Oncogenic Mutations Rewire Signaling Pathways by Switching Protein Recruitment to Phosphotyrosine Sites

Lundby, Alicia; Franciosa, Giulia; Emdal, Kristina B; Refsgaard, Jan C; Gnosa, Sebastian P; Bekker-Jensen, Dorte B; Secher, Anna; Maurya, Svetlana R; Paul, Indranil; Mendez, Blanca L; Kelstrup, Christian D; Francavilla, Chiara; Kveiborg, Marie; Montoya, Guillermo; Jensen, Lars J; Olsen, Jesper V

Published in:
Cell

DOI:
10.1016/j.cell.2019.09.008

Publication date:
2019

Document version
Early version, also known as pre-print

Oncogenic mutations rewire signaling pathways by switching protein complex recruitment to phosphotyrosine sites

Short title: ***Decoding EGFR signaling network architecture by functional annotation of in-vivo phosphotyrosine sites***

Alicia Lundby[1,2,#], Giulia Franciosa[1], Jan C. Refsgaard[1], Kristina B. Emdal[1], Sebastian P. Gnosa[3], Dorte B. Bekker-Jensen[1], Anna Secher[1,3], Svetlana R. Maurya[2], Indranil Paul[1], Blanca L. Mendez[1], Christian D. Kelstrup[1], Chiara Francavilla[1], Marie Kveiborg[3], Guillermo Montoya[1], Lars J. Jensen[1], Jesper V. Olsen[1,#]

Affiliations:

[1] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health and Medical Sciences, Blegdamsvej 3b, DK-2200 Copenhagen, Denmark.

[2] Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[3] Biotech Research and Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

[4] Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Maaloev, Denmark.

[#] Correspondence should be addressed to A.L (alicia.lundby@sund.ku.dk) or J.V.O. (jesper.olsen@cpr.ku.dk)

HIGHLIGHTS

- Outline of EGF dependent *in-vivo* phosphotyrosine signaling in lung tissue
- A peptide based screen identifies proteins recruited to regulated phosphotyrosine sites in lung tissue
- Several somatic mutations affect proline residues within position +3 of regulated phosphotyrosine sites and introduce molecular switches
- A lung cancer oncogenic mutation in EGFR causes aberrant activation of phosphorylation signaling pathways by switching of a recruited protein complex
- Cancer mutations in vicinity of phosphotyrosine sites induce molecular switches that alter protein signaling networks
- Introduction of cutting-edge LC-MS instrumentation and DIA enables scalable, rapid and high-throughput analysis of phosphotyrosine site interactions

ABSTRACT:

Tyrosine phosphorylation regulates multi-layered signaling networks with broad implications in (patho)physiology, but high-throughput methods for functional annotation of phosphotyrosine sites are lacking. To decipher phosphotyrosine signaling directly in tissue samples, we developed a mass spectrometry-based interaction proteomics approach. We measured the *in-vivo* EGF-dependent signaling network in lung tissue quantifying >1000 phosphotyrosine sites. To assign function to all EGF-regulated sites we determined their recruited protein signaling complexes in lung tissue by interaction proteomics. We demonstrate how seemingly harmless mutations near tyrosine residues introduce molecular switches that rewire cancer signaling networks, and we revealed oncogenic properties of such a lung cancer EGFR mutant. To demonstrate the scalability of the approach we performed >1000 phosphopeptide pulldowns, analyzed them by rapid mass spectrometric analysis revealing tissue-specific differences in interactors. Our approach is a general strategy for functional annotation of phosphorylation sites in tissues enabling in-depth mechanistic insights into oncogenic rewiring of signaling networks.

MAIN TEXT

Tyrosine phosphorylation controls physiological signaling networks and represents a mechanism for cells to transiently alter protein function, such as enzymatic activity, protein-protein interactions and protein localization. In absence of signaling events, tyrosine phosphorylation is maintained at low stoichiometric levels (Sharma et al., 2014). Despite its low abundance, it is a central post-translational modification (Hunter and Sefton, 1980; Sefton et al., 1980) that upon deregulation is critically involved in disease, notably cancer. Tyrosine kinases have accordingly become prominent drug targets (Cohen, 2002; Klaeger et al., 2017; Rix and Superti-Furga, 2009; Zhou et al., 2013). For example, oncogenic driver mutations are prevalent in the epidermal growth factor receptor (EGFR) in lung adenocarcinomas (Kandoth et al., 2013), which are targeted with anti-EGFR therapies (Paez et al., 2004; Soria et al., 2018). The physiological importance of tyrosine kinases are recognized, but our knowledge of the molecular consequences of aberrant wiring of phosphotyrosine signaling in pathophysiological states is limited. The signaling response downstream of tyrosine kinases is complex involving multiple layers of signaling waves, which are difficult to measure. The immediate signaling response induces changes in phosphorylation site stoichiometry. Such changes have been analyzed by quantitative mass spectrometry-based phosphoproteomics, which can readily analyze thousands of phosphorylation sites (Francavilla et al., 2013; Olsen et al., 2006; Sharma et al., 2014). For the more abundant serine and threonine phosphorylation responses, some studies have even done so in tissues (Humphrey et al., 2015; Liu et al., 2018; Lundby et al., 2013). Investigating the signaling network in a tissue-specific context is preferential as it provides insight into the mechanisms encoding specificity (Liu et al., 2018). The next signaling layer involves dynamical formation of protein-protein complexes at regulated phosphorylation sites. Site-specific tyrosine phosphorylation predominantly occurs within intrinsically disordered protein regions (IDRs), which are peptide segments without stable tertiary structure (van der Lee et al., 2014). The accessibility of short linear peptide motifs consisting of 3-7 amino acid stretches in IDRs allow peptide-protein interactions to occur (Tompa et al., 2014). Upon phosphorylation, tyrosine residues often function as docking sites for recruitment of adaptor proteins containing SH2 or PTB domains (Pawson and Scott, 1997). This recruitment of adaptor proteins to regulated phosphotyrosine sites wire an assembly of protein complexes. The dynamically recruited protein complexes exert their individual functions by extending the signaling wave with yet another layer. It remains a formidable challenge to assign molecular function, such as recruitment of protein complexes, to phosphorylated residues. An essential step in understanding the complex wiring of phosphotyrosine signaling is to develop methods to detect which phosphotyrosine sites are regulated upon a stimulus, as well as detecting which protein complexes are recruited to the regulated sites. Here we have developed a proteomics based approach to assess i) which tyrosine residues are phosphorylated *in-vivo* upon tyrosine kinase activation, and ii) which interacting

proteins are recruited as readers of the regulated phosphotyrosine sites (Fig. 1a). We present a first depiction of the EGF-dependent phosphotyrosine signaling network *in-vivo* in lung tissue and unravel the protein complexes assembled at EGF-dependent phosphotyrosine sites. We demonstrate how the lung cancer mutation EGFR P1019L induces a switch in adaptor protein interactions at position pY1016, which leads to sustained activation of downstream kinase signaling pathways ultimately resulting in enhanced cell migration and invasiveness. Cancer mutations in vicinity of phosphotyrosine sites are recurrent, and as demonstrated for EGFR P1019L their detrimental effect can be caused by introduction of molecular switches that alter protein signaling networks. To enable rapid and high-throughput analysis of such mutations, and of protein interactions at phosphotyrosine sites in general, we present a strategy based on cutting-edge mass spectrometry instrumentation and data independent acquisition. For several cancer mutations in vicinity of phosphotyrosine sites we show that their functional consequence is introduction of switches in the molecular wiring of recruited protein complexes. Our strategy presents an approach for multilayered tissue-based investigation of phosphotyrosine signaling in general, and for functional assignment of cancer mutations in proximity to phosphotyrosine sites specifically, providing a molecular explanation of their mechanism of action.

RESULTS

Phosphotyrosine signaling activated in lung tissue by EGF stimulation

Amplified EGFR signaling is a common causal oncogenic mechanism in non-small cell lung cancer (Janne et al., 2005; Rikova et al., 2007), but the multi-layered EGF-dependent signaling response in lungs is poorly understood. To evaluate the signaling layer orchestrated by tyrosine phosphorylation *in-vivo*, we designed animal experiments, where two groups of rats received intravenous saline or EGF injections for five minutes, respectively (Fig. 1a). As expected (Zhang et al., 2005), a strong phosphorylation response was elicited and validated by immunoblotting (Fig. 1b). From each set of lungs three samples were prepared: (1) a sample for proteome measurements, (2) titanium dioxide (TiO_2)-enriched phosphopeptides for serine/threonine phosphoproteomics and (3) antibody-enriched phosphotyrosine peptides for tyrosine phosphoproteomics (Rush et al., 2005). All samples were analyzed by high-resolution LC-MS/MS on a Q-Exactive Orbitrap mass spectrometer by single-shot measurements. Raw LC-MS/MS files were processed with MaxQuant (www.maxquant.org) and led to identification of >45,000 unique peptides in the proteome samples (Supplementary Table S1), >13,000 unique serine/threonine phosphorylated peptides (Supplementary Table S2) and more than a thousand quantified unique tyrosine phosphorylated peptides (Supplementary Table S3)

(Sup. Fig. S1a). In previous *in-vivo* phosphoproteomics studies, phosphotyrosines made up approximately 1% of phosphorylation sites (Humphrey et al., 2015; Huttlin et al., 2010; Lundby et al., 2012). Here, using specific antibodies the number of identified phosphotyrosine sites increased 20-fold representing 13% of the total phosphoproteome (Sup. Fig. S1b). Motif analysis of the amino acid sequences proximal to regulated phosphosites confirm that the signaling response is mainly mediated by ERK (PxS/TP sequence motif) and basophilic kinases such as AKT (RxxS sequence motif) (Sup. Fig. S1c).

To evaluate the importance of investigating the signaling response in lung tissue rather than in a cell line, we compared our EGF-dependent lung phosphotyrosine response with corresponding measurements from A549 human non-small cell lung cancer cells (Supplementary Table S4) and HeLa human cervix carcinoma cells (Francavilla et al., 2016b). The tyrosine phosphoproteome of A549 cells resembled that of HeLa cells more than that of orthologous sites in rat lung tissue (Sup. Fig. S1d). For several regulated sites that were conserved between lung tissue and A549 cells, we confirmed their EGF-dependence by immunoblotting (Sup. Fig. S1e-h). Yet, less than half of the proteins with regulated tyrosine phosphorylation sites in rat lung tissue were EGF-dependent in A549 cells. These observations underscore the importance of performing tissue-specific investigations of signaling pathways to ensure physiological relevance of the findings.

All three sample types (proteomes, serine/threonine phosphoproteomes and tyrosine phosphoproteomes) were reproducible between biological replicates with Pearson correlation coefficients $0.89<R<0.99$ for quantile-based normalized peptide MS signal intensities (Sup. Fig. S2a). Based on label-free quantification we identified more than eighty tyrosine residues with increased phosphorylation stoichiometry in response to 5 min EGF stimulation (t-test $p < 0.05$ and fold change > 4, Fig. 1c). The cut-offs applied in this analysis are stringent and correspond to an estimated false discovery rate of 0.002 (Sup. Fig. S2b). Gene Ontology enrichment analysis on the set of proteins harboring the 88 regulated phosphotyrosines couples the response to an overrepresentation of tyrosine kinases and adaptor protein activity, which supports a comprehensive activation of the EGF-dependent signaling pathway (Sup. Fig. S2c). Proteins with regulated phosphotyrosine sites are distributed across the entire protein abundance spectrum, despite EGFR being among the least abundant proteins detected in our lung proteomes (Sup. Fig. S2d). All EGF-dependent phosphotyrosine sites identified in lung tissue are provided in Table 1, where proteins are grouped by their biological function. That is, an *in-vivo* quantitative phosphoproteomics approach allowed us to identify 88 tyrosine residues with increased phosphorylation stoichiometry in lungs in response to 5 minutes EGF stimulation.

Peptide-based interaction screen of regulated phosphotyrosines

Regulated phosphotyrosines often lead to formation of dynamical protein complexes (Blagoev et al., 2003) by presenting docking sites for adaptor proteins containing for instance SH2 domains (Bae et al., 2009). Cellular specificity of phosphotyrosine-binding proteins is achieved through the amino acid sequences surrounding the phosphotyrosine site, the higher order structures and co-expression patterns (Schlessinger and Lemmon, 2003; Songyang et al., 1993). Recruitment of such protein binders represents the second wave in the signaling response. For a few phosphotyrosines, protein binders have been identified in lysates from cell lines by phosphopeptide pulldown strategies with quantitative mass spectrometry as read-out (Boersema et al., 2010; Schulze et al., 2005) or by two-hybrid assays (Petschnigg et al., 2014). To identify protein complex formations assembled at EGF-dependent phosphotyrosines directly in lung tissue, we synthesized peptides corresponding to the 88 regulated phosphotyrosine sites and the six flanking amino acids on either side of each phosphotyrosine site and used these as baits in pulldown experiments in lysates from lungs. Each peptide was synthesized with a biotin tag and a hydrophilic linker. Phosphopeptide pulldowns were performed in a 96-well format based on a strategy (Eberl et al., 2013) that we modified for large-scale interaction proteomics in tissue samples. Pulldowns were performed in triplicates from three sets of lung tissue lysates resulting in a total of 264 peptide pulldowns (Fig. 1d). Interacting proteins were digested on-beads and the 264 pulldown samples were subsequently analyzed by high-resolution LC-MS/MS with 1h gradients.

Web platform with analysis tool for quantitative interaction proteomics data

In pulldown experiments, it is essential to discriminate between specific and unspecific binding partners. For phosphotyrosine sites this is particularly challenging as unspecific interactions occur even in the low micro-molar range (Sharma et al., 2009). To overcome this challenge, we developed an analytical framework that takes into consideration quantitative information from multiple pulldown experiments as well as the abundance of each protein in the proteome. The principle of the strategy is explained here and illustrated in Figure 2a, but extensive details on the analytical strategy and computational framework are provided in the Supplementary Material. To enable others to apply this analysis strategy, we present a web platform, where analyses of quantitative interaction proteomics datasets can be performed.

The analytical framework first normalizes log-transformed protein intensities by median subtraction and calculates the protein variance for each set of pulldowns as a function of the protein intensity (Fig. 2a). We exploit the strong interdependency between MS-based protein intensities and variance to determine a sigmoid function that explains the relationship between the two parameters. A key feature of our analytical

tool is that the statistical test we perform to identify specific interactors is based on these estimated variances. Phosphotyrosine-interacting proteins are frequently missed in classical control experiments due to their low abundance, which hampers accurate quantitation. We solved this missing value problem by imputing values according to a hierarchy, in which median protein intensity measurements across all other pulldown experiments are prioritized over empty bead control experiments, which is prioritized over scaled total proteome measurements. The benefit of this imputation scheme comes from the empirical observation that the abundance of an unspecific background binder in a pulldown strongly correlates with its expression level in the proteome (see Supplementary Material for details). For each protein in a pulldown, a modified Welch's t-test termed significance C is performed using the estimated variance, and the data are represented in a volcano plot. A combined evaluation of the protein p-value and intensity ratio measurement determines whether the protein is classified as a specific interactor or not.

Benchmarking our analysis strategy against the classical approach of phosphopeptide pulldown versus matched unmodified peptide pulldown demonstrates an improved ability to discriminate for unspecific binding of SH2 domain containing proteins to phosphotyrosine sites (Sup. Fig. S3). For example, our strategy reduced the number of significant interactors of EGFR pY1172 from several hundred to only six, among these were the two known interactors of EGFR pY1172, SHC1 and GRB2 (Schulze et al., 2005). Our approach effectively filters the data, highlighting fewer but likely more relevant interactors. Importantly, our strategy has the advantage that it eliminates the necessity of pulldowns using unmodified peptides by exploiting the quantitative information from multiple phosphopeptide pulldowns, thereby reducing the number of peptides that need to be synthesized and tested to half.

To identify which particular protein interactor is the most abundant in each pulldown, and thereby the likely direct interactor, we calculated the fractional stoichiometry of the significant proteins based on their iBAQ intensities (Schwanhausser et al., 2011) from each pulldown (Hein et al., 2015). The fractional stoichiometry can be used to prioritize direct binders among the significant interactors and is provided on a normalized scale next to each volcano plot (Fig. 2a).

Protein complexes recruited to EGF-dependent phosphotyrosine sites in lung tissue

We applied the analytical framework presented above to analyze all of our 264 phosphotyrosine peptide pulldown experiments from lung tissues. This enabled us to determine specific protein complexes recruited to the 88 EGF-regulated phosphotyrosine sites in lung tissue (Supplementary table S5). Volcano plots for all 88 peptide pulldowns performed are provided in Supplementary Figure S4 as well as on the web platform.

Volcano plots for two representative examples including validation experiments by co-immunoprecipitation are shown in Sup. Fig. S5. For each of the recruited protein complexes we calculated the relative stoichiometry of the protein interactors, thereby allowing us to pinpoint the most likely direct interactor. The relative stoichiometries are provided next to each volcano plot. The recruited interaction partners – direct as well as indirect – cover hundreds of proteins (Fig. 2b), which underscores the complexity of the EGF-dependent signaling response beyond regulation of phosphorylation site stoichiometry. Some of the complexes identified are highlighted in figure 2b. We integrated our dataset with information in the most comprehensive protein-protein interaction databases (BioGRID (Stark et al., 2006) and InWeb (Li et al., 2017)) and phosphoprotein resources (PhosphoSitePlus (Hornbeck et al., 2012) and Uniprot (UniProt, 2015)). Only 27 of the 88 sites were known to be EGF-dependent and of the 503 protein-protein interactions we identified, 72 were previously reported but only 16 were known to depend on the specific phosphotyrosine site (Supplementary Table S6). The quantitative dataset of *in-vivo* phosphotyrosine site interactions represented here is the largest resource of EGF-dependent and phosphorylation site-specific protein-protein interactions.

Core protein complexes in EGFR signaling

The EGF response is initiated by dimerization and auto-phosphorylation of EGFR at numerous tyrosine sites, which both lead to an amplification of its kinase activity and creates docking sites for adaptor proteins that wire the assembly of dynamic protein complexes. To visualize the two layers of regulation on EGFR, we depicted all regulated tyrosine sites together with the proteins that we identified as interacting with them (Fig. 2c). Our data show that protein interactors of EGFR are site-specific. For instance, a CBL complex interacts with EGFR pY1069, a CRK/VAV complex interacts with EGFR pY1016, and a SHIP complex interacts with EGFR pY998. Seven EGFR interaction partners were independently confirmed by co-immunoprecipitations to be EGF-dependent (Fig. 2d). We observe that a GRB2/GRAP protein complex interacts with multiple phosphotyrosine sites on the receptor, such as pY1091, pY1109, pY1137 and pY1172. GRB2/GRAP is the key adaptor protein activating the RAS-RAF-MEK-ERK pathway, which controls cell proliferation downstream of EGFR (Francavilla et al., 2016b; Kolch, 2005). The redundancy of GRB2/GRAP complex recruitment to multiple EGFR sites represents a means by which the cell can ensure a robust ERK response and indicates that this signaling axis is the most important part of the EGF response.

To extract information on other key players in the EGF response, we clustered protein interactors identified in at least three different peptide pulldowns (Sup. Fig. S6a). We identified 11 key protein complexes, which given their frequency of co-occurrence are deemed central for the EGF response in lung tissue. For the 74

proteins in these key complexes, we depicted their protein interaction network (Fig. 2e). For several of the interactions shown, we confirmed our MS-based stoichiometries by isothermal calorimetric measurements (Sup. Fig. S6b), as well as the EGF-dependence of the interactions by co-immunoprecipitations (Sup. Fig. S6c-e). From the network representation in Figure 2e, it is evident that the majority of core members of the EGFR signaling network harbor interaction domains, in particular SH2/PTB and SH3, or are kinases or phosphatases. Many of the proteins in the network also contain EGF-regulated phosphorylation sites, highlighting their role in integrating the two signaling layers in the EGF response. This combined representation of regulated tyrosine phosphorylation sites and their interactors highlights the complex core network of signaling proteins involved in the early EGF-dependent response in lung tissue.

Sequence motifs for recruitment of core protein complexes to phosphotyrosine sites

The amino acids flanking phosphotyrosine sites are important determinants of the affinity for individual SH2 domain containing proteins (Kavanaugh et al., 1995; Schlessinger and Lemmon, 2003). To identify sequence preferences for each of the 11 key protein complexes represented in figure 2e, we analyzed the amino acid sequences surrounding the regulated phosphotyrosine sites they interact with and identified overrepresented sequence motifs using IceLogo (Colaert et al., 2009). We found specific sequence motifs for different phosphotyrosine-interacting protein complexes: GRB2/GRAP based complexes preferentially bind to the amino acid sequence motif pYxN, whereas PTPN6/PTPN11 based complexes preferentially binds to pYVxL (Sup. Fig. S6f). These findings underscore the importance of the amino acid sequence in vicinity of phosphotyrosine sites for recruitment of specific protein complexes. Furthermore, the sequence motifs suggest that mutations near a regulated phosphotyrosine site may impact the protein complex recruited to the site.

Oncogenic mutation near phosphotyrosine site causes molecular switch in recruited protein complex

Disease-associated missense mutations can affect short linear motifs in IDRs of proteins and thereby interfere with their functions by disrupting or changing protein interactions (Vacic et al., 2012). Peptide-based proteomic pulldown screens have recently been employed to investigate the impact of mutations in IDRs on protein-protein interactions (Meyer et al., 2018). EGFR P1019L is reported as a lung cancer mutation (Bamford et al., 2004), and it affects a highly conserved amino acid in an IDR of EGFR (Fig. 3a). The P1019L mutation is located three residues downstream from a regulated phosphotyrosine site that interacts with a

CRK complex also harboring VAV and RASA1 (Sup Fig. S4). The CRK complex preferentially binds to a pYxxP sequence (Sup. Fig. S6f), as also reported previously (Miller et al., 2008). Given this knowledge, we hypothesized that the EGFR P1019L patient mutation may hinder the protein complex formation between EGFR and CRK upon activation. We tested the hypothesis by synthesizing a phosphorylated peptide carrying the patient mutation and compared the protein interactors between this peptide to that of the corresponding wildtype peptide. The cancer mutation completely changes the recruited protein complex. The cancer mutation introduce a molecular switch that abolish the CRK interaction and instead leads to binding of a SH2B1/SHIP2 complex (Fig. 3b, Supplementary Table S7). This is consistent with our sequence motif analysis that revealed an amino acid sequence preference of VAV3, RASA1, CRK and CRKL for binding to pYxxP, and a preference of SH2B1, ZAP70, SYK and SHIP for pYxxL (Fig. 3c). To confirm the specific interactions of the patient mutant versus wildtype, we produced and purified SH2 domains of the identified binding partners and performed isothermal titration calorimetric experiments to measure their affinities for the wildtype and the patient-mimicking mutant (Sup. Fig. S7-S8, Supplementary Table S8). All interactions were confirmed to require tyrosine phosphorylation, and high affinities for CRK and CRKL were established for wildtype EGFR pY1016, whereas the affinities of patient mutant P1019L were switched towards SH2B1 and SYK (Fig. 3d-e). That is, the lung cancer mutation P1019L introduce a molecular switch that alters the outcome of phosphorylation at EGFR Y1016.

Oncogenic properties of lung cancer mutation EGFR P1019L

To evaluate the functional consequence of the molecular switch caused by the EGFR P1019L cancer mutation, we created a CRISPR/Cas9 EGFR knockout version of the lung cancer A549 cell line (Sup. Fig. S9) and reintroduced GFP-tagged EGFR in either wildtype or P1019L versions (Sup. Fig. S10a-c). First, to investigate differential interaction partners of the wildtype and P1019L mutant receptor we performed quantitative interaction proteomics of the receptors as a function of EGF stimulation. This experiment confirmed the specific interaction of the wildtype receptor and VAV observed in the peptide pulldown. Our results further demonstrated that GRB2 and SHC1 interacted stronger with the mutant receptor (Sup. Fig. S10d-e). By co-immunoprecipitation experiments, we confirmed the increased interaction with GRB2 and SHC1 for the mutant receptor as well as the specific interaction between SHIP2 and the mutant receptor identified in the peptide pulldown experiment (Fig. 3f). The lipid phosphatase SHIP2 has been reported to enhance EGFR signaling in breast cancer (Prasad, 2009), and we therefore hypothesized that binding of SHIP2 to the mutant receptor may be responsible for the increased interactions with GRB2 and SHC1. We confirmed a SHIP2-mediated interaction between SHC1 and EGFR P1019L by co-immunoprecipitation experiments showing decreased interaction between SHC1 and the mutant receptor in SHIP2 depleted cells (Fig. 3g, Sup. Fig. S10f).

The SHC1-GRB2 complex is the master regulator of ERK activity in EGFR signaling (Bisson et al., 2011; van Biesen et al., 1995). Accordingly, the mutant receptor altered ERK signaling dynamics by changing it from a transient to a sustained response (Fig. 3h, Sup. Fig. S10g). Furthermore, in contrast to the wildtype receptor, activation of the mutant receptor also resulted in sustained AKT phosphorylation dynamics (Sup. Fig. S10h-i). We verified that the sustained ERK signaling dynamics is mediated by the EGFR P1019L interaction with SHIP2, as the dynamics was switched back to a transient response upon SHIP2 depletion (Fig. 3h). We have previously shown that sustained ERK and AKT signaling downstream of EGFR can lead to increased cell migration and proliferation (Francavilla et al., 2016a). To evaluate if the mutant receptor encodes oncogenic properties, we compared the migratory potential of lung cancer cells expressing mutant versus wildtype receptors. Indeed, expression of EGFR P1019L resulted in significantly increased cell migration (Fig. 3i). Similarly, cells expressing mutant receptor also had increased cell proliferation rate (Sup. Fig. S10j). We also found the cell migration properties of EGFR P1019L expressing cells to be SHIP2-dependent (Sup. Fig. S10k). The SHIP2 dependency is likely mediated by its adaptor protein function as inhibition of its catalytic activity does not significantly alter migration rates (Sup. Fig. S10l). To further validate the SHIP2 dependency of the oncogenic properties of EGFR P1019L, we performed a matrigel-based cell invasion assay in EGFR P1019L expressing cells after SHIP2 knock-down. The EGF-dependent invasion potential was significantly impaired in SHIP2-knockdown cells (Fig. 3j). Finally, to verify the EGFR P1019L-enhanced invasion *in-vivo*, we made use of a zebrafish xenograft invasion model (Rouhi et al., 2010). Briefly, wildtype EGFR or EGFR-P1019L expressing cells were DiI-labelled and injected into the perivitelline space of eGFP-transgenic zebrafish embryos 48h post-fertilization in the presence of EGF ligand. Local tumor invasion and dissemination to the tail was quantified 48h later (Fig. 3k). A significantly greater number of cells expressing EGFR P1019L compared to wildtype EGFR invaded locally and disseminated to the tail (Fig. 3l). Accordingly, the *in*-vivo as well as the cell-based experiments validate and confirm the molecular switch introduced by the cancer mutation in vicinity of a regulated EGFR phosphotyrosine site. The mutation changes the recruited interaction complex from a CRK complex to a SHIP2 complex. This change leads to increased binding of SHC1, which ultimately results in a sustained activation of AKT and ERK, which endows the receptor with oncogenic properties (Fig. 3m).

MS method for functional assignment of phosphotyrosines at large scale

The ability to identify protein complexes recruited to regulated phosphotyrosines is important for depicting a more complete picture of the complex signaling networks downstream of tyrosine kinases in general. In light of the oncogenic molecular switch identified, the question arises whether this is furthermore important to investigate as a general mechanism of action in disease contexts. The methodological approach we have

developed represents a general strategy to functionally annotate phosphotyrosine sites. To enable larger scale studies, we optimized the MS-based workflow for high-throughput analyses. In doing so, we used state-of-the-art data-independent acquisition (DIA) in combination with short LC gradients on an Evosep One system. The Evosep system is based on a fundamentally new LC concept (Bache et al., 2018) that significantly increases protein coverage when analyzed on a Q Exactive HF-X mass spectrometer with fast MS/MS scanning capabilities (Kelstrup et al., 2018). These advancements provide the possibility for a scalable methodology, where 60 pulldown experiments can be analyzed in just one day of LC-MS instrument time. To demonstrate the scalability of this method for large-scale DIA-based interaction analyses, we synthesized more than three hundred biotin-tagged peptides, which were selected based on a literature analysis prioritizing known phosphotyrosine sites in signaling proteins such as receptor tyrosine kinases and adaptor proteins. Peptide pulldowns using these 300 peptides were performed in four biological replicates of tissue lysates (Fig. 4a). That is, more than 1,200 peptide based pulldown experiments were performed. These experiments were performed in liver tissue to investigate for potential tissue-specific differences in recruited protein complexes. Bound proteins were digested with trypsin on-bead and resulting peptides analyzed with 21-minute LC-MS gradients. To generate project specific spectral libraries for the DIA matching and identification, we performed deep proteome profiling of the tissue lysates by massive offline high pH chromatography and analyzed each fraction with the same LC-MS setup (Bekker-Jensen et al., 2017). Proteins identified in each individual pulldown were analyzed against all other pulldowns by label-free quantitation to identify significantly-enriched interacting proteins. This resulted in 918 significant interactors for 225 different phosphopeptide baits (Fig. 4b, Supplementary Table S9). As expected, based on the lung pulldown experiments, several SH2 domain containing adaptor proteins were also identified as interactors of multiple phosphopeptides baits in the liver lysates. For example, Ptpn6, Vav2, Grb7 and SH2B1 were identified as interactors in more than fifty different pulldowns. Knowledge of the protein interactors of phosphotyrosine sites in tissues is important for identifying their tissue-specific roles and phenotypes. However, most phosphosite-protein interactions reported have no tissue-specific context. For a subset of baits, we evaluated their interaction partners in both lung and liver tissue lysates. From this data, we evaluated potential tissue-specific compositions of the interacting protein complexes. There are indeed differences in the protein complexes recruited to a particular phosphotyrosine site across tissues. For instance, for EGFR pY1109 we identified a Grb2/Grap/Grap2 complex as the strongest interaction in lung tissue (Fig. 4c), whereas the same site preferentially bound a Grb7/Grb14 complex in liver tissue (Fig. 4d). For this site, the tissue-specific differences in interactors can be explained by their differential proteome abundance. We find that Grb7 is highly expressed in liver tissue but hardly detectable in lung and, conversely, that Grap and Grap2 are expressed in lung tissue but not detected in liver proteomes. This tissue-specific difference in interactors

highlights the importance of performing interaction screens in the appropriate tissue and context-specific lysates (Fig. 4e).

Molecular switches in vicinity of phosphotyrosine sites as mechanism of action of cancer mutations

Identification of the molecular switch introduced by the oncogenic mutation EGFR P1019L led us to ask, whether this could represent a general mechanism of action. We thus set out to address, if molecular switches in protein complex assembly at phopshotyrosine sites could represent a mechanism of action for oncogenic mutations. To probe for additional cancer-specific molecular switches, we included a set of phosphopeptides covering 12 known cancer mutations (Bamford et al., 2004) in the large-scale peptide pulldown experiment. For these experiments we compared the mutated phosphopeptide pulldowns to their corresponding wildtype phosphopeptide experiment. For eight of the twelve cancer mutations tested, we identified switches in their recruited protein complexes (Supplementary Table S10). For example, Gab1 phosphorylated at Y317 specifically pulls down a complex consisting of CrkL and Rasa1, whereas the cancer mutation Gab1 P320S completely abolishes this interaction (Fig. 4f). The importance of proline in the +3 position to the phosphorylated site for the interaction with CrkL was validated by ITC experiments by which we established that the interaction between the wildtype Gab1 pY317 peptide and the SH2 domain of CrkL was 650 nM, whereas no binding was observed with the P320S mutated phosphopeptide (Fig. 4g). Likewise, analyzing the protein interactions of phosphorylated Crk Y136 identified the lipid phosphatase Sacm1l as the main binder, whereas the cancer mutation Crk A134V, two amino acids upstream of Y136, changed the specific interactions to the tyrosine phosphatases, Ptpn6 (SHP-1) and Ptpn11 (SHP-2). This molecular switch was found in both lung and liver tissues. We further identified molecular switches introduced by five EGFR cancer mutations: EGFR N1140S, EGFR P1170S, EGFR P1019S, EGFR N1112S and EGFR N1094Y (Supplementary Table S10). For the cancer mutation EGFR N1094Y, the molecular switch has different outcomes depending on whether or not the tyrosine residue is phosphorylated. That is, cancer mutations near phosphotyrosine sites can indeed introduce molecular switches in recruited signaling complexes.

DISCUSSION

We describe and validate a general proteomics strategy to resolve multiple layers of phosphotyrosine signaling *in-vivo*, including both quantitative analysis of phosphotyrosine sites and systematic examination of the dynamically regulated protein complexes they recruit. The approach thereby enables assignment of function to phosphotyrosine sites on a large scale. Phosphotyrosine interaction screens have successfully been performed using a yeast two-hybrid approach identifying hundreds of phosphotyrosine-dependent

protein complexes (Grossmann et al., 2015), but this approach does not reveal which tyrosine site is responsible for the interaction. Here we based our approach on the principle of phosphopeptide pulldowns in combination with quantitative mass spectrometry (Hanke and Mann, 2009; Schulze et al., 2005) combined with the ability to perform such experiments at a large scale. To identify the protein complexes recruited to a particular regulated phosphotyrosine site, we developed an analytical framework that handles the inherent challenge of a general high affinity of SH2 domain-containing proteins towards phosphotyrosine sites. This is achieved by integration of affinity information from multiple phosphopeptide pulldowns quantitatively analyzing each phosphopeptide pulldown against multiple other phosphopeptide pulldowns and by implementing information from deep proteome measurements. We have made our analytical framework accessible to other researchers via a web-based tool http://pulldown.jensenlab.org.

Our approach represents several important advances. Firstly, we performed the peptide pulldowns in the relevant tissue lysate to establish the molecular interactions between proteins in the appropriate biological setting. This is important to ensure coverage of tissue-specific findings (Lundby et al., 2014). Our comparisons between interaction partners identified in lung and liver tissues further underscores this. Secondly, to the best of our knowledge, our dataset represents the first analysis of the composition of an entire phosphotyrosine protein network activated by a distinct growth factor *in-vivo*. Specifically, it is the first comprehensive and quantitative map of EGFR signaling in lung tissue and it identifies key components of the dynamic protein complexes formed. We demonstrate that 5 min of EGF stimulation *in-vivo* leads to assembly of 11 key signaling protein complexes encompassing 74 different proteins. The nodal proteins in each of the 11 complexes contain SH2 domains and their individual amino acid sequence preferences surrounding the phosphotyrosine sites dictate their binding affinities. This insight explains why single amino acid mutations in the vicinity of phosphorylated tyrosine sites can switch the interacting adaptor protein complexes and thereby clarify the molecular mechanism of oncogenic driver mutations. Such a mechanism has previously been shown for a cancer mutation in the fibroblast growth factor receptor FGFR4 (Ulaganathan et al., 2015), and it represents a molecular mechanism of how cancer mutations can corrupt dynamic transmission properties in signaling pathways, changing cell decisions in a pathological manner (Bugaj et al., 2018). Here, we show that the lung cancer mutation EGFR P1019L introduces a switch in the signaling complex recruited to the receptor at pY1016, changing the signaling response in lung tissue. The phosphoinositol phosphatase SHIP2 is a critical regulator of signaling pathways as it negatively controls phosphatidylinositol-3,4,5-trisphosphate levels (Erneux et al., 2011). In breast cancer cells SHIP2 positively affects EGF-regulated AKT activation and induces EGF-dependent cell proliferation and tumor growth (Prasad, 2009). We found that SHIP2 enhances SHC1 interaction with the mutant receptor in an EGF-dependent manner, which leads to sustained ERK and AKT signaling ultimately increasing cell migration and invasiveness. This can explain the

cellular outcome of the cancer-driving mutation EGFR P1019L. That is, we identify a specific adaptor protein, SHIP2, mediating the switch in signaling cascades responsible for the oncogenic properties of the mutant receptor. This suggests SHIP2 as a potential target for anti-cancer therapy in EGFR mutated lung cancers. We confirmed the oncogenic properties of the mutant *in-vivo* by using a zebrafish xenograft model. From our large-scale efforts beyond EGFR mediated signaling, we identified and characterized eight other cancer mutations that also introduce molecular switches in the recruited protein complex formation upon activation of near-by phosphotyrosine sites. For Gab1, we found that an oncogenic mutation at position 320 abolishes recruitment of CrkL to pY317 upon activation, and for Crk a cancer mutation at position 136 introduces recruitment of phosphatases Ptpn6 and Ptpn11 upon activation of pY136. These oncogenic mutations in vicinity of phosphotyrosine sites thereby introduce molecular switches that alter the ultimate outcome of the signaling response. Delineating such tissue-specific signaling networks is important as they have the potential to offer novel candidate targets for developing alternative therapeutic interventions. We demonstrate the scalability of our approach and present it as a general strategy to evaluate functional effects of cancer mutations near phosphotyrosine sites, as well as a general strategy to investigate regulated protein–protein interactions in a tissue-specific manner.

ACKNOWLEDGEMENTS

The authors thank all lab members for fruitful discussion, Simon Kamenov for assistance with part of the pulldown experiments and Rossana Foti for access to microscopy. Work at The Novo Nordisk Foundation Center for Protein Research (CPR) is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001). The proteomics technology developments applied was part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement: MSmed-686547, EPIC-XS-823839 and ERC synergy grant 810057-HighResCells; Manufacturer Vilhelm Pedersen & wife's Memorial Fund Award (J.V.O.); Sapere Aude and YDUN Grant from The Danish Council for Independent Research (A.L., grant number DFF-4092-00045 and grant number DFF-6110-00166) and The Novo Nordisk Foundation (A.L., grant number NNF15OC0017586). SPG is funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 798716.

AUTHOR CONTRIBUTIONS

AS performed animal experiments shown in Fig. 1A. GF generated mutant cell lines and performed experiments shown in Fig. 2D, 3F-J, Sup. Fig. S1G-H, S5C-D, S6C-E, S9-10. IP generated the EGFR knockout cells and P1019L mutant. BM performed ITC experiments shown in Fig. 3D-E, Fig.4G and Sup. Fig. S7-8, supervised by GM. CF performed western blots shown in Fig. 1B. DBBJ performed all peptide pulldown experiments shown in Sup. Fig. S4, KBE with SRM and DBBJ performed peptide pulldown experiments using DIA approach. CDK provided input for optimization of peptide pulldown experiments shown in Sup. Fig. S4. SPG performed zebrafish experiments in Fig. 3K-L under supervision of MK. JCR and LJJ developed statistical framework and web-interface for analysis of pulldown data. AL performed experiments and analyzed data shown in all remaining figures. AL and JVO conceived the project, designed the experiments, analyzed all MS data, critically evaluated results and wrote the manuscript. All authors provided input for the manuscript.

Table 1

Protein kinases and regulators			Protein phosphatases and regulators			Actin and cytoskelton organization			Membrane ruffles		
EGFR	Y869	2.1*	PRDX1	Y193	2	EPB41L2	Y606	2.9	CLASP2	Y1360	2.5
EGFR	Y998	5.2*	PTPN11	Y546	2.6	DLG2	Y726	2.2	FGD5	Y1450	2.4
EGFR	Y1016	3.5**	PTPRC	Y652	4.1	DLG3	Y368	2.8	FGD5	Y898	2
EGFR	Y1069	6.1*	SLAMF6	Y322	7.5	LMO7	Y133	3.4	FRMD4B	Y871	2.9
EGFR	Y1092	4.2*	**SH3/SH2 adaptor proteins**			SPTAN1	Y942	4.3	**Replication and transciption regulators**		
EGFR	Y1110	6.5	CRK	Y108	3.9	WDR1	Y72	2.2	ERH	Y92	4.2
EGFR	Y1138	6.7	CRK	Y136	3.6	WDR6	Y332	2.6	HNRNPUL1	Y510	2.5
EGFR	Y1172	7.3	CRKL	Y127	2	**Focal adhesion proteins**			PCNA	Y211	3.2
EGFR	Y1197	2.6	GAB1	Y317	2.5	EVL	Y38	2.5	PTBP1	Y126	2
EPHA1	Y782	3.1	GAB1	Y659	4.4	SDCBP	Y92	2.5	**Calcium binding proteins**		
FGR	Y197	2.5	GAB2	Y603	4.6	TNS2	Y652	2.6	CALM1	Y100	4.6
GAREM	Y453	4.3	HCLS1	Y323	3	TNS2	Y668	2.7	CALU	Y47	3.2
LYN	Y244	2.4	PLCG1	Y1253	3.6	**Metabolic processes**			**Endocytic proteins**		
MAPK1	Y185	3	SH2B1	Y55	3.2	ACTB	Y169	2.1	CPNE8	Y423	2.1
MAPK13	Y182	2.6	SHANK3	Y197	2.1	BPGM	Y92	3.6	EHD4	Y451	5.3
MAPK8	Y185	2.4	SHC1	Y423	6.1	CBR1	Y194	3.6	RUFY1	Y314	4.8
PIK3AP1	Y694	2.9	STAM	Y199	5.5	DBI	Y29	4.3	TRAPPC9	Y573	4.8
PIK3R1	Y431	2.9	VAV1	Y826	3.1	EIF2B5	Y578	3.1	**Apoptotic and proteolytic proteins**		
PIK3R1	Y580	3.5	**Small GTPase mediated signaling**			FBP2	Y216	2	EFHD2	Y83	2.1
PRPF4B	Y849	2.2	ARHGDIA	Y133	2.8	PABPC1	Y54	5.5	GSTP1	Y109	2.7
PTK2	Y928	2.2	ARHGEF6	Y715	2.2	PABPC1	Y56	5.5	TJP2	Y1093	2.2
SGK269	Y631	5.6	GPSM1	Y150	3.7	PGAM2	Y96	2.9	USP14	Y285	3.2
TAOK1	Y43	2	RAB10	Y6	2.4	PGM1	Y353	3.1			
TNK2	Y518	3.3	TBC1D9B	Y854	2.7	RPSA	Y139	3.5			
TNK2	Y859	3.5									
ZAP70	Y392	3.3									

Table 1. *List of all proteins with significantly regulated phosphotyrosine sites upon EGF stimulation in lung tissue*. The gene name of the proteins with significantly regulated phosphotyrosines sites are listed along with the site information for the regulated phosphotyrosine site. The label-free $\log_2$-transformed intensity ratio measurements between EGF stimulated and control rats are given. The proteins are grouped according to their known biological functions. The amino acid position of tyrosine phosphorylation sites for EGFR are indicated according to the human amino acid sequence. In five cases ratios measured in non-small lung cancer A549 cells (*) or HeLa cells (**) are provided due to lack of coverage in the rat lung tissue.

Fig. 1

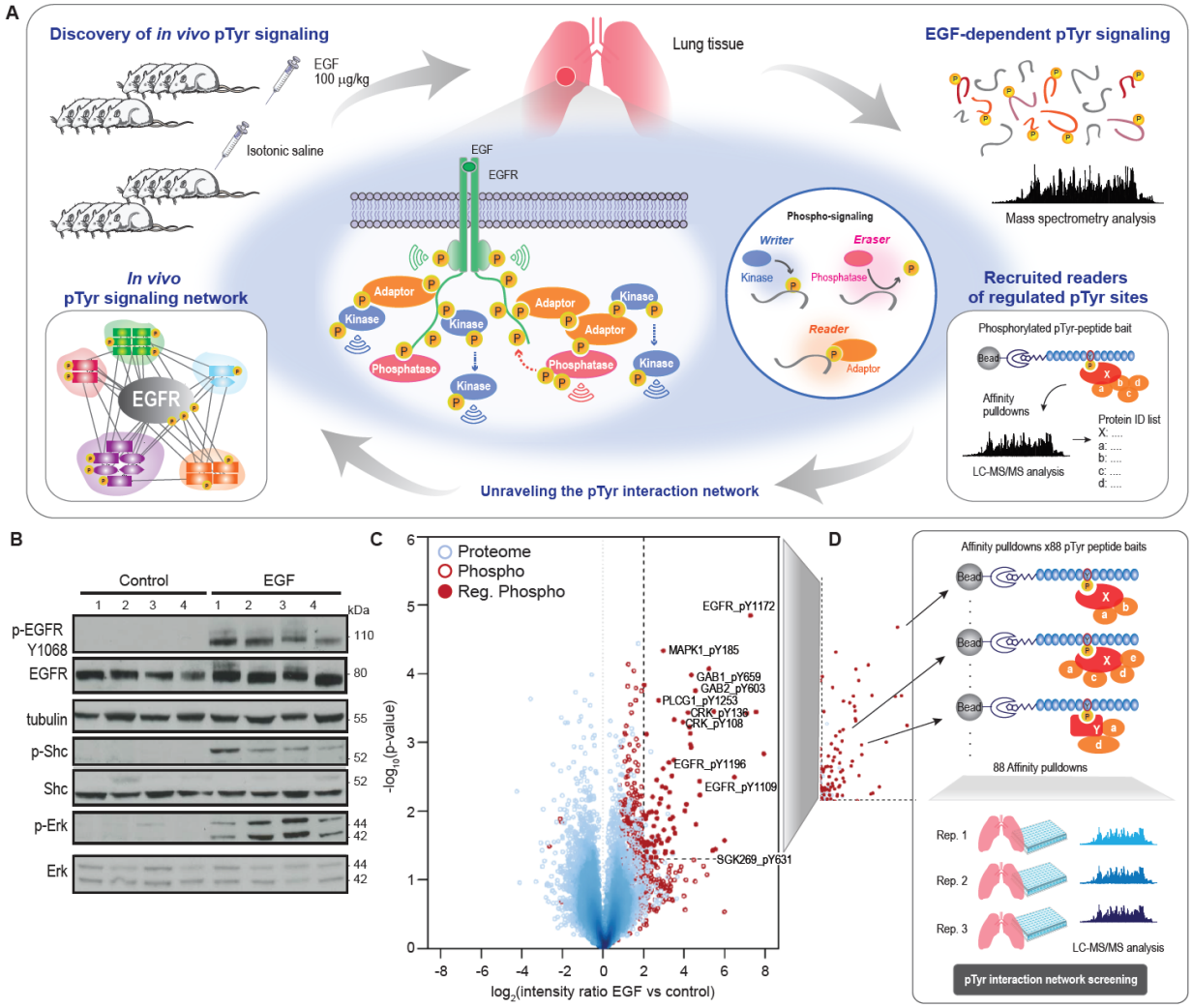

Figure 1: *Quantitative proteomics of EGF-dependent signaling in rat lung tissue*. (A) When a phosphotyrosine signaling response is elicited it results in i) changed phosphorylation site stoichiometry induced by kinases, ii) binding of adaptor proteins to the phopshotyrosine sites and iii) removal of phosphate groups by phosphatases. Here, we performed a MS-based quantitative phosphoproteomics analysis of EGF dependent signaling in lung tissue (two biological replicates of n=4 rats in each group injected with either saline or EGF). Total peptide mixtures, titanium dioxide (TiO_2)-enriched serine/threonine phosphopeptides and antibody-enriched tyrosine phosphopeptides extracted from the lung tissues were analyzed by high-resolution LC-MS/MS. Tyrosine phosphorylation sites regulated upon EGF stimulation were identified, and for each regulated phosphotyrosine site we performed peptide based pulldowns in tissue lysate to identify protein complexes recruited to each site. From these data, we can build the signaling network activated upon EGF stimulation, covering both regulated phosphorylation sites but also the function of regulated phosphotyrosines in terms of their recruited protein complexes. (B) Immunoblots for EGFR, downstream signaling molecules (SHC and ERK) and tubulin as a loading control for four rats in each of the two groups (control versus EGF stimulated). (C) Volcano plot analysis of the phosphotyrosine-proteomics data. Each dot represents a peptide, those in blue are unmodified peptides from the proteome measurements and those in red are tyrosine phosphorylated peptides. Significantly regulated peptides are highlighted as filled circles, and for some of them the gene name and the regulated site is indicated (significance criteria, student's t-test $p<0.05$ and ratio>4 for EGF-stimulated versus control rats, are shown as dashed lines). (D) We synthesized peptides corresponding to the regulated phosphotyrosine residues and their six flanking amino acids for all regulated phosphotyrosine sites. These peptides were used as baits in pulldown experiments in three different lung tissue homogenates in a 96-well plate format. Proteins bound to synthesized peptides were digested on beads and subsequently analyzed by high-resolution LC-MS/MS.

Fig. 2

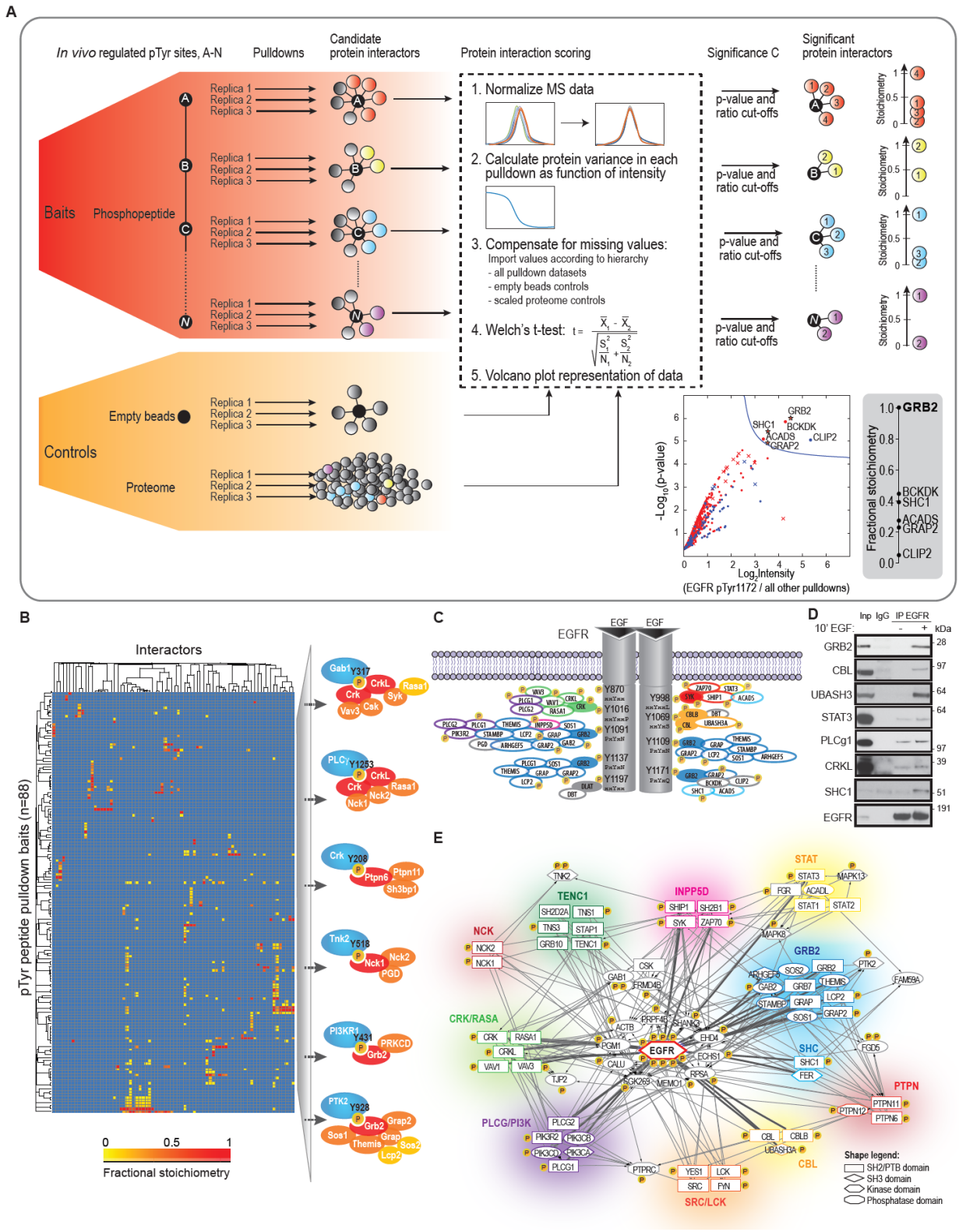

Figure 2: *Protein complexes recruited to regulated phosphotyrosine sites.* (A) Schematic representation of the analytical framework we developed for analysis of pulldown datasets. Bioinformatics analyses of the pulldowns are performed in the web-interface we developed (https://pulldown.jensenlab.org). Protein intensities are normalized and their variances are calculated as a function of intensity. Missing values in control experiments are inferred according to a hierarchy and a modified Welch's t-test is performed for each pulldown. Statistically significant protein interactors for each pulldown are determined based on Significance C and represented

in volcano-plots. Fractional stoichiometries of significant protein interactors are provided, highlighting the likely direct interaction partner. An example result is shown for peptide pulldowns of a peptide flanking the regions of EGFR phosphorylated on tyrosine 1172 using this approach, highlighting GRB2 as direct interaction partner. (B) Clustering of interaction partners identified in the 88 interaction complexes. Proteins are color-coded by their fractional stoichiometry in the complex. For six of the 88 complexes the information contained in the clustering plot is illustrated. (C) The core signaling complex of the in-vivo EGF dependent signaling response in lung tissue. Schematic of EGFR with the tyrosine residues that we identified as significantly regulated upon EGF stimulation is shown. For each site, the proteins identified to significantly interact with the phosphotyrosine residue are depicted together with regulated phosphorylation sites that we identified on these proteins. (D) Western blot validation of EGF-dependent interaction partners of EGFR by co-immunoprecipitation. (E) Protein interaction network of the 11 key protein complexes in the EGFR signaling network are depicted along with protein–protein interactions we have measured. The 11 key protein complexes are color-coded and the shape indicates if the protein contains a SH2, SH3/PTB, kinase or phosphatase domain. P denotes a regulated phosphorylation site.

Fig. 3

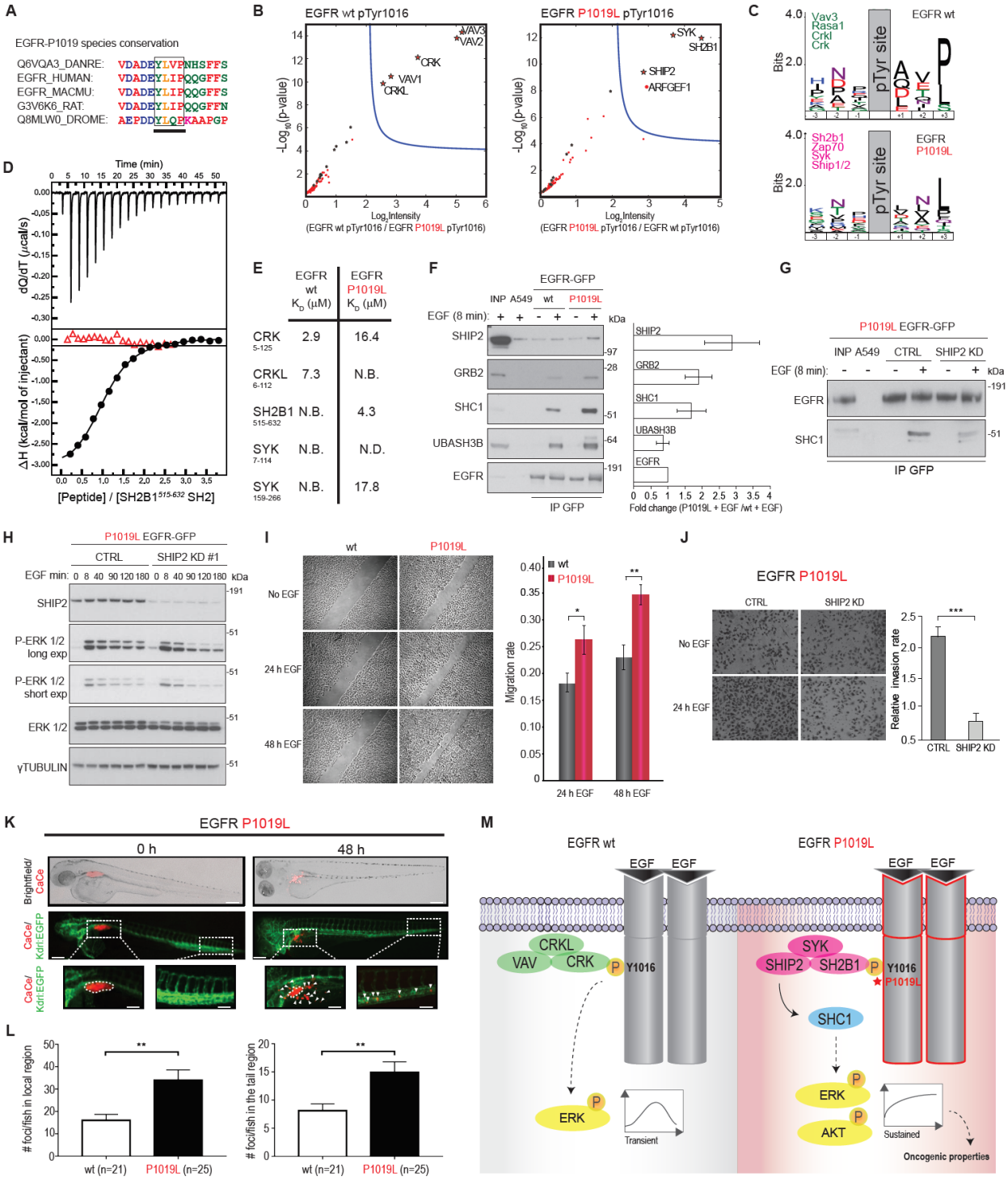
A
EGFR-P1019 species conservation
Q6VQA3_DANRE: VDADEYLVPNHSFFS
EGFR_HUMAN: VDADEYLIPQQGFFS
EGFR_MACMU: VDADEYLIPQQGFFS
G3V6K6_RAT: VDADEYLIPQQGFFN
Q8MLW0_DROME: AEPDDYLQPKAAPGP
B
EGFR wt pTyr1016
EGFR P1019L pTyr1016
-Log10(p-value)
Log2Intensity
(EGFR wt pTyr1016 / EGFR P1019L pTyr1016)
(EGFR P1019L pTyr1016 / EGFR wt pTyr1016)
VAV3
VAV2
CRK
VAV1
CRKL
SYK
SH2B1
SHIP2
ARFGEF1
C
EGFR wt
EGFR P1019L
Vav3
Rasa1
Crkl
Crk
Sh2b1
Zap70
Syk
Ship1/2
pTyr site
Bits
D
Time (min)
dQ/dT (μcal/s)
ΔH (kcal/mol of injectant)
[Peptide] / [SH2B1515-632 SH2]
E
EGFR wt Kd (μM)
EGFR P1019L Kd (μM)
CRK 5-125 2.9 16.4
CRKL 6-112 7.3 N.B.
SH2B1 515-632 N.B. 4.3
SYK 7-114 N.B. N.D.
SYK 159-266 N.B. 17.8
F
EGFR-GFP
INP A549 wt P1019L
EGF (8 min):
kDa
SHIP2
GRB2
SHC1
UBASH3B
EGFR
IP GFP
Fold change (P1019L + EGF /wt + EGF)
G
P1019L EGFR-GFP
INP A549 CTRL SHIP2 KD
EGF (8 min):
kDa
EGFR
SHC1
IP GFP
H
P1019L EGFR-GFP
CTRL
SHIP2 KD #1
EGF min: 0 8 40 90 120 180 0 8 40 90 120 180 kDa
SHIP2
P-ERK 1/2 long exp
P-ERK 1/2 short exp
ERK 1/2
γTUBULIN
I
wt
P1019L
No EGF
24 h EGF
48 h EGF
Migration rate
J
EGFR P1019L
CTRL
SHIP2 KD
No EGF
24 h EGF
Relative invasion rate
K
EGFR P1019L
0 h
48 h
Brightfield/ CaCa
CaCe/ Kdrl:EGFP
L
foci/fish in local region
foci/fish in the tail region
wt (n=21)
P1019L (n=25)
M
EGFR wt
EGFR P1019L
EGF
CRKL
VAV
CRK
Y1016
SYK
SHIP2
SH2B1
P1019L
SHC1
ERK
AKT
Transient
Sustained
Oncogenic properties

Figure 3: *Molecular switch caused by EGFR lung cancer mutation P1019L.* (A) Amino acid sequence conservation for EGFR P1019 site across species. (B) Interaction partners for EGFR phosphorylated at Y1016 (left) and interaction partners for the P1019L lung cancer mutation of EGFR phosphorylated at Y1016 (right). Significant interactors are determined by Significance C and their gene names are displayed and phosphotyrosine-binding domains are indicated with stars. Note CRK interaction with wildtype and SHIP2 interaction with P1019L mutant. (C) Sequence motifs generated from our large-scale pulldown dataset show amino acid sequence preference of YxxP for peptides significantly interacting with CRK, whereas there is an amino acid sequence preference of YxxL for peptides significantly interacting with SHIP. (D) ITC measurements of binding affinities between the SH2 domain of SH2B1 and EGFR peptides covering phosphorylated Y1016 in both wildtype and P1019L-mutant versions. Black circles indicates mutant peptide, whereas red triangles are wildtype. (E) Binding affinities determined by ITC for SH2 domains of CRK, CRKL, SH2B1 and SYK against EGFR peptides covering phosphorylated Y1016 in both wildtype and P1019L mutant versions. (F) Western blot validation of differential EGF-dependent interactors for wildtype and mutant EGFR by co-immunoprecipitation. Parental A549 are used as negative control of the GFP-based pull down experiment. Results are quantified in the bar-graph displaying fold-change of mutant relative to WT (mean ±SEM of two independent experiments, each representing two WT and two P1019L clones, p-values evaluated by two-sided Student's T-test). (G) Western blot evaluation of SHC1 interaction from GFP-based pulldown of EGFR-P1019L in EGF stimulated cells after SHIP2 knock-down. (H) Immunoblots of EGFR and ERK phosphorylation dynamics for wildtype and mutant receptor as function of EGF stimulation. (I) Wound-healing assay to quantify cell migration. Images of wildtype and mutant EGFR expressing cells after EGF stimulation for 0, 24h and 48h. Quantitation of cell migration calculated as the absolute migration rate, where 0 indicates no migration and 1 equals full migration. Quantification is displayed to the right as mean ±SEM of two independent experiments, each including three WT and two P1019L clones, p-values evaluated by two-sided Student's T-test. (J) Mutant EGFR P1019L expressing cells were transfected with Ctrl or SHIP2 siRNA and invasion assay was performed. Representative images of the transwell matrigel-based invasion assay are shown on the left. . EGF-dependent invasion rate is shown on the right as mean ±SEM of two independent experiments, each with a different P1019L clone, p-values evaluated by two-sided Student's T-test. (K) Example of local and tail foci determination in embryonic zebrafish injected with EGFR P1019L expressing cells. (L) Quantification of foci in local and tail regions 48h after injection with wildtype or EGFR P1019L expressing cells. (M) Model representing the molecular switch introduced by the cancer mutation EGFR P1019L. Upon phosphorylation of Y1016, EGFR interacts with a CRK complex which induce a transient ERK and AKT activation response. For the cancer mutant, instead a SHIP2 complex is recruited to the receptor leading to sustained activation of ERK and AKT.

Fig. 4

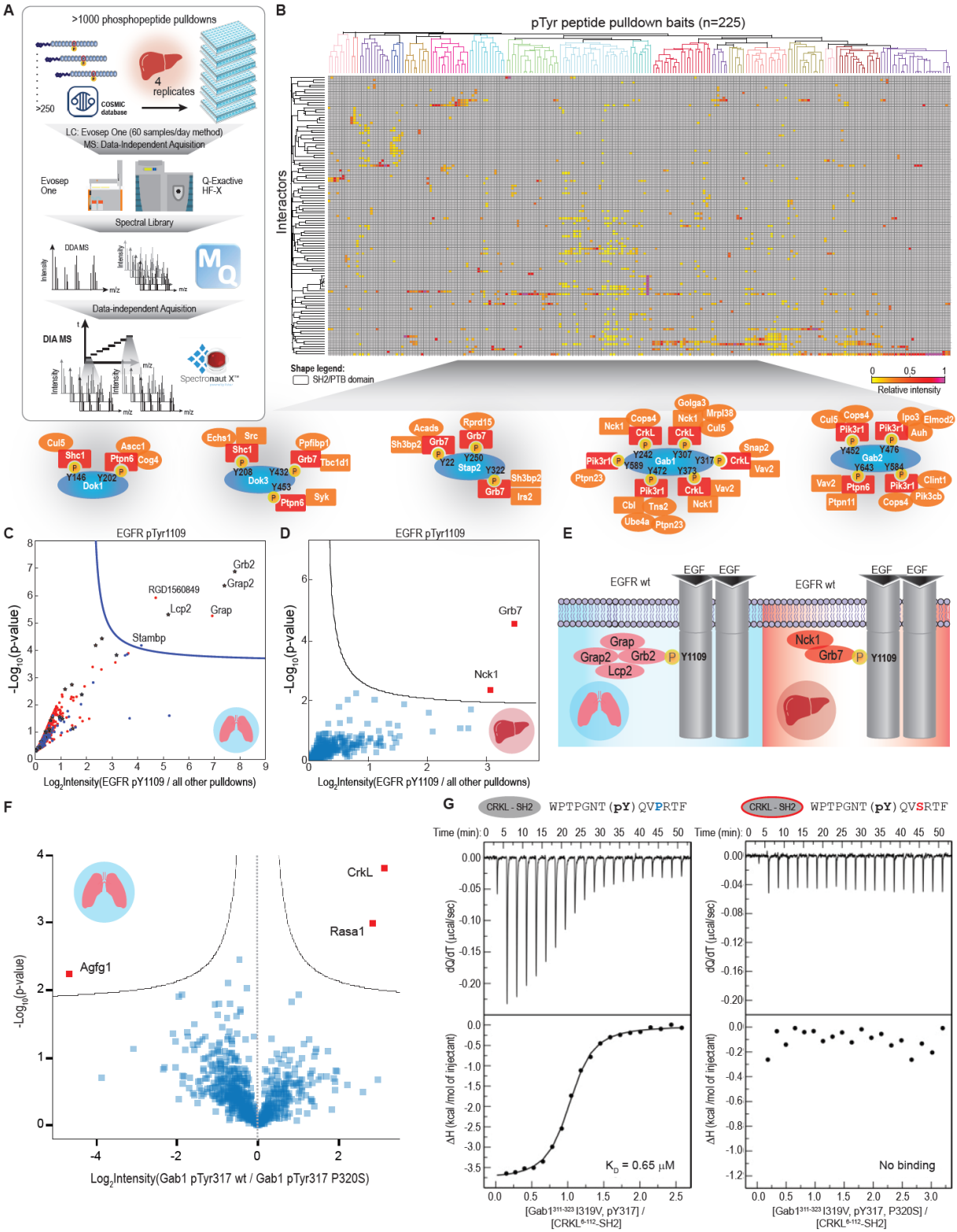
A
>1000 phosphopeptide pulldowns
4 replicates
>250
COSMIC database
LC: Evosep One (60 samples/day method)
MS: Data-Independent Aquisition
Evosep One
Q-Exactive HF-X
Spectral Library
DDA MS
Data-independent Aquisition
DIA MS
Spectronaut X™
B
pTyr peptide pulldown baits (n=225)
Interactors
Shape legend:
SH2/PTB domain
Relative intensity
C
EGFR pTyr1109
-Log10(p-value)
Log2Intensity(EGFR pY1109 / all other pulldowns)
Grb2
Grap2
RGD1560849
Lcp2
Grap
Stambp
D
EGFR pTyr1109
Grb7
Nck1
E
EGFR wt
EGF
Grap
Grap2
Grb2
Lcp2
Y1109
Nck1
Grb7
F
CrkL
Rasa1
Agfg1
Log2Intensity(Gab1 pTyr317 wt / Gab1 pTyr317 P320S)
G
CRKL - SH2
WPTPGNT(pY)QVPRTF
WPTPGNT(pY)QVSRTF
Time (min): 0 5 10 15 20 25 30 35 40 45 50
dQ/dT (µcal/sec)
ΔH (kcal /mol) of injectant
KD = 0.65 µM
No binding
[Gab1311-323 I319V, pY317] / [CRKL6-112-SH2]
[Gab1311-323 I319V, pY317, P320S] / [CRKL6-112-SH2]

Figure 4: *Molecular switches introduced by cancer mutations and tissue dependency of recruited protein complexes.* (A) Experimental outline for large-scale pulldown experiment. More than 300 peptide based pulldowns were performed in four biological replicates in lysates from either liver or lung tissues. The MS-based method was based on data independent acquisition (DIA) and fast online LC chromatography using the Evosep One system coupled to a Q Exactive HF-X orbitrap tandem mass spectrometer. (B) For 225 phosphotyrosine sites, specific interaction partners were identified. These are represented as a clustering profile highlighting the fractional stoichiometry of the interacting proteins. For five example complexes identified, the adaptor proteins recruited are illustrated. (C) In lung tissue, phosphorylation of EGFR Y1109 interacts with GRB2/GRAP/GRAP2 based complex. (D) In liver tissue, the same phosphotyrosine site, EGFR Y1109, interacts with a GRB7 based complex. (E) Model representation of the tissue specific interaction partners of EGFR Y1109 in lung and liver respectively. (F) The Gab1 P320S cancer mutation abolishes interaction at Y317 with CRKL and RASA1. (G) ITC experiments measuring the affinity of GAB1 phosphorylated at Y317, wildtype (left) and P320S (right) for CRKL.

REFERENCE LIST

Bache, N., Geyer, P.E., Bekker-Jensen, D.B., Hoerning, O., Falkenby, L., Treit, P.V., Doll, S., Paron, I., Muller, J.B., Meier, F.*, et al.* (2018). A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. Mol Cell Proteomics *17*, 2284-2296.

Bae, J.H., Lew, E.D., Yuzawa, S., Tome, F., Lax, I., and Schlessinger, J. (2009). The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. Cell *138*, 514-524.

Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R.*, et al.* (2004). The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer *91*, 355-358.

Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., Haldrup, C., Bramsen, J.B., Sorensen, K.D., Hoyer, S., Orntoft, T.F., Andersen, C.L.*, et al.* (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Syst *4*, 587-599 e584.

Bisson, N., James, D.A., Ivosev, G., Tate, S.A., Bonner, R., Taylor, L., and Pawson, T. (2011). Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. Nat Biotechnol *29*, 653-658.

Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., and Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol *21*, 315-318.

Boersema, P.J., Foong, L.Y., Ding, V.M., Lemeer, S., van Breukelen, B., Philp, R., Boekhorst, J., Snel, B., den Hertog, J., Choo, A.B.*, et al.* (2010). In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol Cell Proteomics *9*, 84-99.

Bugaj, L.J., Sabnis, A.J., Mitchell, A., Garbarino, J.E., Toettcher, J.E., Bivona, T.G., and Lim, W.A. (2018). Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. Science *361*.

Cohen, P. (2002). Protein kinases--the major drug targets of the twenty-first century? Nat Rev Drug Discov *1*, 309-315.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nat Methods *6*, 786-787.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Molecular cell *49*, 368-378.

Erneux, C., Edimo, W.E., Deneubourg, L., and Pirson, I. (2011). SHIP2 multiple functions: a balance between a negative control of PtdIns(3,4,5)P(3) level, a positive control of PtdIns(3,4)P(2) production, and intrinsic docking properties. J Cell Biochem *112*, 2203-2209.

Francavilla, C., Papetti, M., Rigbolt, K.T., Pedersen, A.K., Sigurdsson, J.O., Cazzamali, G., Karemore, G., Blagoev, B., and Olsen, J.V. (2016a). Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking. Nature structural & molecular biology *23*, 608-618.

Francavilla, C., Papetti, M., Rigbolt, K.T., Pedersen, A.K., Sigurdsson, J.O., Cazzamali, G., Karemore, G., Blagoev, B., and Olsen, J.V. (2016b). Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking. Nature structural & molecular biology.

Francavilla, C., Rigbolt, K.T., Emdal, K.B., Carraro, G., Vernet, E., Bekker-Jensen, D.B., Streicher, W., Wikstrom, M., Sundstrom, M., Bellusci, S.*, et al.* (2013). Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs. Molecular cell *51*, 707-722.

Grossmann, A., Benlasfer, N., Birth, P., Hegele, A., Wachsmuth, F., Apelt, L., and Stelzl, U. (2015). Phospho-tyrosine dependent protein-protein interaction network. Molecular systems biology *11*, 794.

Hanke, S., and Mann, M. (2009). The phosphotyrosine interactome of the insulin receptor family and its substrates IRS-1 and IRS-2. Mol Cell Proteomics *8*, 519-534.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F.*, et al.* (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712-723.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic acids research *40*, D261-270.

Humphrey, S.J., Azimifar, S.B., and Mann, M. (2015). High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat Biotechnol *33*, 990-995.

Hunter, T., and Sefton, B.M. (1980). Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proc Natl Acad Sci U S A *77*, 1311-1315.

Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villen, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. Cell *143*, 1174-1189.

Janne, P.A., Engelman, J.A., and Johnson, B.E. (2005). Epidermal growth factor receptor mutations in non-small-cell lung cancer: implications for treatment and tumor biology. J Clin Oncol *23*, 3227-3234.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Kavanaugh, W.M., Turck, C.W., and Williams, L.T. (1995). PTB domain binding to signaling proteins through a sequence motif containing phosphotyrosine. Science *268*, 1177-1179.

Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., and Olsen, J.V. (2018). Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. Journal of proteome research *17*, 727-738.

Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.A., Reinecke, M., Ruprecht, B., Petzoldt, S., Meng, C.*, et al.* (2017). The target landscape of clinical kinase drugs. Science *358*.

Kolch, W. (2005). Coordinating ERK/MAPK signalling through scaffolds and inhibitors. Nature reviews. Molecular cell biology *6*, 827-837.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Staerfeldt, H.H.*, et al.* (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods *14*, 61-64.

Liu, J.J., Sharma, K., Zangrandi, L., Chen, C., Humphrey, S.J., Chiu, Y.T., Spetea, M., Liu-Chen, L.Y., Schwarzer, C., and Mann, M. (2018). In vivo brain GPCR signaling elucidated by phosphoproteomics. Science *360*.

Lundby, A., Andersen, M.N., Steffensen, A.B., Horn, H., Kelstrup, C.D., Francavilla, C., Jensen, L.J., Schmitt, N., Thomsen, M.B., and Olsen, J.V. (2013). In vivo phosphoproteomics analysis reveals the cardiac targets of beta-adrenergic receptor signaling. Sci Signal *6*, rs11.

Lundby, A., Rossin, E.J., Steffensen, A.B., Acha, M.R., Newton-Cheh, C., Pfeufer, A., Lynch, S.N., Olesen, S.P., Brunak, S., Ellinor, P.T.*, et al.* (2014). Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. Nat Methods *11*, 868-874.

Lundby, A., Secher, A., Lage, K., Nordsborg, N.B., Dmytriyev, A., Lundby, C., and Olsen, J.V. (2012). Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nature communications *3*, 876.

Meyer, K., Kirchner, M., Uyar, B., Cheng, J.Y., Russo, G., Hernandez-Miranda, L.R., Szymborska, A., Zauber, H., Rudolph, I.M., Willnow, T.E.*, et al.* (2018). Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. Cell *175*, 239-253 e217.

Miller, M.L., Hanke, S., Hinsby, A.M., Friis, C., Brunak, S., Mann, M., and Blom, N. (2008). Motif decomposition of the phosphotyrosine proteome reveals a new N-terminal binding motif for SHIP2. Mol Cell Proteomics *7*, 181-192.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635-648.

Paez, J.G., Janne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J.*, et al.* (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science *304*, 1497-1500.

Pawson, T., and Scott, J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075-2080.

Petschnigg, J., Groisman, B., Kotlyar, M., Taipale, M., Zheng, Y., Kurat, C.F., Sayad, A., Sierra, J.R., Mattiazzi Usaj, M., Snider, J.*, et al.* (2014). The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. Nat Methods *11*, 585-592.

Prasad, N.K. (2009). SHIP2 phosphoinositol phosphatase positively regulates EGFR-Akt pathway, CXCR4 expression, and cell migration in MDA-MB-231 breast cancer cells. Int J Oncol *34*, 97-105.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y.*, et al.* (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell *131*, 1190-1203.

Rix, U., and Superti-Furga, G. (2009). Target profiling of small molecules by chemical proteomics. Nat Chem Biol *5*, 616-624.

Rouhi, P., Jensen, L.D., Cao, Z., Hosaka, K., Lanne, T., Wahlberg, E., Steffensen, J.F., and Cao, Y. (2010). Hypoxia-induced metastasis model in embryonic zebrafish. Nat Protoc *5*, 1911-1918.

Rush, J., Moritz, A., Lee, K.A., Guo, A., Goss, V.L., Spek, E.J., Zhang, H., Zha, X.M., Polakiewicz, R.D., and Comb, M.J. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol *23*, 94-101.

Schlessinger, J., and Lemmon, M.A. (2003). SH2 and PTB domains in tyrosine kinase signaling. Sci STKE *2003*, RE12.

Schulze, W.X., Deng, L., and Mann, M. (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol *1*, 2005 0008.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Sefton, B.M., Hunter, T., Beemon, K., and Eckhart, W. (1980). Evidence that the phosphorylation of tyrosine is essential for cellular transformation by Rous sarcoma virus. Cell *20*, 807-816.

Sharma, K., D'Souza, R.C., Tyanova, S., Schaab, C., Wisniewski, J.R., Cox, J., and Mann, M. (2014). Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell reports *8*, 1583-1594.

Sharma, K., Weber, C., Bairlein, M., Greff, Z., Keri, G., Cox, J., Olsen, J.V., and Daub, H. (2009). Proteomics strategy for quantitative protein interaction profiling in cell extracts. Nat Methods *6*, 741-744.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J.*, et al.* (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767-778.

Soria, J.C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K.H., Dechaphunkul, A., Imamura, F., Nogami, N., Kurata, T.*, et al.* (2018). Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. The New England journal of medicine *378*, 113-125.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic acids research *34*, D535-539.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A million peptide motifs for the molecular biologist. Molecular cell *55*, 161-169.

Ulaganathan, V.K., Sperl, B., Rapp, U.R., and Ullrich, A. (2015). Germline variant FGFR4 p.G388R exposes a membrane-proximal STAT3 binding site. Nature *528*, 570-574.

UniProt, C. (2015). UniProt: a hub for protein information. Nucleic acids research *43*, D204-212.

Vacic, V., Markwick, P.R., Oldfield, C.J., Zhao, X., Haynes, C., Uversky, V.N., and Iakoucheva, L.M. (2012). Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS computational biology *8*, e1002709.

van Biesen, T., Hawes, B.E., Luttrell, D.K., Krueger, K.M., Touhara, K., Porfiri, E., Sakaue, M., Luttrell, L.M., and Lefkowitz, R.J. (1995). Receptor-tyrosine-kinase- and G beta gamma-mediated MAP kinase activation by a common signalling pathway. Nature *376*, 781-784.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T*., et al.* (2014). Classification of intrinsically disordered regions and proteins. Chem Rev *114*, 6589-6631.

Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., and White, F.M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol Cell Proteomics *4*, 1240-1250.

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J., and Mohammed, S. (2013). Toward a comprehensive characterization of a human cancer cell phosphoproteome. Journal of proteome research *12*, 260-271.