated from the peptide solutions at room temperature and transferred to 2 mL Eppendorf tubes for lyophilization. Then resultant peptide powders were individually dissolved in 2 mL Eppendorf tubes with anhydrous DMSO and immobilized on Sepharose beads (NHS-activated Sepharose 4 Fast Flow, 17-0906-01, GE Healthcare) at a coupling density of 1% (0.2 $\mu\text{mol/mL}$ bead) with triethylamine (T-0886, Sigma-Aldrich) for 20 h. Subsequently, the remaining activated sites on the beads were blocked with ethanolamine for 20 h. After washing with 10 volumes of DMSO and 30 volumes of ethanol, the matrix was stored at 4°C in pure ethanol until required (Figure 12).

For the negative control, an affinity matrix comprised of 24 random peptides was similarly processed. Beads blocked by ethanolamine were also used as an additional negative control to enable removal of all potential background proteins that bind to the surface of beads. The peptide matrix DDR1_pY484 (REPPPpYQEPRP) was used as a positive control to monitor the reproducibility of the affinity enrichment between batches.

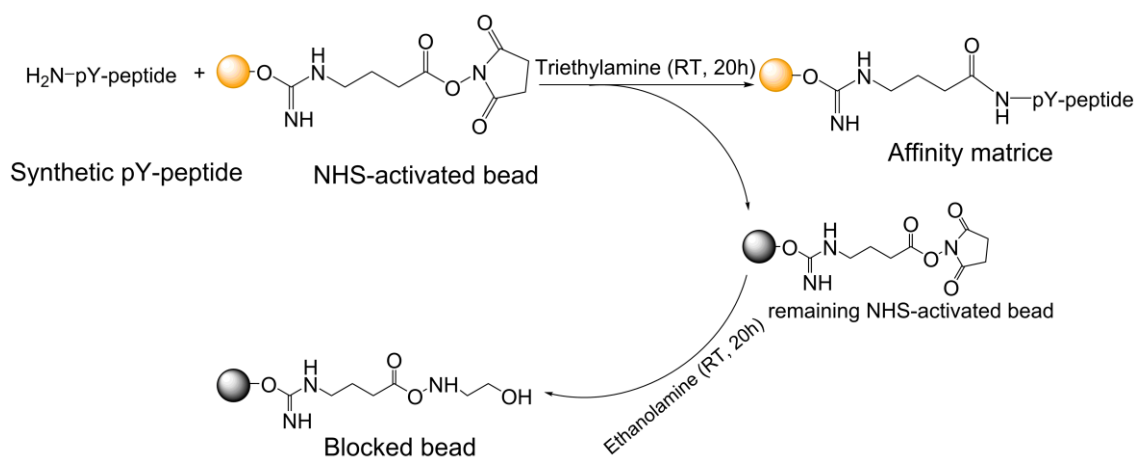


Figure 12 | Generation of pY-peptide affinity matrices via a coupling reaction with NHS-activated beads

2.5 Affinity enrichment (AE) assay

Four cell lysates were clarified by ultracentrifugation at 167,000×*g* for 20 min at 4°C, mixed in equal proportions and then diluted with 1× CP buffer, phosphatase inhibitor I, II, III, and protease inhibitors. One milliliter cell lysate (5 mg/mL) was distributed equally in a 96-well filter plate and incubated with 35 μL settled beads (equal to 7 nmol phosphotyrosine or random peptide) in an end-over-end rotator for 60 min at 4°C. After extensive washing with 3 mL 1× CP buffer with 0.4% NP-40 and 2 mL 1× CP buffer with 0.2% NP-40, bound proteins were incubated with 40 μL 2× NuPAGE LDS sample buffer (Invitrogen) containing 50 mM dithiothreitol (DTT) at 50°C for 30 min. Afterward, eluted proteins were centrifuged into a 96-well plate prior to in-gel digestion (Figure 13).

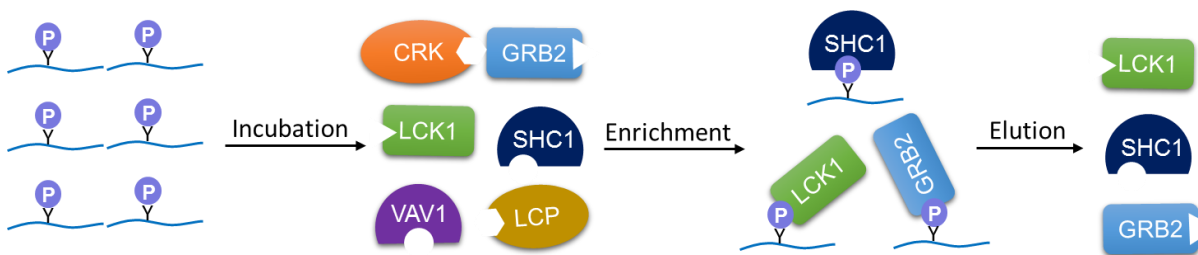


Figure 13 | Affinity enrichment assay using synthetic pY-peptides as baits to enrich proteins from cell lysates

2.6 Competition pull-down assays

Once dissolved in DMSO, each 'free compound' (peptide) was added to 1 mL cell lysate (5 mg/mL) at concentrations of 1 μ M and 10 μ M. As a control, 5 μ L DMSO was added to another 1 mL cell lysate. The pre-incubation procedure in the 2 mL deep-well plate was performed for 1 h at 4°C. The mixture was then incubated with the corresponding affinity matrix in a 96-well filter plate for 1 h at 4°C. After the washing and elution steps as described in the previous section, the flow-through lysate from the DMSO-treated wells were incubated again with the same immobilized peptides for another 1 h at 4°C. This provided the data to calculate the correction factor (r) to estimate the dissociation constant (K_d) between the ligand and protein (Figure 14). For the wild-type and isoform pY-peptides, the competition assay was usually performed between the same peptide sequences. For a subset of the mutation-gain tyrosine peptides and hot-spot mutant peptides, the competition assay was performed in 3 ways: (1) mutant pY-peptide vs. mutant pY-peptide; (2) mutant pY-peptides vs. wild-type form; and (3) wild-type peptide vs. wild-type peptide.

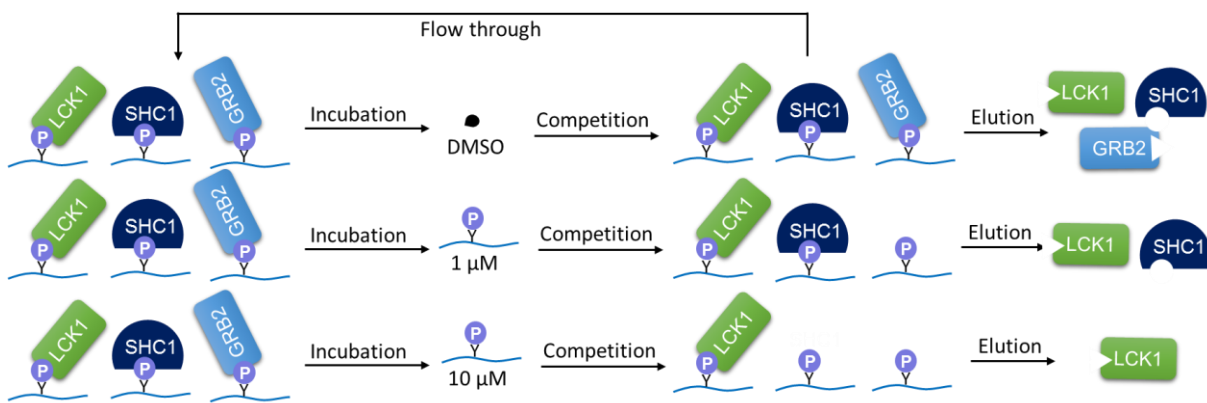


Figure 14 | Peptide competition assay. Two doses of peptides are used to compete for the interaction to the proteins with affinity matrix. DMSO is used as negative control, and a PDPD is performed to calculate the depletion factor.

2.7 In-gel protein digestion

2.7.1 pY-peptide affinity enrichment samples

After cysteine residues were alkylated with 0.8 mg chloroacetamide (CAA) at room temperature in the dark for 30 min, half of the sample was concentrated in a 10-well 4–12% Bis-Tris NuPAGE gel (Invitrogen) at 200 V for 5 min. The gel was then stained with Coomassie Brilliant Blue (Roti® Blue) (Roth, Germany) for 20 min and destained with 25% ethanol and 0.1% formic acid (FA) for 1~2 h, respectively, until the proteins bands were visualized. In-gel digestion with trypsin was performed according to a standard laboratory protocol by a technician in the lab, and the lyophilized samples were stored at -20°C until analysis by nLC-MS/MS.

2.7.2 Cell line full proteome

For each the 4 cell lysates, 400 µg protein was individually denatured with 2× NuPAGE LDS sample buffer (Invitrogen) containing 50 mM dithiothreitol (DTT) and boiled at 50°C for 30 min. After cysteine residues were alkylated with 0.8 mg CAA at room temperature in the dark for 30 min, 50 µg of the sample in 8 duplicates were concentrated in a 10-well 4–12% Bis-Tris NuPAGE gel (Invitrogen) at 200 V for 5 min. The gel was stained with Coomassie Brilliant Blue (Roti® Blue) (Roth, Germany) for 30 min and destained by standard procedures until the protein were visualized. In-gel digestion with 1 µg trypsin per excised band was performed according to a standard laboratory protocol by a technician in the lab, and the lyophilized samples were stored at -20°C before peptide fractionation.

2.8 Peptide fractionation

Tryptic peptides for full proteome profiling were dissolved in 200 μ L 10 mM NH_4OAc and one quarter was loaded onto an LC system with a Thermo Acclaim Trinity P1 column (mixed mode, 3 μ m, 150 mm \times 2.1 mm i.d.) (ThermoFisher Scientific, USA) and a fractionation collector (FC 203B, Gilson). Mobile phase A and B were 10 mM ammonium acetate (pH 4.7) and 10 mM ammonium acetate in 95% acetonitrile (pH 5.3), respectively. Peptide separation was performed with a gradient of 0–2 min, 15% B; 2–30 min, 15%–50% B; 30–31 min, 50%–84% B; 31–34 min, 84% B; 34–36 min, 84%–50% B; 36–38 min, 50% B. Flow rates were as follows: 0–30 min, 200 μ L/min; 30–31 min, 200–300 μ L/min; 31–34 min, 300 μ L/min; 34–36 min, 300–400 μ L/min; 36–38 min, 400–20 μ L/min. From 2 to 34 min, 32 fractions of 1 min intervals were collected into a 96-well plate. Samples were lyophilized prior to analysis by nLC-MS/MS⁸⁸.

2.9 LC-MS/MS analysis

2.9.1 Interactome screening and full proteome profiling

Tryptic peptides from the AE-MS assay were dissolved in 20 μ L 0.1% formic acid (FA), and half was analyzed by nLC-MS/MS using a nanoLC–Ultra (Eksigent) system coupled to an LTQ–Orbitrap Elite mass spectrometry (ThermoFisher Scientific). Similarly, all 32 fractions of the tryptic peptides from the 4 cell lines were dissolved in 0.1% FA, and 1/3 were analyzed on the same nLC-MS/MS system.

Samples were firstly loaded onto a trap column (C18, 3 μ m, 20 mm \times 100 μ m i.d.) in 0.1% FA at a flow rate of 5 μ L/min for 10 min, followed by chromatographic separation on an analytical column (C18, 3 μ m, 40 cm \times 75 μ m i.d.) heated at 50°C. A linear gradient was performed from 2%–32% solvent B (0.1% FA and 5% DMSO in 100% ACN) at a flow rate of 300 nL/min for 90 min. Peptides were sprayed into the Orbitrap Elite via a stainless steel emitter (40 mm, OD 1/32", 2.2 kV, ThermoFisher Scientific) connected to the end of the analytical column. The mass spectrometer was operated in positive-ion data-dependent acquisition (DDA) mode with a capillary temperature of 275°C. An inclusion list of SH2 or PTB domain-containing protein was utilized to improve detection and sensitivity. Full scan MS¹ spectra (m/z 360–1300) were acquired in the Orbitrap with a resolution of 30,000 and the top 15 ions were fragmented (resolution: 15,000) in the HCD cell at 30% normalized collision energy after accumulating an ion target value of up to 2E4. Samples were analyzed in sequences with a blank analysis between each to avoid peptide carry-over into subsequent samples.

2.9.2 Competition assay

For the tryptic peptides from the competition assay, 1/4 of a subset of the samples was analyzed on an LC-MS system comprised of a Dionex HPLC (NCS_3500RS) connected to an Orbitrap HF-X mass spectrometer (ThermoFisher Scientific, USA). The configuration of the analytical column, trap column, flow rates, column temperature, and buffers was identical as described above, except a 24 min linear gradient from 4–32% solvent B was used to separate the tryptic peptides before MS analysis. The analytical column was then flushed with 80% solvent B for 2 min and 2% solvent B for 2 min at 500 nL/min. MS¹ spectra (m/z 360–1300) were acquired in the Orbitrap with a resolution of 30,000 and an AGC target of 3E6. The top 20 ions were fragmented (resolution: 15,000) in the HCD cell at 28% normalized collision energy after accumulating a target value of up to 1E5 and a maximum injection time (IT) of 25 ms.

All other samples were measured on an Orbitrap Fusion Lumos (ThermoFisher Scientific, USA) coupled to a Dionex HPLC (NCS_3500RS) and 1/2 the sample was injected. The HPLC configuration for the buffers, flow rates, trap column, and column temperature was the same as the information given above, except a laser-pulled fused silica emitter integrated into the analytical column was used. Additionally, a 20 min linear gradient from 4%-32% buffer B was implemented and a spray voltage of 2 kV was used. The Orbitrap Fusion Lumos was operated in DDA mode from m/z 360–1,300 and an MS¹ resolution of 60,000. The AGC target was 4.0E5 with a maximum IT of 50 ms. The top 20 ions were selected for fragmentation at a resolution of 15,000 with a collision energy of 30%. Fragment ions were accumulated up to an AGC value of 1E5 with a maximum IT of 25 ms.

2.10 Protein identification and quantification

All raw files from the AE-MS assay, including the blocked bead and random peptides, and the full proteome were processed and searched with MaxQuant (version 1.5.2.8) to identify and quantify the proteins. Excluding the function 'matching-between-runs', the parameters given below were used throughout. To avoid error, 'matching-between-runs' was applied separately to the two different datasets.

In addition to the naming nomenclature for the peptides that was described earlier, an MS plate number and p number were assigned to each raw file for the search. The letters 'NC' representing 'negative control' were used to name the blocked beads (BB) and random peptides (RP). 'PC' was added to annotate the 'positive control' DDR1_pY484.

Data was searched against UniProt human database (v22.07.13) with a 1% FDR. Carbamidomethylated cysteine residues was a fixed modification, and acetyl (protein N-term) and oxidation (M) were variable modifications. Trypsin/P was selected as the specific enzyme, and two miss cleavage sites were allowed. Peptide mass tolerance was 60 and 4.5 ppm for first- and second-pass searches, respectively; and the MS/MS tolerance was 100 ppm. The intensity-based absolute quantification (iBAQ) was used to quantify the proteins at a 1% false discovery rate (FDR). With a 20 min alignment window and a 0.7 min match window, the 'match-between-runs' function was applied to rescue misidentified peptides. All other database search settings were the default values.

The raw files generated from the competition assay were divided into two subsets depending on the mass spectrometer used. The search was performed with the same parameters as described above, excluding the following: peptide tolerance was 20 ppm for the first-pass search and 20 ppm for FTMS MS/MS matching. Every four raw files generated with the same peptide matrix were grouped, and LFQ was selected for protein quantification. The minimum ratio of peptides was 2, and the option of 'separate LFQ in parameter groups' was selected to avoid unexpected errors. The 'match-between-runs' window was 5 min and applied to all raw files. Due to the limited number of parameter group (20 groups) in MaxQuant, raw files were searched separately and the results combined for subsequent plotting.

2.11 Data analysis

2.11.1 Data normalization and estimation of enrichment factors

After the removal of a protein identified as 'only identified by site', 'reverse' and 'potential contaminant', the iBAQ intensity of each quantified protein was \log_{10} transformed and used to normalize the data in a two-step procedure. Firstly, all the affinity enrichment experiments of pY-peptide were median centered. Secondly, the variance of each experiment was estimated using the interquartile range (IQR) and then aligned across all experiments. These computing steps stabilize the variances and retain the robustness of the data set with respect to keeping the outliers in an individual experiment.

The average iBAQ intensity of each protein in the full proteome data set acquired from the four cell lines was used as the baseline for protein abundance⁸⁹. Moreover, missing values were computed with a constant of 3.5, which is less than the lowest iBAQ intensity that can be detected.

The ability of a pY-peptide enriching a protein was estimated using an enrichment factor (EF), where a high EF is indicative of a high-binding affinity between a protein and a peptide. With the biological assumption that the pY-peptides can specifically recruit proteins and the fact that the bead matrix alone can also randomly enrich non-specific proteins (bead proteomes) irrespective of the presence of pY-peptides, a local polynomial regression (LOESS) was used to fit a regression that can predict protein abundance in affinity enrichment experiments using the baseline protein abundance. Points (proteins) closely around the regression line are not enriched by the pY-peptide, whereas the proteins high above the regression line are more likely to be enriched by the peptide (Figure 15).

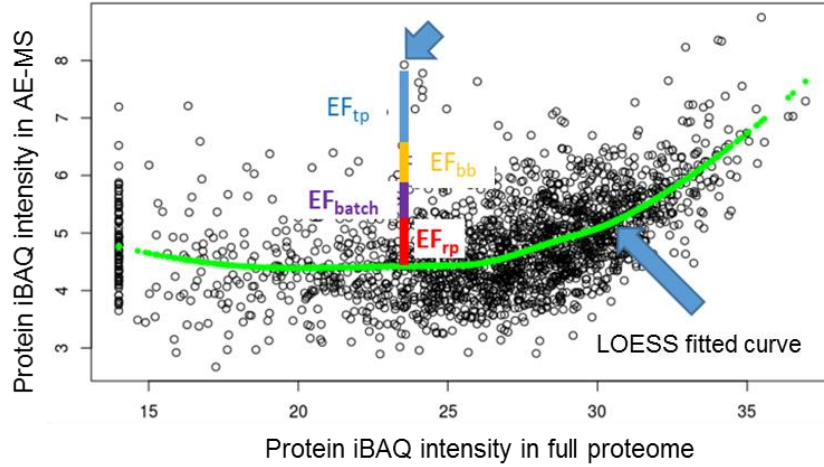


Figure 15 | The enrichment factor estimation of interactors of pY-peptide. Local Polynomial Regression (LOESS) was used to fit a regression that predicted protein abundance in pY-peptide enrichment experiments (green curve). The enrichment factor for each protein was estimated by considering the effect of blocked beads, random peptides, and batch effects.

Formally, the observed enrichment factor (EF_{obs}) is defined as the vertical distance from the protein abundance in the affinity enrichment experiment to the fitted LOESS curve, *i.e.*:

$$EF_{obs}^p = A_r^p - loess(B^p)$$

EF_{obs}^p is the observed EF of protein p enriched by a specific RTK pY-peptide r ; A_r^p is the normalized iBAQ intensity of protein p enriched by pY-peptide r ; B^p is the mean of the normalized iBAQ intensity of protein p in the full proteome of the four cell lines, indicating the baseline expression of protein p ; Last, $loess(\cdot)$ is the function calculating loess-fitted value for a given B^p .

According to the experimental design of this study, it is assumed that the observed EF can be decomposed as a linear combination of multiple, separated EFs, *i.e.*:

$$EF_{obs} = EF_{tp} + EF_{bb} + EF_{rp} + EF_{batch}$$

In this formula, the EF_{bb} is the EF of blocked beads. For a protein p , it is estimated as:

$$EF_{bb}^p = \bar{A}_{bb}^p - loess(B^p)$$

where \bar{A}_{bb}^p is the median of normalized iBAQ intensities of protein p across 35 affinity enrichment experiments of blocked beads.

The EF_{rp} is the EF of random peptides. For a protein p , it is estimated as:

$$EF_{rp}^p = \bar{A}_{rp}^p - loess(B^p)$$

where \bar{A}_{rp}^p is the median of normalized iBAQ intensities of protein p across 24 affinity enrichment experiments of random peptides.

The EF_{batch} accounts for the potential batch effects inherited from the different plates used in the mass spectrometry experiments. For protein p in plate n, it is calculated as:

$$EF_{batch_n}^p = A_{batch_n}^p - loess(B^p)$$

where $A_{batch_n}^p$ is the 25% quantile of normalized iBAQ of protein p in batch n. This term is introduced based on the assumption that it is unlikely that all RTK peptides in the same batch (normally 40-96 affinity enrichment experiments) have the same binding proteins. Therefore, if a protein is non-differentially identified in all affinity enrichment experiments in the same batch, this is more likely to be a result of a batch effect rather than a true binder.

Lastly, the EF_{tp} indicates the EF of RTK peptides that are of interest and can be calculated by removing EF of blocked beads, random peptides, and batches:

$$EF_{tp} = EF_{obs} - EF_{bb} - EF_{rp} - EF_{batch}$$

For a more convenient comparison in the following steps, the EF_{tp} value e was transformed to a \log_2 scale. Proteins with the $EF_{tp} \geq 2$ and more than one identified peptide are defined as significant interactors.

The script for the data normalization and filtering were performed by Dr. Chen Meng (Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany) (unpublished method).

2.11.2 Database mapping

The newest version of data from the four databases: PhosphoSitePlus²⁴ (date 30 November 2018), BioGRID²¹ (Homo_sapiens-3.5.168.tab2), and PepSpotDB²⁶ (kindly provided by Prof.Dr. Cesareni), ProteomicsDB²³ were downloaded for comparison. Through UniProt ID and sequence identity of RTKs, all pY-residues from PhosphoSitePlus that were identified by either LTP (low-throughput) or HTP (high-throughput) approaches, observed in cell lines, annotated as regulatory sites, evidenced in disease development, and/or known to have interactors, were all extracted and mapped to the interactome data set from this study. Duplicate values in the data set of 'disease-related' were combined. In the end, the information based on pY-residues was mapped for further analysis. The same approach was applied to the data from PepSpotDB.

Regarding the data from the BioGRID, the 'Official Symbol Interactor A' and 'Official Symbol Interactor B' were mapped to the names of RTKs, respectively. The data were then combined, and the dimeric interaction of the RTKs removed. Duplicate interactions from different types of experiments were combined and information on interaction at the protein level, *i.e.*, RTK-proteins, was integrated.

Using the first UniProt ID of the interactors in the category 'Majority protein ID', the expression levels (quantified in MS¹) in cell lines were retrieved from ProteomicDB. When both co-expressed in the cell lines, the pair-wise correlation coefficients of the RTKs and interactors were calculated. For each interactor and RTK, the three common cell lines with the highest protein intensity were listed and integrated into the data matrix.

The script for the database mapping was performed by Dr. Chen Meng (Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany).

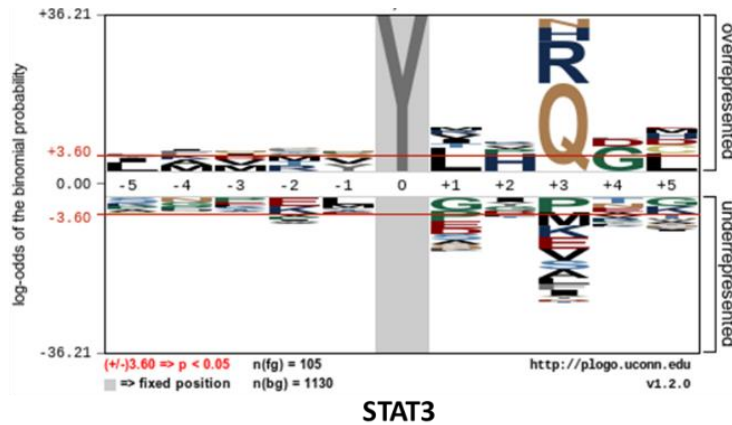
2.11.3 Motif analysis

The unmodified pY-peptide sequences (Y replaces pY) of each significant interactor were motif analyzed by pLogo (v1.2.0, <https://plogo.uconn.edu>)⁹⁰. Using the default setting, the analyzed sequences were input as the foreground, and all the sequences of the RTKs with 11 amino acids for the AE-MS assay were set as the background. The residue Y was set as the fixed amino acid at position 0 to generate the motif with the single letter codes of the amino acids (Table 2).

Table 2 | Amino acids and the corresponding one letter codes

| Amino acid | One letter code | Amino acid | One letter code |
|---------------|-----------------|---------------|-----------------|
| alanine | A | leucine | L |
| arginine | R | lysine | K |
| asparagine | N | methionine | M |
| aspartic acid | D | phenylalanine | F |
| cysteine | C | proline | P |
| glutamic acid | E | serine | S |
| glutamine | Q | threonine | T |
| glycine | G | tryptophan | W |
| histidine | H | tyrosine | Y |
| isoleucine | I | valine | V |

Critical residues were manually annotated with log-odds of the binomial probability after the corresponding residues. For example, when the protein binding to the motif logo is given as Figure 16, the annotation is presented as [x][x][x][x][x][pY][L(5.54)][H(4.15)][Q(20.57)/R(9.94)/P(-3.65)][G(5.98)][L(4.15)].



[x][x][x][x][x][pY][L(5.54)][H(4.15)][Q(20.57)/R(9.94)/P(-3.65)][G(5.98)][L(4.15)]

Figure 16 | The pLogo of STAT3 (upper) with annotation information (bottom). The red horizontal bars on the pLogo correspond to $p=0.05$. The motif is presented as the pY in the center (position 0) with amino acids at ± 5 position. Each amino acid with its log-odds of the binomial probability (value in the round bracket) is put in the square bracket. A slash is used to separate the amino acids, if they are critical in the same position. An amino acid with positive value in the round bracket suggests it is critically important to the position, whilst the residue with negative value indicates it is excluded. The letter x denotes any natural amino acids.

2.11.4 Interactome network assembly

All interactome networks were created in Cytoscape (version 3.4.0)⁹¹. The bait (pY-peptide or RTK sub-family) and prey (interactor) are represented as a blue ellipse and an orange rectangle, respectively. The layout of the network is organized with the overlapped interactors in the middle and unique interactors on the left or right panel.

2.11.5 Competition assay

Every two-dose competition assay was performed as four different but simultaneous experiments (10 μ M, 1 μ M, DMSO, and PD of PD). Comprised of three data points (10 μ M, 1 μ M, and DMSO), LFQ values for every interactor with each pY-peptide was used to plot the dose-dependent values. Data from the PD of PD was then applied to correct for the depletion factor (r):

$$r = \text{Intensity (PDPD)} / \text{Intensity (DMSO)}^{92}$$

where Intensity are both LFQ intensity.

Next, the IC_{50} could be estimated⁹³ as:

$$IC_{50} = \frac{Res}{1 - Res} * Conc_{Res}$$

where Res represents the lowest ratio of LFQ intensity between one of the treatments (10 μ M or 1 μ M) and the DMSO control, and the $Conc_{Res}$ is the corresponding concentration.

Then, the K_d (dissociation constant) was calculated⁹⁴ as:

$$K_d = IC_{50} * r$$

Finally, the number of peptides, unique peptides and MS/MS spectral counts were computed together with the competition bar plot to assist in determining the interactors. All the plots were output as a pdf file for visualization.

The script for this part of data modeling and plotting were performed by Dr. Mathias Wilhelm (Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany) (unpublished method).

Chapter 3 Results and discussion

As a systematic phosphotyrosine interactome study of RTKs, two major experiments: an AE-MS assay and a competition pull-down assay were performed to identify significant interactors. The data set from AE-MS was integrated with four publicly-available databases: PhosphoSitePlus, PepSpotDB, BioGRID, and ProteomicsDB to cross-validate the novel interactors with or without known binding domains to pY-peptides. Meanwhile, motif analysis was employed to search for specific peptide sequences that bind to the interactors. Novel interactors were validated by the competition pull-down assay, and the dissociation constant (K_d) and IC_{50} values were calculated.

3.1 Analysis of the synthetic peptides

For the AE-MS and competition assays, 1,243 peptides (including 24 random peptides) were designed for synthesis by Fmoc chemistry. A peptide was considered successfully synthesized when the presence of a peak in the MALDI-TOF spectrum corresponded to the correct molecular mass. An example is shown below (Figure 17).

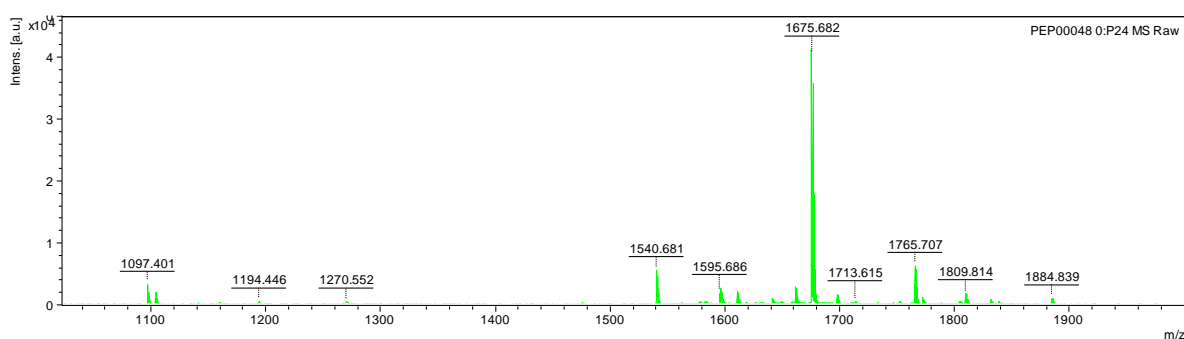


Figure 17 | MALDI-TOF spectrum of pY-peptide DDR1_pY484, where the peak at m/z 1675.682 is closest to the calculated monoisotopic peptide mass of 1675.759 Da

In addition, sodium (Na), potassium (K) and even sodium plus potassium (Na+K) adduct peaks were frequently observed in the MALDI-TOF spectra, with m/z values approximately 22, 38, or 60 Da higher than the theoretical peptide masses. An example of KIT_pY570 (1593.659 Da) is shown below (Figure 18).

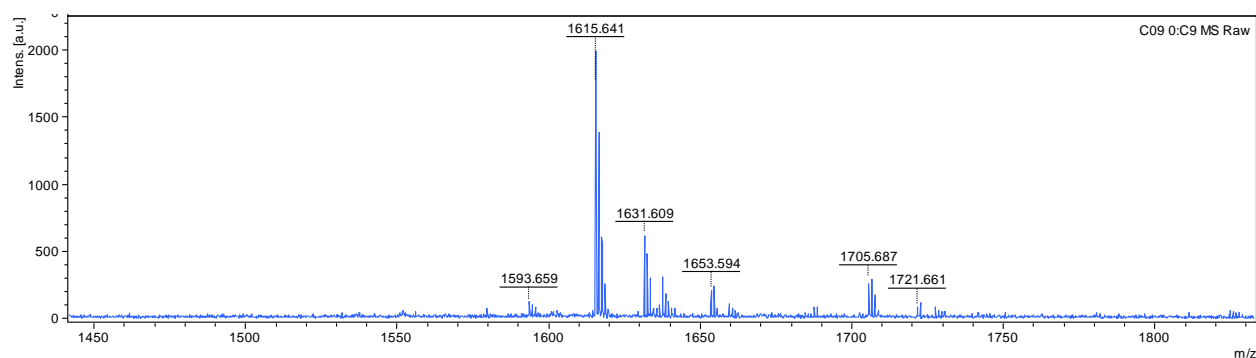


Figure 18| MALDI-TOF spectrum of pY-peptide KIT_pY570. The presence of a peak at m/z 1,593.659 corresponded to the expected mass. Additionally, there are peaks corresponding to the peptide with a sodium, potassium and double sodium plus potassium adducts at m/z 1,615.641, m/z 1,631.609, m/z 1653.594, respectively.

In addition, the spectra of peptides containing a methionine residue, *e.g.*, DDR1_pY740, often showed an unusual feature whereby a peak of approximately -48 Da than the expected peptide mass was observed. The mechanism of this loss is not yet resolved, but the assumption was

that this was generated from random fragmentation (decomposed carboxymethylated methionine) in the MALDI-TOF or even rearrangement of the side chain in the peptides. To further confirm the existence of this contamination product, another mass spectrometer with an electrospray ionization (ESI) source was used to analyze the DDR1_pY740 peptide. Interestingly, the target peptide was observed with the charge state of +2, while the unknown peak was only weakly apparent (Figure 19). In the end, this meant that the influence of the unknown product on the purity of the target peptides was negligible. Therefore, the other peptides that showed the same pattern in the MALDI-TOF spectra were regarded as successfully synthesized.

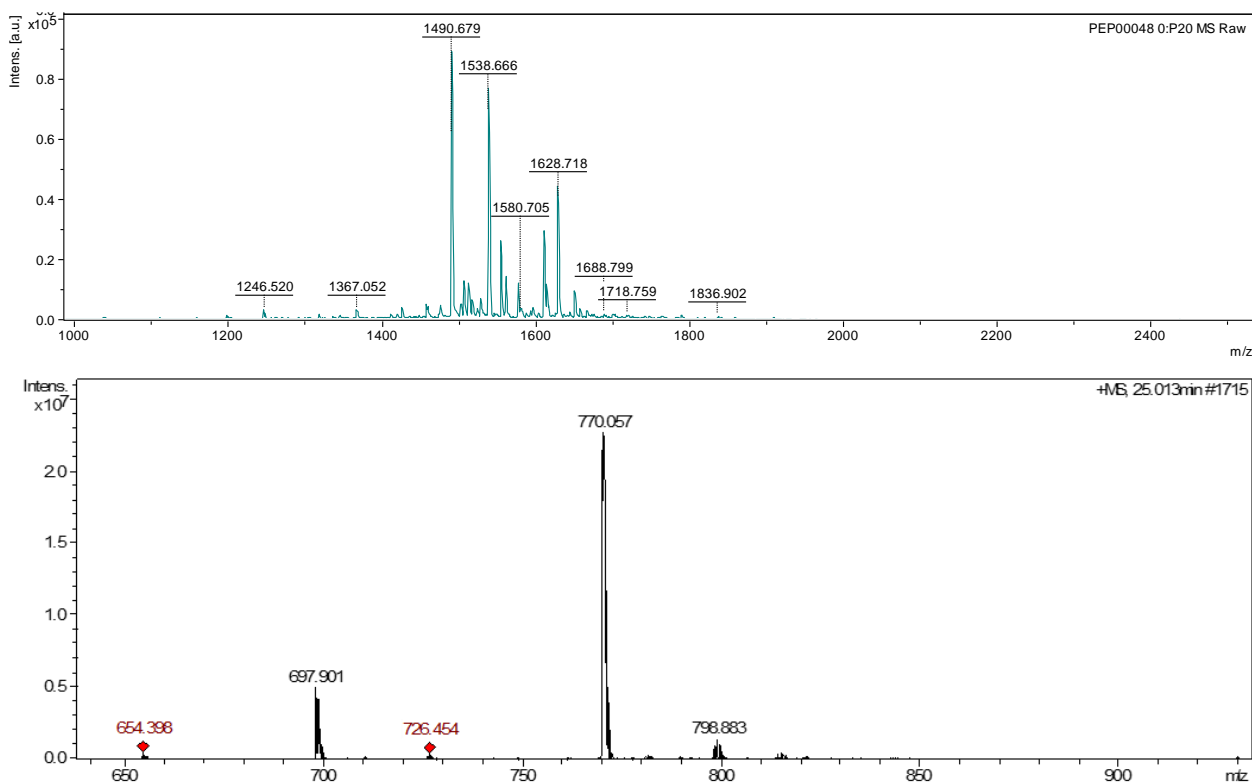


Figure 19 MALDI-TOF (upper panel) and ESI (lower panel) spectrum of pY-peptide DDR1_pY740. The unknown peak (m/z 1,490.679) in the MALDI-TOF spectrum was absent in the ESI spectrum, indicating that the effect on target peptide purity was negligible.

According to the MALDI-TOF spectrum, the synthesis of the random peptides (in theory containing over 64 trillion sequences) was also considered as complete (Figure 20). Briefly, the masses of the broad synthetic peak ranged from 1,000 to 2,000 Da. This falls into the region of theoretical molecular masses ranging from 860 to 2,281 Da. The theoretical m/z was calculated

by assuming extreme peptide compositions exist, *i.e.*, containing either 11 identical serine or 11 tryptophan with SGS linkers. Therefore, the random peptide matrix could be meaningful negative controls for removing the background proteins that interact with random non-phosphotyrosine peptides.

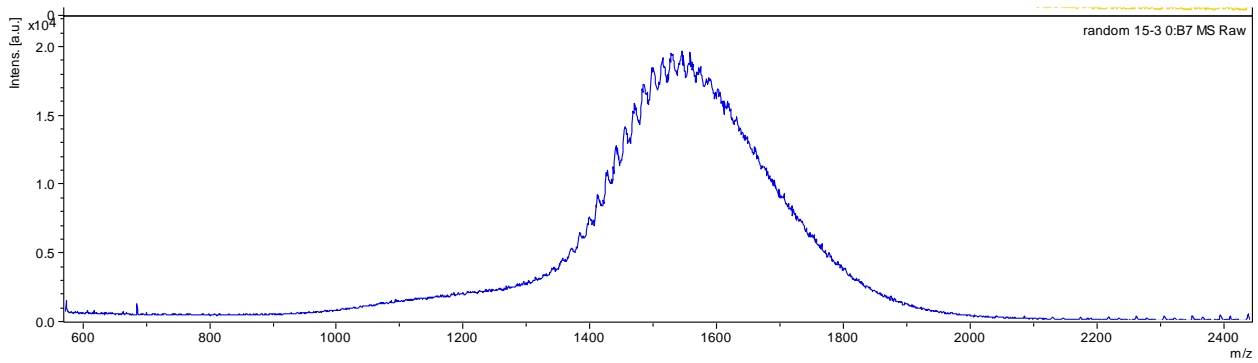


Figure 20 | MALDI-TOF-MS spectrum of the random synthetic peptides.

3.2 Global profiling of the RTK interactome by AE-MS

To enable the identification of true interactors from a large number of background proteins in a large-scale screen, high-quality controls are imperative. Two sets of negative controls, including blocked beads and random peptides, were utilized to enrich background proteins that can bind to the surface of the Sepharose beads or random peptide sequences. By applying the pipeline shown in Figure 21, the significant interactors of the pY-peptides could be identified and globally quantified.

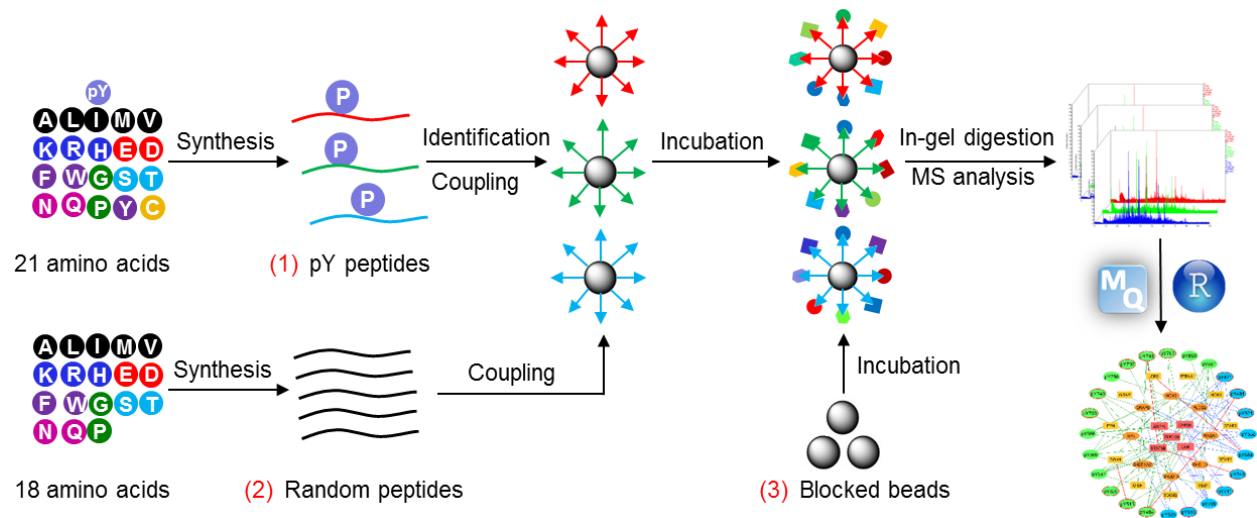


Figure 21 | Screening workflow of the phosphotyrosine-mediated interactome of RTKs using affinity enrichment-mass spectrometry (AE-MS)

3.2.1 Identification and quantification of significant interactors

Before statistical filtering, the raw interactome of the 1,144 pY-peptides consisted of 6,862 quantified proteins. Amongst these, 114 were SH2 or PTB domain-containing proteins that consist of several protein isoforms from 97 unique genes. With known interacting domains, these proteins have iBAQ intensities of approximately 9.4%.

In the affinity enrichment experiment, it is quite common for the beads to enrich many non-specific binders. These background proteins are the so-called 'bead proteome'⁸⁹ that originate from the random binding or precipitation of proteins onto the surface of the Sepharose beads. Therefore, the affinity matrix of blocked beads was introduced to account for a sub-set of non-specific interactors. From 35 blocked-bead experiments, a total of 5,398 proteins were quantified. To further increase the specificity of the interactors to pY-peptides, 24 'affinity matrix of random peptides' were also included to remove non-specific interactors. This set of negative controls is theoretically composed of 64 trillion random peptides and from this, 4,825 proteins were enriched. For analyzing the data, integrating these two sets of negative controls could minimize the influence of the background proteins, and thus increase the significance of the true interactors.

Since the SH2 or PTB domain-containing proteins are known interactors of pY-peptides, the proportion of SH2/PTB domain proteins before and after removing the non-specific EFs was evaluated as a reference point. Two steps were introduced to filter the data set. Firstly, proteins that were quantified by only one peptide were removed. Then a cut-off of EF value ≥ 2 was applied to remove insignificant interactors. This filtering led to the quantification of 86 different SH2 or PTB domain-containing proteins (75 genes) from a total of 2,184 unique interactors. Encouragingly, the proportion of the interaction between pY-peptides and SH2/PTB domain-containing proteins was increased to 28.1% and a total iBAQ intensity of 66% (Figure 22, Supplementary File 3).

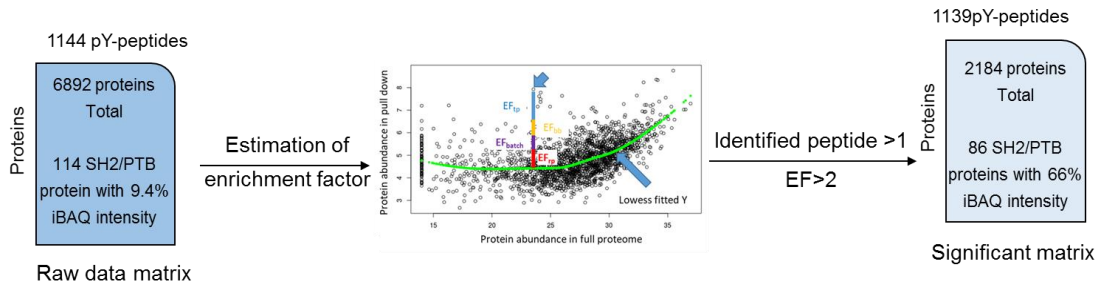


Figure 22 | Statistical analysis and filtering to identify significant interactors of pY-peptides. The raw interactome data set consisted of 114 SH2 or PTB domain-containing proteins with an iBAQ intensity of 9.4%. After removing proteins identified with only one peptide and $EF_{tp} < 2$, a significant interactome matrix was generated with 86 SH2 or PTB domain-containing proteins showed an iBAQ intensity of 66%.

As a proof-of-concept, this stringent filtering enabled the quantification of 20 specific interactors (containing an SH2 or PTB domain) of 11 pY-peptides in DDR1 that were validated by a competition assay⁹⁵. Amongst the six binders that were not quantified, four showed the non-specific binding pattern to the pY-peptides reported in the literature. Moreover, the other two interactors displayed a rather weak competition in the peptide concentration range of 1 to 10 μ M. This discovery illustrated the ability of the filtering step to remove non-specific binders, with a minimal loss of weak binders. In total, this study could quantify 57 interactors of the SH2 or PTB domain of DDR1, thus demonstrating the existence of additional signaling cascades for this receptor.

According to such data cross-evaluation, the interactome network for each RTK quantified by AE-MS could be constructed with less error using the filtering criteria. Also, a comparison of the interactomes of RTKs to reveal biological function similarities and differences is feasible.

3.2.2 Domain analysis of interactors

All significant interactors were submitted to DAVID (<https://david.ncifcrf.gov/>)^{96,97} for domain analysis across the three databases: INTERPRO, PFAM, and SMART. It was clearly shown (Figure 23) that the SH2 domain was the most significantly-enriched protein structure. The other structures were either potential novel binding domains or domains from associated protein complex members.

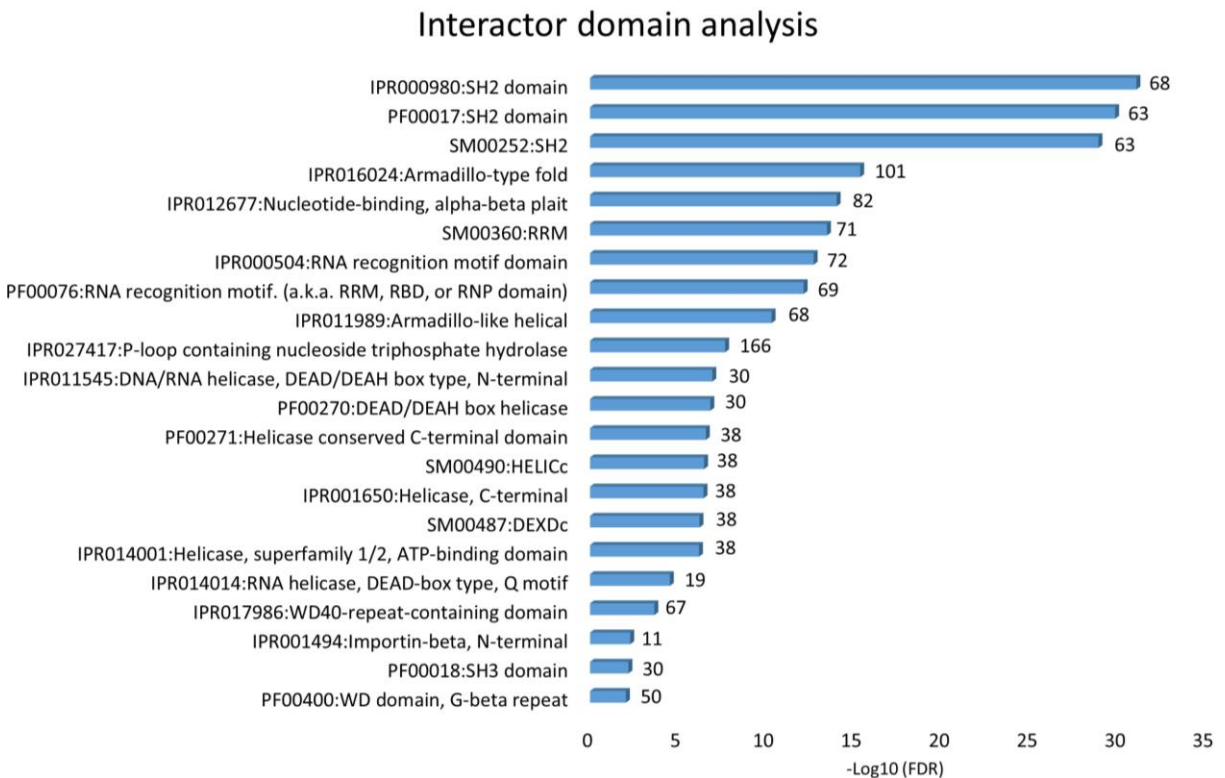


Figure 23 | Protein domain analysis using three databases. The number to the right of each bar represents the number of interactors identified in the category. Only the domain categories with FDRs ≤ 0.01 are shown.

3.3 Analysis of the SH2 or PTB domain-containing interactors

The SH2 or PTB domain-containing proteins dominated the iBAQ intensity for the significant interactome matrix, and the domain analysis addressed this significance across all other interactors. Therefore, further analysis and discussion focused on these interactors.

3.3.1 Binding specificity and motif analysis

The frequencies of the significant interactors containing SH2 or PTB domains was first analyzed to reveal the binding redundancies and specificities of the phosphotyrosine residues. As shown in Figure 25, the protein STAT5B, PIK3R3, and PLCG1 are predominantly enriched across the 1,144 pY-peptides. This is not surprising and provides an explanation as to why these are always identified as interactors in current publicly-available databases.

It is known that RTK activation can trigger well-studied signaling pathways such as PI3K/AKT/mTOR, JAK/STAT, or the recruitment of PLCG1 to produce the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3). Interestingly, these redundant interactors still display a feature that interacts with the pY-peptide exclusive to other redundant interactors. For example, the STAT5B proteins tend to aggregate on the right panel of the heat map, where the other redundant proteins PIK3R3 and PLCG1 display different binding preferences. Therefore, despite non-specific interactions recognizing both N-terminal and C-terminal residues (Figure 24), there are still some specificities to the pY-peptides amongst these versatile interactors. These non-specific interactions could occur in a second binding pocket located in, or outside, the SH2 domains as has been previously observed^{55,56}.

From the middle to the bottom panel of the heat map, interacting partners display increasing binding specificities towards phosphotyrosine residues, which implies unique roles in cell signaling cascades. For instance, the E3 ubiquitin-protein ligase CBLB could specifically interact with 6 pY-peptides (Figure 24). It is known to modulate the degradation of RTKs as part of the negative regulation machinery for the tyrosine phosphorylation of RTKs.

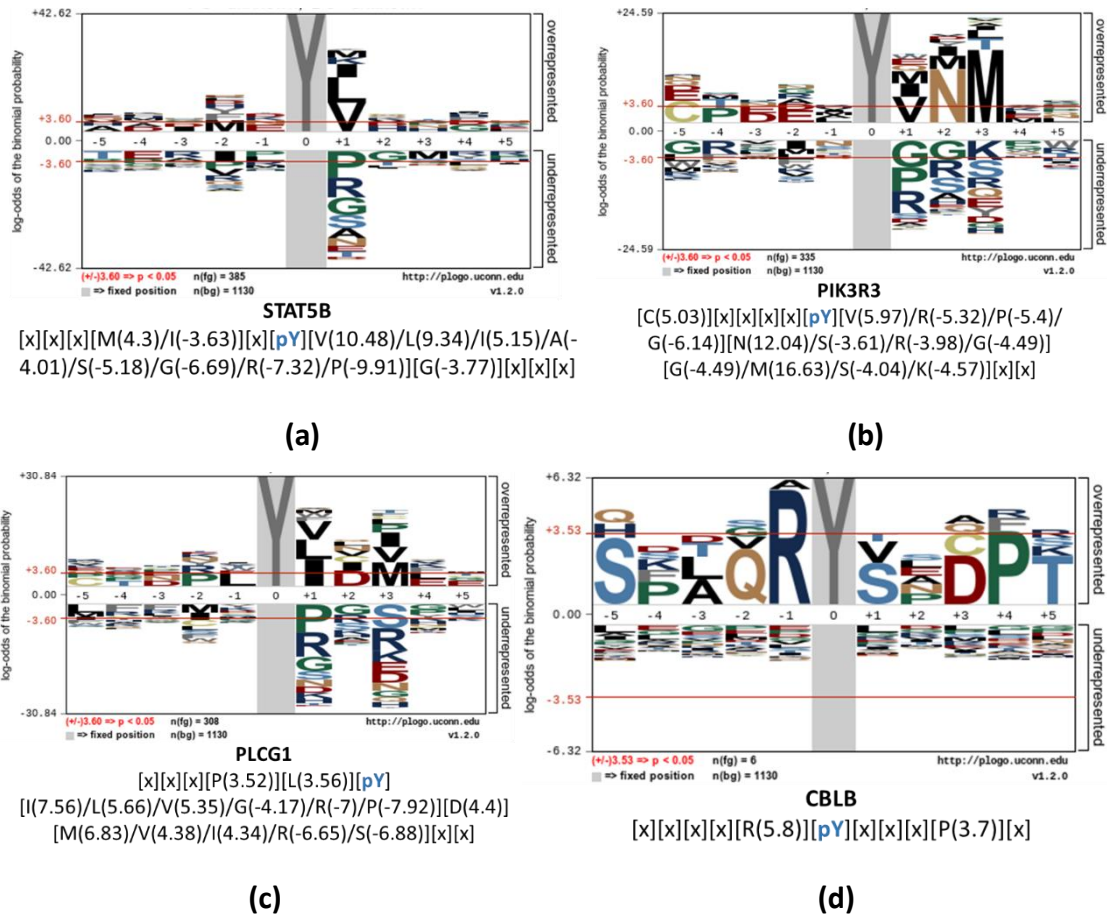


Figure 24 | Binding motifs of STAT5B (a), PIK3R3 (b), PLCG1 (c), and CBLB (d). STAT5B, PIK3R3, and PLCG1 are the redundant interactors that recognize both the N-terminal and C-terminal residues of pY-peptides. CBLB is a specific interactor to the pY-peptides with R at +1 position.

Overall, it can be seen that the PTB domain-containing proteins (including proteins with PTB and PID domains) are more specific binders of pY-peptides in RTKs (motifs are shown in Supplementary File 4). The exception is SHC1. This interactor is more versatile because it is comprised of an N-terminal PTB domain and a C-terminal SH2 domain that both can interact with pY-peptides. It is known that the PTB domain can interact with both pY-peptides and phospholipids, showing a low specificity of ligand binding⁶³. The SH2 domain would be available for this interaction only when the tyrosine residue on the central proline-rich collagen homology 1 (CH1) domain is phosphorylated⁹⁸. Therefore, it can be concluded that the specificity of this interactor would be regulated by the SH2 domain that is utilized to recruit the pY-peptides.

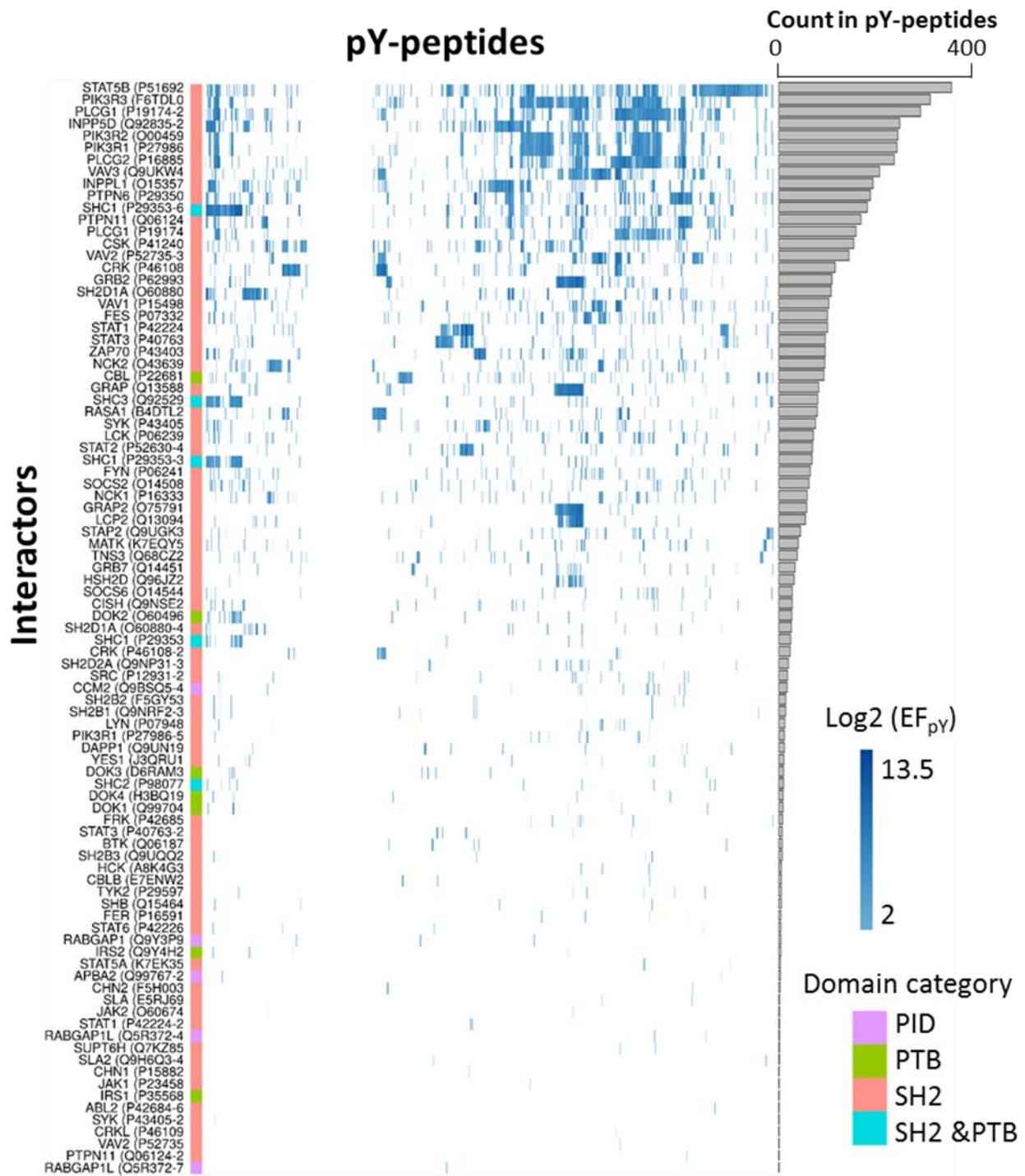


Figure 25 | The interactome of the SH2 or PTB domain-containing proteins to pY-peptides in RTKs. From top to bottom, the specificity of the interaction increases as binding frequency (right panel) decreases. Proteins containing different domains are annotated with the colored bar on the right side of the protein names (left panel).

It is known that the SH2 domain usually recognizes the C-terminal amino acids of the pY-peptides^{49–54}. For instance, the well-recognized binder GRB2 (growth factor receptor-bound protein 2) has been identified as a redundant interactor of 120 unique pY-peptides. These pY-peptides not only possess the consensus motif [pY][x][N]⁹⁹ but also display the binding feature where the N at +1, H and P at -2 positions are critical residues. This motif is also observed in pY-peptides that bind GRB7, where additionally the Y at the +2 position is the statistically significant residue (Figure 26).

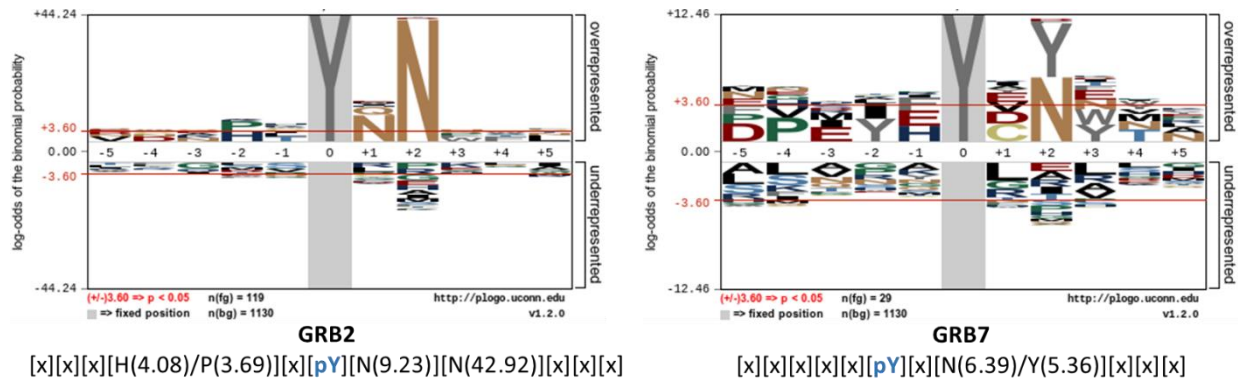


Figure 26 | Motif analysis of GRB2 (left) and GRB7 (right).

Another two adaptors related to GRB2, GRAP, and GRAP2 also have an SH2 domain. These consistently interact with the same motif. Interestingly, GRAP2 only binds to these residues without exception, whilst GRAP still has two other critical residues N and L at positions +1 and +5 (Figure 27).

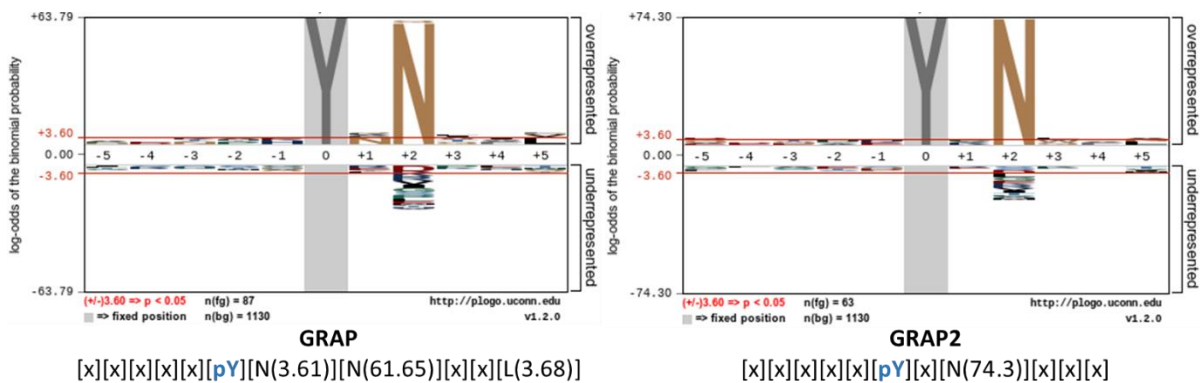


Figure 27 | Motif analysis of GRAP (left) and GRAP2 (right)

One of the most exciting groups of interactors is the STAT family. These play a central role in the JAK-STAT signaling pathway when STAT proteins are phosphorylated. Following homo- or

hetero-dimerization, these enter the nuclei to interact directly with DNA and subsequently regulating cellular processes. The enrichment of this group of proteins demonstrated that such a signaling event could occur even when JAK is not present because the kinase domain of RTKs can also phosphorylate the STAT proteins and promote activity. Amongst the 6 STAT proteins, STAT5B is a ubiquitous interactor, followed by STAT3, STAT1, STAT2, STAT5A, and STAT6. Motif analysis revealed that STAT5B tends to prefer the hydrophobic residues V, L, I at position +1. To some extent, this explains the versatility in binding to different pY-peptides. Encouragingly, the binding motif of STAT2 was determined as [x][x][x][x][x][pY][x][x][R][x][x], and STAT1 displayed the same binding feature to this motif. STAT3, however, displayed a different binding pattern to pY-peptides that possess the motif [x][x][x][x][x][pY][x][x][Q/R][x][x] (Figure 28). Currently, a comprehensive study on the dimerization of STAT proteins has not yet been published, although heterodimers of STAT1-STAT2 and STAT1-STAT3 have been found in response to cytokine stimulation¹⁰⁰. From the perspective of motif analysis in this AE-MS assay, 72.7% of the STAT2 and 43.7% of the STAT3 had a tendency to interact with the same pY-peptides, whilst 57.3% STAT1 and 60% STAT3 also presented the same interaction to pY-peptides. This overlap could indicate the potential existence of heterodimers STAT1-STAT2 and STAT1-STAT3. Meanwhile, the potential existence of novel heterodimers of STAT2-STAT3 was suggested, where 53% STAT2 and 33.3% STAT3 displayed the same interaction to pY-peptides. Nevertheless, the possibility that the STAT proteins did not contain dimers in the AE-MS assay is feasible. As such, the overlap could be a consequence of the homologous SH2 domain.

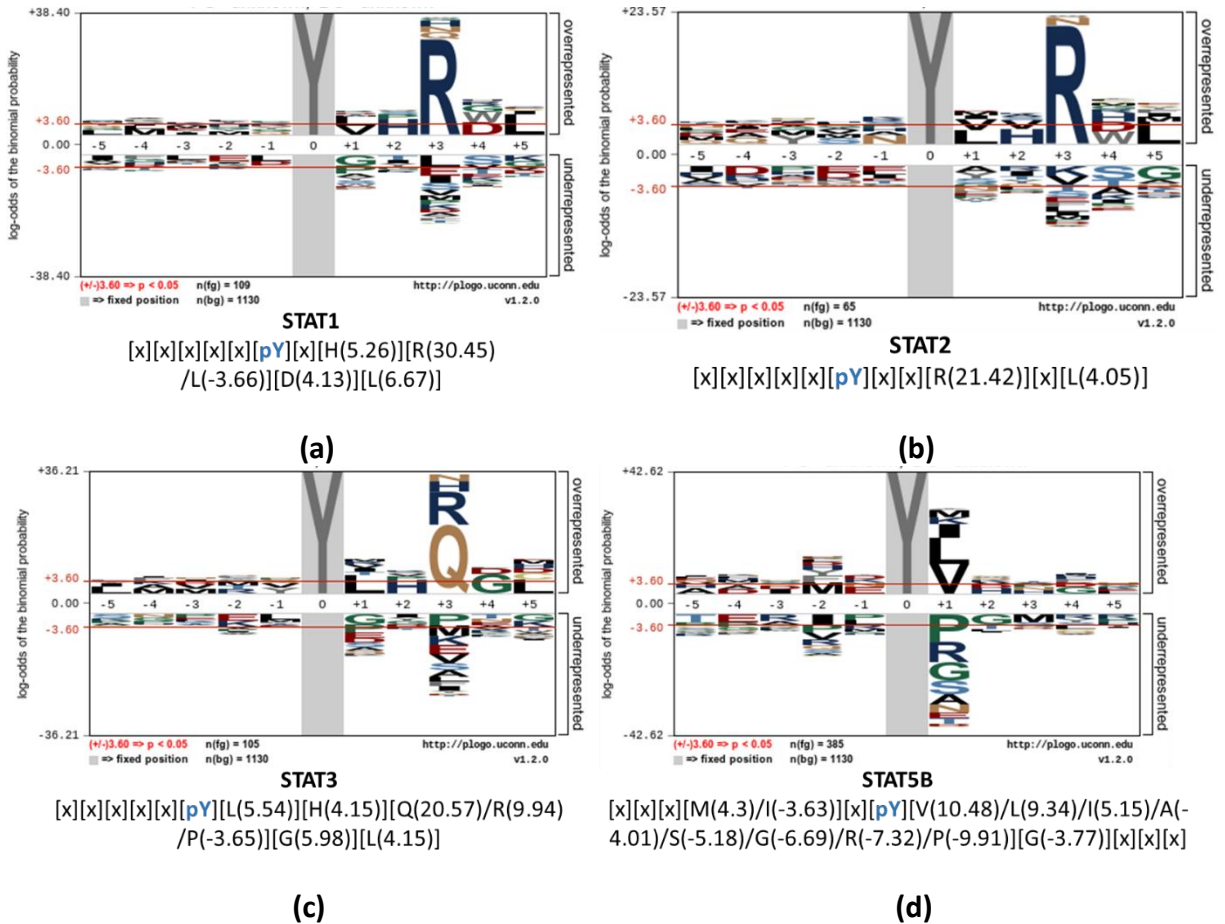


Figure 28 | Motif of STAT1 (a), STAT2 (b), STAT3 (c), and STAT5 (d). Overlapping motifs between STAT1, STAT2, and STAT3 could indicate the existence of heterodimers.

It is true that the pY-peptides in this study also prefer binding to PTB domains via the N-terminal residue. For instance, the protein DOK2 that has the motif feature [x][x][N][P][x][pY] in this study. Although the three isoforms of the adaptor SHC1 were broadly-enriched in this study, these did not display a markedly different preference for binding to pY-peptides. Nevertheless, it was noteworthy to observe that all these motifs suggest the importance of residues on both the N- and C-terminus of the pY residues. This feature again addresses the binding of pY-peptides to the N-terminal PTB domain and C-terminal SH2 domain of the proteins in a different, yet specific, manner (Figure 29).

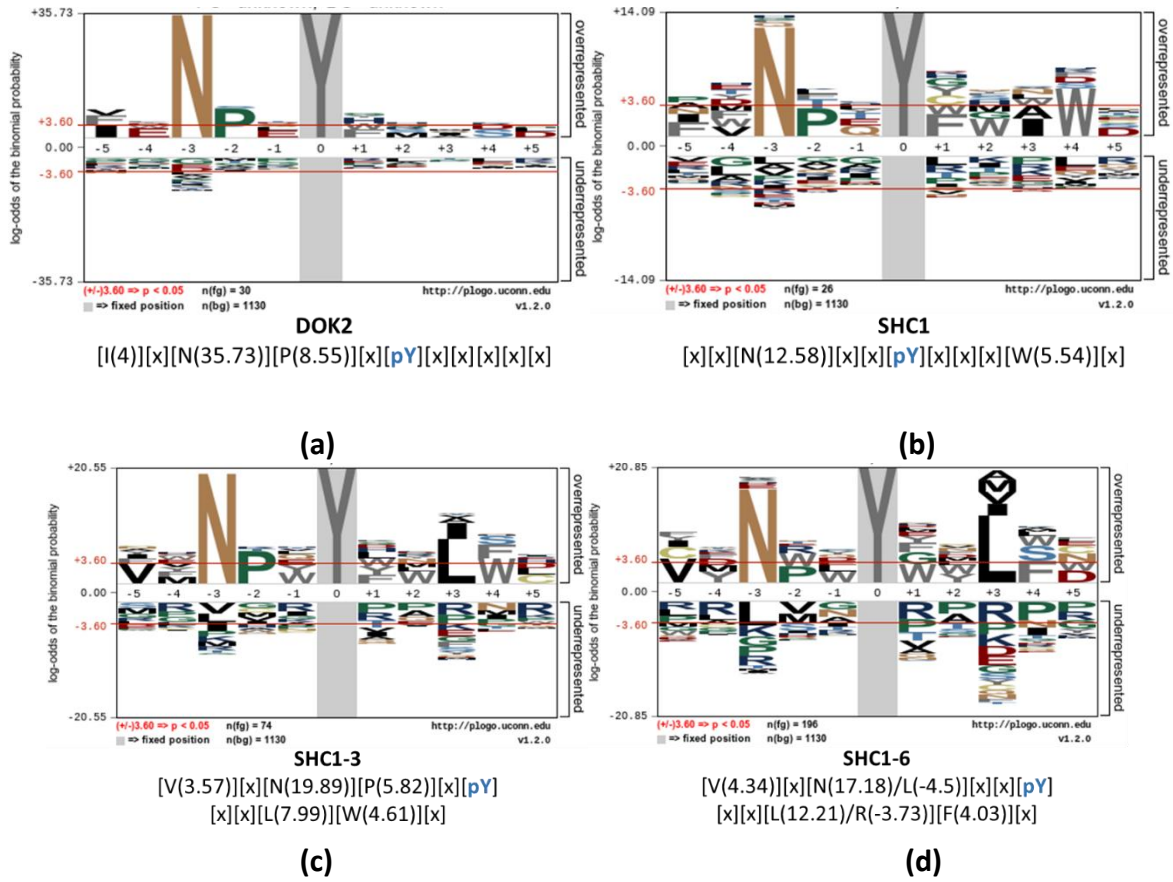


Figure 29 | Motif of DOK2 (a), SHC1 (b), SHC1-3 (c), and SHC1-6 (d)

Overall, the binding specificities of the SH2 or PTB domain-containing proteins are diverse, and the motifs are summarized in the Supplementary File 4.

3.3.2 Comparison with public interactome databases

Systematic comparison of the interactome data set with the public databases from different perspectives could provide scientists with a deeper insight into the phosphotyrosine-centric interaction network. To this end, the pY interactome was firstly mapped to the database BioGRID²¹. Although this database lacks information on the relationship between binding domains and post-translationally-modified sites, it can still provide an overview of all the possible protein-protein interactions.

Overall, the AE-MS in the study not only recapitulated 69 SH2/PTB domain-containing proteins in BioGRID, but also six unreported interactors were discovered: HSH2D, STAT6, DOK3, TYK2, BTK, and RABGAP1L. Interestingly, AE-MS could quantify most of the known interactors of well-studied RTKs such as the EGFR (Figure 30, left panel). From 34 interactors containing SH2/PTB domain quantified in AE-MS, 30 proteins overlapped with the known interactome of EGFR and four are novel interactors. For RTKs that have not been well-studied, for instance, AATK, a further 31 interactors were significantly enriched from the cell lysates in the AE-MS assay. Moreover, one of the reasons for the low number of interactors of AATK in BioGRID could be because the novel interactors and pY do not frequently co-exist in the cell. It is also possible, however, that AATK phosphorylation has not been well-studied because of less research involvement or limitations of available technology to harvest this membrane protein.

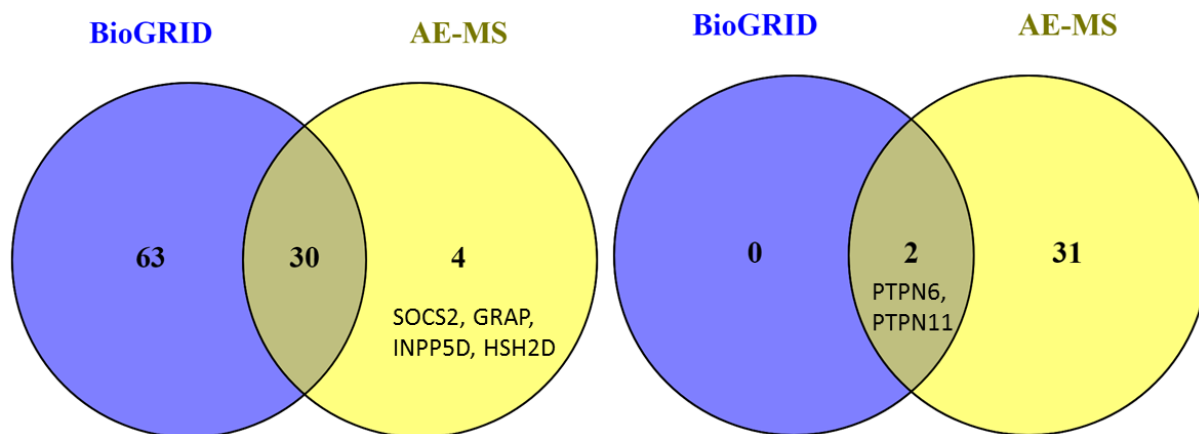


Figure 30 | Comparison of SH2/PTB domain-containing interactors of EGFR (left) and AATK (right) with BioGRID. Interactors of EGFR in AE-MS overlapped to a high degree with BioGRID. For AATK, AE-MS identified the only two known interactors in the databases plus a further 31 new interactors.

Due to limitations of BioGRID, many studies have engaged experiments to learn the interaction between peptides and interacting domains/proteins. One of the most promising recent studies, PepSpotDB, has released the interaction landscape of 70 recombinant SH2 domains from 66 genes against 6,057 synthetic peptides²⁶. From these, 251 are intracellular pY-peptides of RTKs. A comparison to the SH2 interactome landscape was performed to cross-evaluate these two different approaches to identify typical interactors that contain SH2 domains.

Overall, AE-MS successfully recovered 177 (31.8%) interactions between 59 pY-peptides and 33 SH2 domains (29 proteins) (Figure 31). The AE-MS study also showed the capacity to interact with 13 proteins that were not successfully expressed in PepSpotDB. These additional interactors that were not selected for the screening in PepspotDB are composed of 11 SH2 domains and 5 PTB domains. This discovery highly supports the notion that the successfully-recovered proteins could be direct interactors and suggests the advantage that AE-MS provides in identifying novel interactors without the requirement of target domain expression. By interacting with PTB domain-containing proteins in an AE-MS assay, the interactome landscape of RTKs is now not only limited to SH2 domain-containing proteins.

Moreover, with the current increase in newly-discovered phosphotyrosine residues, PepSpotDB is no longer comprehensive. Nevertheless, the database could be supplemented with our AE-MS data set.

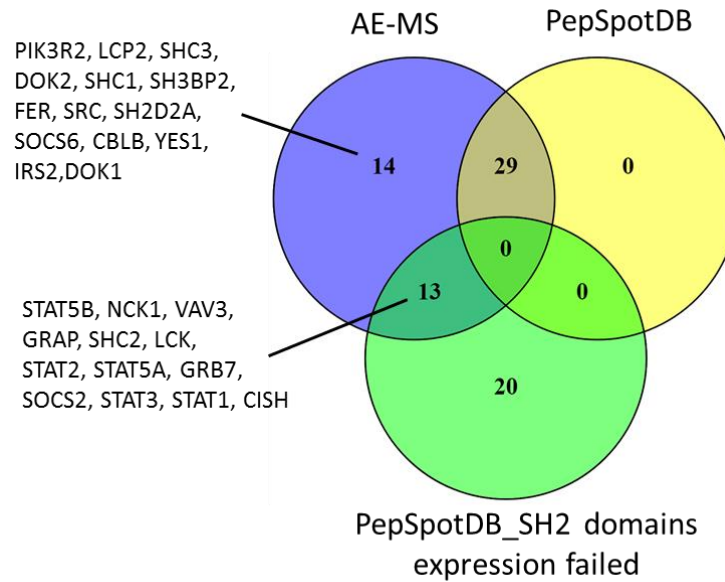


Figure 31 | The overlap of SH2 domain-containing interactors in AE-MS assay with PepSpotDB. For further comparison, the number of failed expressed SH2 domain is shown in green.

As a database for studying protein PTMs, PhosphoSitePlus® (PSP)²⁴ currently has approximately 480 intracellular phosphotyrosine residues of RTKs that have been characterized by mass spectrometry or antibody recognition. This number covers about half of the wild-type tyrosine residues (48%) of RTKs. From these, 53.3%, 27.7%, and 19% have been localized to the protein kinase domain, the C-terminus, and the juxtamembrane domain, respectively. Furthermore, 28 tyrosine residues are classified as disease-promotion markers, whilst 216 pY residues are involved in the regulation of RTK activity. Further analysis revealed that 154 pY-peptides have been recorded to interact with 46 unique genes of proteins (excluding isoform and non SH2 domain-containing proteins) in 288 interactions. These interactors were primarily identified by the co-IP together with residue mutation in the full-length receptor. This mutation-based investigation method is, however, restricted to the utilization of antibodies. As such, not all interactors can be identified at the same time point. Therefore, a comprehensive blueprint of these phosphotyrosine sites cannot be provided. Compared to the retrieved interactors from PSP, the AE-MS data set not only enriched the same interactors but also displayed the ability to recruit more potential true binders.

Overall, the AE-MS assay recovered 176 interactions (61.1%) from 95 pY-peptides and 25 interactors in PSP. Conversely, 191 regulatory pY-peptides without interactors in PSP were

found to possess 1,455 interactions with 63 proteins in the AE-MS dataset. Of these, 116 interactions between 27 proteins and 45 pY-peptides were also observed in the PepSpotDB database. Interestingly, there are still 20 disease-related phosphotyrosine residues where no interactors have been determined. AE-MS provided 103 possible interaction with 38 proteins, plus 28 additional binders to the other eight pY-peptides with known interactors. Again, 19 novel interactions to 6 pY-peptides were validated by PepSpotDB.

A further annotation to the ProteomicsDB was performed to provide expression levels for the RTKs and interactors. Together with the other three databases, this work would enable discovery of the probability of a binding event occurring in cells. For instance, the receptor INSR interacts with the guanine nucleotide exchange factor VAV3. According to BioGRID, this interaction has been previously shown by affinity capture-western blot. Information in PSP and the AE-MS data revealed that INSR_pY1149 is responsible for this binding. As reported in PSP, the interaction is relevant to the development of diabetes mellitus. Information from PSP suggests the INSR_pY1149 residue is present in Jurkat cells, and the ProteomicsDB data indicates that INSR and VAV3 are co-expressed in this cell lines. The protein VAV3 is known to possess multiple functions in the cell including regulation of the B cell receptor signaling pathway. It is possible that VAV3 could similarly influence the T cell receptor signaling pathway. This hypothesis holds true because it has been proven that the upregulation of INSR and the following T cells activation is required for efficient adaptive immunity¹⁰¹. More interestingly, another residue INSR_pY1185 was also shown to interact with a second guanine nucleotide exchange factor, VAV1. Moreover, the co-expression of this interactor and the pY residue has been validated in Jurkat cells by both PSP and ProteomicsDB. Therefore, uncovering the effect of these two binding events in Jurkat cells would be an interesting future perspective.

Overall, integrating all the data from the AE-MS assay with the four databases has created an impressively comprehensive blueprint of all the possible interactors of pY-peptides in RTKs. Therefore, it could drive new biological hypotheses and discovery in the future.

3.3.3 Interactome of wild-type tyrosine residues

It would be interesting to determine if the biological relationship of receptors fell in the same subfamily of RTKs. This could be ascertained by comparing proximal interactome similarities and differences. Encouragingly, the AE-MS assay could provide the most comprehensive data set for this purpose. Several examples are given below.

Two homologous RTKs, discoidin domain receptor 1 and 2 (DDR1 and DDR2) share 51% sequence similarity and possess 15 and 14 tyrosine residues in the intracellular domain, respectively. Both display distinct preferences for ligand recruitment and activation kinetics compared to other RTKs¹⁰². DDR1 prefers the basement membrane collagen type IV, and DDR2 accepts the ligands of collagen type II and type X¹⁰³⁻¹⁰⁵. Evidence from traditional biochemistry methods so far only suggested that stimulation of DDR1 with collagen would induce more diverse signaling pathways such as PI3K/ATK, Ras/ERK/MAPK, and STAT5A/B. Moreover, for DDR2, a different direction in signaling transduction is preferred and is related to the recruitment of the kinase SRC and adaptor SHC proteins¹⁰⁶. A complete understanding, however, of the signaling pathways is still absent and challenging. For example, the different binding specificities to the activation ligands would induce distinct signaling directions. Moreover, signaling is cell/tissue type-specific and context-dependent¹⁰⁶. Therefore, resolving all the potential interactors of the phosphotyrosine residues would firstly provide access and direction to understand the signal transduction pathways in detail.

The phosphotyrosine-mediated interactome of DDR1 has been systematically profiled²⁸, but not DDR2. By using a similar approach, construction of the DDR2 interactome would enable the discovery of similarities and differences in the proximal interactome and potential signaling pathways of these two receptors. Employing a faster and more sensitive mass spectrometry, the AE-MS study has achieved this goal. From the AE-MS data, a much more comprehensive picture of the DDR1 interactome consisting of 57 proteins was generated; with only the loss was of two low-affinity interactors. Comparison between the two receptors revealed that DDR1 and DDR2 shared 22 (59.5%) interactors (Figure 32). This indicates that the proteins exist as a core signaling hub. Interestingly, DDR2 specifically prefers recruitment of adaptor proteins (GRAP2, GRB2, and GRAP), tyrosine kinases (ZAP70 and LCP2). The DDR1 counterpart showed

more diversity, and included binding to numerous signal regulators and transcription factor. To some extent, the signaling preferences of DDR1 and DDR2 could be explained by these unique interactors, although these are not completely matched to each.

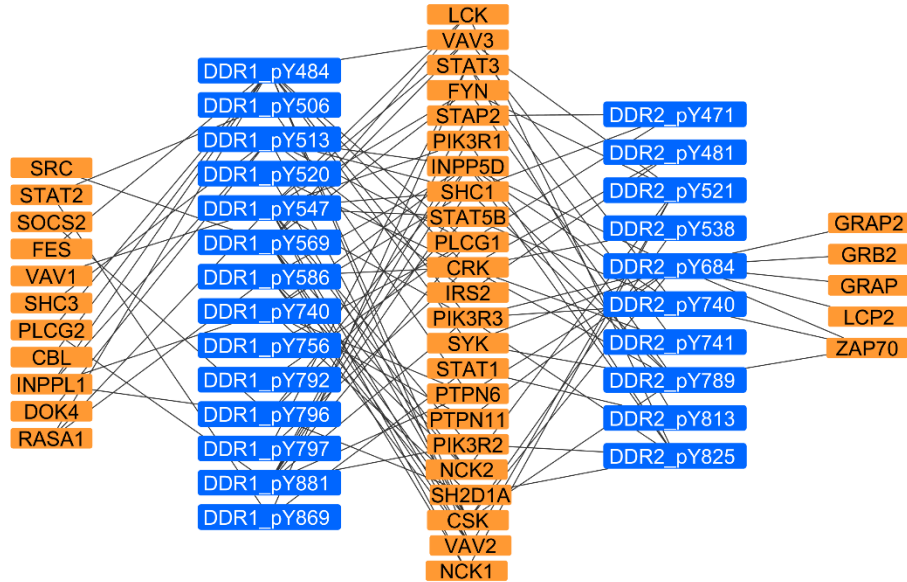


Figure 32 | Comparison of interactome of DDR1 and DDR2. The pY-peptides and interactors are shown in the blue and orange rectangle, respectively. The edge between them indicates the interaction. Apart from the overlapping interactors (middle panel), DDR1 tends to trigger more diverse signaling cascades by recruiting proteins with multiple functions (left panel). Conversely, DDR2 prefers signaling pathways with adaptor proteins and kinases (right panel).

In a similar manner, the interactome of another RTK sub-family, ErbB, has been previously profiled²⁷. Only ten unique proteins were shown to interact with all 41 pY-peptides. The adaptors GRB2 and SHC were identified as the dominant interactors, which is not as diverse as the databases BioGRID, PepSpotDB, and PhosphoSitePlus. Therefore, it was necessary to profile the interactome more comprehensively by AE-MS. As a result, the assay recovered 66.7%, 77.8%, 90%, and 75% of the interactions for EGFR, ERBB2, ERBB3, and ERBB4, respectively. The tyrosine phosphatase PTPN11 was the least identified interactor in the corresponding AE-MS pY-peptides, but was frequently identified as an interactor with other pY residues of this family. This inconsistency could imply that the interactions of PTPN11 to the reported sites are of low affinity in the AE-MS assay. Despite this observation, the number of interactions to 83 wild-type pY-peptides from these four receptors was dramatically increased from 54 to 643 (Figure 33). Most of the interactions were determined as novel, except 73 and 60 of them were validated by

the databases of PhosphoSitePlus and PepSpotDB, respectively. In the end, the comprehensive interactome network of the EGFR family has been updated by this AE-MS study. This systematic interactome of the ErbB family significantly contributes to understanding the relationship of these receptors in various diseases such as cancer. For example, from 5 cancer-relevant regulators, EFGR_pY1092 (usually associated with the non-small cell lung cancer (NSCLC)²⁴) was identified as interacting with 18 proteins. From these, 14 interactors were novel to this AE-MS assay. As such, this could provide a new direction to learning more about this residue in cancer. Indeed, half of these novel interactors co-express with EGFR in cell lines in ProteomicsDB, and the EFGR_pY1092 residue is ubiquitously identified in the PSP database. Therefore, it would be an excellent starting point to evaluate the function of this phosphotyrosine residue in other cancers including NSCLC.

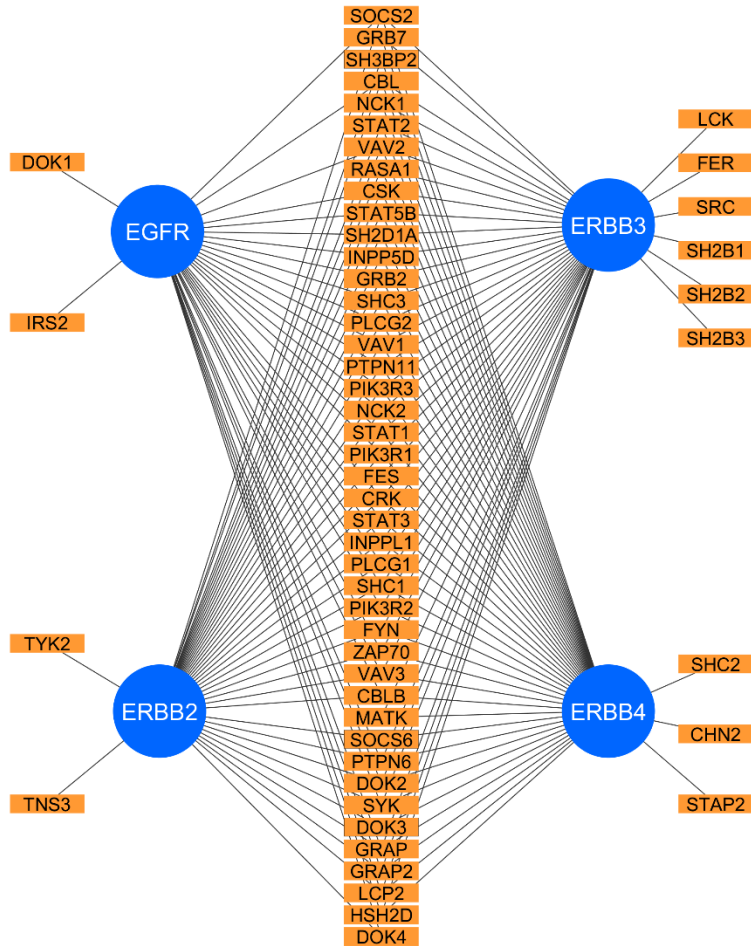


Figure 33 | The updated interactome of the ErbB family with wild-type pY peptides by AE-MS. Four receptors of ErbB family and their interactors are shown in the blue ellipse and orange rectangle, respectively. The edge between them indicates the interaction. These four receptors display the shared (middle panel) and unique (left or right panels) interactors, indicating the core signaling hub in the cell as well as the specific interaction cascades.

Besides the ErbB family that is highly connected to disease development, the fibroblast growth factor receptors (FGFR) are also involved in various cancers, and there are several clinical kinase inhibitors for cancer treatment¹⁰⁷. This subfamily consists of 4 members, FGFR1, FGFR2, FGFR3, and FGFR4, that share 45.4% sequence similarity⁵. In total, there are 63 different phosphotyrosine residues, with seven shared between two or three proteins. The redundant residues no doubt demonstrate aspects of identical signaling capacities when activated. Currently, however, there is insufficient information concerning the specific interactors of the phosphotyrosine residues. For the first time, the AE-MS assay provides a comprehensive

phosphotyrosine interactome of the FGFR family to aid in understanding potential signaling cascades.

Overall, the four receptors can recruit the same 22 proteins; and excluding FGFR1, each also has specific interactors (Figure 34). Deciphering these interactors would drive a deeper understanding of shared or unique signaling roles in health or diseases. For instance, it was found that the N-terminal mutation of G388R would expose a membrane-proximal cytoplasmic site FGFR4_pY390 for the recruitment of STAT3 (signal transducer and activator of transcription 3). Then the binding event would enhance the STAT3 signaling in breast and lung cancers¹⁰⁸. Not surprisingly, the wild-type FGFR4_pY390 also interacted with STAT3 in the AE-MS assay. This same binding between mutant and wild-type could be explained by the STAT3-binding motif that only recognizes the C-terminal R/Q at the +3 position. Therefore, it was proven that STAT3 is a direct interactor of FGFR4_pY390. Furthermore, this interaction was also observed with FGFR1_pY677 and FGFR2_pY680, indicating potential consistent roles in cell signaling. More interestingly, STAT1 was observed to uniquely interact with this site on FGFR4. Would STAT1 elicit the same function as STAT3 in cancer? Further experiments on interaction and function could be performed in the future.

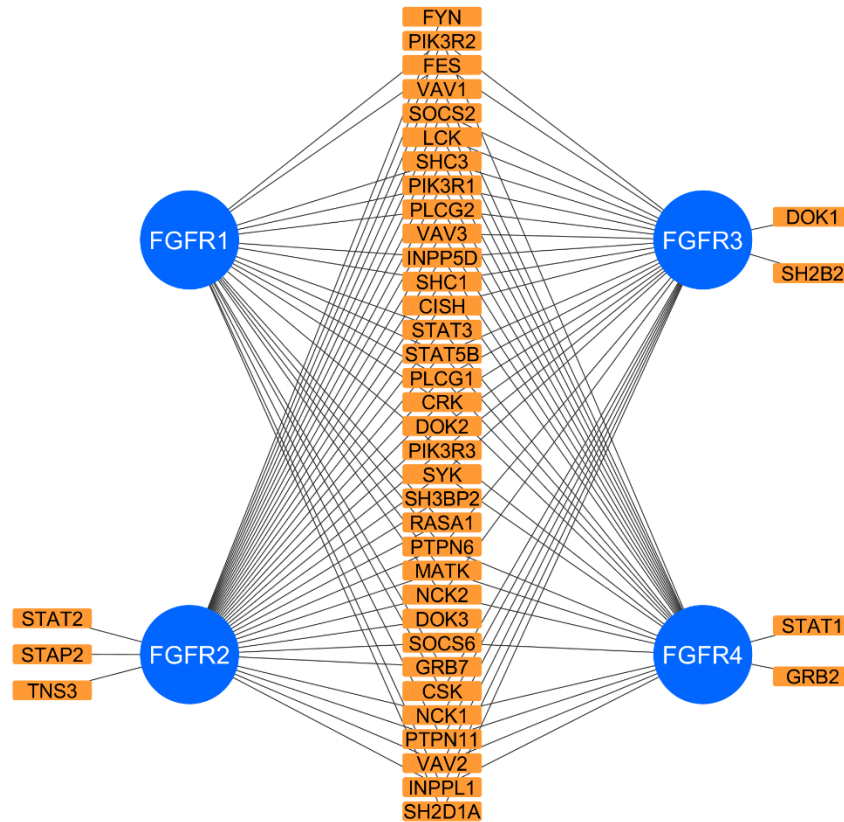


Figure 34 | Interactome comparison between FGFR1, FGFR2, FGFR3, and FGFR4. Four receptors of FGFR family and their interactors are shown in the blue ellipse and orange rectangle, respectively. The edge between them indicates the interaction. Except FGFR1, the other three receptors all have unique interactors.

It is common knowledge that a protein from a single gene can exist as several subtypes, called ‘isoforms’. This phenomenon is also true for the RTK family, especially for receptors that are ubiquitously expressed in different tissues. With variations in the sequence of a receptor, the tyrosine residues could also differ from one another, and thereby have entirely distinct functions at different localizations. One example is the interactome of NTRK3_pY555 and NTRK3_pY557, which are known to localize in the C-terminus of the isoform 2 of the receptor replacing the canonical sequence of NTRK3_pY558. Protein mapping has revealed that the canonical NTRK3 is widely-expressed but mainly in nervous tissue, and isoform two is expressed at a higher level in adult brain than fetal brain (<https://www.uniprot.org/uniprot/Q16288#expression>)⁵. A possible answer to differing expression levels is differences in the interactome. Indeed, the AE-MS assay revealed that with this change in the tail of the receptor, the interactors of NTRK3_pY558 would be completely

switched from GRB2, INPP5D, and PTPN6 to GRB7, PIK3R2, and TNS3 on NTRK3_pY555 (Figure 35). The full proteome of brain¹⁰⁹ revealed that the isoform NTRK3 is co-expressed with GRB2, PIK3R2, and TNS3. This means that the signaling switch possibly exists. Similarly, pY557 could significantly reverse the interactor to the other 15 SH2/PTB domain-containing proteins. The full proteome profiling¹⁰⁹ of human brain revealed that the interactors NCK2, PLCG1, NCK1, PIK3R1, PIK3R2, FYN, and TNS3 are co-expressed with isoform 2 of NTRK3. Therefore, it is possible to trigger the relevant signaling pathway in the brain via the pY residues of the isoform. In the end, the NTRK3 could execute the specific functions in the brain.

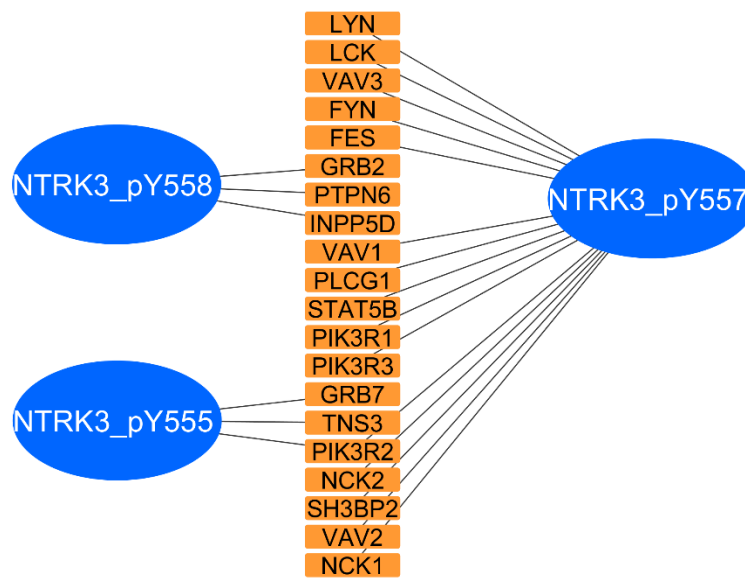


Figure 35 | Interactome comparison of three isoform pY-peptides of NTRK3. The pY-peptides and interactors are shown in the blue ellipse and orange rectangle, respectively. The edge between them indicates the interaction. The completely reversed interactors in the pY-peptides of isoform NTRK3 suggests the different signaling capacities.

3.3.4 Interactome of hot-spot mutations

To date, 16 amino acid hotspot mutations that flank functional phosphotyrosine residues in RTKs have been discovered. All are clinically-linked to cancer development. For example, the hot-spot mutation of D769Y flanking ERBB2_pY772 is known as a gain-of-function mutation³¹. Moreover, the ERBB2-targeted inhibitor Neratinib³² shows the potency to treat cancer with this ERBB2 alteration. The molecular mechanism of how this mutation attenuates or enhances the existing signaling cascades is not yet evidenced, nor is the inhibitor treatment. Deciphering the interactome of this hot-spot mutation would aid in understanding the underlying mechanism and ideally drive the development of drug treatment.

Taking the ERBB2 as an example. There are three hot-spot mutations in the N-termini of ERBB2_pY772: D769Y, D769H, and D769N (Figure 36). When substitution of D769Y occurs, the ERBB2_pY772 can lose seven interactors, including the kinase ZAP70. Meanwhile, the mutant can gain three interactors: CBL, PTPN6, and SYK. If D at position 769 is replaced by a positive residue H, the interactome of wild-type ERBB2_pY772 can lose the four interactors that are also lost with ERBB2_pY772_D769Y. Moreover, a new interactor VAV1 is gained by the mutation of ERBB2_pY772_D769H. A similar pattern was observed with the occurrence of ERBB2_pY772_D769N. The docking proteins DOK2 and DOK4, and the scaffold protein TNS3 are gained.

Also, there are two residues that mutate in the C-terminus of ERBB2_pY772: V777L and V777M. Although these two mutations are four amino acids distant to ERBB2_pY772, ERBB2_V777L can still recruit the new interactors FES and TNS3. Overall, this loss and gain of interactors can provide an explanation as to how problems in ERBB2 signaling can occur. This hypothesis should be experimentally validated in the future.

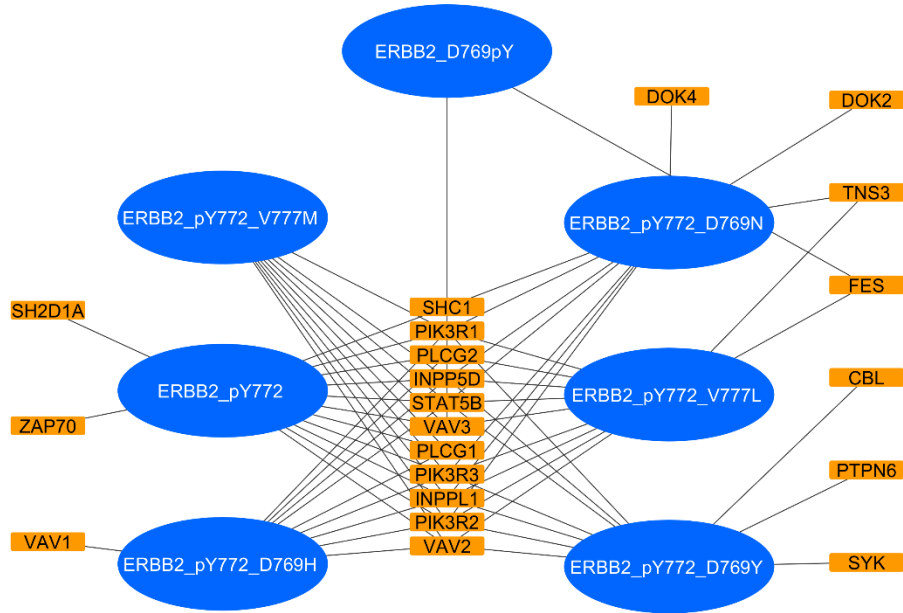


Figure 36 | Interactome comparison of wild-type and hot-spot mutations in ERBB2. The pY-peptides and interactors are shown in the blue ellipse and orange rectangle, respectively. The edge between them indicates the interaction. Hot-spot mutation at N-terminal and C-terminal of pY-peptides alter the interactome of wild-type ERBB2_pY772.

3.3.5 Interactome of mutation-gain tyrosine residues

There are 204 missense mutations frequently observed in 49 RTKs. Here, tyrosine residues replace other amino acids; referred to as mutation-gain tyrosine (MGY) in this study. Moreover, a few have been identified as associated with cancer development and treatment³¹. For example, the ERBB2_pY772 mentioned above has a mutation of D769Y, and the influence of this alteration on the interactome has been shown. Another hypothesis about the influence of such a mutation is that this MGY could act as a docking site to trigger the signaling pathway when phosphorylated in the cell. This theory was proven by the interactome of ERBB2_D769pY in the AE-MS assay, which could interact with the kinase FES (Figure 36). Therefore, profiling the interactome of all the phosphorylated MGY residues would also be a great help in revealing the potential influences on the wild-type interactome.

It was found that the DDR2 possesses 10 MGY residues, while DDR1 only contains one. Interactome analysis depicted that these DDR2 mutant residues possible induce the same signaling cascades as DDR1 by interacting with four same interactors. Moreover, the MGY residues of DDR2, can recruit two novel signal regulators HSH2D and SH2D2A (Figure 37). Overall, with MGY residues, DDR2 tends to be more versatile in recruiting new interactors than wild-type DDR2. Conversely, the MGY residues of DDR1 could only recruit two common interactors with the wild-type DDR1 (VAV2 and VAV3). This overlap suggests that the existing signaling pathways of VAV2 and VAV3 could be amplified by mutant DDR1.

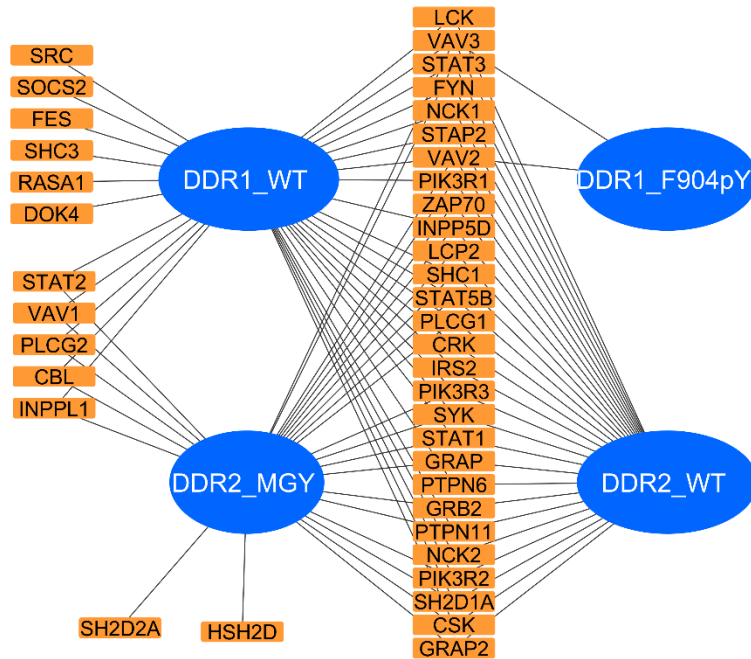


Figure 37 | pY-peptide interactome of wild-type DDR1, wild-type DDR2, and mutant DDR2. The wild-type (WT) and mutant (MGY) receptors are shown in the blue ellipse, and the interactors are presented with orange rectangle. The edge between them indicates the interaction. MGY residues of DDR2 recruit two more new interactors than wild-type DDR1 and DDR2.

Another RTK sub-family, the ErbB family, also possesses 18 MGY residues. ERBB3 and ERBB4 contain the most diverse mutations across various amino acids. Compared to the WT interactome of the ErbB family, the MGY mutant interactome possesses 37 shared binders and two additional interacting proteins: CISH and RABGAP1L. Moreover, the two interactors are all from MGY residues of ERBB3, which could illustrate the completely different signaling cascades triggered by the mutant ERBB3 (Figure 38).

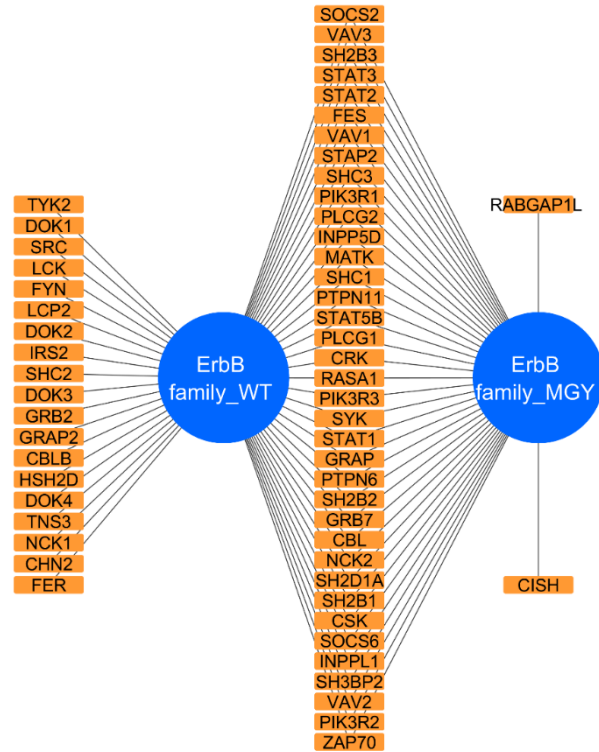


Figure 38 | pY-peptide interactome comparison between wild-type (blue ellipse on the left) and mutant (blue ellipse on the right) ErbB family. The edges between them indicate the interaction. The MGY residues of ErbB family receptor are capable of recruiting same interactors to wild-type receptors, with two more binders (orange rectangle on the right).

Last but not least, a similar comparison was also performed for the FGFR family. It was shown that the FGFR family of receptors generated an MGY interactome where 15 of the same interactors as the WT interactome were recruited. Meanwhile, three novel proteins: SRC, CBL, and ZAP70 were exclusively bound via the contribution of three MGY residues in FGFR2 (Figure 39).

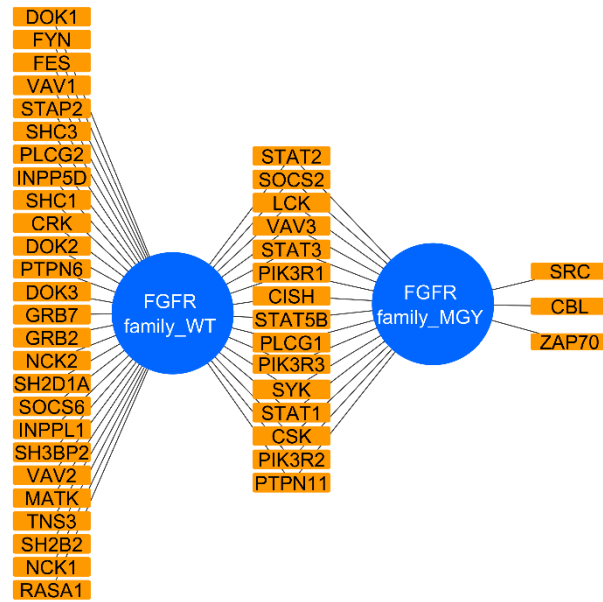


Figure 39 | pY-peptide interactome comparison between the mutant (blue ellipse on the right) and wild-type (blue ellipse on the left) FGFR family. The MGY residues of FGFR family interact with three new interactors (orange rectangle on the right)

Overall, this is the first time of profiling the MGY interactome of RTKs by AE-MS. These encouraging results could be specific resources in the future for the cancer research where mutations are frequently observed. Validation of this interactome at the full-length protein level could be performed by employing gene-editing tools such as CRISPR-Cas9 together with co-IPs. After that, drug treatment could be performed to further investigate potencies, which could promote new treatment strategies. All the other RTKs interactome network were constructed in Supplementary File 5.

3.4 Non-SH2 or PTB domain-containing interactors

Apart from the SH2 or PTB domain-containing proteins, 2,098 proteins that shared 34% of the iBAQ intensity were identified as potential interactors of the pY-peptides of RTKs. Amongst these atypical interactors, 197 PIK3CA, 359 PIK3CB, and 184 PIK3CD are known protein complex members of PIK3R1, PIK3R2, and PIK3R3. The GRB2 complex members SOS1 and SOS2 were also identified together with GRB2 83 times. Regardless of these known protein complex members, more effort was spent on searching for new interactors of the pY-peptides in RTKs. Moreover, encouragingly, several proteins were identified as interactors that bind to specific motifs. These interactors included, but were not limited to, the representative examples below.

3.4.1 Interactors with HYB or C2 domain

The HYB domain of CBLL1 is known to interact with pY-peptides; however, whether this interaction occurs with pY-peptides from RTKs has not been reported. Encouragingly, the AE-MS assay proved the existence of this interaction. Albeit only four wild-type pY-peptides significantly recruited CBLL1. This low-frequency of interaction could be the reason why this has not been previously observed with conventional biochemistry approaches. Among these binding peptides, three of them harbor an E at position +3. To some extent, this matches the description in the literature⁷⁰ that acidic residues at position +2 and +3 mainly contribute to binding to HYB domains. With this discovery, the function of the interaction between RTKs and CBLL1 would be worthy of future exploration.

Not surprisingly, another interacting domain of pY-peptides, *i.e.*, the C2 domain of PRKCD and PRKCQ, was significantly enriched in 18 and 8 AE-MS assays, respectively. Three pY-peptides interacted with both PRKCD and PRKCQ, and both contain an R at the +1 position. Further motif analysis, however, revealed that PRKCQ preferentially interacts with pY-peptides containing an R residue at the C-terminus of the peptides, while PRKCD tended to interact with the motif [x][x][x][x][C][pY][x][L][x][x][x] (Figure 40).

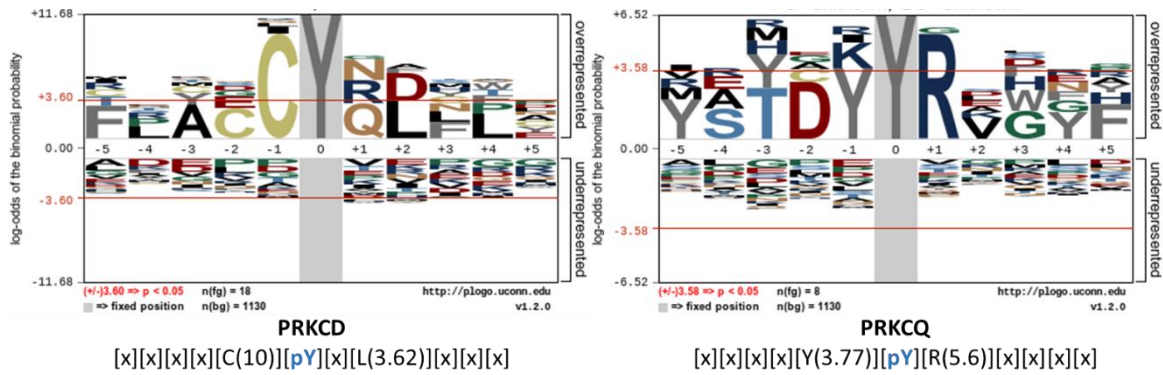


Figure 40 | Motifs binding to PRKCD (left) and PRKQC (right). PRKCD prefers binding to pY-peptide with C at -1 position while PRKQC is biased to R at +1 position of the peptide.

The other 7 C2 domain-containing proteins (excluding proteins with an SH2/PTB domain), such as DOCK2, were also enriched by 45 pY-peptides in the AE-MS assay. These peptides, however, do not contain a significant binding motif.

3.4.2 Interactors with PTP or PTP-like domain

The experimental buffers used in the AE-MS assay contain a higher concentration of the three phosphatase inhibitors. Thus, it was hypothesized that the catalytic motif in the phosphatases was temporarily inhibited. Therefore, the binding between the pY-peptides and the temporarily-inactivated phosphatase domain could be established under these experimental conditions.

As a proof-of-principle, 37 phosphatases were significantly enriched by 237 pY-peptides. In 193 AE-MS assays, ubiquitin-associated and SH3 domain-containing B (UBASH3B) was the most-enriched protein. It is known that UBASH3B can display phosphatase activity towards EGFR, subsequently downregulating the receptor⁵. Therefore, the observation of the interaction between UBASH3B and phosphotyrosine residues implies a potential, direct interaction. Motif analysis suggested that two structurally-similar residues, F and Y at -2 position and another F at +2 position, are the critical amino acids for the recruitment of this phosphatase. A similar event also occurs with the recruitment of protein tyrosine phosphatase non-receptor type 12 (PTPN12), that was found to bind to the motif with I and W specifically located at the C-terminus (Figure 41).

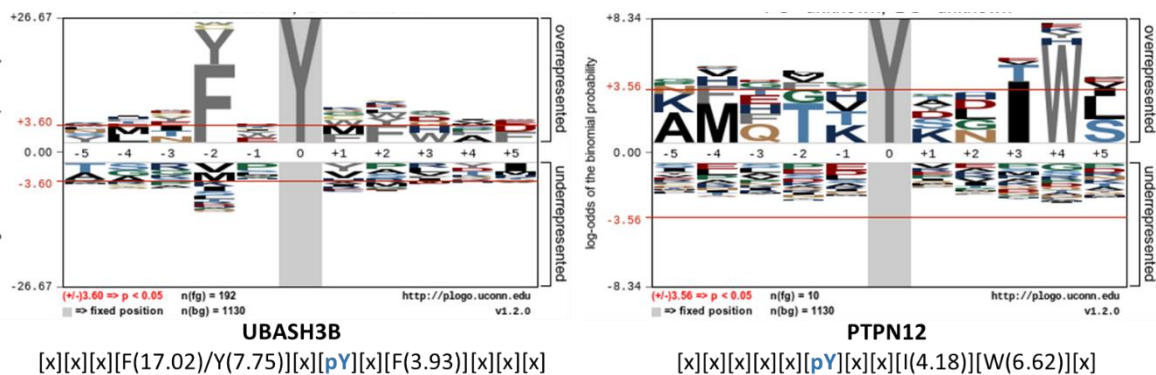


Figure 41 | Motifs binding to UBASH3B (left) and PTPN12 (right)

Protein tyrosine phosphatase non-receptor type 2 (PTPN2) was only enriched by 3 pY-peptides and no feature motif was identified. Conversely, PTPN1 was identified as an interactor of 18 pY residues with a specific motif. This motif, however, is significantly different from the two phosphatases above (Figure 42). Such a discovery indicated the ability of phosphatases to regulate different phosphotyrosine residues.

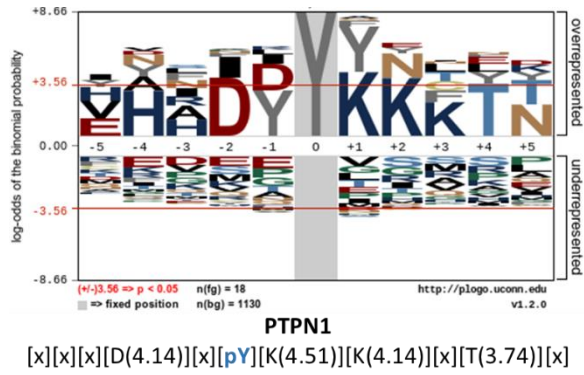


Figure 42 | PTPN1 prefers binding to pY-peptides with K at C-terminus and D at N-terminus

As protein tyrosine phosphatases with SH2 domains, PTPN6 and PTPN11 could also interact with the pY-peptides via a similar mechanism to these phosphatases. The two interactors, however, both suggested different binding motifs to the three phosphatases above. PTPN6 and PTPN11 could both bind to 122 of the pY-peptides, and uniquely interact with 84 and 63 pY-peptides, respectively. It is known that they both contain an N-terminal and a C-terminal SH2 domain, and these preferentially bind different motifs. The discovery of versatile binding motifs in the AE-MS assay confirmed the existence of binding to two different domains (Figure 43).

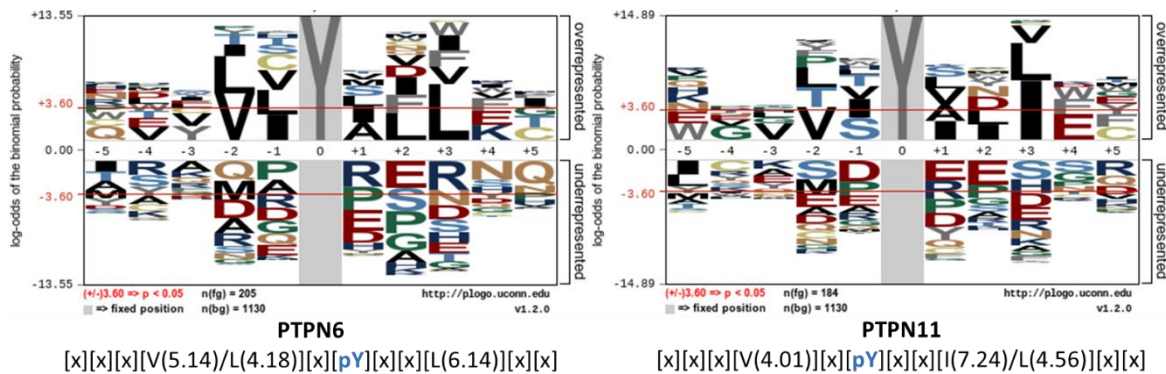


Figure 43 | Binding motifs of PTPN6 (left) and PTPN11 (right). The various binding patterns indicated interactions between the N-terminal and C-terminal SH2 domain exist in these two proteins.

In the end, the interaction of these five phosphatases with pY-peptides was constructed as shown in Figure 44.

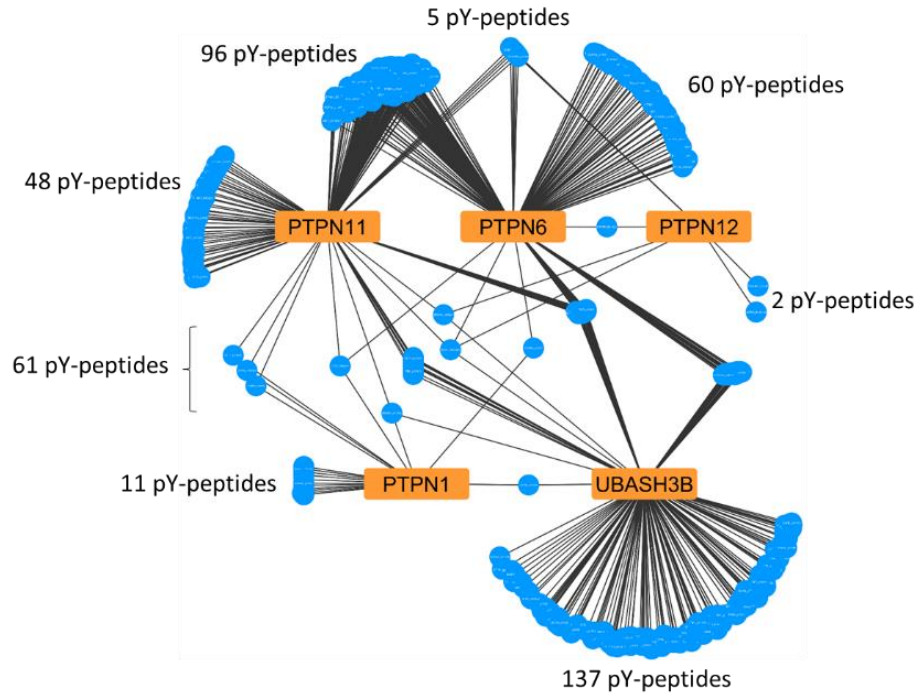
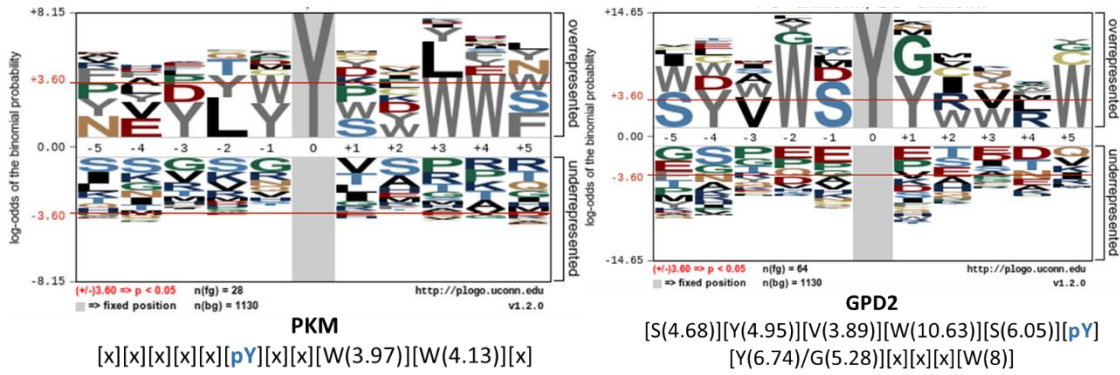


Figure 44 | Interactome network of protein tyrosine phosphatases (orange rectangles) and pY-peptides (blue ellipses) of RTKs. Protein tyrosine phosphatases display the pattern to interact with their pY-peptide clusters, especially UBASH3B has a greatest number of unique binding pY-peptides.

3.4.3 Other unclassified interactors

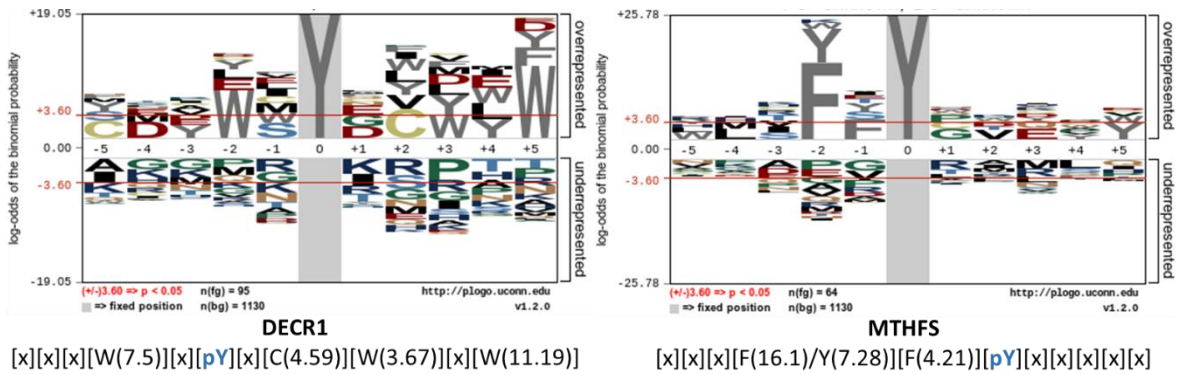
Not surprisingly, the interaction between PKM2 and pY-residues was significantly identified in this study. In disagreement to the known binding motif where E/D in position -4 to -1 is critical for the interaction, sequence analysis in this AE-MS study defined novel motifs where W at +3 and +4 position is essential for binding. Notably, LanC-like protein (LANCL1), 2,4-dienoyl-CoA reductase (DECR1), prostaglandin E synthase 2 (PTGES2), 5-formyltetrahydrofolate cyclo-ligase (MTHFS), and glycerol-3-phosphate dehydrogenase (GPD2) were all frequently and significantly enriched by 105, 95, 86, 68, and 65 pY-peptides, respectively. Although a known interacting region does not exist, all specifically bind to the phosphotyrosine motifs (Figure 45). Interestingly, the motifs that bind to GPD2, DECR1, and PTGES2 are similar to one another. Thus indicating similar specificities to the pY-peptides or that these are shared protein complex members. More interestingly, the enrichment of DECR1, MTHFS, and PTGES2 are in agreement with a study where SILAC was used to identify significant interactors to pY residues³⁹. Therefore, it would be interesting to draw a broader picture describing the real binding sites in these interactors after further validation.

Some validation experiments to confirm the binding of these new interactors with the pY-peptides have been performed using the competition assay discussed in the next section.



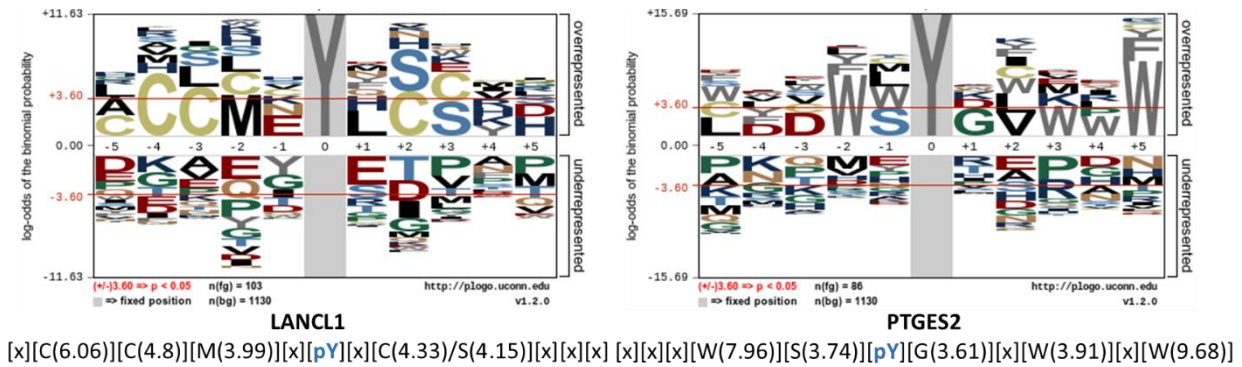
(a)

(b)



(c)

(d)



(e)

(f)

Figure 45 | Motif binding to PKM2 (a), GPD2 (b), DECR1 (c), and MTHFS (d), LANCL1 (e), and PTGES2 (f). The GPD2, DECR1, and PTGES2 displayed similar binding motifs, suggesting similar specificities to pY-peptides or that these are components of the same complex.

3.5 Validation of the interactome by competition assay

After completing the comprehensive interactome of RTKs by AE-MS, the novel interactors of 59 pY-peptides were evaluated by a competition assay using two doses of pY-peptides as the ‘free compound’ (Figure 46). The competition assay not only increases the quality and certainty of the data and the entire interactome as a whole; but also provides a deeper biological understanding of health and disease.

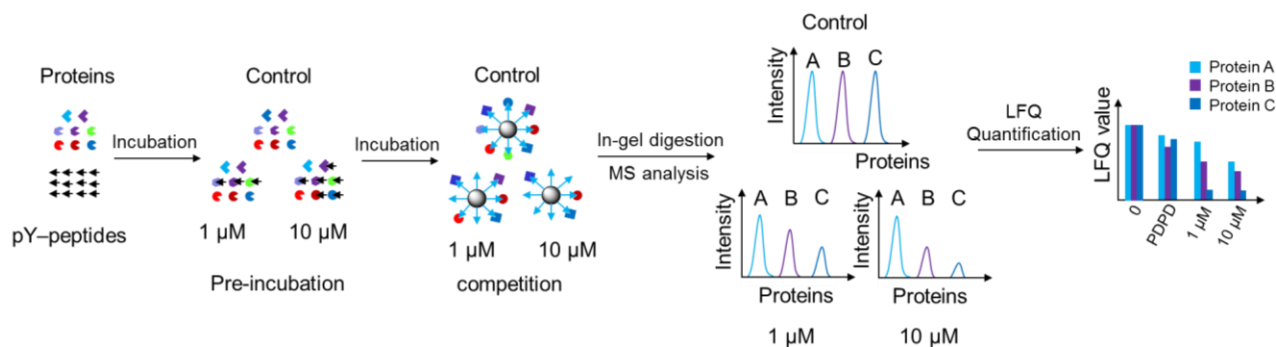


Figure 46 | Workflow of the competition assay for 59 pY-peptides. Two concentrations (1 μM and 10 μM) of the pY-peptide and one DMSO control were incubated with the corresponding pY-peptide matrix. The bound proteins were eluted and digested in gel and analyzed by nLC-MS/MS. With the competition fraction of each peptide against itself (the affinity matrix) revealed, the IC_{50} and the K_d could be calculated for evaluation of the affinity.

3.5.1 Interactor-centric selection of representative pY-peptides

From the perspective of the interacting proteins, 59 pY-peptides were selected for the competition assay. Depending on whether the interactors contained the SH2/PTB domain or not, peptides were grouped into two categories. The first category of pY-peptides to evaluate SH2/PTB domain-containing interactors consisted of 5 groups: (1) peptides that have not been previously studied, but annotated as regulators in the PhosphoSitePlus database; (2) peptides that do not have any annotated interactors in the BioGRID database; (3) peptides from the AE-MS assay that interacted with proteins that were contradictory to reported literature; (4) hotspot mutant peptides that significantly enriched distinct interactors compared to the wild-type peptides; and (5) peptides derived from mutation-gain tyrosine residues that enriched novel interactors compared to all of the wildtype pY-peptides interactors. The corresponding wild-type peptides without pY residues were also utilized to compare interactome alteration.

For the second category, peptides were selected when interactors met the following criteria: (1) interactors contain known domains (*e.g.*, C2 or Hakai) or a new domain (*e.g.*, PTP); (2) interactors are enriched despite the lack of confirmation from the literature.

Ultimately, this competition data set could at least provide an explanation for the following six novel categories of finding; thus enabling deeper insights into the RTK interactome mediated by phosphotyrosine residues.

3.5.2 Proof of concept

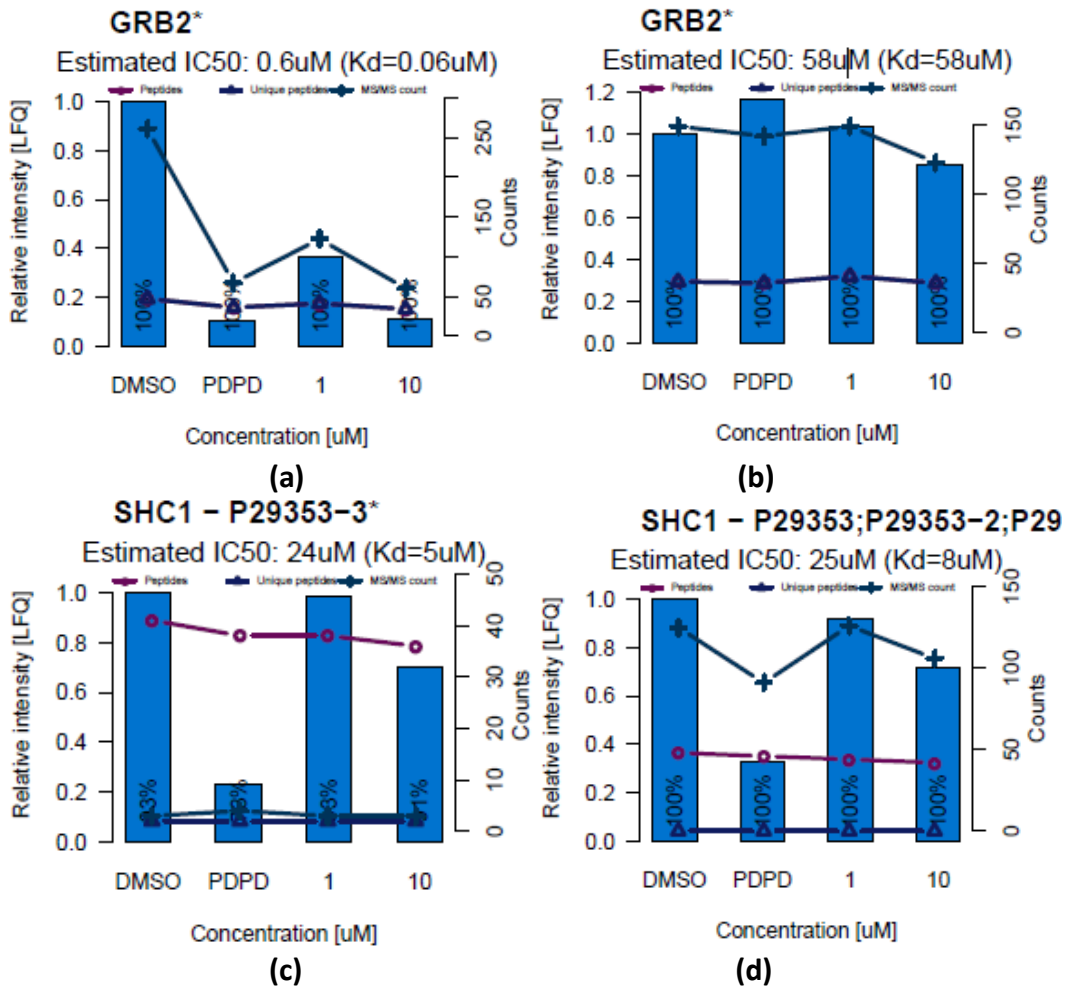
With the application of the competition assay, 44% of the interaction between 59 pY-peptides and their interactors in the AE-MS was validated with estimated IC_{50} and K_d values. Besides, these validated interactors occupied approximately 90% of the iBAQ intensities, which, in turn, demonstrated the high quality of the AE-MS (Supplementary File 6 and 7). Moreover, deeper analysis found the competition assay not only enabled evaluation of the interaction affinity similarities and differences between known interactions, but also validation of the novel interactors. In the end, such data can lead to new biological discoveries.

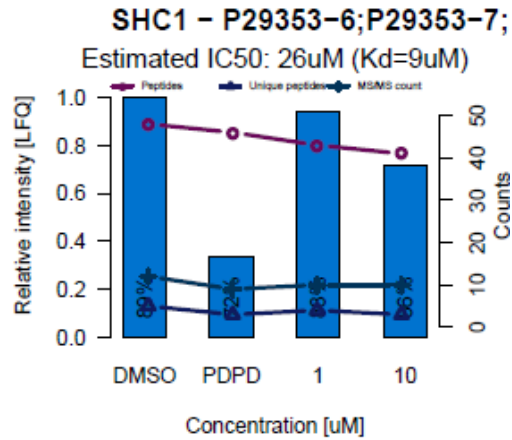
As an example, consider the three phosphotyrosine residues of FLT4. Previously, it has been documented that pY1230 and pY1231 (both bind to GRB2) together with pY1337 (interacting with SHC1) contribute to promoting endothelial cell proliferation, migration, and survival²⁰. By applying a competition assay, the existence of these interactions was successfully validated (Figure 47). Importantly, the competition assay revealed that GRB2 interacts more strongly with FLT4_pY1230 (lower K_d) than with FLT4_pY1231. This is understandable due to the sequence differences between the pY-peptides. The sequence of FLT4_pY1230 matches the GRB2-binding motif [x][x][x][x][x][pY][x][N][x][x][x], that is reported not only in the literature⁹⁹ but was also observed in the AE-MS assay. Shifting the critical residue N at the +2 position in the sequence of FLT4_pY1231 to the +1 position significantly decreased the binding affinity to GRB2 by approximately 1,000 fold. Thus, the competition assay could readily distinguish the interaction of GRB2 with the two different pY-peptides with different affinities. Biologically, this observation suggested that FLT4_pY1230 contributes to the signaling pathway significantly more than FLT4_pY1231.

Similarly, it is known from the literature that FLT4_pY1337 can bind to SHC1. This interaction was also confirmed in our AE-MS and competition assays; although three isoforms of SHC1 were identified. As described in section 3.3.1, there was no significant difference in the binding motif for these three SHC1 proteins. In agreement with the motif analyses; the competition assay revealed that the interaction between FLT4_pY1337 and the three isoforms of SHC1 have similar affinities (5, 8, and 9 μ M). Furthermore, the competition assay also validated the recruitment of additional novel interactors such as GRAP, GRAP2, HSH2D, INPP5D, and SHC3

(Supplementary File 6 and 7) to these three pY-peptides with similar affinities. These could be further examined in the same cell type to discover new cell effects in the future.

Overall, these findings confirmed that the competition assay is a feasible approach to validate and confirm interactions and discover high-affinity interacting proteins that bind to the pY-peptides.





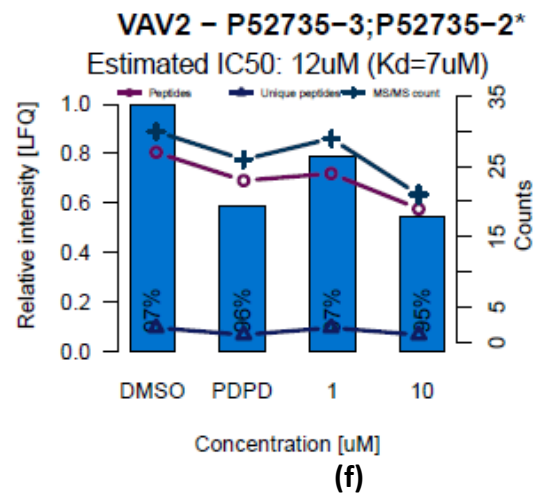
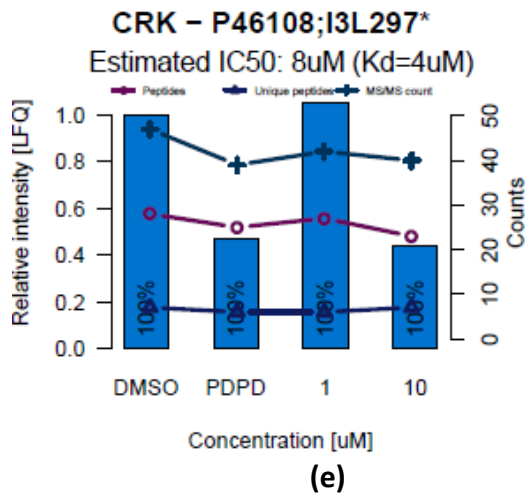
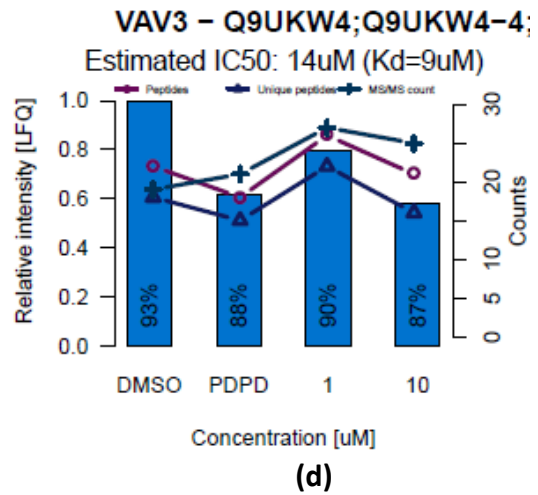
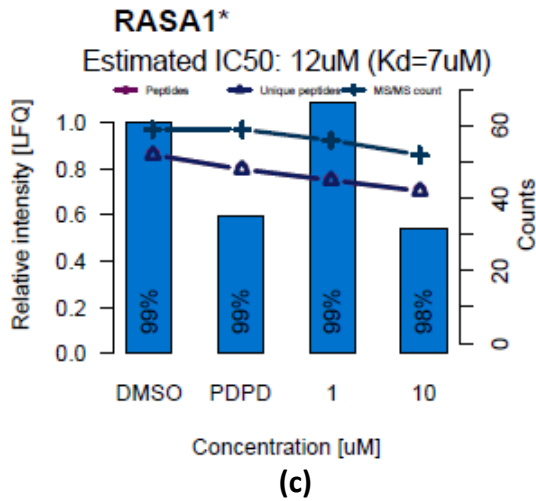
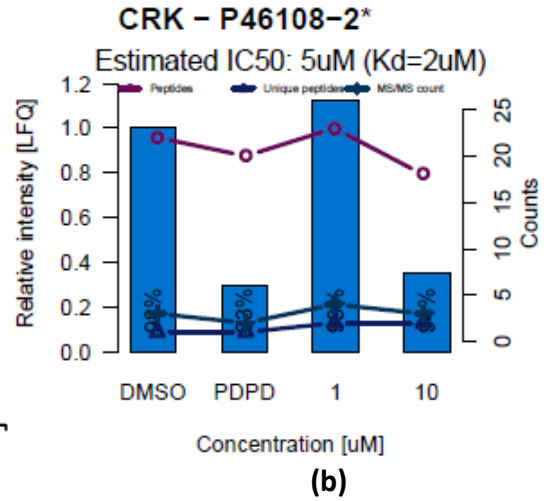
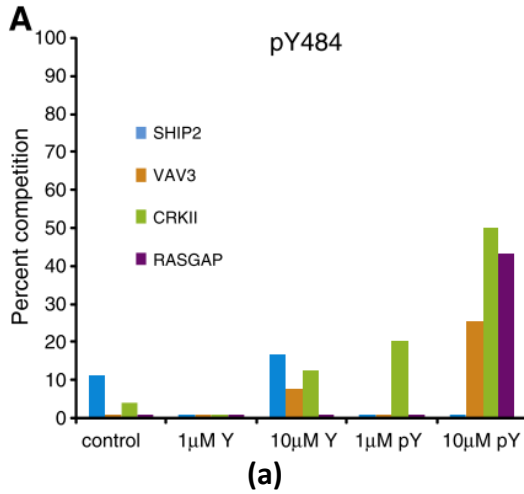
(e)

Figure 47 | Validation of known interactors of FLT4_pY1230, 1231, and 1337. FLT4_pY1230 (a) recruited GRB2 with approximately 1,000 times higher affinity than FLT4_pY1231 (b). The interaction between FLT4_pY1337 and the three isoforms of SHC1 implied there is no preferential bias towards the pY-peptide to interact with SHC1 under the given experimental conditions (c-e).

3.5.3 Re-elucidation of known peptide-protein interactions

As documented in the published interactome of DDR1⁹⁵, a competition assay revealed that the three specific proteins RASA1, VAV3, and CRK-II are recruited when DDR1_pY484 is phosphorylated⁹⁵. In our competition assay, all three interactors were successfully quantified with a similar competition fraction at two concentrations (1 and 10 μ M) of the pY-peptide (Figure 48a-34d). Also, the SH2 domain-containing proteins VAV2, CSK, and CRK were validated as novel interacting proteins of DDR1_pY484 (Figure 48e-g); thereby expanding the interactome of DDR1. Moreover, the pull-down of the pull-down experiment in this competition assay strengthened the value of the interaction by providing an estimated dissociation constant (K_d). The potential roles of these new interactors in the DDR1 signaling pathways could be studied further.

Overall, the competition assay provided a feasible strategy to validate known and novel interactors. The assay could be applied further in this study to investigate affinities to other novel interactors.



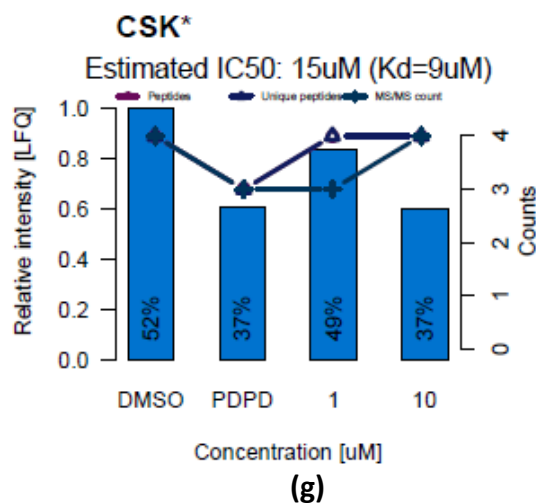
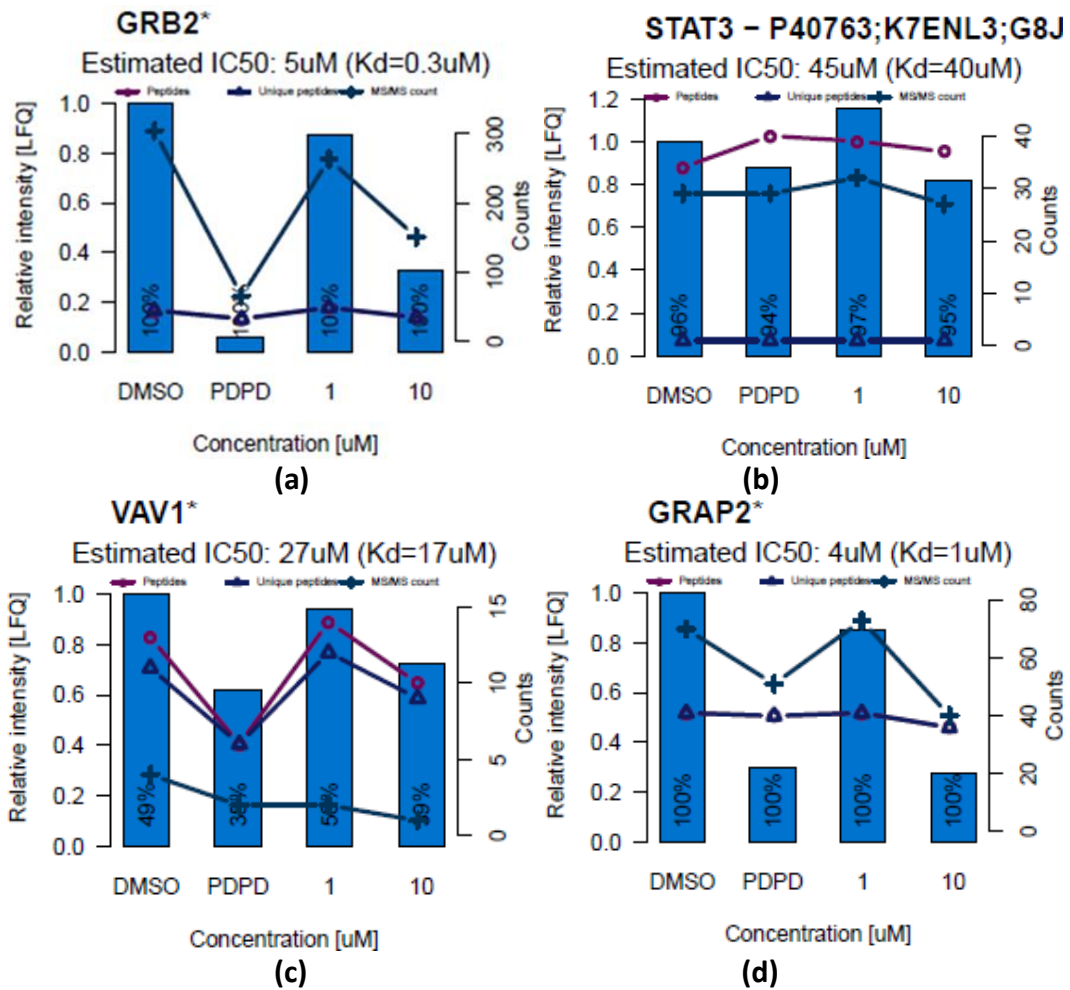


Figure 48 | Competition assay of DDR1_pY484 validated reported interactors (a-d) and novel interactors (e-g). (a) adapted from reference²⁸. (b-g) Interactors were validated by a dose-dependent competition assay, where the LFQ intensities for DMSO, 1 and 10 μ M were used to estimate the IC₅₀ and PDPD to calculating the K_d. The bar plots were integrated with the scatter plots to show the number of identified peptides, unique peptides, and MS/MS spectra. Further ranking of the protein abundance in each pull-down was shown at the base of each bar, *e.g.*, 99% means the LFQ value of that protein was in the top 1%.

The well-studied residue EGFR_pY1092 was included as an inter-laboratory positive control to examine the data set in this investigation. This residue has been frequently identified by a low-throughput (LTP) methods using an antibody (114 times), and high-throughput (HTP) approaches using proteomics (370 times). EGFR_pY1092 is known to interact with five proteins (CBL, GRB2, RASA1, PTPN6, and STAT3)²⁴ in the PSP database. It has also been highlighted as relevant to the development of non-small cell lung cancer/adenocarcinoma (NSCLC). From the five known interactors, only GRB2 and STAT3 were quantified in the AE-MS assay; but an additional 16 SH2/PTB domain-containing proteins were identified. Competition assays were performed for the 5 known and 16 novel interactors. The assay confirmed the data for GRB2 and STAT3, but the three absent proteins were still not observed. Moreover, 14 of the 16 novel interactors were also validated. The novel binders VAV1 and GRAP2 have been evidenced as true interactors in PepSpotDB²⁶ (Figure 49). Such cross-validation provided strong support that the AE-MS assay is a reliable approach to identify more potential interactors. Furthermore, AE-MS provides the possibility of exploring the relationship between EGFR_pY1092 and cancers such as NSCLC. For instance, the novel interactor VAV1 bound to EGFR_pY1092 with a higher

affinity (lower K_d) than the known interactor STAT3. In the ProteomicsDB database, VAV1 is co-expressed with STAT3 in NSCLC. Learning the role of proteins such as VAV1 in NSCLC could potentially advance the treatment of this cancer. Furthermore, the same principle could be applied to study the other new interactors with low K_d values. For example, the role of LCP2 and VAV3 in other cancers using the cell lines OVCAR-4 and Caki-1 cell, respectively.



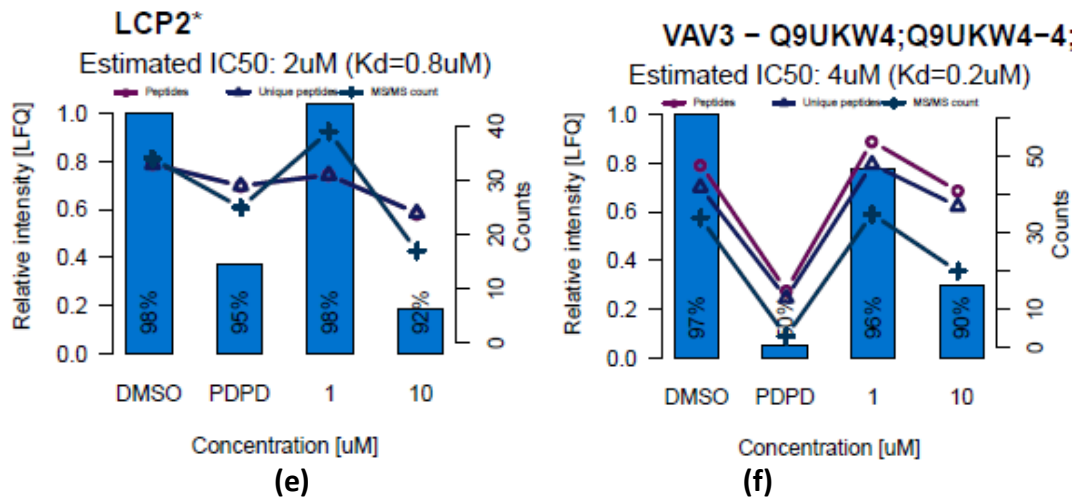


Figure 49 | A subset of validated interactors of EGFR_pY1092. From the PhosphoSitePlus database, GRB2 and STAT3 are known interactors (a-b); whilst the other proteins are novel interactors (c-f).

3.5.4 Beyond the known interactors of phosphotyrosine residues

The EPHA8_pY616 has been shown to interact with the SH2 domain-containing protein FYN and this leads to a modification of cell adhesion in mice¹¹⁰. FYN was successfully elucidated in our competition assay with an estimated K_d value of 3 μ M (Figure 50). As human and mouse have identical protein sequences flanking this tyrosine residue in the ± 5 position, it naturally suggests a similar signaling event via FYN can be triggered in humans following phosphorylation of the tyrosine residue.

The SH2 domain-containing protein SLA is known to interact with the phosphorylated juxtamembrane tyrosine of RTKs. Via the C-terminal region, SLA then interacts with CBL and engages CBL to ubiquitinate the RTKs¹¹¹. Interestingly, the competition assay validated that this residue genuinely interacts with SLA (Figure 50); thus indicating a possible mechanism of downregulating RTKs. Furthermore, the interaction of EPHA8_pY616 to SLA could be confirmed by sequence analysis. The pY-peptide contains YEEP at the C-terminus, which has the same residues at position +2 and +3 of the conserved binding motif [pY]xDP¹¹¹.

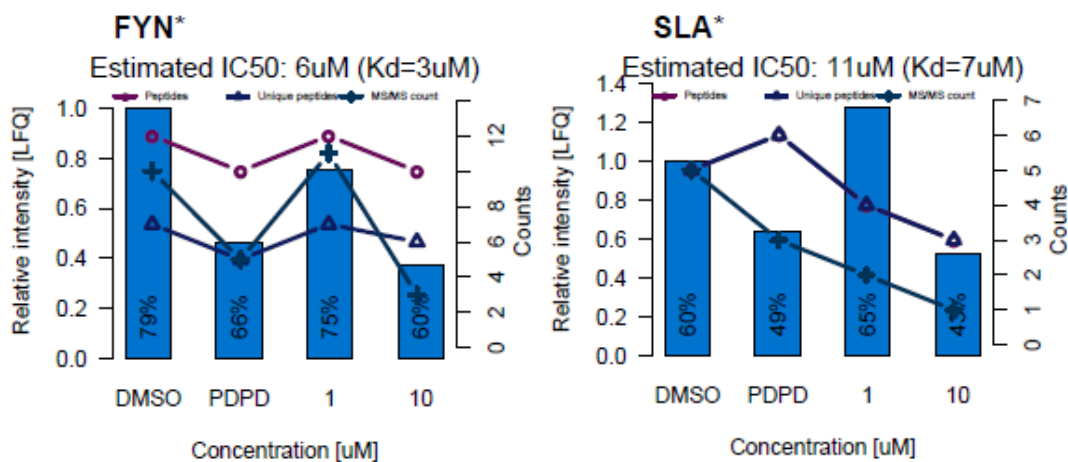


Figure 50 | Validation of known (left) and novel (right) interactors of EPHA8_pY616.

Furthermore, the AE-MS study also identified novel interactors that were not known to interact with specified RTKs in the BioGRID database. Compared to the AE-MS assay in this study, for example, the interactome of FLT1 in BioGRID was lacking LCP2. The residue FLT1_pY1213 is known to augment cell secretion¹¹² and bind three interactors in the PhopshoSitePlus database, plus two unknown proteins with molecular masses of 27 and 74 kDa, respectively¹¹³. Moreover, the entire kinase activity of FLT1 can be directly blocked in the FLT1_Y1213F cell¹¹⁴. All the

evidence emphasized the importance of FLT1_Y1213F in cell responses and deciphering a comprehensive interactome is necessary to fully understand the function of the protein. Encouragingly, the competition assay not only confirmed all three known interactors plus the novel interactor LCP2; but also an additional 16 novel proteins (Figure 51).

Furthermore, as evidenced by the molecular masses of 25.206 and 73.452 Da, respectively, the two unknown interactors from the literature are now defined as GRB2 and THEMIS. THEMIS is not known to contain a domain that can interact with a phosphotyrosine residue. Thus, it was concluded that THEMIS forms a complex with GRB2.

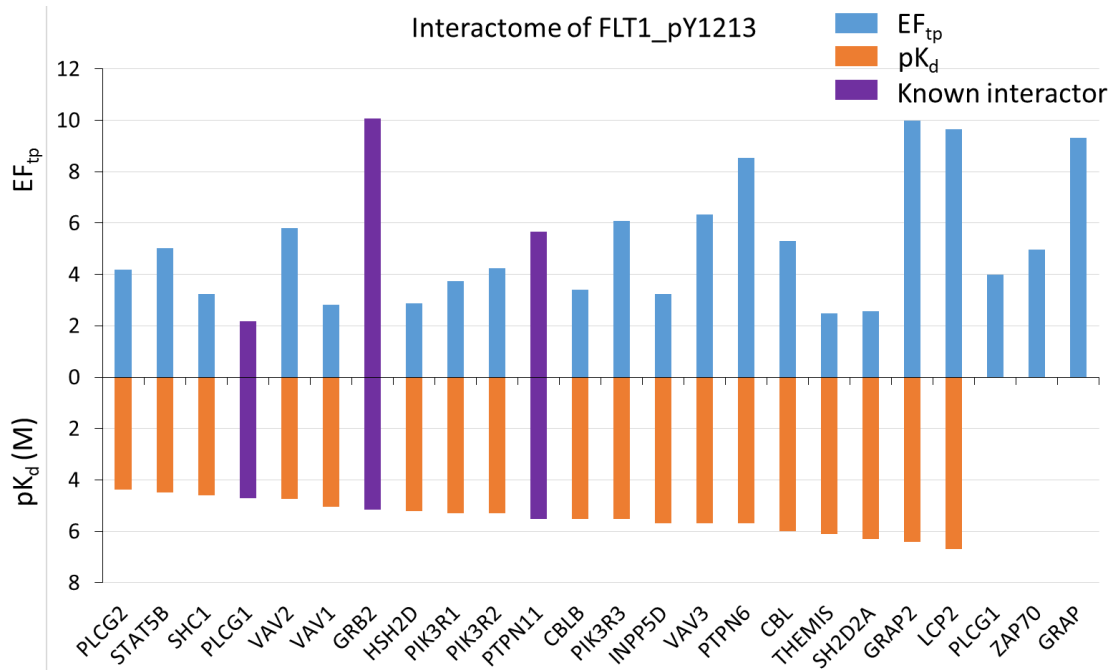


Figure 51 | Validation of the interactors of FLT1_pY1213. The upper bars represent the enrichment factors (EF_{tp}) of the interactors identified in AE-MS assay, and the lower bars are the pK_d ($-\log_{10} K_d$, in the unit of M, higher values represent a higher affinity) of the validated interactors in the competition assay. The known interactors are shown in purple. The last three interactors without pK_d values were those not observed by competition assay after treatment with pY-peptides.

3.5.5 Identification of inconsistent interactors of phosphotyrosine residues

According to previous studies, the phosphotyrosine residues FLT3_pY572 and FLT3_pY793 can both recruit the adaptor protein GRB10 in response to ligand stimulation. FLT3_pY793 has a stronger affinity than FLT3_pY572¹¹⁵.

GRB10 was expressed in the full proteome data generated in this study with an iBAQ intensity of 8.9-30.7 (log₂ scale), and displayed a medium abundance with an iBAQ intensity of 18.6 (log₂ scale). Despite this, however, both the AE-MS and competition assays disagreed with this information for both residues. Instead, other SH2 or PTB domain-containing proteins were identified as significant interactors (Figure 52). This inconsistency could be due to a number of factors. Firstly, the affinities of the interactions may be too low, and thus the limited binding sites in the pY-peptide matrix are fully-occupied by other interacting proteins with stronger binding affinities. Secondly, it is feasible that the interaction between phosphotyrosine residues in the full-length FLT3 and GRB10 can only occur with the assistance of the other residues close to the binding site. Thirdly, the data from the literature could have originated from overexpression of the GRB10 via a transfection experiment. This could potentially increase the likelihood of the interaction with the phosphotyrosine sites although the affinity is low. Finally, the association of GRB10 with these two phosphotyrosine residues is an indirect interaction. Consequently, the interaction between GRB10 and the pY-peptides could not be confirmed in this study.

The conflicting finding in the study here reminds us that more significant interactors of the FLT3_pY572 could be other proteins such as STAT1, STAT3, and STAP2. Compared to FLT3_pY572, FLT3_pY793 could recruit different functional interactors including INPPL1, INPP5D, and VAV2 to trigger different signaling cascades (Figure 52).

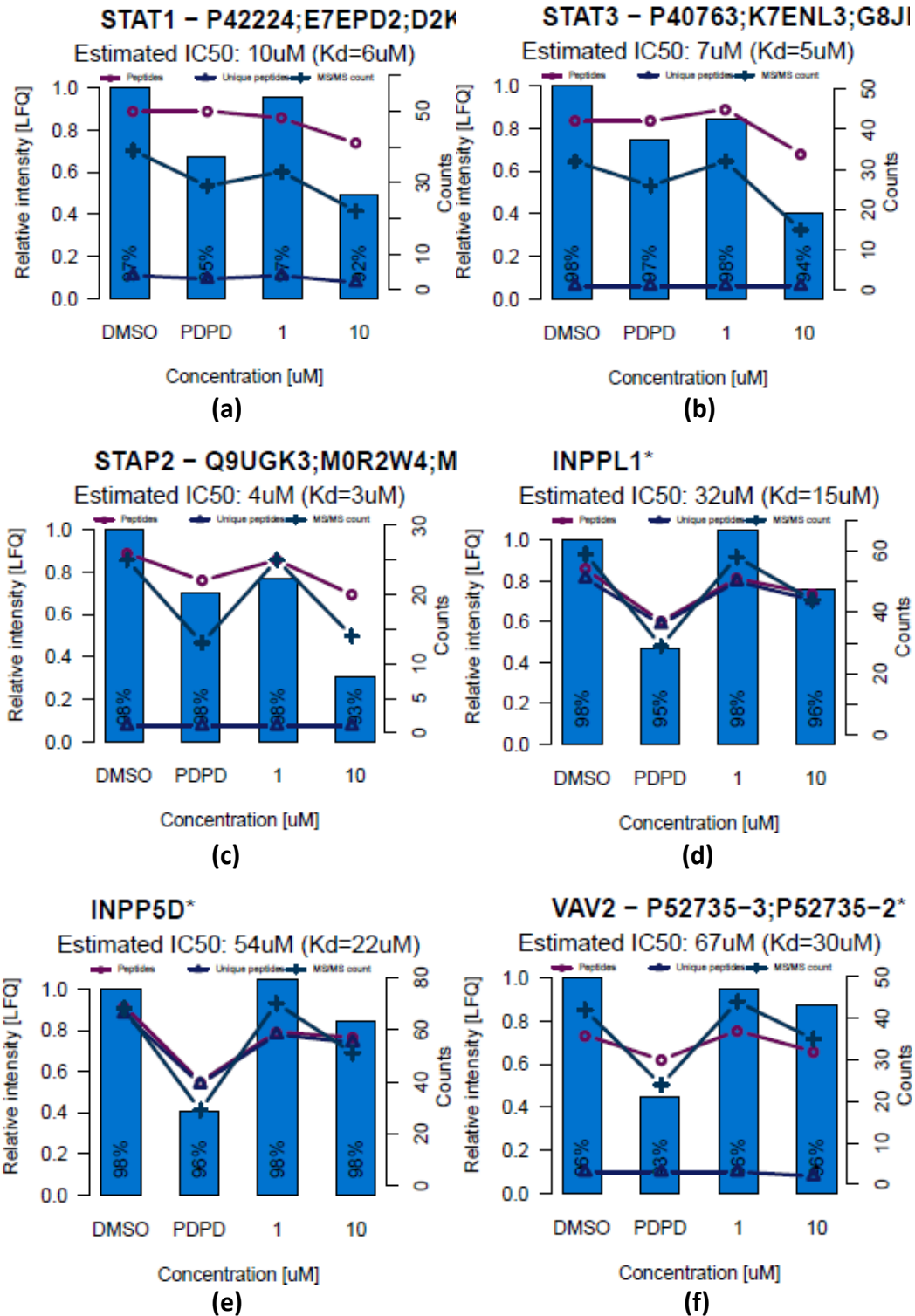


Figure 52 | Validation of inconsistent interactors of FLT3_pY572 (a-c) and FLT3_pY793 (d-f).

A similar disagreement with respect to interactors was also observed for KIT_pY936. Using the recombinant SH2 domain of GRB7 and KIT_pY936 peptide, a competition assay reported a selective interaction¹¹⁶. Interestingly, GRB7 was quantified with an iBAQ intensity of 16.1 (log₂ scale) in the full proteome data set and was enriched by several pY-peptides in this study. Moreover, the binding of GRB7 to the other two pY-peptides (FLT3_pY591 and FLT3_pY599, see Supplementary File 6 and 7) was validated by the competition assay. Therefore, the loss of interaction of GRB7 to KIT_pY936 could occur because of the same reasons as described for GRB10. Regardless, this contradictory finding again addresses the point that the GRB7-related signaling pathway could be shifted to a direction that is regulated by, *e.g.*, GRB2, LCP2, INPP5D or CSK (Figure 53).

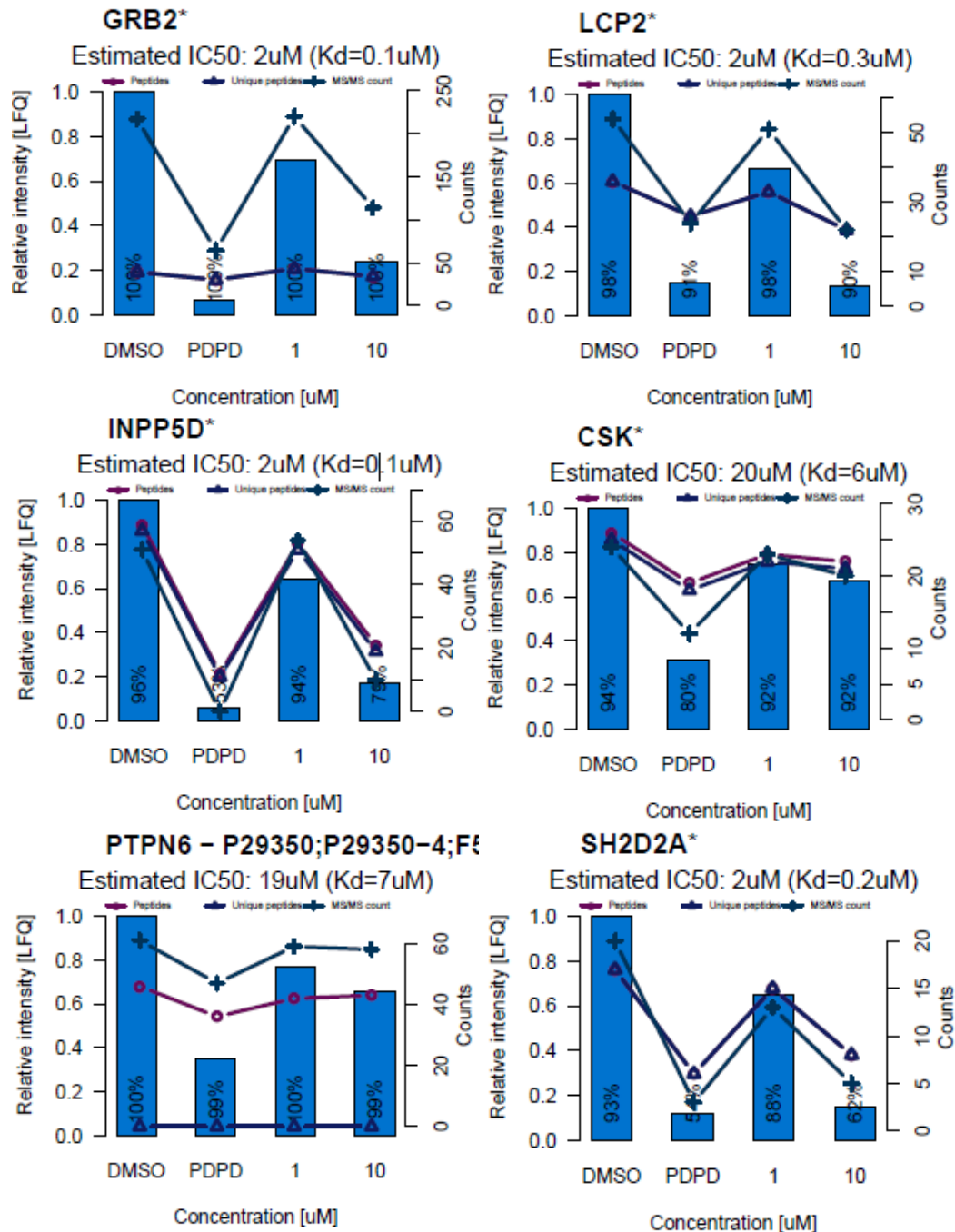


Figure 53 | Validation of contradictory interactors of KIT_pY936. Unlike the literature where GRB7 was identified, the AE-MS assay and competition assay both showed interactions with other novel interactors.

Last but not least, a contradictory finding was also observed for FLT4_pY1063. In the literature, CRKI and CRKII have been elucidated as interactors of this residue to trigger survival signaling events¹¹⁷; however, these proteins were not enriched by FLT4_pY1063 (Figure 54). The

competition assay confirmed the interaction of FLT4_pY1063 with VAV3 and STAT5B, and indeed, these proteins could mediate the reported signaling event.

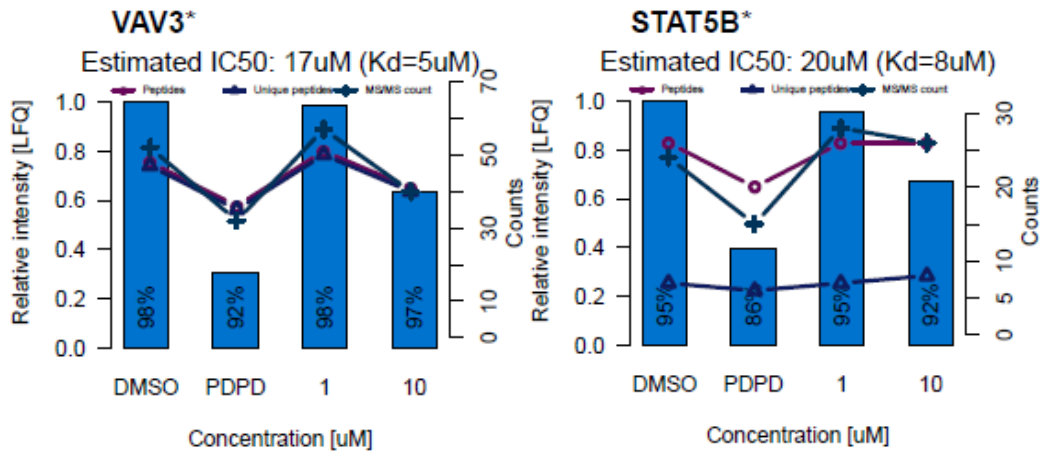


Figure 54 | Significant interactors of FLT4_pY1063 that are contradictory to literature.

3.5.6 Exploration of unreported interactors of phosphotyrosine residues

Currently, not all phosphorylated tyrosine residues of RTKs are known to function in specific cell signaling pathways; even when the consequences of the signaling have been well-studied. The AE-MS study could provide reference interactors to explore where signaling cascades are initiated, and a follow-up competition assay would further validate this information. For instance, the residue EPHA3_pY602 is known to respond to ligand binding. Entirely reversing such a detailed cellular response can be only achieved by mutating the peptide to an inactive state together with EPHA3_pY779¹¹⁸. Identifying the specific interactors of these two residues would undoubtedly provide the molecular mechanism of this event. In the AE-MS assay, variously identified interactors such as the kinases LCK and FYN, and guanine nucleotide exchange factor VAV2 and VAV3 were recruited to the EPHA3_pY602 with K_d values ranging from 0.2 to 7 μ M (Figure 55). These interactors suggested that terminating the cellular response of EPHA3 by mutating the EPHA3_pY602 to an inactive state could result from blockage of the associated cell signaling pathway. Unfortunately, the competition assay with EPHA3_pY779 was not performed because of difficulties in obtaining the purified peptide. Therefore, a worthwhile future perspective would be to investigate the affinity of the interactors of EPHA3_pY779.

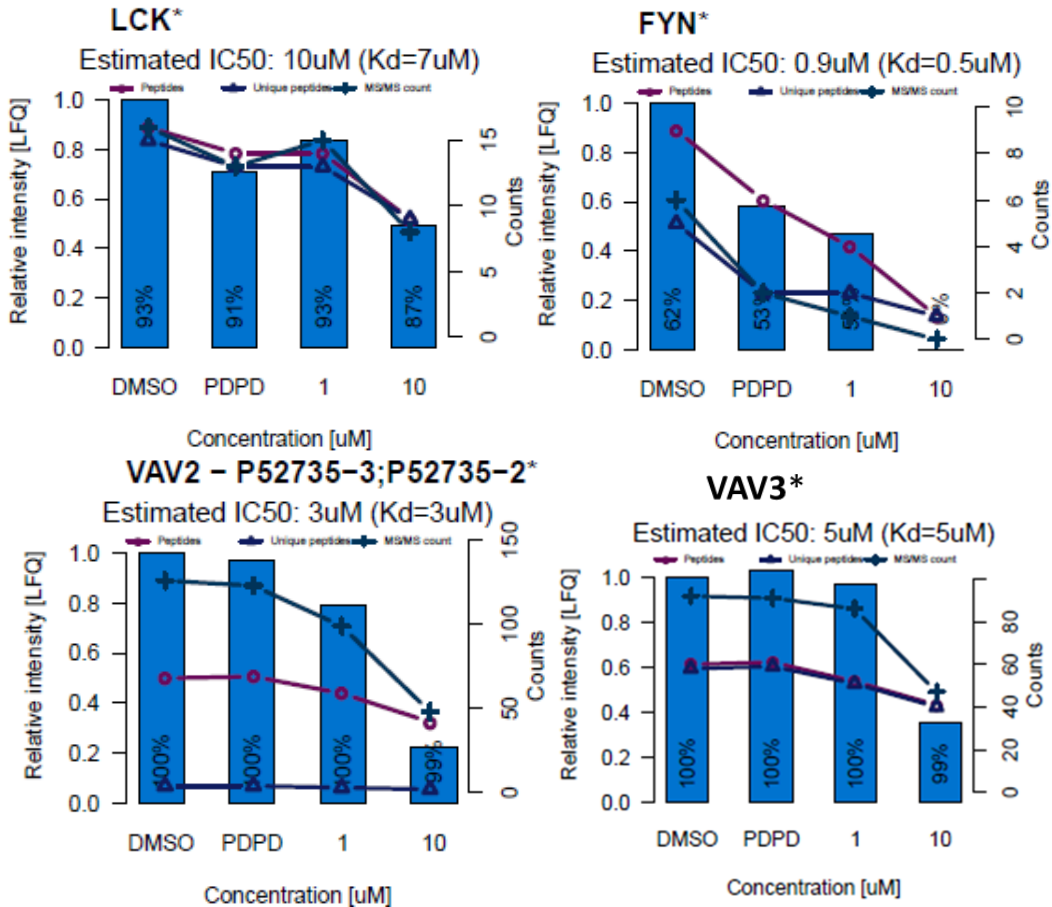


Figure 55 | A subset of validated interactors of EPHA3_pY602, that could be explored further in the future to reveal the mechanisms responding to ligand binding.

Another exciting finding was the confirmation of STAT3 and STAT1 interaction with ALK_pY1586 with a promising K_d value of 1 and 3 μM , respectively (Figure 56). A previous study has indicated that the activation of this site enhances an inflammatory reaction in a breast cancer cell line, and the initiation of the JAK/STAT signaling pathway¹¹⁹. Determining STAT3 and STAT1, and not the other STAT family proteins, as direct interactors could significantly contribute to the drug treatment of breast cancer. In neuroblastoma, this residue is known to be continuously phosphorylated in the mutated and truncated versions of ALK¹²⁰, but not in the wild-type version. Identification of STAT3 and STAT1 as interactors also suggests the existence of a similar pathway in neuroblastoma. Moreover, the adaptor protein GRB2 also showed an interaction with this site. With data from this AE-MS study, it is feasible that additional signaling pathways in neuroblastoma could be resolved in the near future.

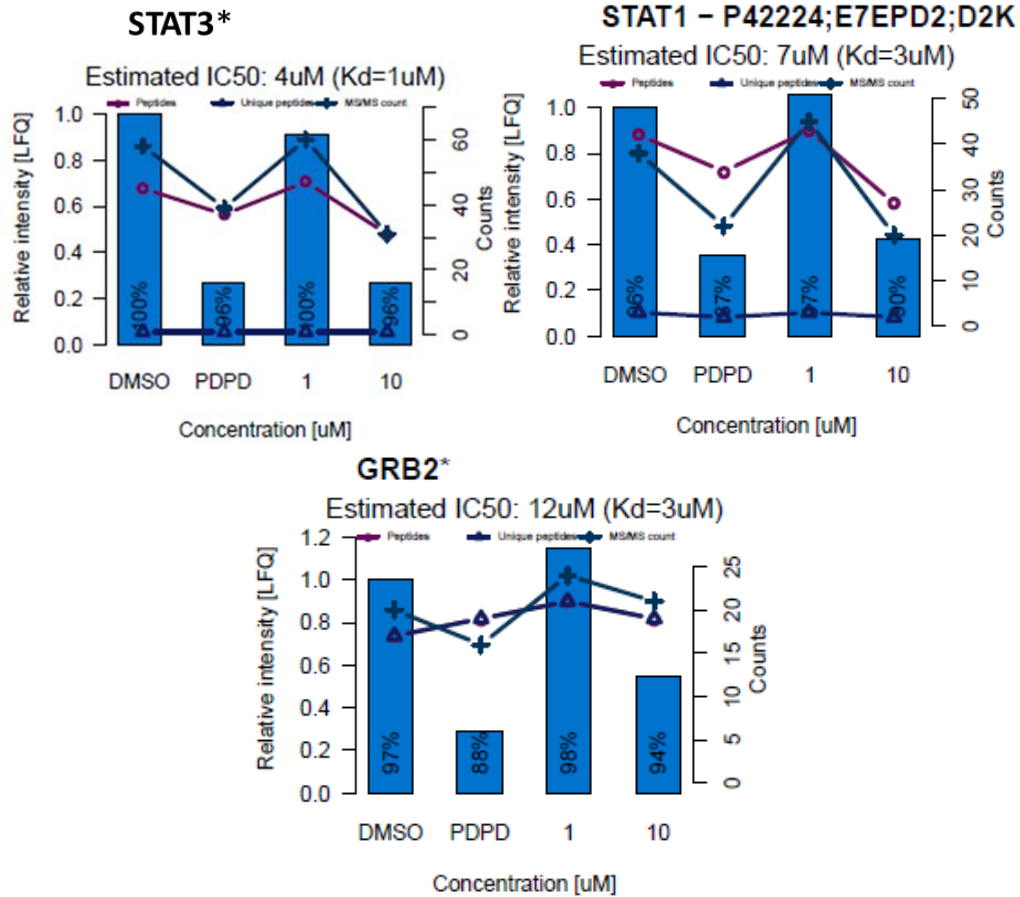
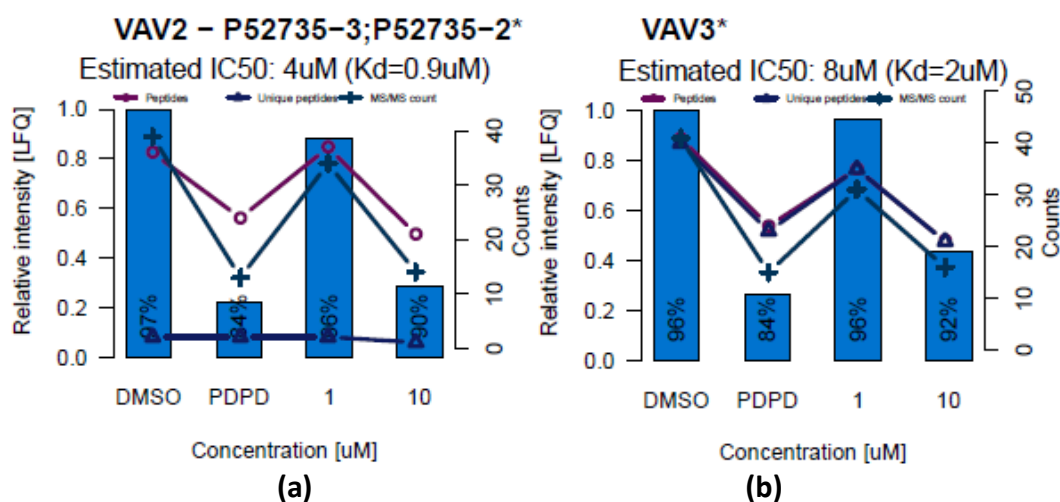


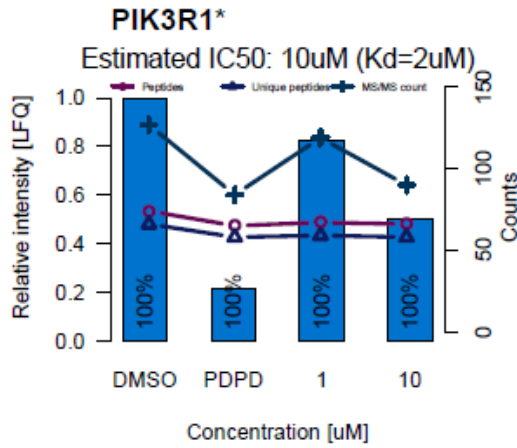
Figure 56 | STAT3 and STAT1 are true interactors of ALK_pY1586 that contributes to JAK/STAT signaling in breast cancer. Together with GRB2, they could also potentially influence the development of neuroblastoma where the ALK mutation causes constitutively-active ALK_pY1586.

3.5.7 The potential function of mutation-gain tyrosine residues

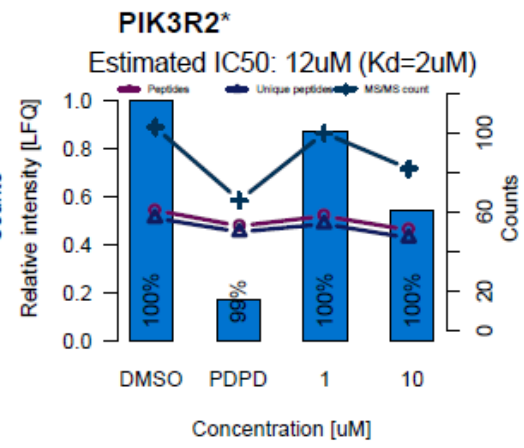
One of the effects of mutating a protein sequence is that this can result in the alteration of the interactome, signaling pathways and cellular responses. Some mutations can lead to the introduction of a tyrosine residue into an RTK; and subsequently, an association with drug resistance. For example, the mutation of FLT3_D835Y leads to constitutively-activated FLT3 in acute myelogenous leukemia patients with either Quizartinib or Sorafenib treatment^{29,33,34}. Revealing the protein interactors of such missense mutations, referred to in this study as mutation-gain tyrosine (MGY), would be a vital step in providing a direction for drug targeting in the future.

The mutation FLT3_D835Y was taken as an example. Firstly, the competition assay demonstrated that the pY-peptide FLT3_D835pY could interact with VAV2, VAV3, PIK3R1, PIK3R2, and PIK3R3 in the K_d range 0.9 – 2 μ M (Figure 57). An additional competition assay between FLT3_D835pY (affinity matrix) and FLT3_D835 (free peptide) confirmed that the wild-type peptide FLT3_D835 does not compete with the pY-peptide FLT3_D835pY for these interactors. Moreover, a third competition assay with the wild-type peptide FLT3_D835 also revealed that no binding occurred with these five interactors. Overall, the evidence indicated that these interactions are specific to the pY-peptide FLT3_D835pY. Further *in vitro* experiments could be performed to evaluate this interaction and the subsequent effect on cellular responses.

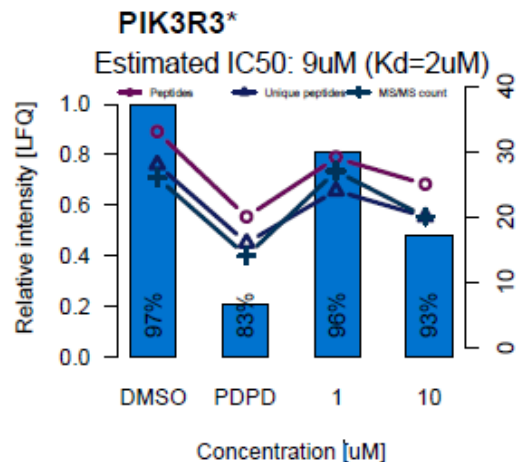




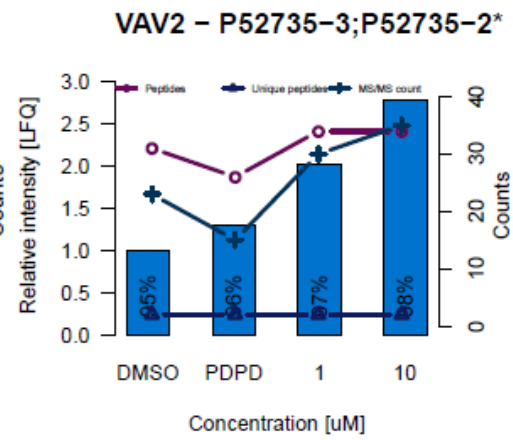
(c)



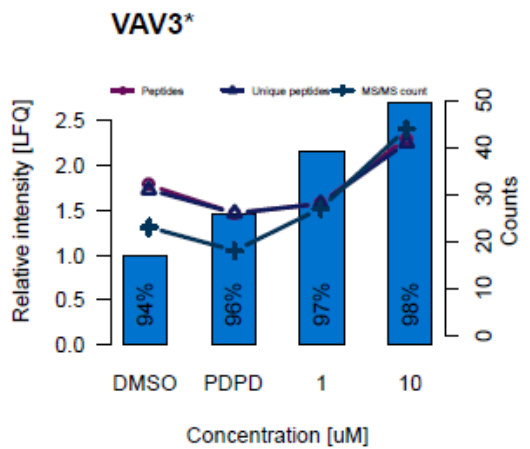
(d)



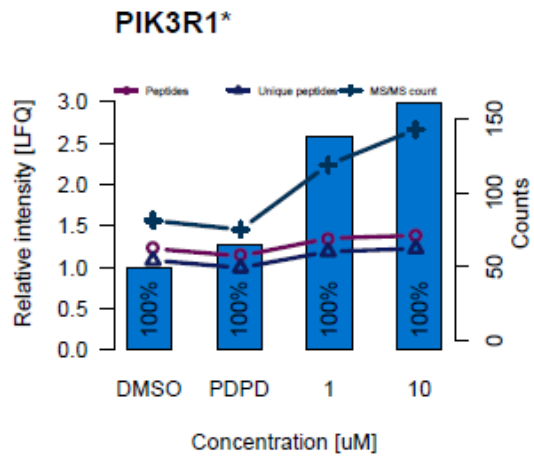
(e)



(f)



(g)



(h)

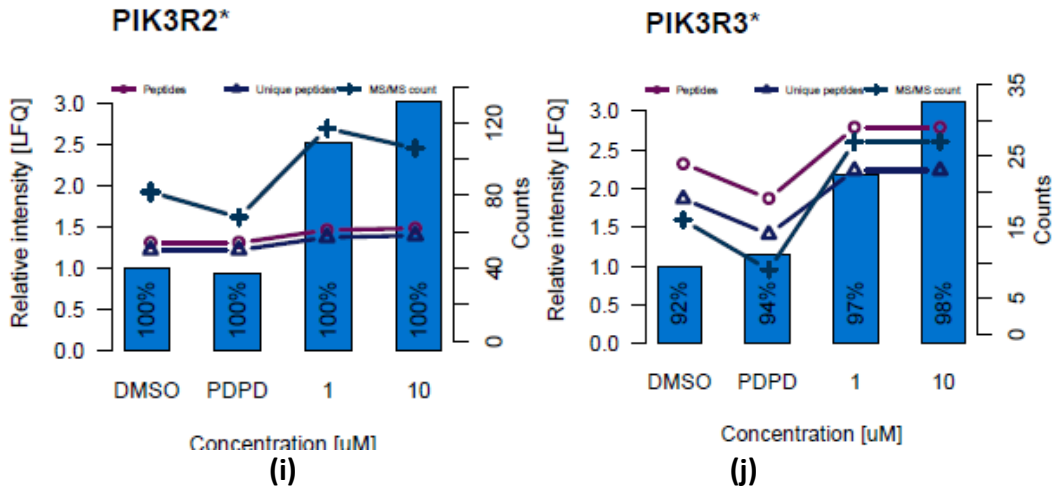


Figure 57 The competition assay indicated a specific interaction of FLT3_D835pY with five interactors (a-e), compared to competition using pY-peptide matrix against the wild-type peptide (f-j).

Currently, the mutation of FGFR2_C808Y is predicted to produce a medium effect on the receptor³¹, although the full impact is not clear. Structural analysis suggests that the residue is located in the loose region of the C-terminus of the receptor. This is further indicative of an increased chance of phosphorylation followed by recruitment of functional molecules to trigger the signaling pathway. The AE-MS and competition assays both showed that this mutation-gain tyrosine residue can interact with STAT1 and STAT2. Further competition using the pY-peptide matrix against the wild-type peptide revealed that these interactions had K_d values that were approximately 100 times higher. The wild-type matrix versus the wild-type peptide neither showed enrichment nor competition.

Altogether, the data indicated that STAT1 and STAT2 are specific interactors of FGFR2_C808pY. These interactions can trigger different pathways for the FGFR2 receptor because the wild-type residues of FGFR2 do not interact with these two STAT proteins (Figure 58).

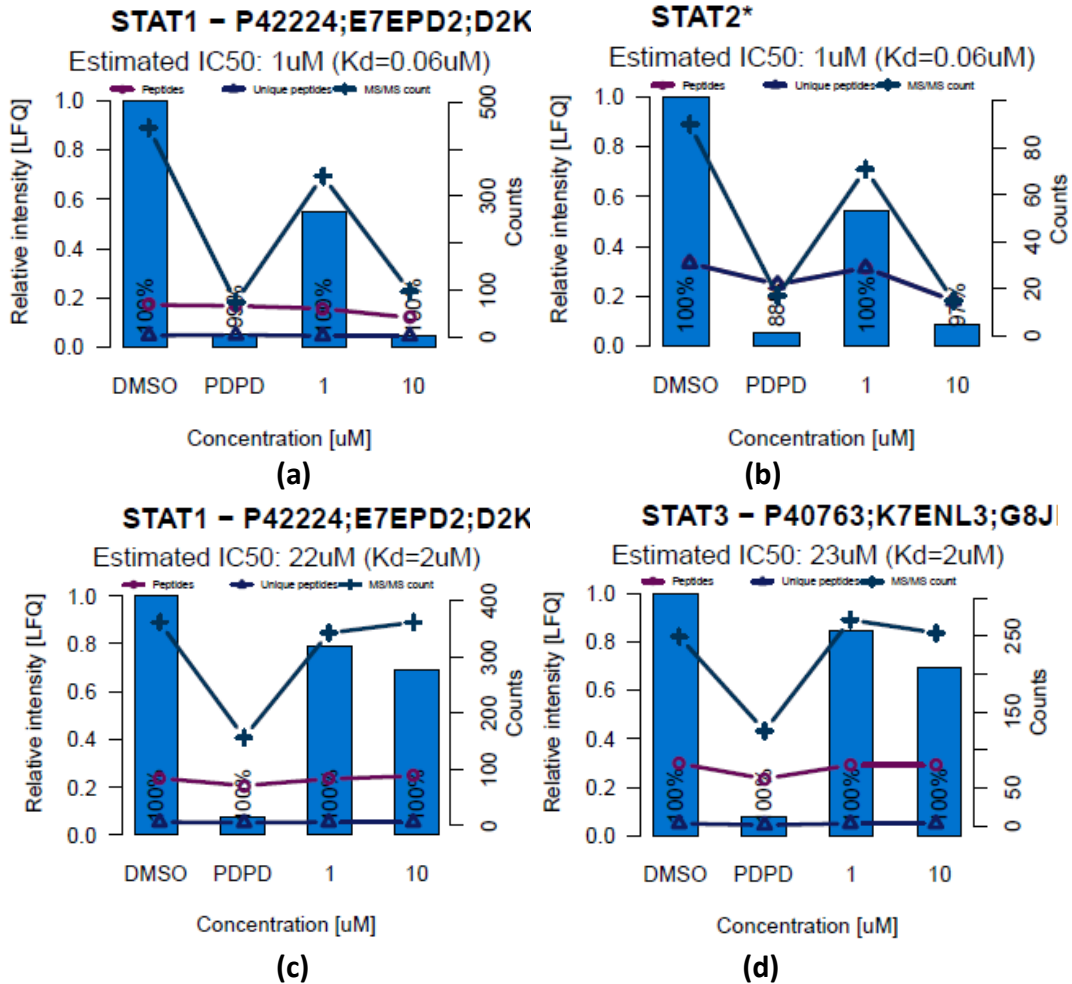


Figure 58 | Competition assay where FGFR2_C808pY against itself (a-b) and FGFR2_C808pY vs FGFR2_C808 (c-d)

Apart from this interaction, a similar picture was also observed with other mutant peptides (Supplementary File 6 and 7). Overall, the competition assay validated significant interactors of phosphorylated MGY peptides in RTKs. From this data, considerably more effort could be spent to learn the biological impact of cell signaling in cancer.

3.5.8 Beyond the SH2 or PTB domain-containing interactors

In the competition assay, several non-SH2/PTB domain-containing proteins also displayed a competition pattern. Subsequently, interactions of atypical proteins with versatile phosphotyrosine residues in RTKs was confirmed. This information provides a resource for not only structurally validating such binding, but also to explore the potential roles of such atypical interactors in the cell signaling of RTKs.

For example, the protein CBLL1 (containing an HYB domain) interacted with ERBB4_pY1284 in a dose-dependent manner. The estimated K_d value is 6 μ M. This is much lower than 90% of the proteins with an SH2 or PTB domain, and includes known interactors of SHC1 (22 μ M). Such a discovery indicated the rather strong affinity of ERBB4. Further investigation of the function of CBLL1 recruitment to the ERBB4 in the signaling cascades is warranted (Figure 59).

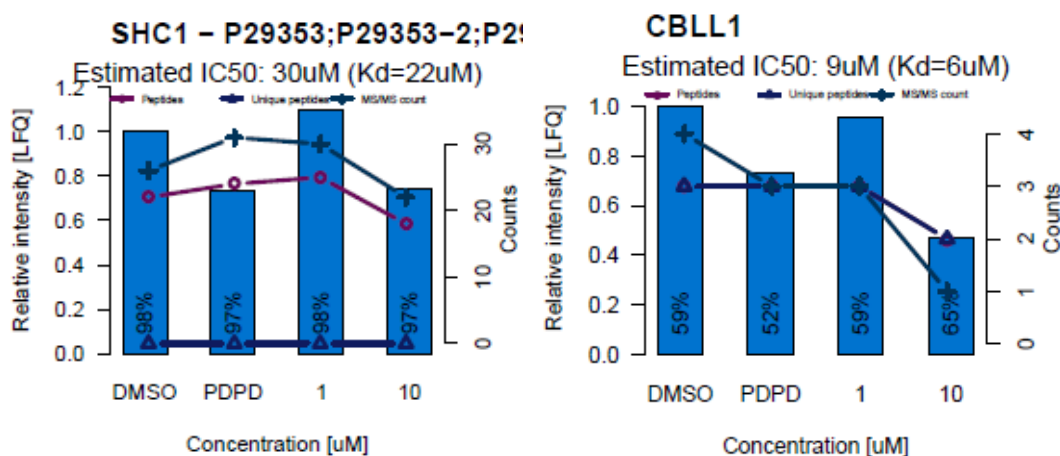


Figure 59 | Dose-dependent binding of SHC1 and CBLL1 to ERBB4_pY1284. CBLL1 displayed a higher affinity than the known interactor SHC1.

With similar affinities, the C2 domain of PRKCD and PRKCQ was also validated to interact with ALK_pY1278, NTRK3_pY710, and NTRK1_pY681. This similarity in interaction could be because of the shared pattern in the C-terminal sequences of the proteins, with R in the +1 position. Similar to the CBLL1 interacting pY-peptides, most of the interactions were stronger than the SH2 or PTB domain-containing proteins. Moreover, the three phosphotyrosine residues were identified in the PhosphoSitePlus database; although without confirmation of the interactors. Therefore, further investigation concerning the roles of these proteins in the cellular signaling of RTKs would be of interest (Figure 60).

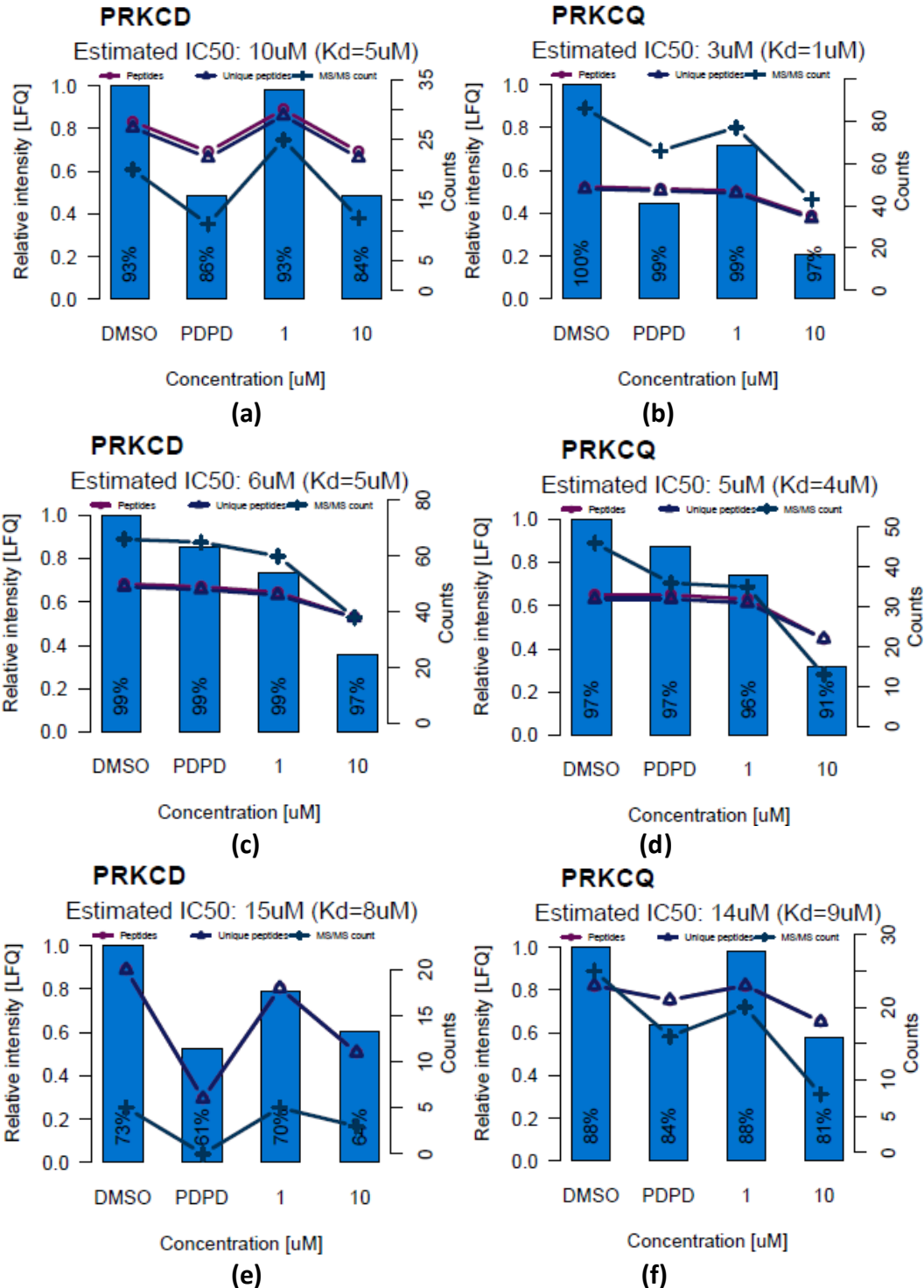


Figure 60 | Competition assay of PRKCD and PRKCQ against ALK_pY1278, NTRK3_pY710 and NTRK1_pY681. Affinities are similar to one another, and are a consequence of the residue R in the +1 position.

The phosphatase UBASH3B showed a dose-dependent competition profile for 8 from 11 pY-peptides. K_d values ranged from 3 to 34 μ M (Figure 61). As the motif analysis showed, UBASH3B

prefers to bind to pY-peptides with either Y or F at the -2 position. All eight pY-peptides contained one of these two motifs. The remaining three peptides that did not show a reasonable competition profile did not have this pattern. Deeper analysis indicated that the affinities of the pY-peptides were influenced by the -1 position. Peptides with the motif [x][x][x][Y][x][pY][x][x][x][x][x] usually had a lower affinity if residue x was an E (with a hydrophilic side chain), *i.e.*, FLT3_pY599. This situation was reversed for pY-peptides with an [x][x][x][F][x]Y[x][x][x][x][x] motif, where a lower affinity was observed when the larger residue Y was located between F and Y (*e.g.*, IGF1R_pY1281).

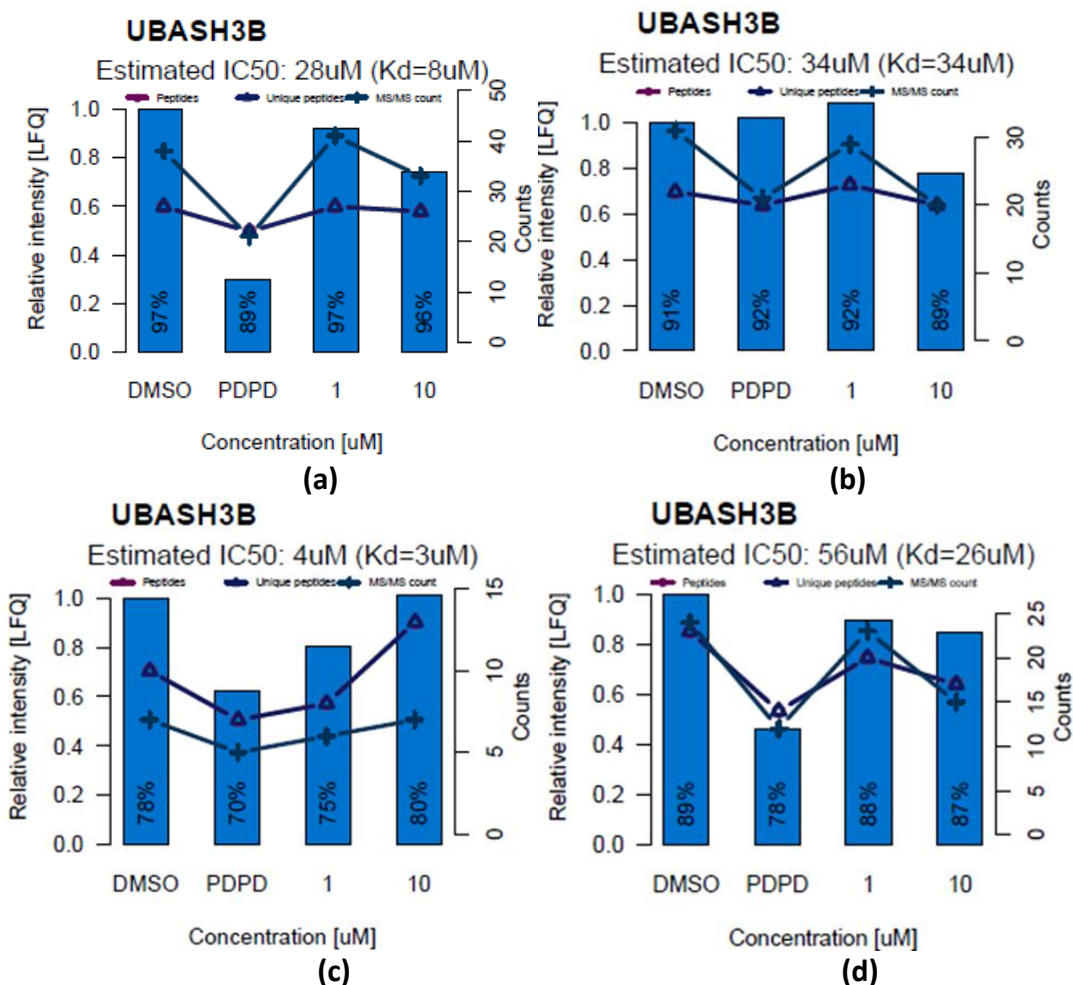


Figure 61 | Competition binding of UBASH3B to FLT3_pY591 (DNEYFpYVDFRE), FLT3_pY599 (FREYEpYDLKWE), FLT3_pY793 (LLCFApYQVAKG) and IGF1R_pY1281 (EVSFYpYSEENK). pY-peptides with the motif [x][x][x][Y][x][pY][x][x][x][x][x] were unfavorable when x = E (hydrophilic side chain), whilst peptides with the motif [x][x][x][F][x]Y[x][x][x][x][x] were unfavorable when x = Y (large residue).

From 7 pY-peptides, the sequence that most strongly bound MTHFS (IGF1R_pY1281) displayed a pattern that also completely matched the binding motif of MTHFS. Another pY-peptide from FLT3_pY793 that possesses the same motif did not show a reasonable competition profile against the protein. Sequence alignment suggested that the residue at the -1 position influences the affinities. This observation is similar to the pY-peptides that bind to UBASH3B, but with a reversed effect. For a strong affinity, UBASH3B prefers peptides without the Y at this position, whilst MTHFS interact strongly with the pY-peptides with Y at the same position (Figure 62). This phenomenon demonstrated that these pY-peptides interact with both UBASH3B and MTHFS, but with different affinities. Moreover, these different features in the binding domain could exist.

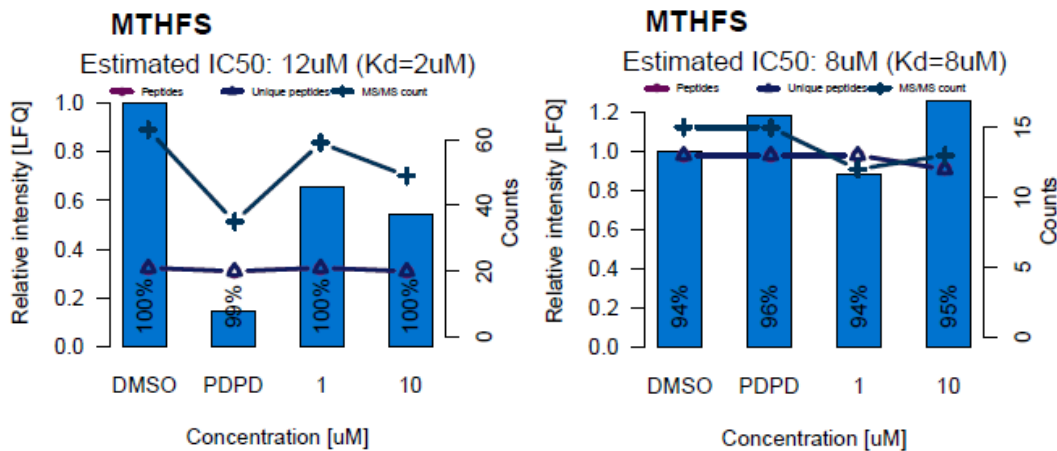


Figure 62 | Competition assay of MTHFS against IGF1R_pY1281 (left) and FLT3_pY793 (right)

Without a known binding domain, the protein DECR1 also showed a competition profile against 8 pY-peptides with various K_d values. Only a few, however, showed strong affinities. The competition profile for the strongest pY-peptide is given in Figure 63.

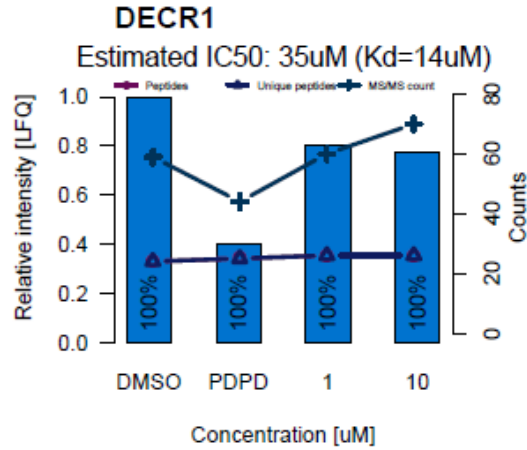


Figure 63 | Competition assay of LMTK2_476 showed the strongest affinities to DECR1 than the other pY-peptides.

Chapter 4 General discussion

As the next generation of challenge in the proteomics field approaches, protein-protein interaction networks have been pinpointed as the critical step in understanding the function of the cell proteome. Uncovering the pY interactome network of RTKs could significantly sharpen the picture on how the residues exert function.

With the development of the protein-array or peptide-array technologies, the interactome screening of phosphotyrosine residues can nowadays be performed in a routine manner^{25,26,61}. However, these assays often only focus on the interaction of pY-peptides to the recombinant SH2 or PTB domains in an artificial environment. Besides, the array is not feasible when the expression of the recombinant domains cannot successfully achieve. For instance, the work of PepSpotDB²⁶ was intended to express 99 SH2 domain but 29 failed mainly due to their poor solubility. The AE-MS approach provides another alternative to profile the interactome in an endogenous condition without any prior modification on SH2 or PTB domains. Taking advantage of AE-MS, a few pY interactomes of RTKs, including DDR1 and ErbB family receptors, have already been profiled against the proteome of the cell line or placenta^{27,28}. However, these studies mainly focused on the interactions between SH2 or PTB domain-containing proteins. Further effort on identifying the other interactors was not enough. To date, researches have defined several novel phosphotyrosine recognition domains or proteins without the known domains³⁵⁻³⁹, whose interactions to the RTKs are not confirmed systematically. Therefore, the task to reveal all the phosphotyrosine residues interactome of RTKs are not accomplished, and the current interactome pictures are not updated yet. By employing mass spectrometry with better performance and meaningful negative controls (blocked beads and random peptides) to account for specific interactors, this AE-MS study has achieved the goal of profiling all the all significant binders including typical interactors with SH or PTB domains, atypical interactors with or without known domains.

Interestingly, this systematic work displays a higher recovery of interaction in database PhosphoSite (61.1%)²⁴ than PepspotDB (31.8%)²⁶. To some extent, it implies the interaction between pY-peptide and interactor, not recombinant SH2 domain, is closer to the real interaction between two full-length proteins. This discovery demonstrates the flexibility of

using the peptide mimetics for the systematic screening work. Moreover, apart from the agreement to several interactions curated in PhosphoSitePlus, this AE-MS study has expanded the interactome of functional phosphotyrosine residues with novel interactors. It is a promising start to exploit their biological meaning in health and diseases. Furthermore, with the validation experiment by a competition assay, some contradictory findings were confirmed. It is likely that these contradictory interactors such as GRB7 in the study¹¹⁶ were the indirect or weak binders, because it was validated in the other competition assays of FLT3_pY591 and FLT3_pY599 with K_d values of 4 and 6 μM , respectively. Nevertheless, further work could perform to validate these contradictory discoveries and in the end, interactome and signaling pathway could be revised.

Overall, this work has expanded the knowledge on conventional binders of pY-peptides and provides a new direction to explore atypical interactors in the next step of research.

4.1 Exploration of interactors

With the achievement of RTKs interactome mediated by phosphotyrosine residues, this AE-MS study provides the endless possibilities to discover new biology and expand the knowledge of RTKs signaling.

For example, only 75 (47.5%) SH2 or PTB domain-containing proteins displayed the high enrichment to the pY-peptides of RTKs in the AE-MS. Where are the other SH2 or PTB domain-containing proteins? Reasons as to why the other proteins in this group did not show a strong enrichment are discussed below. Moreover, some possible works to validate them could be carried out in the future.

(1) Throughout this AE-MS study, a washing step was performed following enrichment. Thus, the affinity between the pY-peptides and absent proteins may have been too weak to be retained. (2) The binding domain of absent proteins preferentially interacts with phosphotyrosine residues not located within the RTK family. This is evidenced that around 6000 pY-peptides not from RTK family showed various affinities to the other proteins with SH2 domains in the PepspotDB database²⁶. Moreover, the number of phosphotyrosine residues of RTKs only account for around 2% of the identified phosphotyrosine residues in human proteome²⁴, indicating the huge number of protein-protein interaction mediated by the phosphotyrosine residues in the other proteins. (3) The missing proteins prefer the indirect binding to the pY-peptides via the interaction to the docking platform created by the direct interactors, but the interaction was too weak to retain under the experimental conditions. For instance, the SH2 domain-containing protein ABL (rarely enriched in AE-MS study) employ its Src homology-3 (SH3) domain to interact with the proline-rich loop of another SH2 domain-containing protein CRK-II^{121,122}. This, in turn, also suggests some identified interactors in the AE-MS study might not be the direct binders to the pY-peptides, which, however, could not be distinguished easily. (4) These SH2 or PTB domains are only accessible after an activation or inhibition process that occurs via a post-translation modification (PTM) or protein-protein interaction. For example, the interaction between the SH2 domain of SHC1 and pY-peptides was regulated by the phosphorylation of tyrosine residues on the central proline-rich collagen homology 1 (CH1) domain is phosphorylated⁹⁸. (5) Reversely, these missing interactions could

be modulated by the PTM of the amino acids surrounding the pY residues in a manner analogous to other protein interaction domains, such as PDZ domains¹²³.

In order to examine these assumptions, it is flexible to apply the same AE-MS approach with competition assay to validate the interactors with known binding motifs firstly. After that, modification of the peptides sequences (baits) to evaluate the interaction affinities and specificities could be performed.

Despite the question of 'missing' interactors, this AE-MS study provides another direction to exploit phosphotyrosine interactors. For example, scientists are searching for an alternative approach to enrich pY-peptides instead of using phosphotyrosine antibodies such as 4G10 and P-Tyr-100. Such an effort has been made to create 'superbinders' that are mutated from the natural SH2 domain of GRB2 and SRC. These 'superbinders' have shown to have incredibly low selectivity and high efficiency for pY-peptides¹²⁴. In the AE-MS study, the STAT5B, PLCG1, and PIK3R1/2/3 displayed much higher binding diversities (low selectivity) than the GRB2 and SRC, which indicates they potentially are the excellent resources for such a biochemical application.

Also, the known atypical interactors such as PRKCD, PRKCQ (C2 domain)^{71,72}, CBLL1 (HYB domain)⁷⁰, and PKM³⁵ have been identified as the interactors of the pY-peptides of RTKs by the AE-MS and competition assay in this study. This is a decent beginning for further investigation of their biological impact to the RTKs signaling. Perturbation experiments to stimulate or inhibit the RTKs phosphorylation, along with the co-IP method, could be performed to monitor their interactions in the next step. Then the multilayered proteomics strategy¹²⁵ is suitable to find out the dynamic change of RTK interactome, ubiquitinome, phosphoproteome, and late proteome in response to treatment.

Regarding the novel interactors with temporarily inactive PTP domain (*e.g.*, UBASH3B) and proteins without known binding domain such as MTHFS and DECR1, their interaction could be further confirmed *in vitro* firstly by the techniques of isothermal titration calorimetry (ITC)⁹⁸ or X-ray crystallography³⁵ using purified pY-peptides and recombinant proteins.

Furthermore, as illustrated above, when a mutation appears in the sequence flanking the phosphotyrosine residues, the signaling cascades can alter the binding strength of the interactors or detour these in another direction. Similarly, the mutation is also frequently

observed in the binding domains of phosphotyrosine residues³¹. For instance, 74 proteins with an SH2 domain were found to have 324 amino acids mutations. Of these, PTPN11, PIK3R1 and PLCG2 contained the highest number of mutations. A similar mutation pattern in the PID domain of phosphatases was also observed.

Interestingly, the AE-MS study also identified the vulnerable proteins PIK3R1 and PLCG2 as the highly redundant interactors to the pY-peptides of RTKs. Because the AE-MS study employed cell lysate from 4 cancer cell lines for the interactor enrichment, the predominantly identification of these two proteins could be resulted from the mutation in their SH2 domain. Further exploration of the mutation present in these susceptible proteins could be performed in the future, which could further explain the dominant signaling pathway in cancers. Therefore, unlocking the signaling relationship between pY and mutant interactors will undoubtedly aid in improving the resolution of disease.

Last but not least, one of the protein clusters that was significantly enriched in the pY interactome was the B or T cell-related kinases ZAP70, LCP2, and SYK. The full proteome data suggested that several RTKs such as PDGFRB, IGF1R, and INSR are expressed in the Jurkat and MV4-11 cell lines. High expression of these RTKs could mean that there is still a signaling pathway yet to be identified in these types of cell and the function thereof is awaiting discovery. Whether these receptor tyrosine kinases are contributing to the function of the cell is unknown. Therefore, from a biological point of view, the exploitation of this data set would be promising. Overall, this pY-peptide interactome is a rich and valuable resource to inspire new hypotheses and drive discoveries.

4.2 Exploration of pY-peptides

The pY interactome in this study could be an exceedingly useful supplementary resource for other applications besides regular RTKs signaling research.

Firstly, by revealing the proximal pY-peptides interactome of RTKs, more effective drugs could be precisely used to disable the interaction either by inhibiting the tyrosine phosphorylation or blocked the interacting domains. The EGFR_pY1092 in this study has been found to be versatile in interacting with 14 novel interactors, and of these, the VAV1 was found to possess a higher affinity than the known interactor STAT3 (Figure 49). In NSCLC, both EGFR and VAV1 are co-expressed in the database ProteomicsDB. This phosphotyrosine residue has also been observed in multiple NSCLC experiments in the PhosphoSiePlus database. This implied that continued treatment of NSCLC could begin from this interaction when the FDA-approved EGFR inhibitor. Erlotinib no longer shows the required potencies against non-small cell lung cancer (NSCLC) due to the drug resistance¹²⁶. Encouragingly, it has been shown that the inhibitor azathioprine can target VAV1 in pancreatic tumors¹²⁷. Therefore, the disruption of the interaction between EGFR_pY1092 and VAV1 in NSCLC with azathioprine would be feasible. Furthermore, the combination of the EGFR inhibitor and the VAV1 inhibitor will also be possible in NSCLC treatment in the future.

Secondly, this study has revealed that pY residues of RTKs harbor vast numbers of unique interactors. These pY-peptides would be a useful resource for the development of peptide drugs, *i.e.*, peptidomimetics. It is feasible to develop an inhibitor derive from the pY-peptide to target the disease-related proteins with SH2 or PTB domains. Related efforts have been underway for years¹²⁸. For example, scientists utilized the high-affinity STAT3-binding peptide pYLPQTV derived form Interleukin-6 receptor subunit beta (IL6ST) to target STAT3 and then received the potent inhibitors with IC_{50} 0.15 μ M¹²⁹. After that, a further design on the drug-like analogues resulted in a more potent inhibitor with low IC_{50} (0.069 μ M)¹³⁰. Interestingly, the starting peptide sequence pYLPQTV falls into the STAT3 binding motifs. There are four pY-peptides in the competition assay match to this motif as well, among which the peptide (pYLPQYP) from FGFR2_C808pY displayed the strongest affinity with K_d 0.07 μ M (Figure 64).

This observation implies the flexibility to use the comprehensive AE-MS dataset to aid in the design of peptidomimetics inhibitors in the future.

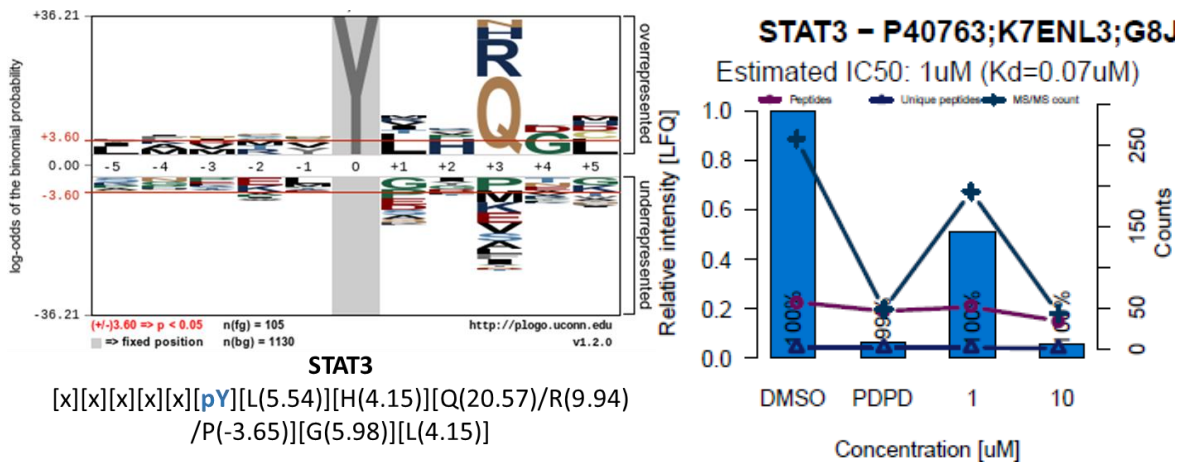


Figure 64 | Binding motif of STAT3 (left) and competition assay of STAT3 on FGFR2_C808pY (right).

Moreover, the versatile pY-peptides with redundant interactors can contribute to the drug screening in the future. Several promising works^{26,131} has constructed the affinity matrix to enrich the SH2 domain-containing proteins from the cell lysates. The optimal combination of the pY-peptides in the AE-MS assay with those matrices could improve the coverage of SH2 domain-containing proteins and further the proteins with PTB domains. In the end, the optimal matrix could act as the unselective baits to evaluate the potency of the inhibitors in competition assays. This effort could help the identification of the specific and unspecific targets of the inhibitors, and further promote faster drug development.

Chapter 5 Outlook

With the accomplishment of an RTK interactome mediated by phosphotyrosine in this study, it would naturally require some further experiments, such as co-IP, to validate the novel findings. The perturbation experiments with the ligands or drug and the competition assay with free corresponding pY peptides could be performed. After that, monitoring the dynamics of the tyrosine phosphorylation levels and the PTM influences of the downstream interactors, changes in protein localization, and the cell signaling outcome could be investigated.

Moreover, it is also worthy of employing the CRISPR-Cas 9 approach to resolve the influence of some mutation-gain tyrosine residues or hot-spot mutation on diseases such as cancer. Then the perturbation of the interactors could be performed to determine if the relevant signaling triggered by the particular phosphotyrosine residues could be regulated. In the end, these studies would contribute to drug discovery and drug re-purposing.

Another valuable aspect of this AE-MS study is the integration of multiple information from four databases. For instance, by mapping to PhosphoSitePlus and ProteomicsDB, the possibility of the pY interaction in the specific cell lines or tissue is visualized. With more cell full proteomes and phosphotyrosine sites resolved in the future, this data set would be a greater resource to predict possible existing cell signaling pathway in diseases.

Furthermore, with the great performance, the AE-MS could be applied to screen the interactome of the peptides with doubly- or triply-phosphorylated tyrosine residues, or various combinations of phosphotyrosine, phosphoserine, phosphothreonine, or even the other post-translation modifications such as ubiquitination and acetylation. For example, phosphoserine and phosphothreonine residues are frequently-identified in cell lines, as these are both significantly more abundant than phosphotyrosine residues. It has been shown that the protein 14-3-3 can specifically interact with phosphoserine peptides containing two motifs¹³². However, what the specificities of these are with respect to recruiting other proteins is yet to be determined. Through the introduction of other high-throughput MS-based techniques, *e.g.*, SILAC or TMT, with on-site membrane peptide synthesis, such screening could be performed effectively in the future.

Abbreviations

| | | | |
|-------|--|-----------|--|
| ACN | acetonitrile | LANCL1 | LanC-like protein |
| AE-MS | affinity enrichment-mass spectrometry | LC-MS/MS | liquid chromatography-tandem mass spectrometry |
| AGC | automatic gain control | LFQ | label-free quantification |
| ATCC | American Type Culture Collection | LOESS | local polynomial regression |
| CAA | chloroacetamide | LTP | low-throughput |
| CH1 | collagen homology 1 | MALDI | matrix-assisted laser desorption/ionization |
| co-IP | co-immunoprecipitation | MALDI-TOF | matrix assisted laser desorption ionization-time of flight |
| DDA | data-dependent acquisition | MGY | mutation-gain tyrosine |
| DECR1 | 2,4-dienoyl-CoA reductase | MS1 | precursor mass |
| DMSO | dimethyl sulfoxide | MTHFS | 5-formyltetrahydrofolate cyclo-ligase |
| DSPs | dual-specificity phosphatases | NH4OAC | ammonium acetate |
| DTT | dithiothreitol | NH4OH | ammonium hydroxide |
| EF | enrichment factor | nLC-MS/MS | nano-liquid chromatography-tandem mass spectrometry |
| ERK | mitogen-activated protein kinase | NSCLC | non-small-cell lung cancer |
| ESI | electrospray ionization | PD | pull-down |
| FA | formic acid | PDB | Protein Data Bank |
| FBP | fructose-1,6-bisphosphate | PH | pleckstrin homology |
| FBS | fetal bovine serum | PID | phosphotyrosine interaction domain |
| FDR | false discovery rate | PKM2 | pyruvate kinase muscle isozyme M2 |
| Fmoc | fluorenylmethyloxycarbonyl chloride | PSP | PhosphoSitePlus |
| GFP | green fluorescent protein | PTB | phosphotyrosine-binding domain |
| GPD2 | glycerol-3-phosphate dehydrogenase | PTGES2 | prostaglandin E synthase 2 |
| GST | glutathione S-transferase | PTM | post-translation modification |
| HCCA | α -cyano-4-hydroxycinnamic acid | PTP | protein tyrosine phosphatase |
| HCD | higher-energy collisional dissociation | PTPN2 | protein tyrosine phosphatase non-receptor type 2 |
| HPLC | high performance liquid chromatography | PTPN1 | protein tyrosine phosphatase non-receptor type 1 |
| HSM | hotspot | pY | phosphotyrosine |

| | | | |
|-----------|---|------------|---|
| HTP | high-throughput | RPTPs | receptor protein tyrosine phosphatases |
| HYB | Hakai-tyrosine binding | RTK | receptor tyrosine kinase |
| iBAQ | intensity-based absolute quantification | SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| IC_{50} | half-maximal inhibitory concentration | SH2 domain | Src homology-2 |
| IF | isoform | SILAC | stable isotope labeling by amino acids in cell culture |
| IL-2R | interleukin-2 receptor | TFA | trifluoroacetic acid |
| IQR | interquartile range | TMT | Tandem Mass Tag |
| IT | maximum injection time | WT | wild-type |
| K_d | dissociation constant | Y2H | yeast two-hybrid |

References

1. Hanks, S., Quinn, A. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science (80-.)*. **241**, 42–52 (1988).
2. Yarden, Y. & Ullrich, A. Growth Factor Receptor Tyrosine Kinases. *Annu. Rev. Biochem.* **57**, 443–478 (1988).
3. Schlessinger, J. Signal transduction by allosteric receptor oligomerization. *Trends Biochem. Sci.* **13**, 443–447 (1988).
4. Williams, L. Signal transduction by the platelet-derived growth factor receptor. *Science (80-.)*. **243**, 1564–1570 (1989).
5. Bateman, A. *et al.* UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
6. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
7. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
8. Schlessinger, J. & Ullrich, A. Signal Transduction by Receptors with Tyrosine Kinase Activity. *Cell* **61**, 203–212 (1990).
9. Hirsch, F. R. *et al.* Epidermal Growth Factor Receptor in Non–Small-Cell Lung Carcinomas: Correlation Between Gene Copy Number and Protein Expression and Impact on Prognosis. *J. Clin. Oncol.* **21**, 3798–3807 (2003).
10. Nicholson, R. ., Gee, J. M. . & Harper, M. . EGFR and cancer prognosis. *Eur. J. Cancer* **37**, 9–15 (2001).
11. Ohsaki, Y. *et al.* Epidermal growth factor receptor expression correlates with poor prognosis in non-small cell lung cancer patients with p53 overexpression. *Oncol. Rep.* **279**, 3662–3669 (2000).
12. Furdui, C. M., Lew, E. D., Schlessinger, J. & Anderson, K. S. Autophosphorylation of FGFR1 kinase is mediated by a sequential and precisely ordered reaction. *Mol. Cell* **21**, 711–717 (2006).
13. Favelyukis, S., Till, J. H., Hubbard, S. R. & Miller, W. T. Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. *Nat. Struct. Biol.* **8**, 1058–1063 (2001).
14. Honegger, A. M., Kris, R. M., Ullrich, A. & Schlessinger, J. Evidence that autophosphorylation of solubilized receptors for epidermal growth factor is mediated by intermolecular cross-phosphorylation. *Proc. Natl. Acad. Sci.* **86**, 925–929 (1989).
15. Till, J. H. *et al.* Crystal Structure of the MuSK Tyrosine Kinase: Insights into Receptor Autoregulation. *Structure* **10**, 1187–1196 (2002).
16. Ward, C. W., Lawrence, M. C., Streltsov, V. A., Adams, T. E. & McKern, N. M. The insulin and EGF receptor structures: new insights into ligand-induced receptor activation. *Trends Biochem. Sci.* **32**, 129–137 (2007).
17. Clayton, A. H. A. *et al.* Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis. *J. Biol. Chem.* **280**, 30392–30399 (2005).
18. Gadella, T. W. J. & Jovin, T. M. Oligomerization of epidermal growth factor receptors on

- A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation. *J. Cell Biol.* **129**, 1543–1558 (1995).
19. Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **103**, 211–225 (2000).
 20. Salameh, A. Direct recruitment of CRK and GRB2 to VEGFR-3 induces proliferation, migration, and survival of endothelial cells through the activation of ERK, AKT, and JNK pathways. *Blood* **106**, 3423–3431 (2005).
 21. Stark, C. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
 22. Marshall, C. J. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
 23. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
 24. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
 25. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
 26. Tinti, M. *et al.* The SH2 Domain Interaction Landscape. *Cell Rep.* **3**, 1293–1305 (2013).
 27. Schulze, W. X., Deng, L. & Mann, M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* **1**, E1–E13 (2005).
 28. Lemeer, S. *et al.* Phosphotyrosine mediated protein interactions of the discoidin domain receptor 1. *J Proteomics* **75**, 3465–3477 (2012).
 29. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 1–7 (2018). doi:10.1093/nar/gky1015
 30. Dibb, N. J., Dilworth, S. M. & Mol, C. D. Switching on kinases: Oncogenic activation of BRAF and the PDGFR family. *Nat. Rev. Cancer* **4**, 718–727 (2004).
 31. Gauthier, N. P. *et al.* MutationAligner: A resource of recurrent mutation hotspots in protein domains in cancer. *Nucleic Acids Res.* **44**, D986–D991 (2016).
 32. Bose, R. *et al.* Activating HER2 Mutations in HER2 Gene Amplification Negative Breast Cancer. *Cancer Discov.* **3**, 224–237 (2013).
 33. Smith, C. C. *et al.* Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature* **485**, 260–263 (2012).
 34. CH, M. *et al.* Sorafenib treatment of FLT3-ITD(+) acute myeloid leukemia: favorable initial outcome and mechanisms of subsequent nonresponsiveness associated with the emergence of a D835 mutation. *Blood* **119**, 5133–5143 (2012).
 35. Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–186 (2008).
 36. Morishige, M. *et al.* GEP100 links epidermal growth factor receptor signalling to Arf6 activation to induce breast cancer invasion. *Nat. Cell Biol.* **10**, 85–92 (2008).
 37. Park, S. *et al.* Regulation of RKIP binding to the N-region of the Raf-1 kinase. *FEBS Lett.* **580**, 6405–6412 (2006).
 38. Tavel, L. *et al.* Ligand Binding Study of Human PEBP1/RKIP: Interaction with Nucleotides and Raf-1 Peptides Evidenced by NMR and Mass Spectrometry. *PLoS One* **7**, e36187 (2012).
 39. Christofk, H. R., Wu, N., Cantley, L. C. & Asara, J. M. Proteomic Screening Method for

- Phosphopeptide Motif Binding Proteins Using Peptide Libraries. *J. Proteome Res.* **10**, 4158–4164 (2011).
40. Sadowski, I., Stone, J. C. & Pawson, T. A Noncatalytic Domain Conserved among Cytoplasmic Protein-Tyrosine Kinases Modifies the Kinase Function and Transforming Activity of Fujinami Sarcoma Virus P130gag-fPs. *Mol. Cell. Biol.* **6**, 4396–4408 (1986).
 41. Mayer, B. J., Hamaguchi, M. & Hanafusa, H. A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272–275 (1988).
 42. Anderson, D. *et al.* Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. *Science (80-.)*. **250**, 979–982 (1990).
 43. Liu, B. A. *et al.* The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. *Sci. Signal.* **4**, 1–18 (2011).
 44. Colicelli, J. ABL Tyrosine Kinases: Evolution of Function, Regulation, and Specificity. *Sci. Signal.* **3**, re6-re6 (2010).
 45. Songyang, Z. & Cantley, L. C. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci.* **20**, 470–475 (1995).
 46. Waksman, G. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: Crystal structures of the complexed and peptide-free forms. *Cell* **72**, 779–790 (1993).
 47. Waksman, G. *et al.* Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**, 646–653 (1992).
 48. Kaneko, T., Joshi, R., Feller, S. M. & Li, S. S. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. *Cell Commun. Signal.* **10**, 32 (2012).
 49. Liu, B. A. *et al.* The SH2 Domain-Containing Proteins in 21 Species Establish the Provenance and Scope of Phosphotyrosine Signaling in Eukaryotes. *Sci. Signal.* **4**, ra83-ra83 (2011).
 50. Liu, B. A. *et al.* The Human and Mouse Complement of SH2 Domain Proteins—Establishing the Boundaries of Phosphotyrosine Signaling. *Mol. Cell* **22**, 851–868 (2006).
 51. Huang, H. *et al.* Defining the Specificity Space of the Human Src Homology 2 Domain. *Mol. Cell. Proteomics* **7**, 768–784 (2008).
 52. Kaneko, T. *et al.* Loops Govern SH2 Domain Specificity by Controlling Access to Binding Pockets. *Sci. Signal.* **3**, ra34-ra34 (2010).
 53. Marengere, L. E. M. *et al.* SH2 domain specificity and activity modified by a single residue. *Nature* **369**, 502–505 (1994).
 54. Kaneko, T., Sidhu, S. S. & Li, S. S. C. Evolving specificity from variability for protein interaction domains. *Trends Biochem. Sci.* **36**, 183–190 (2011).
 55. Poy, F. *et al.* Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell* **4**, 555–561 (1999).
 56. Bae, J. H. *et al.* Structural basis underlying a novel mechanism for control of receptor tyrosine kinase signal selectivity. *Cell* **138**, 514–524 (2009).
 57. Blaikie, P. *et al.* A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *J. Biol. Chem.* **269**, 32031–4 (1994).
 58. Kavanaugh, W. & Williams, L. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science (80-.)*. **266**, 1862–1865 (1994).

59. Uhlik, M. T. *et al.* Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J. Mol. Biol.* **345**, 1–20 (2005).
60. Stolt, P. C. *et al.* Origins of peptide selectivity and phosphoinositide binding revealed by structures of Disabled-1 PTB domain complexes. *Structure* **11**, 569–579 (2003).
61. Smith, M. J., Hardy, W. R., Murphy, J. M., Jones, N. & Pawson, T. Screening for PTB Domain Binding Partners and Ligand Specificity Using Proteome-Derived NPXY Peptide Arrays. *Mol. Cell. Biol.* **26**, 8461–8474 (2006).
62. DiNitto, J. P. & Lambright, D. G. Membrane and juxtamembrane targeting by PH and PTB domains. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1761**, 850–867 (2006).
63. Rameh, L. E. *et al.* A comparative analysis of the phosphoinositide binding specificity of pleckstrin homology domains. *J. Biol. Chem.* **272**, 22059–22066 (1997).
64. Alonso, A. *et al.* Protein Tyrosin Phosphatase in the Human Genome. *Cell* **117**, 699–711 (2004).
65. ANDERSEN, J. N. A genomic perspective on protein tyrosine phosphatases: gene structure, pseudogenes, and genetic disease linkage. *FASEB J.* **18**, 8–30 (2004).
66. Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* **7**, 833–846 (2006).
67. Wishart, M. J., Denu, J. M., Williams, J. A. & Dixon, J. E. A single mutation converts a novel phosphotyrosine binding domain into a dual-specificity phosphatase. *J. Biol. Chem.* **270**, 26782–26785 (1995).
68. Haynie, D. T. & Ponting, C. P. The N-terminal domains of tensin and auxilin are phosphatase homologues. *Protein Sci.* **5**, 2643–2646 (1996).
69. Fujita, Y. *et al.* Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat. Cell Biol.* **4**, 222–231 (2002).
70. Mukherjee, M. *et al.* Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. *EMBO J.* **31**, 1308–1319 (2012).
71. Benes, C. H. *et al.* The C2 Domain of PKC δ Is a Phosphotyrosine Binding Domain. *Cell* **121**, 271–280 (2005).
72. Stahelin, R. V. *et al.* Protein Kinase C θ C2 Domain Is a Phosphotyrosine Binding Module That Plays a Key Role in Its Activation. *J. Biol. Chem.* **287**, 30518–30528 (2012).
73. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
74. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10**, 2763–2788 (2009).
75. Grossmann, A. *et al.* Phospho-tyrosine dependent protein-protein interaction network. *Mol. Syst. Biol.* **11**, 794–794 (2015).
76. Yao, Z. *et al.* A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. *Mol Cell* **65**, 347–360 (2017).
77. Auerbach, D. & Stagljar, I. in *Proteomics and Protein-Protein Interactions* 19–31 (Springer US). doi:10.1007/0-387-24532-4_2
78. Deane, C. M., Salwiński, Ł., Xenarios, I. & Eisenberg, D. Protein Interactions. *Mol. Cell. Proteomics* **1**, 349–356 (2002).
79. Van Der Geer, P. *Analysis of protein-protein interactions by coimmunoprecipitation. Methods in Enzymology* **541**, (Elsevier Inc., 2014).

80. Heiss, E. *et al.* Identification of Y589 and Y599 in the juxtamembrane domain of Flt3 as ligand-induced autophosphorylation sites involved in binding of Src family kinases and the protein tyrosine phosphatase SHP2. *Blood* **108**, 1542–1550 (2006).
81. Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol. Cell. Proteomics* **14**, 120–135 (2015).
82. Hanke, S. & Mann, M. The Phosphotyrosine Interactome of the Insulin Receptor Family and Its Substrates IRS-1 and IRS-2. *Mol. Cell. Proteomics* **8**, 519–534 (2009).
83. Ostresh, J. M., Winkle, J. H., Hamashin, V. T. & Houghten, R. A. Peptide libraries: Determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers* **34**, 1681–1689 (1994).
84. Labadia, M. E., Ingraham, R. H., Schembri-King, J., Morelock, M. M. & Jakes, S. Binding affinities of the SH2 domains of ZAP-70, p56lck and Shc to the ζ chain ITAMs of the T-cell receptor determined by surface plasmon resonance. *J. Leukoc. Biol.* **59**, 740–746 (1996).
85. Du, H., Fuh, R.-C. A., Li, J., Corkan, L. A. & Lindsey, J. S. PhotochemCAD: A Computer-Aided Design and Research Tool in Photochemistry. *Photochem. Photobiol.* **68**, 141–142 (1998).
86. Dixon, J. M., Taniguchi, M. & Lindsey, J. S. PhotochemCAD 2: A Refined Program with Accompanying Spectral Databases for Photochemical Calculations¶. *Photochem. Photobiol.* **81**, 212–213 (2007).
87. Gill, S. C. & Von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data [published erratum appears in *Anal Biochem* 1990 Sep;189(2):283]. *Anal. Biochem.* **182**, 319–326 (1989).
88. Yu, P. *et al.* Trimodal Mixed Mode Chromatography That Enables Efficient Offline Two-Dimensional Peptide Fractionation for Proteome Analysis. *Anal. Chem.* **89**, 8884–8891 (2017).
89. Trinkle-Mulcahy, L. *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **183**, 223–239 (2008).
90. O’Shea, J. P. *et al.* PLogo: A probabilistic approach to visualizing sequence motifs. *Nat. Methods* **10**, 1211–1212 (2013).
91. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
92. Médard, G. *et al.* Optimized chemical proteomics assay for kinase inhibitor profiling. *J. Proteome Res.* **14**, 1574–1586 (2015).
93. Kuzmič, P. *et al.* High-throughput screening of enzyme inhibitors: Automatic determination of tight-binding inhibition constants. *Anal. Biochem.* **281**, 62–67 (2000).
94. Daub, H. Quantitative proteomics of kinase inhibitor targets and mechanisms. *ACS Chem. Biol.* **10**, 201–212 (2015).
95. Lemeer, S. *et al.* Phosphotyrosine mediated protein interactions of the discoidin domain receptor 1. *J. Proteomics* **75**, 3465–3477 (2012).
96. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
97. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**,

- 1–13 (2009).
98. George, R., Schuller, A. C., Harris, R. & Ladbury, J. E. A Phosphorylation-Dependent Gating Mechanism Controls the SH2 Domain Interactions of the Shc Adaptor Protein. *J. Mol. Biol.* **377**, 740–747 (2008).
 99. Zhou, S. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778 (1993).
 100. Ihle, J. N. The Stat family in cytokine signaling. *Curr. Opin. Cell Biol.* **13**, 211–217 (2001).
 101. Fischer, H. J. *et al.* The Insulin Receptor Plays a Critical Role in T Cell Function and Adaptive Immunity. *J. Immunol.* **198**, 1910–1920 (2017).
 102. Leitinger, B. in *Int Rev Cell Mol Biol* **310**, 39–87 (2014).
 103. Leitinger, B., Steplewski, A. & Fertala, A. The D2 Period of Collagen II Contains a Specific Binding Site for the Human Discoidin Domain Receptor, DDR2. *J. Mol. Biol.* **344**, 993–1003 (2004).
 104. Shrivastava, A. *et al.* An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol Cell* **1**, 25–34 (1997).
 105. Vogel, W., Gish, G. D., Alves, F. & Pawson, T. The Discoidin Domain Receptor Tyrosine Kinases Are Activated by Collagen. *Mol. Cell* **1**, 13–23 (1997).
 106. Fu, H.-L. *et al.* Discoidin Domain Receptors: Unique Receptor Tyrosine Kinases in Collagen-mediated Signaling. *J. Biol. Chem.* **288**, 7430–7437 (2013).
 107. Chae, Y. K. *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. *Oncotarget* **8**, 16052–16074 (2017).
 108. Ulaganathan, V. K., Sperl, B., Rapp, U. R. & Ullrich, A. Germline variant FGFR4 p.G388R exposes a membrane-proximal STAT3 binding site. *Nature* **528**, 570–574 (2015).
 109. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
 110. Choi, S. & Park, S. Phosphorylation at Tyr-838 in the kinase domain of EphA8 modulates Fyn binding to the Tyr-615 site by enhancing tyrosine kinase activity. *Oncogene* **18**, 5413–5422 (1999).
 111. Wybenga-Groot, L. E. & McGlade, C. J. RTK SLAP DOWN: The emerging role of Src-like adaptor protein as a key player in receptor tyrosine kinase signaling. *Cell. Signal.* **27**, 267–274 (2015).
 112. Raghu, H. *et al.* UPA and uPAR shRNA inhibit angiogenesis via enhanced secretion of SVEGFR1 independent of GM-CSF but dependent on TIMP-1 in endothelial and glioblastoma cells. *Mol. Oncol.* **6**, 33–47 (2012).
 113. Ito, N. *et al.* Identification of Vascular Endothelial Growth Factor Receptor-1 Tyrosine Phosphorylation Sites and Binding of SH2 Domain-containing Molecules. *J Biol Chem* **273**, 23410–23418 (1998).
 114. Ito, N., Huang, K. & Claesson-Welsh, L. Signal transduction by VEGF receptor-1 wild type and mutant proteins. *Cell. Signal.* **13**, 849–854 (2001).
 115. Kazi, J. U. FLT3 signals via the adapter protein Grb10 and overexpression of Grb10 leads to aberrant cell proliferation in acute myeloid leukemia The adaptor protein Grb10 plays important roles in mitogenic signaling. However, its roles. *Mol. Oncol.* **7**, 402–418 (2013).
 116. THÖMMES, K., LENNARTSSON, J., CARLBERG, M. & RÖNNSTRAND, L. Identification of Tyr-703 and Tyr-936 as the primary association sites for Grb2 and Grb7 in the c-Kit/stem cell

- factor receptor. *Biochem. J.* **341**, 211 (1999).
117. Galvagni, F. *et al.* Endothelial cell adhesion to the extracellular matrix induces c-Src-dependent VEGFR-3 phosphorylation without the activation of the receptor intrinsic kinase activity. *Circ. Res.* **106**, 1839–1848 (2010).
 118. Shi, G., Yue, G. & Zhou, R. EphA3 functions are regulated by collaborating phosphotyrosine residues. *Cell Res.* **20**, 1263–1275 (2010).
 119. Robertson, F. M. *et al.* Presence of anaplastic lymphoma kinase in inflammatory breast cancer. *Springerplus* **2**, 1–12 (2013).
 120. Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455**, 967–970 (2008).
 121. Anafi, M., Rosen, M. K., Gish, G. D., Kay, L. E. & Pawson, T. A potential SH3 domain-binding site in the Crk SH2 domain. *J. Biol. Chem.* **271**, 21365–21374 (1996).
 122. Donaldson, L. W., Gish, G., Pawson, T., Kay, L. E. & Forman-Kay, J. D. Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. *Proc. Natl. Acad. Sci.* **99**, 14053–14058 (2002).
 123. Ikenoue, T., Inoki, K., Zhao, B. & Guan, K. L. PTEN acetylation modulates its interaction with PDZ domain. *Cancer Res.* **68**, 6908–6912 (2008).
 124. Bian, Y. *et al.* Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. *Nat. Chem. Biol.* **12**, 1–10 (2016).
 125. Francavilla, C. *et al.* Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking. *Nat Struct Mol Biol* **23**, 608–618 (2016).
 126. Tang, J., Salama, R., Gadgeel, S. M., Sarkar, F. H. & Ahmad, A. Erlotinib resistance in lung cancer: Current progress and future perspectives. *Front. Pharmacol.* **4 FEB**, 1–9 (2013).
 127. Razidlo, G. L. *et al.* Targeting Pancreatic Cancer Metastasis by Inhibition of Vav1, a Driver of Tumor Cell Invasion. *Cancer Res.* **75**, 2907–2915 (2015).
 128. Kraskouskaya, D., Duodu, E., Arpin, C. C. & Gunning, P. T. Progress towards the development of SH2 domain inhibitors. *Chem. Soc. Rev.* **42**, 3337–3370 (2013).
 129. Ren, Z., Cabell, L. A., Schaefer, T. S. & McMurray, J. S. Identification of a high-affinity phosphopeptide inhibitor of Stat3. *Bioorganic Med. Chem. Lett.* **13**, 633–636 (2003).
 130. Mandal, P. K., Ren, Z., Chen, X., Xiong, C. & McMurray, J. S. Structure–Affinity Relationships of Glutamine Mimics Incorporated into Phosphopeptides Targeted to the SH2 Domain of Signal Transducer and Activator of Transcription 3. *J. Med. Chem.* **52**, 6126–6141 (2009).
 131. Höfener, M., Heinzlmeir, S., Kuster, B. & Sewald, N. Probing SH2-domains using Inhibitor Affinity Purification (IAP). *Proteome Sci.* **12**, 41 (2014).
 132. Yaffe, M. B. *et al.* The Structural Basis for 14-3-3:Phosphopeptide Binding Specificity. *Cell* **91**, 961–971 (1997).

Acknowledgment

With the completion of this thesis, I would like to greatly appreciate all the lab members who offered me so much helps during my Ph.D. study at the Chair of Proteomics and Bioanalytics in the Technical University of Munich in Germany, and all my friends who supported me and experienced much fantastic time with me in Freising and Munich.

First of all, I would like to deliver my appreciation to my great supervisor: Prof. Dr. Bernhard Kuster, not just because of his help during this staying of four and a half years, but also his acceptance of me, an Analytical Chemist without the Proteomics experience, as his Ph.D. student in the beginning. Besides, he provided me several opportunities to attend the national and international conferences to expand my knowledge. Secondly, I want to thank the China Scholar Council for supporting my four years of Ph.D. study. Without this financial resource, I do not think I could make my dream of studying in a top group in the world come true. Thirdly, the tremendous appreciation is for two of my colleagues: Dr. Chen Meng and Dr. Mathias Wilhelm, who collaborate with me in the data analysis and all the meaningful discussion. Moreover, I would like to thank our technician Mr. Andreas Klaus for his splendid in-gel digestion work, Mrs. Andrea Hubauer for the instruction of cell culture and peptide synthesis, Mrs. Krötz-Fahning for her lab-work instruction, and two secretaries Mrs. Gabriele Kröppelt and Mrs. Silvia Rötzer for their help in the document-related work.

Last but not least, I also want to thank my parents for the spiritual supporting and encouragement during this particular period and the excellent understanding of my life.

Then I would like to thank all of my examiners Prof. Dr. Dieter Langosch (TUM), Prof. Dr. Bernhard Kuster (TUM) and Prof. Dr. Pascal Falter-Braun (Helmholtz Zentrum München/Ludwig-Maximilians-Universität München) for being to organize my Ph.D. defense.

Curriculum Vitae

Runsheng Zheng

runsheng.zheng@tum.de

Mobile: +49-015902606175

Date of birth: 27th of April, 1988

Education

September 2014 – present, **Technical University of Munich, Germany**

Thesis: Profiling the phosphotyrosine interactome of receptor tyrosine kinases

Chair of Proteomics and Bioanalytics, Prof. Bernhard Kuster

September 2011– June 2014, **Guangzhou University of Chinese Medicine, China**

Master of Medicine

Thesis: A high-throughput method for the determination of aflatoxin B₁, B₂, G₁, G₂, ochratoxin A and sterigmatocystin in traditional Chinese medicines by high-performance liquid chromatography-tandem mass spectrometry

The Research Centre of Chinese Herbal Resource Science and Engineering, Prof. Hui Xu

September 2007 – June 2011, **Guangzhou University of Chinese Medicine, China**

Bachelor of Science

Thesis: An Ultraviolet-visible spectroscopy method for the evaluation and determination of saccharides in *Morinda Officinalis*

College of Traditional Chinese Medicine

Languages

Chinese (native), English (fluent)

Extracurricular activities

Chinese Martial Art and hiking

Conference contribution and publication

Conferences

1. 66th American Society for Mass Spectrometry Conference (San Diego/USA, 2018)
Topic: Profiling the phosphotyrosine interactome of Receptor Tyrosine Kinases by affinity purification-mass spectrometry
Presentation type: **poster**
2. 16th Human Proteome Organisation World Congress (Dublin/IRL, 2017)
Topic: Profiling the phosphotyrosine interactome of Receptor Tyrosine Kinases
Presentation type: **oral**
3. Proteomics Forum (Potsdam/DE, 2017)
Topic: The phosphotyrosine mediated interactome of Receptor Tyrosine Kinases
Presentation type: **oral presentation in the section “Young Investigator Award”**
4. 13th German Peptide Symposium (Erlangen/DE, 2017)
Topic: Profiling the phosphotyrosine interactome of Receptor Tyrosine Kinases
Presentation type: **oral and poster**
5. 15th Human Proteome Organisation World Congress (Taipei/CN, 2016)
Topic: Profiling the phosphotyrosine interactome of Receptor Tyrosine Kinases
Presentation type: **oral and poster**

Publications

1. **Pharmacoproteomic characterization of human colon and rectal cancer.** Martin Frejno, Riccardo Zenezini Chiozzi, Mathias Wilhelm, Heiner Koch, **Runsheng Zheng**, Susan Klaeger, Benjamin Ruprecht, Chen Meng, Karl Kramer, Anna Jarzab, Stephanie Heinzlmeir, Elaine Johnstone, Enric Domingo, David Kerr, Moritz Jesinghaus, Julia Slotta–Huspenina, Wilko Weichert, Stefan Knapp, Stephan M Feller, and Bernhard Kuster. *Molecular Systems Biology*, 2017 Nov 3. DOI: 10.15252/msb.20177701.
2. **Trimodal mixed mode chromatography enables efficient offline 2D peptide fractionation for proteome analysis.** Peng Yu, Svenja Petzoldt, Mathias Wilhelm, Daniel Zolg, **Runsheng Zheng**, Xuefei Sun, Xiaodong Liu, Günter Schneider, Andreas F. Huhmer, and Bernhard Kuster. *Analytical Chemistry*, 2017 August 1. DOI: 10.1021/acs.analchem.7b01356.

Appendix

Supplementary File 1: SH2 or PTB domain containing proteins of human

Supplementary File 2: All wild-type and mutant pY-peptides of RTKs

Supplementary File 3: Significant interactors in AE-MS after filtering

Supplementary File 4: Motif of significant interactors in AE-MS

Supplementary File 5: Interactome comparison of subfamily RTKs

Supplementary File 6: Interactor evaluation by competition assay

Supplementary File 7: Plots of competition assay results