Reconstructing kinase network topologies from phosphoproteomics data reveals cancer associated rewiring

Maruan Hijazi [1], Ryan Smith [2], Vinothini Rajeeve [1], Conrad Bessant [2,3] and Pedro R. Cutillas [1,3]*

[1] Signalling & Proteomics Group, Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, EC1M 6BQ, United Kingdom

[2] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End, E1 4NS, United Kingdom

[3] The Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, United Kingdom

* Correspondence address:

Pedro R. Cutillas
Barts Cancer Institute
Queen Mary University of London
John Vane Science Centre,
Charterhouse Square,
London EC1M 6BQ
United Kingdom
Tel: +44 (0)20 7882 8266
Fax: +44 (0)20 7882 3891
Email: p.cutillas@qmul.ac.uk

Understanding how oncogenic mutations rewire regulatory protein networks is important for rationalizing the mechanisms of oncogenesis and for individualizing anti-cancer treatments. We report a chemical phosphoproteomics method to elucidate the topology of kinase signaling networks in mammalian cells. We identified >6,000 protein phosphorylation sites that can be used to infer > 1,500 kinase-kinase interactions and devised algorithms that can reconstruct kinase network topologies from these phosphoproteomics data. Application of our methods to primary acute myeloid leukemia (AML) and breast cancer tumors quantified the relationship between kinase expression and activity, and enabled the identification of hitherto unknown kinase network topologies associated with drug resistant phenotypes or specific genetic mutations. Using orthogonal methods we validated that *PIK3CA* wild-type cells adopt MAPK-dependent circuitries in breast cancer cells and that the kinase TTK is important in AML. Our phosphoproteomic signatures of network circuitry can identify kinase topologies associated with both phenotypes and genotypes of cancer cells.

Understanding the circuitry and topology of signaling networks driven by protein and lipid kinases is important to rationalize the molecular basis of signal transduction and how this process is subverted in disease [1]. An established approach to infer kinase network topology and circuitry entails the use of systematic literature mining methods to compile data on the

relationships between enzymes in signaling cascades [2, 3]. Overlaying phosphoproteomic data onto these literature-mined networks offers a useful starting point for analyzing the impact that experimental perturbations have on signal transduction [4]. Although informative, particularly when used in combination with logic modelling [5, 6], due to pathway popularity and reagent availability, literature-mined networks are biased towards well-known kinases and pathways, thus restricting the opportunity for new discovery [7]. In addition, since they use data from different cell types and organisms, such networks consist of composite topologies that are not always representative of how signaling may be wired in a given cell population. Aside from *in silico* methods, experimental approaches for kinome analysis entail measuring kinase expression [8, 9], their phosphorylation [10, 11], enzymatic activity [12, 13], or a combination of these[14, 15], without revealing new information on the kinase-kinase relationships that define signaling cascades and network circuitry. Because of these limitations, comparative systematic analysis of biochemical network topologies has so far been restricted to the analysis of metabolic, protein-protein interaction and transcription factor networks [16, 17].

To address the question of how genetic mutations rewire kinase networks [18], and how these circuitries in turn regulate cancer phenotypes, here, we first developed a chemical phosphoproteomics approach to systematically identify markers of kinase network circuitry. By comparing the selectivity profiles of a set of kinase inhibitors with their effects on cellular phosphoproteomes, we identified 6,206 kinase-phosphosite (K-P) relationships for 103 kinases and 1,508 edges (kinase-kinase relationships). Comparative analysis of these network circuitry markers modelled differences in topologies across tumors and uncovered hitherto uncharacterized connections in these cancer networks. Our study represents a unique resource to investigate the relationships between kinase network topology and the phenotypes and genotypes of cancer cells.

Results

Expectancy of being downstream target (EBDT) algorithm assigns phosphosites to upstream kinases

Most kinase inhibitors inhibit a large number of kinases making it difficult to derive conclusions on K-P relationships from phosphoproteomic analysis of cells treated with kinase inhibitors [19, 20]. To overcome this limitation, we developed an algorithm, named expectancy of being downstream target (EBDT), which utilizes information on kinase inhibitor selectivity to determine the most probable kinase(s) acting upstream of phosphorylation sites present in phosphoproteomics data (**Fig. 1a**, see Supplementary Notes).

To investigate the validity of the approach, we profiled the target selectivity of 20 kinase inhibitors against 403 kinases. These 20 compounds inhibited 135 kinases by >50% (**Fig. 1b**, left panel). To complement these data, we also compiled selectivity data from vendors and the literature [21] for 40 compounds that inhibited 312 kinases (**Supplementary Dataset 1**), 163 of which had unique *in vitro* inhibition patterns across the profiled compounds (**Fig. 1c**) with no bias in terms of kinase families represented (**Fig. 1d**). Despite employing different assays, selectively data from DiscoverX (this study) and literature sources were comparable for compounds included in both datasets (**Supplementary Fig. 1a**). In parallel, we profiled the phosphoproteomes of three different cell lines (HL60, MCF7 and NTERA2) treated with the same compounds for which we had target specificity data. These experiments were performed in biological and technical replicate and required 794 LC-MS/MS runs. In total, we quantified 23,181 unique phosphorylation sites across all runs using previously reported methods [22, 23] (**Supplementary Dataset 2, Supplementary Fig. 1b, c**). We focused the analysis on phosphopeptides with a FDR of identification <0.01 and CV (of control) <50%. The number of phosphosites that met these stringent criteria were 9,326, 9,416, and 12,591, for HL60, MCF7 and NTERA2 cell lines, respectively (**Supplementary Fig. 1d**). To assist the analysis, we created a relational database (named ChemPhoPro) with web interface from which these phosphosites may be browsed (**Supplementary Fig. 2**).

We next used the EBDT algorithm (**Fig. 1a**, see Methods section for details) to assign each of the quantified phosphorylation sites to kinases acting upstream, leading to the identification of 2,706, 3,305, and 3,097 phosphosites (for HL60, MCF7 and NTERA2 cell lines) as putative downstream targets (PDTs) of 103 kinases **(Supplementary Fig. 1e-f**). In total, 6,206 unique PDTs were assigned to upstream kinases in at least one cell line (**Supplementary Dataset 3**).

PDTs represent a source of K-P relationships that capture well-known K-P associations. Furthermore, motif analysis showed the expected associations between kinases and phosphorylation motifs **(Supplementary Dataset 4)**, as illustrated in **Supplementary Fig. 3a** for AKT1 and PKCI (which enriched motifs with basic residues at N-term), and CDKs (SP motifs with basic residues at C-term), consistent with the substrate specificity of these kinase groups [24], [25]. Similarly, comparison with computationally predicted kinase-substrate relationships using NetworKIN highlighted the predicted kinase in the PDT sets (**Supplementary Fig. 3b**), while comparison with published sources of K-P relationships showed that 548 and 375 sites in the PDT dataset were also present in PhosphoSitePlus and Signor databases of kinase-substrate relationships, respectively (**Fig. 1e**). Of note, 5,575 sites identified in this study as being PDT were not present in Signor or PhosphoSitePlus (**Fig. 1e**), and we identified PDTs for poorly represented kinases in previous repositories (**Supplementary Fig. 4a).** In total**,** we mapped 19,410 K-P relationships, which is a considerable larger coverage than those provided by PhosphoSitePlus and Signor (10,220 and 7,237, respectively, **Supplementary Fig. 4b**). Thus, our study significantly extends our knowledge of K-P relationships.

Since PDTs are readily detectable by mass spectrometry (they were all identified with this technique), we reasoned that PDTs would allow inferring kinase activity from phosphoproteins experiments with greater depth than when using other repositories as the source of K-P relationships. To test this, we measured the phosphoproteomes of a cell line (P31/Fuj) after treatment with 6 different kinase inhibitors (**Supplementary Fig. 4c**). This cell line is unrelated

to the panel used to construct the PDT dataset, thus it allowed us to assess the approach in an independent model. Out of the 8,100 unique phosphopeptides identified in this experiment, 1,746 were present in the PDT database, whereas 403 and 228 sites could be mapped to PhosphoSitePlus and Signor, respectively (**Fig. 1f**). These phosphosites were downstream of 82, 58 and 51 kinases in PDTs, PhosphoSitePlus, and Signor, respectively (**Fig. 1f**) with the average phosphosite per kinase being about 7 times greater when using PDTs as source of K-P relationships (**Fig. 1f**). Thus, overall, when used as the source of Kinase Substrate Enrichment Analysis (KSEA, Refs [12, 26]), our PDT dataset significantly extends the coverage of kinase activity measurements that may be obtained from the interrogation of phosphoproteomics data (**Supplementary Fig. 4d** and **Supplementary Dataset 5**).

PDT signatures allow comparative analysis of signaling network topology

Our study provides the community with a rich set of kinase activity markers that are readily measurable by mass spectrometry. However, unlike published repositories of direct kinase-substrate relationships, phosphosites in PDTs comprise direct and indirect substrates of kinases. We reasoned that this feature would allow us to investigate the relationship between the 103 kinases for which we identified PDTs. Using the PI3K/AKT/MTOR pathway as an example, **Fig. 2a** illustrates our approach to investigate kinase-kinase relationships and thus reconstruct networks from these data**;** 43 sites were identified to be PDTs of both PI3Kα (gene *PIK3CA*) and mTOR, 38 sites were PDTs of both PI3Kα and AKT1/2 and 25 of both mTOR and AKT1/2. This data can be represented in network format (**Fig. 2b,** edges in these networks may be conceptualized as signaling axes). Application of this approach to all other kinases with PDTs reconstructed a network consisting of 1,508 edges between 103 kinase nodes with weight ≥ 5 sites (**Fig. 2c**). Each of these edges was defined by 24 phosphorylation sites on average (**Fig. 2d**). The degree distribution of networks generated using high PDT threshold values showed left-skewed (power law) distributions (**Fig. 2e**), which were different to those obtained from randomly generated networks (**Fig. 2f**). Degree distributions that resemble power law distributions are typical of empirical networks, such as the internet and protein

interaction [27, 28]. In addition to well-known kinase-kinase associations, these data uncover a wealth of hitherto unexplored signaling axes and define a kinase network with unprecedented depth. The phosphosites defining network edges are provided as a community resource in **Supplementary dataset 6** and in the ChemPhoPro website (**Supplementary Fig. 2**).

As with the analysis of motifs in PDTs, we found an enrichment of the expected motifs and NetworKIN-predicted kinases in the identified edges, consistent with known kinase substrate specificities (**Supplementary Fig. 5a, Fig. 5c** and **Supplementary Dataset 4**). To confirm the biological relevance of the identified K-P relationships in a functional manner, we obtained time course phosphoproteomics data from Wilkes *et al* [6] and quantified PDT group enrichment (using the principles of KSEA [12, 26]) as a function of treatment with IGF-I or EGF, two agonists with well-characterized kinetic effects on cell signaling [29]. **Fig. 2g** illustrates the quantification approach with sites in the AKT1/2.PIK3CA PDT group, and **Fig. 2h, Supplementary Fig. 5b** show different signaling axes containing MAPK and PI3K nodes that increased in global phosphorylation with kinetics consistent with previous studies [29, 30]. In addition, experiment using cells treated with kinase inhibitors (**Supplementary Fig. 4c**) showed a decreased in enrichment of axes containing the expected kinases (**Supplementary Fig. 5d).** Together, these data strongly suggest that our groups of phosphorylation sites that define network axes/edges are bona fide markers of kinases activities linking the defined network edges.

Identification of network circuitries in cancer cells of distinct drug resistant phenotypes

We next investigated the biological relevance and potential of our PDT library to be used for comparative analysis of kinase network topology. To this end, we used edge enrichment to reconstruct kinase networks in three cell lines (**Fig. 3a-c**), from which we obtained values of centrality for each kinase node (betweenness and degree, **Fig. 3d-f**). Previous studies have shown that, in general, node centrality values are correlated with their importance in maintaining network integrity [27, 28]. In our analysis, kinase nodes in the PI3K pathway showed high centrality values in MCF7 cells relative to NTERA2, whereas MAPK1-3 (ERK) centrality

values were higher in the NTERA2 network (**Fig. 3g-h** and **Supplementary Fig. 6a**). Of note, these data were consistent with the analysis of phosphorylation sites markers of PI3K and MAPK pathways (**Supplementary Fig. 6b**). These observations predicted a key role of PI3K/AKT/MTOR in regulating MCF7 biology, whereas the MAPK pathway was expected to regulate signaling in NTERA2. Consistent with the prediction, NTERA2 was highly sensitive to MEK inhibition, whereas MCF7 was more sensitive to PI3K blockade (**Fig. 3i**). These data suggest that the networks measured using our set of PDTs markers of circuitry encode information on signaling activity and cell behavior.

To investigate the biological relevance of the network circuitries encoded in our PDT dataset in a larger and more clinically relevant set of cancer models, we measured our markers of network topology in 36 primary acute myeloid leukemia (AML) biopsies. Phosphoproteomics and drug sensitivity data for these models have been recently published [15]. As reported before [15], primary AML cells showed highly heterogeneous responses to treatments with inhibitors of MEK1/2, FLT3/PKC, PAK isoforms, P38 and CK2 (**Fig. 4a**). To identify edges in the network that may determine drug responses, we constructed partial least square (PLS) models trained using 18 randomly selected cases and tested using the remaining 18 cases. The accuracy of drug sensitivity prediction ranged from 20% to 45% (**Fig. 4b-e** and **Supplementary Fig. 7**), thus showing good model fits. Edges connecting kinase nodes known to be targeted by the compounds contributed to the models (**Fig. 4b-e** right panels, **Supplementary Fig. 8**), thus providing confidence on the models' biological relevance. PLS models using other metrics of signaling activity also predicted responses to kinase inhibitors with relatively high accuracy. However, the same approaches did not accurately predict overall patient survival (**Supplementary Fig. 7**), suggesting that values of kinase pathway activity and network circuitry specifically encode information on responses to drugs that disrupt kinase networks.

Impact of proteomic and genomic factors on network topology formation

We next investigated the proteomic and genomic factors that may determine kinase activity and network topology. We first measured the extent of association between kinase expression, kinase activity and edge weights in three different primary cancer datasets for which good quality phosphoproteomics and protein data are available [15, 31, 32]. The analysis of PAK1 across 69 ovarian cancers (**Fig. 5a-c**) illustrates the approach that we undertook; while the associations between individual PAK1 phosphosite intensities and PAK1 activity (as estimated by PAK1 PDT enrichment) were not significant (**Fig. 5a**), PAK1 estimated activity correlated with the sum of its phosphopeptide signals (**Fig. 5b**, top), but interestingly, not with unmodified PAK1 expression (**Fig. 5b**, middle). The ratio of phosphokinase to total kinase correlated to the same extent as un-normalized values (**Fig. 5b**, bottom, and **5c**). Extension of this analysis to all identified kinases showed that while the correlation between estimated kinase activity (by PDT enrichment) and protein expression was not significant (just 39 out of 88 kinases showed positive correlation), phosphokinase expression and estimated kinase activity was strongly associated, with 92% (73/79) of these being positively correlated (**Fig. 5d**, left). Similarly, the correlation between phosphokinase expression and edge enrichment was positive for 86% (1,194 out of 1,386) of these, and statistically different to the association between kinase expression and edge enrichment ($p<2.2E-16$ by Wilcoxon test, **Fig. 5d**, right panel). Analysis of AML and breast cancer (BC) datasets also showed greater correlation between kinase activity with phosphokinase expression than with protein expression (**Supplementary Fig. 9a**) and was consistent across all three cancer models tested (**Fig. 5e**).

The number of kinase substrates used to quantify kinase activity did not determine the association between phosphokinase expression and activity (**Supplementary Fig. 9b**). Interestingly, the number of phosphorylation sites on a kinase was proportional to the correlation between kinase activity and their phosphorylation extent (**Supplementary Fig. 9c**) but this factor did not contribute to the association between kinase expression and activity (**Supplementary Fig. 9d**). These data suggest that kinases that are phosphorylated at several

sites are regulated by overall phosphorylation to a greater extent than kinases that are phosphorylated on fewer sites.

To investigate the elements of kinase network topology associated to common genetic mutations, we assessed kinase edge enrichment as a function of genetic mutations in BC and AML biopsies. Primary AML cases positive for *NRAS* G12D mutations enriched kinases downstream of the canonical RAS pathway, including MAP2K1, PAK1, PIK3CA and AKT1 (**Supplementary Fig. 10a-c**), thus providing further evidence of biological relevance. TTK showed high centrality values in the *NRAS* associated AML network. To confirm the involvement of TTK in regulating AML biology, we treated an AML cell panel with CFI-402257, a highly specific TTK inhibitor at the concentrations used in this study [33]. This compound had a greater impact on cell death of six different AML models than inhibitors against kinases of current clinical and preclinical interest in this disease (**Supplementary Fig. 10d-e**), suggesting that TTK regulates AML cell death.

As for primary BC, *TP53*-mutated and triple negative cases (TNBC) enriched kinases associated with cell cycle progression (CDK2, CDK9 and CSNK1E), and with oncogenic transformation (ARAF, PDGFRB) (**Fig. 6a, Supplementary Dataset 7)**. Cases positive for activating *PIK3CA* mutations (in helical domain, HD), which are common in this malignancy [34] [35], enriched several PIK3CA/AKT/MTOR axes, thus providing a positive control for the analysis. Of note, *PIK3CA* HD mutations had a greater impact on PI3K and AKT centrality values than *PIK3CA* kinase domain (KD) mutations (**Fig. 6b-c**), consistent with the distinct biochemical and clinical relevance of these mutations [36, 37]. To confirm these findings, we took advantage of colorectal cancer (CRC) cell lines isogenic for *PIK3CA* mutations on HD or KD domains [35] (**Fig. 6d, Supplementary Fig. 11a, and Supplementary Dataset 8**). *PIK3CA* mutations activated PI3K/AKT signaling in these cells thus confirming the relevance of the models (**Fig. 6e, Supplementary Fig. 11b, Supplementary Fig. 13**). Phosphoproteomics-based network analysis showed that, consistent with the data from primary BC, *PIK3CA* HD mutations had greater impact on the network than cells with KD mutations (**Fig. 6f**). Despite

the different genetic backgrounds between primary BC and CRC cell lines, these *PIK3CA* mutant models showed similar enrichment of specific PI3K and MAPK nodes (**Fig. 6c**). Interestingly, cases positive for *PIK3CA* mutations showed a decrease in the enrichment of MAPK-containing edges (**Fig. 6c**) and the anti-correlation between PI3K and MAPK signaling observed across the 86 BC cases was confirmed in *PIK3CA* mutant isogenic cell lines (**Supplementary Fig. 11c**). These data suggest that the signaling network switches from PI3K to MAPK-dependent circuitries in *PIK3CA* WT cells. To investigate the functional relevance of the observed signaling switch, we tested the impact of PI3K and MAPK pathway inhibitors in reducing cell viability of the isogenic models. Consistent with the predictions of the models, we found that *PIK3CA* MUT cells were more sensitive to PI3K inhibition whereas trametinib (to inhibit the MAPK pathway) preferentially killed *PIK3CA* WT cells (**Fig. 6g** and **Supplementary Fig. 11d**).

Discussion

Our chemical phosphoproteomics approach to systematically identify K-P relationships takes advantage of the well-known promiscuity of most small-molecule kinase inhibitors [38,21]. The EBDT algorithm is based on the premise that phosphorylation sites inhibited in intact cells with the same pattern as the *in vitro* inhibition fingerprint for a given kinase are expected to be downstream of such kinase. Several lines of evidence support the quality of K-P assignments using this approach, including the observations that downstream targets of given kinases were enriched in phosphorylation motifs associated to such kinases' consensus substrate specificity, and that these behave as expected in cells treated with activators and inhibitors of kinase signaling (**Fig. 2, Supplementary Figs. 4 and 5** and **Supplementary Datasets 5** and **6**).

Phosphoproteomics measures kinase activities that have occurred in intact cells in a high-content systematic manner; in this single study we identified >19,564 K-P relationships between 6,206 phosphosites,103 kinases and 1,508 network edges, making this the largest

(to our knowledge) single dataset of experimentally determined K-P relationship to date. Sites in PDT groups are all readily quantifiable by MS and can thus be easily monitored across as many experimental conditions as required. This is an important feature because although databases of kinase-substrate relationships exist [39, 40], these are biased towards well studied kinases [7] and most sites detected in standard phosphoproteomics experiments are not present in these databases [12, 14].

Our approach complements methodologies to identify direct kinase substrates [41-44] because the purpose of the EBDT is to obtain phosphosite groups comprising both direct kinase substrates and indirect downstream targets, from which kinase activities and network circuitries may be inferred. The EBDT algorithm is however restricted to the identification of downstream targets of kinases possessing unique inhibition fingerprint patterns. Nevertheless, given the vast number of kinase inhibitors available [45], unique inhibition fingerprints may already exist for most, if not all, kinases expressed in cells.

Previous studies have investigated the contribution of transcription and translation to the regulation of gene expression [46], but less is known about the global contribution of protein expression and phosphorylation to the regulation of overall enzymatic activities. Activation loop phosphorylation is a well-known mechanism by which kinase catalysis is regulated [47] and the association between single phosphorylation sites and kinase activity has been investigated before systematically [13]. Here we found several kinases for which the sum of their phosphosite intensities correlated with their activity (**Fig. 5**), suggesting that global phosphorylation may be an understudied mechanism of kinase activity regulation.

Our markers of circuitry highlighted network components associated to recurrent genetic mutations in primary BC and AML (**Fig. 6** and **Supplementary Fig. 10**). The anti-correlation between PI3K and MAPK edge enrichment in primary BC tumors and in isogenic CRC cell lines suggests that therapies that target PI3K signaling may have unintended effects in

activating MAPK pathways and this may explain the relatively poor efficacy of single-agent PI3K inhibitors for BC treatment [48, 49].

Acknowledgements

This work was primarily funded by B.B.S.R.C. (BB/M006174/1). Barts and The London Charity (297/2249), CRUK (C15966/A24375 and C16420/A18066) and QMUL's Life Science Initiative also contributed to funding.

Author contributions

M.H. designed and conducted experiments, analyzed data and edited manuscript; R.S. performed bioinformatic experiments, analyzed data and edited manuscript; V.R. performed mass spectrometry experiments; C.B. conceived study, performed bioinformatic experiments, analyzed data, and edited manuscript; P.R.C. conceived study and EBDT approach, designed experiments, performed bioinformatic experiments, analyzed and interpreted data, prepared figures, wrote manuscript.

Competing financial interests

P.R.C. is co-founder of Kinomica Ltd. The other authors declare no competing financial interests.

References

1. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
2. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353-D361 (2017).
3. Fabregat, A. et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142 (2017).
4. Sacco, F., Perfetto, L. & Cesareni, G. Combining Phosphoproteomics Datasets and Literature Information to Reveal the Functional Connections in a Cell Phosphorylation Network. *Proteomics* **18**, e1700311 (2018).
5. Tape, C.J. et al. Oncogenic KRAS Regulates Tumor Cell Signaling via Stromal Reciprocation. *Cell* **165**, 910-920 (2016).
6. Terfve, C.D., Wilkes, E.H., Casado, P., Cutillas, P.R. & Saez-Rodriguez, J. Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data. *Nat Commun* **6**, 8033 (2015).
7. Invergo, B.M. & Beltrao, P. Reconstructing phosphorylation signalling networks from quantitative phosphoproteomic data. *Essays Biochem* (2018).
8. Donnella, H.J. et al. Kinome rewiring reveals AURKA limits PI3K-pathway inhibitor efficacy in breast cancer. *Nat Chem Biol* **14**, 768-777 (2018).
9. Lachmann, A. & Ma'ayan, A. KEA: kinase enrichment analysis. *Bioinformatics* **25**, 684-686 (2009).
10. Johnson, H. et al. Molecular characterization of EGFR and EGFRvIII signaling networks in human glioblastoma tumor xenografts. *Mol Cell Proteomics* **11**, 1724-1740 (2012).
11. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-305 (2002).
12. Casado, P. et al. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Science signaling* **6**, rs6 (2013).
13. Ochoa, D. et al. An atlas of human kinase regulation. *Molecular systems biology* **12**, 888 (2016).
14. Drake, J.M. et al. Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell* **166**, 1041-1054 (2016).
15. Casado, P. et al. Proteomic and genomic integration identifies kinase and differentiation determinants of kinase inhibitor sensitivity in leukemia cells. *Leukemia* **32**, 1818-1822 (2018).
16. Singh, R., Xu, J. & Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12763-12768 (2008).
17. Sharan, R. & Ideker, T. Modeling cellular machinery through biological network comparison. *Nature biotechnology* **24**, 427-433 (2006).
18. Creixell, P. et al. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* **163**, 202-217 (2015).
19. Pan, C., Olsen, J.V., Daub, H. & Mann, M. Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* **8**, 2796-2808 (2009).
20. Alcolea, M.P., Casado, P., Rodriguez-Prados, J.C., Vanhaesebroeck, B. & Cutillas, P.R. Phosphoproteomic analysis of leukemia cells under basal and drug-treated conditions identifies markers of kinase pathway activation and mechanisms of resistance. *Molecular & cellular proteomics : MCP* **11**, 453-466 (2012).
21. Klaeger, S. et al. The target landscape of clinical kinase drugs. *Science* **358** (2017).

22. Wilkes, E. & Cutillas, P.R. Label-Free Phosphoproteomic Approach for Kinase Signaling Analysis. *Methods Mol Biol* **1636**, 199-217 (2017).
23. Wilkes, E.H., Terfve, C., Gribben, J.G., Saez-Rodriguez, J. & Cutillas, P.R. Empirical inference of circuitry and plasticity in a kinase signaling network. *Proc Natl Acad Sci U S A* **112**, 7719-7724 (2015).
24. Alessi, D.R., Caudwell, F.B., Andjelkovic, M., Hemmings, B.A. & Cohen, P. Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. *FEBS Lett* **399**, 333-338 (1996).
25. Ubersax, J.A. & Ferrell, J.E., Jr. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8**, 530-541 (2007).
26. Hernandez-Armenta, C., Ochoa, D., Goncalves, E., Saez-Rodriguez, J. & Beltrao, P. Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* **33**, 1845-1851 (2017).
27. Barabási, A.-L.s. & Pósfai, M.r. Network science. (Cambridge University Press, Cambridge, United Kingdom; 2016).
28. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **411**, 41-42 (2001).
29. Ciaccio, M.F., Wagner, J.P., Chuu, C.P., Lauffenburger, D.A. & Jones, R.B. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat Methods* **7**, 148-155 (2010).
30. Vincent, A.M. & Feldman, E.L. Control of cell survival by IGF signaling pathways. *Growth Horm. IGF Res.* **12**, 193-197 (2002).
31. Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55-62 (2016).
32. Zhang, H. et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166**, 755-765 (2016).
33. Mason, J.M. et al. Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 3127-3132 (2017).
34. Bachman, K.E. et al. The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer biology & therapy* **3**, 772-775 (2004).
35. Samuels, Y. et al. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561-573 (2005).
36. Zhao, L. & Vogt, P.K. Helical domain and kinase domain mutations in p110alpha of phosphatidylinositol 3-kinase induce gain of function by different mechanisms. *Proc Natl Acad Sci U S A* **105**, 2652-2657 (2008).
37. Barbareschi, M. et al. Different prognostic roles of mutations in the helical and kinase domains of the PIK3CA gene in breast carcinomas. *Clin Cancer Res* **13**, 6064-6069 (2007).
38. Davies, S.P., Reddy, H., Caivano, M. & Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem J* **351**, 95-105 (2000).
39. Hornbeck, P.V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research* **43**, D512-D520 (2015).
40. Perfetto, L. et al. SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res* **44**, D548-554 (2016).
41. Carlson, S.M. et al. Large-Scale Discovery of ERK2 Substrates Identifies ERK-Mediated Transcriptional Regulation by ETV3. *Science signaling* **4** (2011).
42. Shah, K., Liu, Y., Deirmengian, C. & Shokat, K.M. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci U S A* **94**, 3565-3570 (1997).
43. Xue, L., Geahlen, R.L. & Tao, W.A. Identification of direct tyrosine kinase substrates based on protein kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* **12**, 2969-2980 (2013).

44. Cohen, P. & Knebel, A. KESTREL: a powerful method for identifying the physiological substrates of protein kinases. *Biochem J* **393**, 1-6 (2006).
45. Ferguson, F.M. & Gray, N.S. Kinase inhibitors: the road ahead. *Nat Rev Drug Discov* **17**, 353-377 (2018).
46. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
47. Adams, J.A. Activation loop phosphorylation and catalysis in protein kinases: is there functional evidence for the autoinhibitor model? *Biochemistry* **42**, 601-607 (2003).
48. Juric, D. et al. Phosphatidylinositol 3-Kinase alpha-Selective Inhibition With Alpelisib (BYL719) in PIK3CA-Altered Solid Tumors: Results From the First-in-Human Study. *J Clin Oncol* **36**, 1291-1299 (2018).
49. Sarker, D. et al. First-in-human phase I study of pictilisib (GDC-0941), a potent pan-class I phosphatidylinositol-3-kinase (PI3K) inhibitor, in patients with advanced solid tumors. *Clin Cancer Res* **21**, 77-86 (2015).
50. Zhang, H. et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166**, 755-765 (2016).

Figures

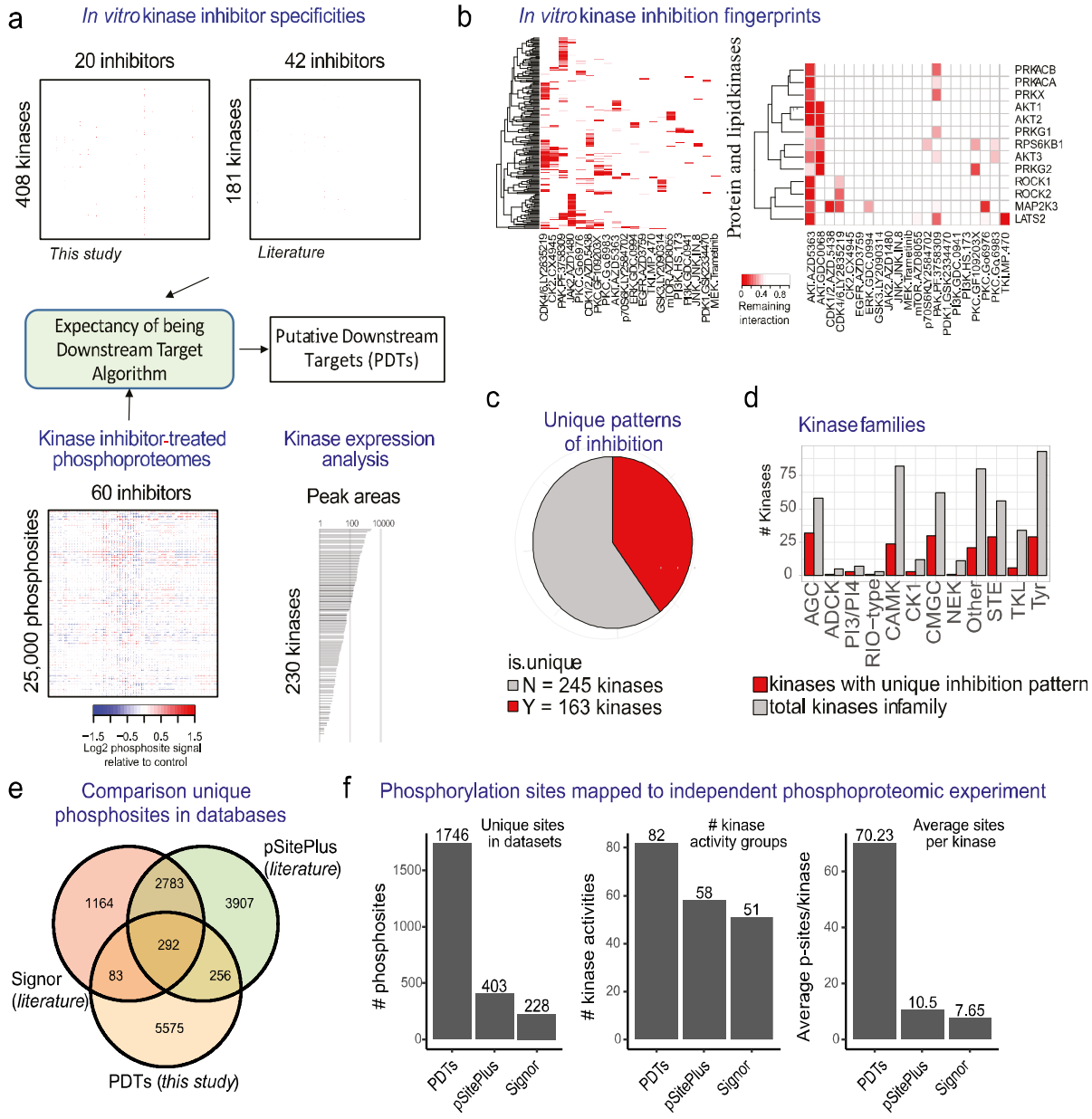

Figure 1. Profiling kinase inhibitor fingerprints to assign phosphorylation sites to upstream kinases

(**a**) Overview of the approach to identify kinase putative downstream targets. The effects of inhibitors with well characterized *in vitro* kinase specificity are matched to their impact on cellular phosphoproteomics. The analysis is restricted to kinases expressed in the cell line under investigation and to kinases with unique inhibitor fingerprints. (**b**) Inhibitor fingerprints

were determined using the DiscoverX platform for 20 kinase inhibitors, complemented with data from the literature. Examples of kinase inhibitor fingerprints for all (left) and a selection (right) of kinases using results from the DiscoverX screen. (**c**) Number of kinases showing unique kinase inhibitor fingerprints. (**d**) Number of kinases with unique inhibitor fingerprints as a function of kinase family. (**e**) Number of phosphorylation sites in the named databases of kinase-phosphosite relationships and their overlap. (**f**) Sites in the named databases were mapped to an independent phosphoproteomics experiment (see Supplementary Fig. 4 for details). Bar plots show the number of unique phosphorylation sites mapped to the different databases (left), number of kinases for which at least 3 phosphosites were mapped (middle) and mean of phosphosites mapped per kinase (right).

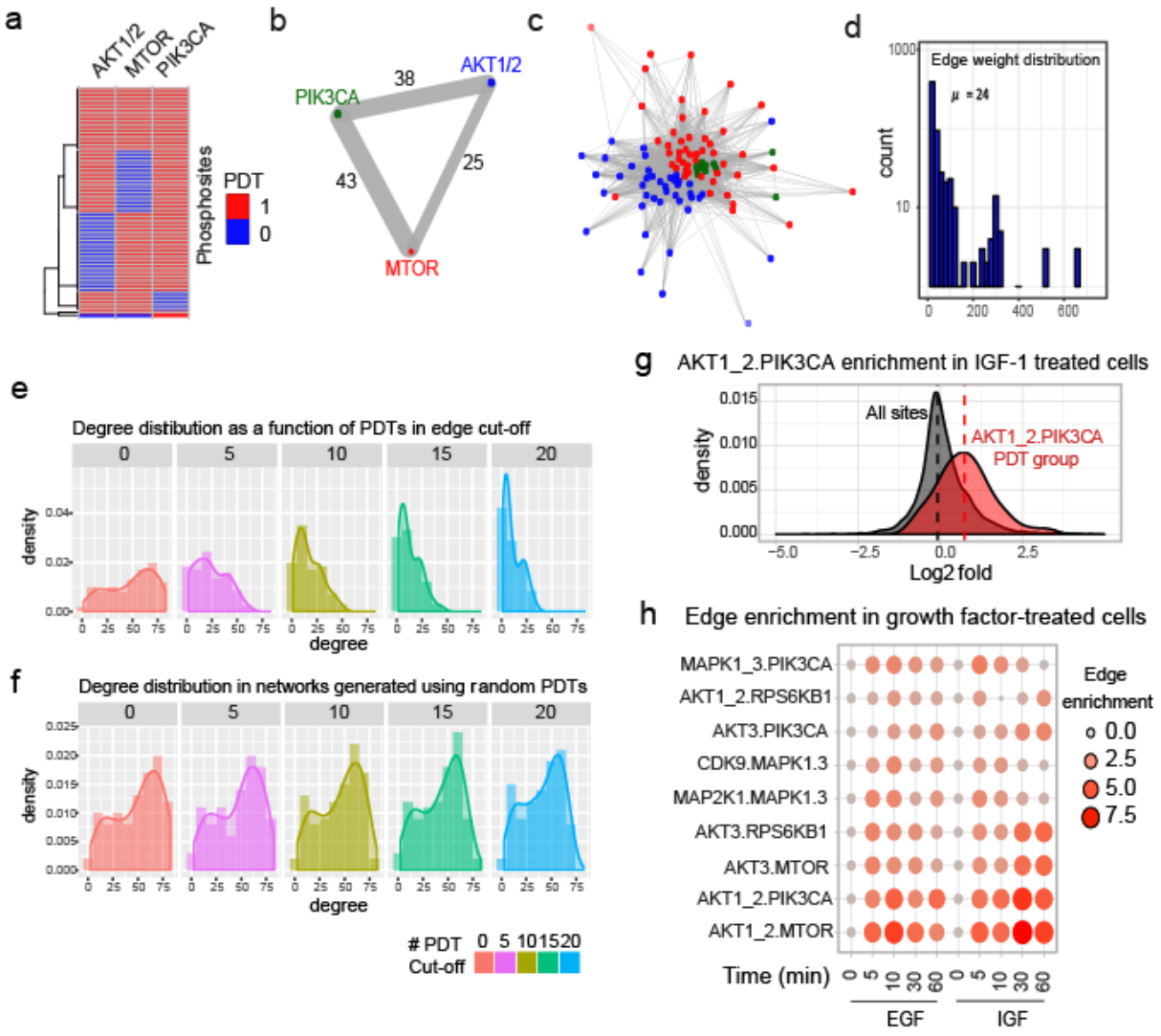

Figure 2. Empirical inference of kinase network circuitry from downstream kinase targets

(**a**) Illustration of approach to identify markers of network edges for PI3K, MTOR and AKT1/2 kinase nodes. Each row denotes a single phosphorylation site. Sharing of PDT (in red) allows determining markers of kinase-kinase relationships. Data is the compilation of K-P relationships found in the cell lines. (**b**) Edges weights between kinase nodes are defined as the number of common PDTs. (**c**) Network reconstructed from combined PDT data. As is common in the representation of biological pathways, edges represent interactions between kinases but these are not necessarily direct physical interactions as they may be modulated by other molecules. (**d**) Distribution of edge weights for the 103 kinases that form the network.

(**e**) Degree distribution of identified kinase network as a function of weight cut-off. (**f**) As in e but networks were also constructed from randomly assigned PTDs to network edges. (**g**) Distribution of fold changes of phosphopeptides that belong to the AKT1_2.PIK3CA axis in comparison to all phosphopeptides fold changes. (**h**) Enrichment of kinase axes as a function of treatment with EGF or IGF-I for the indicated times.

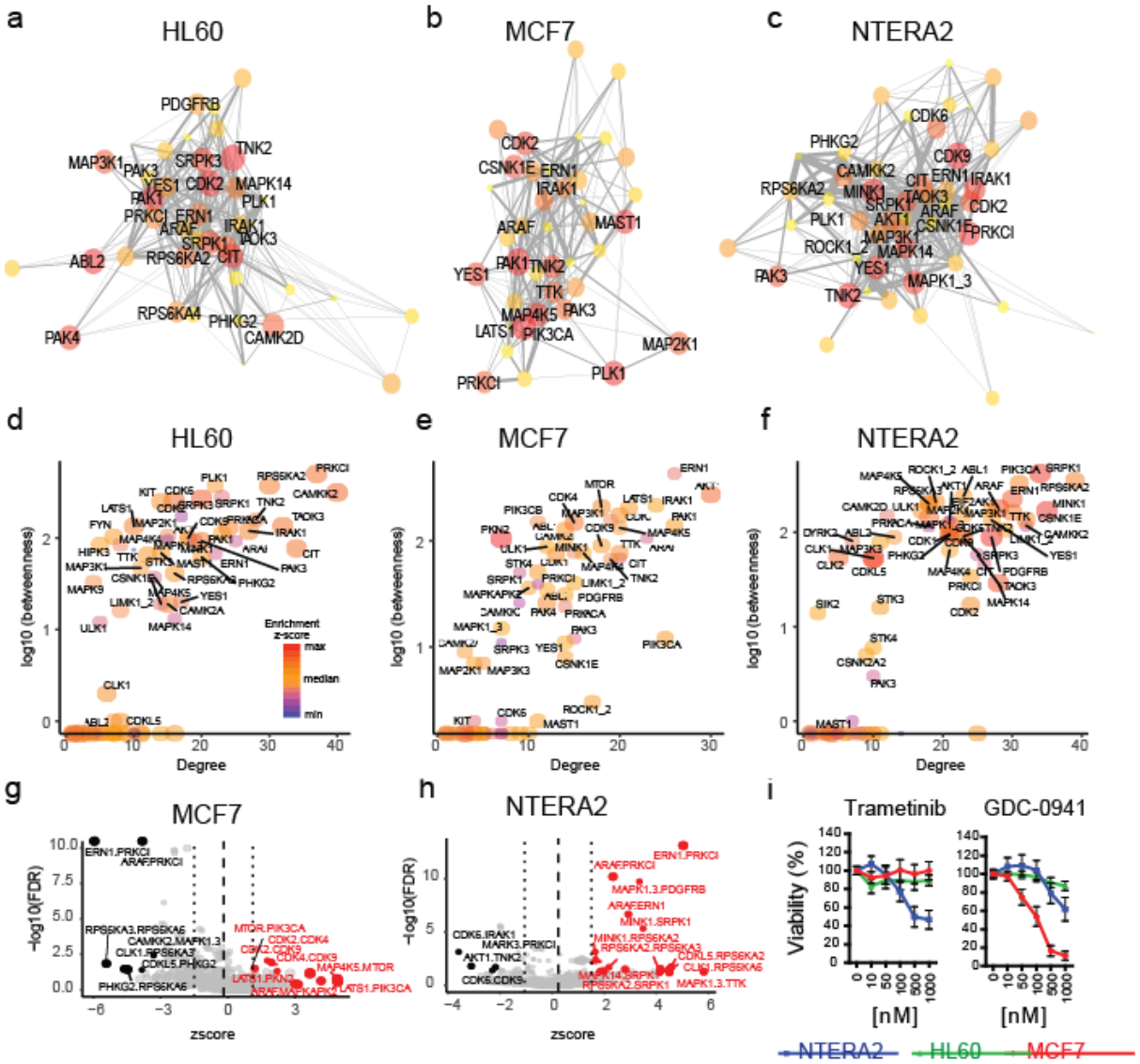

Figure 3. Comparative analysis of network circuitries identifies cell specific kinase network topologies

The figure illustrates approaches to compare kinase signaling circuitry across cellular models. A visual approach is shown in (**a-c**) and mathematical approaches in (**d-f**). The networks were constructed from phosphoproteomics data obtained from the mean of four independent experiments. (**a, b, c**) Kinase nodes showing greater values of centrality are labelled. (**d, e, f**) Betweenness and degree measures of centrality of kinase nodes in the networks shown in a-c panels. (**g, h**) Enrichment of kinase axes in MCF7 and NTERA2 cells; z-scores were calculated using the PAGE method and p-values using a two-sided Kolmogorov-Smirnov test.

FDR was calculated by adjusting p-values for multiple testing using the Benjamini-Hochberg method. These statistics were applied to the mean phosphoproteomics data of four independent experiments. For clarity not all edges are labelled. (**i**) Cell viability as a function of treatments with the named inhibitors at the concentrations shown. Data points are mean ± SD (n = 3 independent experiments). Trametinib is a MEK1/2 (MAP2K1 and MAP2K2) inhibitor and GDC-0941 inhibits PI3K.

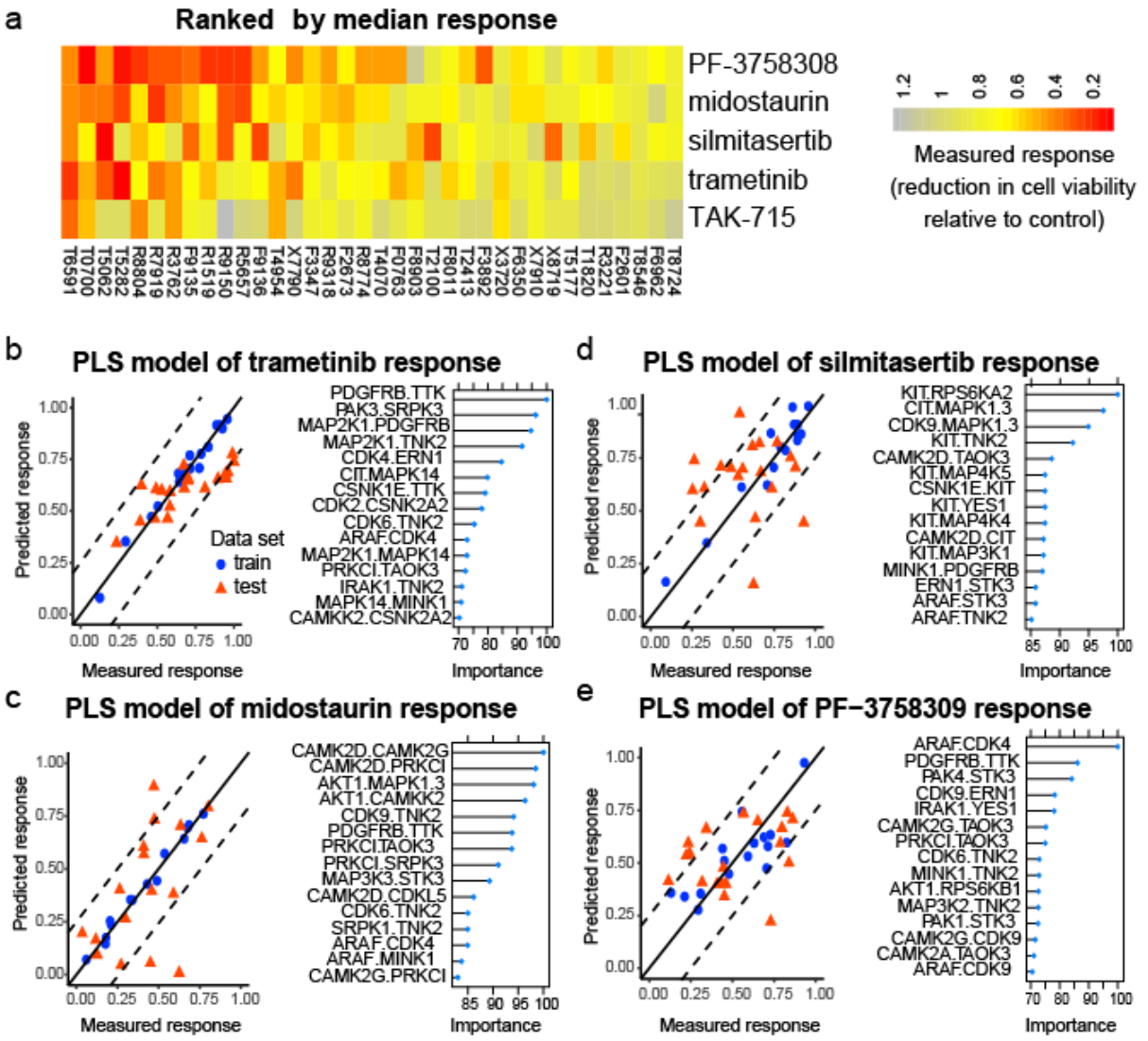

Figure 4. Multivariate regression models reveal kinase network topologies associated to drug responses

Responses of primary AML cells to treatment with four different kinase inhibitors was modelled using kinase network edge enrichment data, calculated for 36 primary AML cases using their published phosphoproteomes [15]. (**a**) Heat-map of drug response data calculated as reduction of cell viability as a function of treatment relative to DMSO control. (**b, c, d, e**) Partial least squares (PLS) models were constructed to associate network edges to drug response profiles without bias. PLS models, trained using data from 18 randomly selected AML cases, were tested in the remaining 18 cases. Left graphs compare predicted with actual responses. Right graphs illustrate the contribution (importance) of the top 15 edges to the models.

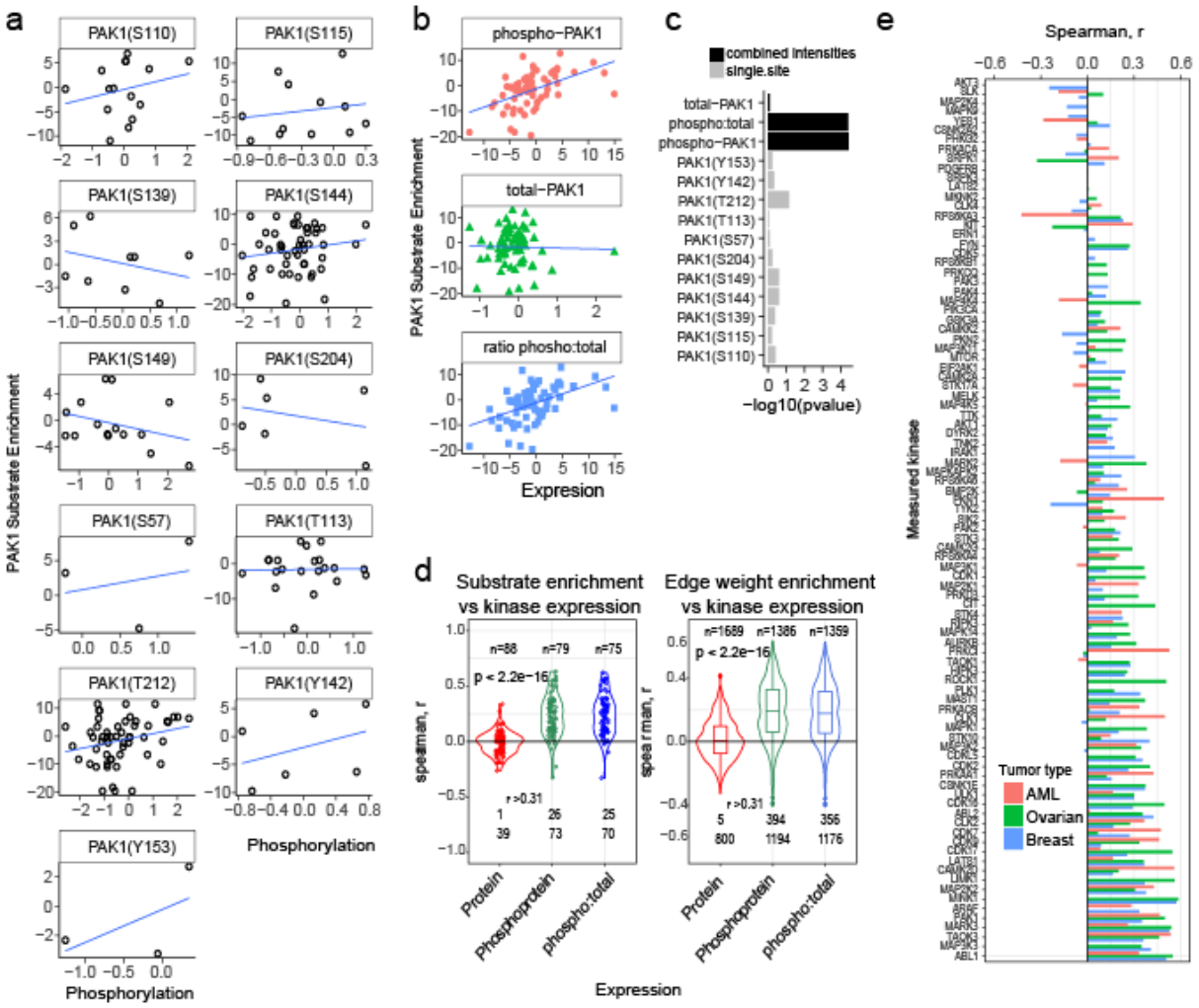

Figure 5. Contribution of kinase expression and phosphorylation to kinase activity and network topologies

(**a**) Phosphorylation extent of the named PAK1 residues versus PAK1 estimated activity across 69 ovarian tumors [50] (each data point represents the quantification of the named phosphosites in a tumor and thus, because of missing data points in [50], sample sizes are phosphosite specific). Note that not all peptides were identified in all samples. (**b**) As in (a) but correlation was between estimated PAK1 kinase activity and PAK1 protein expression (sum of unmodified PAK1 peptides), phosphorylation extent (sum of PAK1 phosphopeptides), or their ratio. (**c**) Statistical significance (by Spearman Rank test) of the associations shown in (a) and (b) (n = as in a, b). (**d**) Systematic analysis of association between kinase activity and kinase expression or phosphorylation. Numbers at the bottom of violin plots refer to positively

correlated edges (spearman $r > 0$, top row) or those with spearman $r > 0.31$ ($p<0.01$, bottom row). Violin plots show the kernel density of the data, in which probability is related to the width of the plot. Boxplots within violin plots (right graph) show the median, interquartile ranges, and max-min ranges. Statistical significance of differences between the spearman rank values for each analysis was calculated using a two-sided Kruskal-Wallis test. (**e**) Association between extent of kinase phosphorylation and estimated kinase activity across AML, breast and ovarian primary cancers (n = 36 for AML data, 83 for breast cancer and 69 for ovarian cancer).

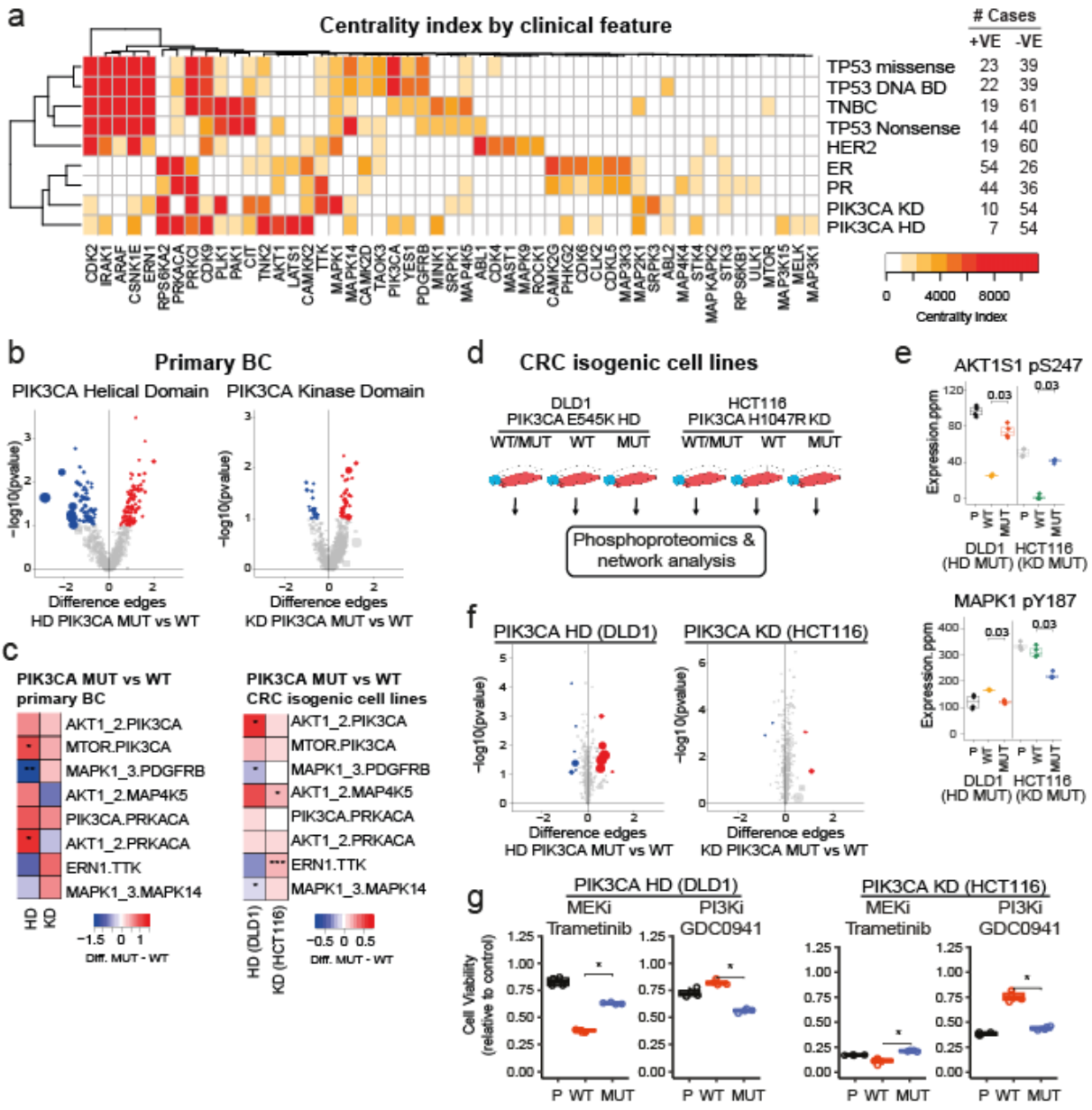

Figure 6. Network analysis of breast cancer identifies impact of genetic mutations on kinase network circuitries

(**a**) A centrality index was calculated as the product between betweenness and degree values for network nodes associated with the mutations shown. The phosphoproteomics data from which kinase topologies were determined were obtained from Mertins *et al* [31]. Sample sizes are shown in right-hand side table. (**b**) Difference in kinase network edge enrichment between breast cancer biopsies positive for *PIK3CA* mutations relative to wild-type cases. Fold difference was calculated as average in network edge weight in mutant positive tumors (n =

10 and 7 for KD and HD, respectively) minus the average in wild-type tumors (n = 54). P-values were calculated using the Wilcoxon test (two-sided). (**c**) Enrichment of edges in primary BC with HC or KD mutations relative to WT cases and CRC cell lines isogenic for *PIK3CA* mutations. (**d**) Experimental design for network analysis of CRC cell lines isogenic for *PIK3CA* mutations. (**e**) Analysis of phosphorylation sites makers of PI3K/AKT and MAPK pathway activities in *PIK3CA* mutant isogenic cell lines. AKT1S1 is also known as PRAS40 and MAPK1 gene encodes ERK2. Boxplots show median and interquartile ranges. P-values were calculated using the Wilcoxon test (two-sided) without adjustments for multiple comparison (n = 2 independent experiments performed in technical replicate). (**f**) Difference in edge enrichment in CRC cell lines as a function of the named mutations. Differences and p-values were calculated using Wilcoxon test (two-sided) as in (b) (n = 2 independent experiments performed in technical replicate). See **Supplementary Fig. 11** for further details. (**g**) The named cell lines were treated with 1 μM of the indicated compounds and cell viability measured relative to vehicle (DMSO) treated cells. Boxplots show median and interquartile ranges. P-values were calculated using a two-sided Wilcoxon test (n = 4 technical replicates). * $p < 0.05$.

Online Methods

Cell culture and inhibitor treatment

MCF7, NTERA2, DLD1 and HCT116 cells were routinely cultured using DMEM (10% FBS, 1% penicillin/streptomycin) at 37 ^{o}C in a humified atmosphere containing 5% CO_2. The culture medium for AML suspension cell lines (HL60, P31/Fuj, Kasumi-1, MV4-11, NOMO-1 and THP-1 cells) was RPMI 1640 (supplemented with 10% FBS, 1% penicillin/streptomycin). Each kinase inhibitor was diluted to 10,000 times the desired concentration for treatment using DMSO (vehicle). For phosphoproteomics experiments, each compound was then added to the cell population at a 1:1000 dilution in the culture medium. Cells were incubated with 1-10 μM of compound for one hour prior to lysis. The isogenic cell lines DLD1 and HCT116 with WT or mutant *PIK3CA* were provided by Dr Victor E. Velculescu and have been extensively characterized in previous publications [35, 51]. MCF7 (HTB-22), NTERA2 (CRL-1973), HL60 (CCL-240), Kasumi-1 (CRL-2724), THP-1 (TIB-202) and MV4-11 (CRL-9591) were purchased from ATCC. P31/Fuj was obtained from JCRB Cell Bank (JCRB0091). NOMO-1 was obtained from the DSMZ collection (ACC-542).

Cell lysis

After cell counting, MCF7, NTERA2, DLD1 and HCT116 cells were seeded in 100 cm^2 Petri dishes ($0.25x10^6$,$0.45x10^6$, $0.05x10^6$ and $0.2x10^6$cells/mL, respectively) and maintained in an incubator overnight at 37 ^{o}C and 5% CO_2. Cells were washed twice with cold PBS supplemented with 1 mM Na_3VO_4 and 1 mM NaF and lysed in 500 uL urea buffer (8 M urea in 20 mM in HEPES pH 8.0 supplemented with 1 mM Na_3VO_4, 1 mM NaF, 1 mM $Na_4P_2O_7$ and 1 mM sodium β-glycerophosphate). For HL60 and P31/Fuj, 10 mL of cell suspension were seeded in T25 flasks ($0.6x10^6$ and $1x10^6$ cells/mL, respectively) and maintained in an incubator overnight at 37 ^{o}C and 5% CO_2. Cell were harvested by centrifugation at 500 x*g* at 4 ^{o}C for 5 min, washed twice with cold PBS supplemented with 1 mM Na_3VO_4 and 1 mM NaF,

snap frozen and stored at -80 ºC until further processing. Cell pellets were lysed in urea buffer and cell lysates were homogenized by sonication for 10 cycles of 30s on 35s off (Diagenode Bioruptor® Plus, Liege, Belgium). Insoluble material was removed by centrifugation at 16,000 x*g* for 15 min at 4 ºC and protein in the cell extracts was quantified by bicinchoninic acid (BCA) analysis (Thermo Fisher Scientific; cat 23225).

Sample preparation for proteomics and phosphoproteomics analysis

For phosphoproteomic analyses, we used published methods [52-54] with some modifications. Drug treatment experiments were performed twice (biological duplicate) and each experiment was analyzed in technical duplicate). Briefly, 350 µg of protein were reduced and alkylated by sequential incubation with 10 mM DTT and 17 mM iodoacetamyde for 1h. The urea concentration was diluted to 2 M with 20 mM HEPES (pH 8.0) and 100 µL of conditioned trypsin beads (50% slurry of TLCK-trypsin) (Thermo Fisher Scientific; cat. 20230) conditioned with 3 washes of 20 mM HEPES (pH 8.0) were added and the samples incubated for 16 hours at 37 ºC with agitation. Trypsin beads were removed by centrifugation at 2,000 x*g* for 5 min at 4 ºC.

Following trypsin digestion, peptide solutions were desalted using 10 mg OASIS-HLB 1cc cartridges (Waters; cat. WAT094225). Briefly, OASIS cartridges were accommodated in a vacuum manifold (-5 mmHg), activated with 1 mL acetonitrile (ACN) and equilibrated with 1.5 mL washing solution (1% ACN, 0.1% TFA). After loading the samples, cartridges were washed twice with 0.75 mL of washing solution. Peptides were eluted with 250 µL of glycolic acid buffer 1 (1 M glycolic acid, 50% ACN, 5% TFA) for phosphoproteomic analysis or with 250 µL of ACN solution (30% ACN, 0.1% TFA) for proteomics analysis. Eluted peptides were, dried in a speedvac (RVC 2-25, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany) and stored at -80 ºC.

Phosphopeptides were enriched from total peptide digests using titansphere TiO_2 beads (GL Sciences; cat. 5020-75010) as previously described with some modifications [23]. Sample

volumes were normalized to 0.5 mL using glycolic acid buffer 2 (1 M glycolic acid, 80% ACN, 5% TFA), and 25 µL of TiO_2 beads (50% slurry in 1% TFA) were added to the peptide mixture, which was then incubated for 5 min at room temperature with agitation and centrifuged for 30s at 1,500 x*g*. For each sample, 80% of the supernatant was transferred to fresh tubes and stored in ice; the remaining 20% was used to re-suspend the bead pellets and these were loaded into an empty prewashed PE-filtered spin tips (Glygen; cat TF2EMT) and packed by centrifugation at 1,500 x*g* for 3 min. After loading the remaining volume of the supernatant by centrifugation at 1,500 x*g* for 3 min onto the packed spin tips, these were sequentially washed with 100 µL of glycolic acid buffer 2, ammonium acetate buffer (100 mM ammonium acetate in 25% ACN) and 10% ACN by centrifugation for 3 min at 1,500 x*g*. For phosphopeptide recovery, the addition of 50 µL of 5% ammonium water followed by centrifugation for 3 min at 1,500 x*g* was repeated 4 times. Eluents were snap frozen in dry ice, dried in a speed vac and peptide pellets stored at -80 ^{o}C.

Mass Spectrometry

For phosphoproteomics, peptide pellets were re-suspended in 18 µL of reconstitution buffer (3% ACN, 0.1% TFA, containing 20 fmol/µL digested yeast enolase) and 5 µL were loaded onto an LC-MS/MS system consisting of a Dionex UltiMate 3000 RSLC directly coupled to an Orbitrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). For proteomics, pellets were re-suspended in reconstitution buffer at a final concentration of 0.5 µg/µL and 2 µL were injected in the LC-MS/MS system, using mobile phases A (3% ACN; 0.1% FA) and B (100% ACN; 0.1% FA). Peptides were trapped in a µ-precolumn (Acclaim PepMap µ-Precolumns; cat. 160454) and separated in an analytical column (Acclaim PepMap 100 C18; cat. 164261). The following parameters were used: 3% to 23% B gradient for 60 min (phosphoproteomics) or 120 min (proteomics) at a flow rate of 0.3 µL/min. Samples were shuffled in the LC-MS/MS autosampler before loading.

As they eluted from the nano-LC system, peptides were infused into the online connected Q-Exactive Plus system operating with a 2.1s duty cycle. Acquisition of full scan survey spectra (m/z 375-1,500) with a 70,000 FWHM resolution was followed by, data-dependent acquisition in which the 15 most intense ions were selected for HCD (higher energy collisional dissociation) and MS/MS scanning (200-2,000 m/z) with a resolution of 17,500 FWHM. A 30 s dynamic exclusion period was enabled with an exclusion list with 10 ppm mass window. Overall duty cycle generated chromatographic peaks of approximately 30 s at the base, which allowed the construction of extracted ion chromatograms (XICs) with at least 10 data points. The mass spectrometry phosphoproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [55] partner repository with the dataset identifier PXD015943.

Peptide identification from tandem mass spectrometry data

Mascot Daemon 2.5.0 was used to automate peptide identification from MS data. Peak list files (MGFs) from RAW data were generated with Mascot Distiller v2.5.1.0 and loaded into the Mascot search engine (v2.5) **[56]**. The searches were performed against the SwissProt Database (SwissProt_Sep2014_2015_12.fasta) with a FDR of ~1% and the following parameters: 2 trypsin missed cleavages, mass tolerance of ±10 ppm for the MS1 scans and ±25 mmu for the MS/MS scans, carbamidomethyl of Cys as a fixed modification, PyroGlu on N-terminal Gln and oxidation of Met as variable modifications. For phosphoproteomics experiments Phosphorylation on Ser, Thr, and Tyr was also included as variable modifications.

Peptide quantification from MS1 data

Label-free quantification was performed as described before [12, 23, 54]. Pescal software [57] was used to automate the generation of extracted ion chromatograms (XICs) for all the peptides identified in at least one sample across all samples, after applying retention time alignment based on common peptide identifications in pair-wise run to run comparisons. Pescal functionality is similar to that provided by other packages [58-60]. XICs were constructed with ± 7

ppm and ± 1.5 min mass and retention time windows, respectively. Peak area values of 2 technical replicates and two biological replicates per condition were averaged after normalization to total sample peak area intensity.

Kinase Enrichment using ATP/ADP Desthiobiotin

Kinase enrichment with desthiobiotin-ATP and –ADP probes was performed using the ActivX kinase enrichment kits (Thermo Fisher Scientific; cat. 88310) according to the manufacturer's instructions. Eluents were dried in a speed vac and stored at -80 ^{o}C. Peptide pellets were resuspended in 20 µL of reconstitution buffer (20 fmol/µL digested yeast enolase in 3% ACN, 0.1% TFA) and 5 µL were loaded onto the LC-MS/MS system.

Western blotting

Cells were washed twice with ice-cold PBS (supplemented with 1 mM Na_3VO_4 and 1 mM NaF). Cells were scraped and the resulting lysate collected in 150 µL of RIPA buffer (50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1 mM Na_3VO_4, 1 mM NaF, 1X protease inhibitor cocktail, 1 mM PMSF and 0.5 µM okadaic acid. Cell lysates were sonicated for 10 minutes and centrifuged at 16,000 x*g* for 15 min at 4 ^{o}C. 55-60 µg of protein extract were analyzed in 4-12% precast commercial gels (NuPAGE Novex 4-12% Bis-Tris Midi Gel 1.0 mm). The buffer used for protein electrophoresis was NuPAGE MOPS SDS running buffer 20x. 10 mM DTT and NuPAGE LDS sample buffer 4x were used to prepare the samples. Electrophoresis was run at room temperature using a constant voltage. After electrophoresis, gels were washed in transfer buffer (10% methanol, 0.1% NuPAGE Antioxidant diluted in NuPAGE transfer buffer 20x) for 10 min. Then, the proteins separated were transferred to a nitrocellulose membrane (iBlot Gel Transfer Stacks Nitrocellulose) for 13 minutes, applying a constant voltage. After blocking for non-specific binding, the membranes were incubated for 14 h at 4 ^{o}C with primary antibody. Antibodies against GAPDH or β-Actin were used to quantify and normalize protein expression. Following this, after several washes membranes were incubated with secondary antibody

against mouse or rabbit immunoglobulin conjugated with peroxidase. Finally, membranes were incubated for 1 min with 1X SuperSignal West Pico ECL solution, which afforded a chemiluminiscence reaction. Antibody affinity was then visualized using a ChemiDoc system and the bands were quantified using Image Studio Lite (v5.2).

Cell Viability assay

AML suspension cell lines (HL60, P31/Fuj, Kasumi-1, MV4-11, NOMO-1 and THP-1 cells) were seeded in 96-well plates and treated with vehicle or 10 to 1000 nM of the indicated inhibitor for 72 hours. Trypsin was used for cell dissociation from adherent cell lines (MCF7, NTERA2, DLD1 and HCT116). Cells were stained with Guava ViaCount reagent (Merck Millipore; cat. 4000-0040) according to the manufacturer's instructions. Viability was measured with a Guava PCA cell analyzer (Guava Technologies Inc, CA, USA) to generate flow cytometry data, which were analyzed using CytoSoft (v2.5.7). Compound testing was performed in four technical replicates and three biological replicates. Viability and cell death values were averaged and expressed relative to vehicle control.

Kinase inhibitor selectivity datasets

The selectivity of a set of 20 kinase inhibitors was profiled using the DiscoverX platform [61]. Compounds were screened at 200 nM. Selectivity data for other compounds were obtained from the literature [21, 62]. Inhibitor selectivity data was normalized on a scale of 0 to 1 (where 0 is inhibition and 1 is no inhibition) with the expression $(x_i - \min(x)) / (\max(x) - \min(x))$, where x_i is the inhibitor data point for a given kinase and x is the vector of inhibition values for all kinases for a given compound. An inhibition value of 0 was given to kinase inhibition data points in cases when compounds are reported to inhibit such kinase but no actual degree of inhibition is reported in the literature (e.g., vendor information).

Bioinformatics

Peptides were parsed from Mascot DAT files using Python scripts and compared to data obtained from searches against decoy databases. Probability of false discovery (False discover rate, FDR) was determined for each peptide by comparing data to decoy searches at several thresholds of Mascot expectancy and mass accuracy windows. FDR values for each phosphopeptides are shown in **Supplementary Dataset 2**. The signals of phosphopeptides containing the same phosphorylation site were added and quantile normalized.

To apply the EBDT algorithm we selected > 50% *in vitro* inhibition as threshold for determining kinase inhibitor fingerprints, and two-fold decrease of phosphosite intensity as a function of kinase inhibitor treatment (with unadjusted $p<0.02$) as heuristic thresholds of significance as these thresholds led to false discovery <0.05 (as estimated with the analysis of randomly permuted data). After parameter tuning (**Supplementary Fig. 12**), final PDTs were assembled by selecting sites with kinase-site ratios >0.75 for kinases with rank >0.8. Highly homologous kinases, such as AKT1 and AKT2, and MAPK1 and MAPK3 (also known as ERK2 and ERK1, respectively) were grouped into single entities.

In addition to being submitted to PRIDE, the data is reported in a website that integrates prior knowledge from a range of sources including UniProt, and experimental data newly generated in this study from DiscoverX assays and phosphoproteomic experiments. To manage these disparate data, we created a relational database in SQLite (3.24.0) containing tables describing proteins (including kinases), perturbagens, substrates (including PDTs) and the relationships between these entities. The database was populated using bespoke software implemented in Python (3.7.0) that imports data, resolves issues of inconsistent nomenclature, and formats the data for insertion into the database. This software was subsequently extended to insert results of our analysis, notably the identified PDTs, into the database. A front end for the database was created using Python, Flask (1.0.2) and other libraries to allow the community to browse the database (see **Supplementary Fig. 2**).

Networks were constructed and analyzed using the iGraph R package (v. 1.2.2). Partial least squares (PLS) was performed using the caret R package [63] (v6.0-80) with resampling model building using randomly chosen 18 samples (LOOCV train control method) and RMSE metric. Model performance was tested using the remaining 18 samples.

Statistics

Statistical analyses were performed in R and Python. Unpaired, two-tail Student's t-test was used to assess significance in phosphoproteomics data. Where applicable, p-values were adjusted for multiple testing using Benjamini-Hochberg method.

Kinase and network edge enrichment was assessed by comparing the distributions of fold-changes (in Log2 scale) of phosphopeptides belonging to PDTs or network edge groups against the distribution of all phosphopeptide fold-changes [9, 12]. This was achieved by calculating z-scores (using the PAGE method [12, 23, 64], to evaluate magnitude of enrichment) and p-values (using the Kolmogorov-Smirnov test [65] to assess statistical significance), which were adjusted using the Benjamini-Hochberg method.

Motif enrichment in network edge groups was calculated as Log2[(*a/b*)/(*x/y*)] where: *a,* number of phosphorylation sites in edge group with a given motif; *b*, number of phosphorylation sites in edge group; *x*, total number of phosphorylation sites with motif; *y*, total number of phosphorylation sites. Enrichment of empirically determined phosphorylation sites in predicted substrates for kinases was done by matching to NetworKIN dataset version 3.1 [66]. Enrichment was calculated as (*a/b*)/(*x/y*), where *a*, number of PDTs for a given kinase in NetworKIN/Netphorest group; *b*, total number of PDTs for same kinase; *x*, number of NetworKIN/Netphorest substrates for kinase group; *y*, total number of NetworKIN/Netphorest substrates.

Life Sciences Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article

Data Availability

The mass spectrometry phosphoproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [55] with the dataset identifier PXD015943. Submission details:

Project Name: Chemical phosphoproteomics reveals kinase network topologies associated to genotypes and phenotypes of cancer cells.

Project accession: PXD015943

Project DOI: Not applicable

Figures with associated raw data: Figure 1, 2 and 6.

Additional availability from ChemPhoPro.org

Code Availability

Code is available from the GitHub repository: https://github.com/CutillasLab/ebdt

Methods-only References

51. Rajeeve, V., Pearce, W., Cascante, M., Vanhaesebroeck, B. & Cutillas, P.R. Polyamine production is downstream and upstream of oncogenic PI3K signalling and contributes to tumour cell growth. *The Biochemical journal* **450**, 619-628 (2013).
52. Gruhler, A. et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* **4**, 310-327 (2005).
53. Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. & Jorgensen, T.J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **4**, 873-886 (2005).
54. Montoya, A., Beltran, L., Casado, P., Rodriguez-Prados, J.C. & Cutillas, P.R. Characterization of a TiO(2) enrichment method for label-free quantitative phosphoproteomics. *Methods* **54**, 370-378 (2011).
55. Vizcaino, J.A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic acids research* **44**, D447-456 (2016).
56. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
57. Cutillas, P.R. & Vanhaesebroeck, B. Quantitative profile of five murine core proteomes using label-free functional proteomics. *Molecular & cellular proteomics : MCP* **6**, 1560-1573 (2007).
58. Tsou, C.C. et al. IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Molecular & cellular proteomics : MCP* **9**, 131-144 (2010).
59. Bateman, N.W. et al. Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Molecular & cellular proteomics : MCP* **13**, 329-338 (2014).
60. Lawrence, R.T., Searle, B.C., Llovet, A. & Villen, J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat Methods* **13**, 431-434 (2016).
61. Fabian, M.A. et al. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* **23**, 329-336 (2005).
62. Elkins, J.M. et al. Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat Biotechnol* (2015).
63. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1-26 (2008).
64. Kim, S.Y. & Volsky, D.J. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* **6**, 144 (2005).
65. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
66. Horn, H. et al. KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods* **11**, 603-604 (2014).