Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities

John D Lapek Jr, Patricia Greninger, Robert Morris, Arnaud Amzallag, Iulian Pruteanu-Malinici, Cyril H Benes & Wilhelm Haas

The formation of protein complexes and the co-regulation of the cellular concentrations of proteins are essential mechanisms for cellular signaling and for maintaining homeostasis. Here we use isobaric-labeling multiplexed proteomics to analyze protein co-regulation and show that this allows the identification of protein–protein associations with high accuracy. We apply this ‘interactome mapping by high-throughput quantitative proteome analysis’ (IMAHP) method to a panel of 41 breast cancer cell lines and show that deviations of the observed protein co-regulations in specific cell lines from the consensus network affects cellular fitness. Furthermore, these aberrant interactions serve as biomarkers that predict the drug sensitivity of cell lines in screens across 195 drugs. We expect that IMAHP can be broadly used to gain insight into how changing landscapes of protein–protein associations affect the phenotype of biological systems.

The proteome forms a link between genotype and phenotype, and its exploration provides a wealth of information about the molecular mechanisms that regulate cellular events. Mass spectrometry has evolved as the key technology to characterize the proteome by measuring protein abundances, post-translational modifications, as well as interactions between proteins. The interactions of a protein reveal its functional network, and the mapping of all protein–protein interactions in a cell—the interactome—as well as their dynamics, will offer unique insights into biological systems and their reaction to perturbations. Major efforts are underway to generate global protein–protein interaction maps using the yeast two-hybrid (Y2H) assay or protein affinity purification coupled to mass spectrometry (AP-MS).

However, generating a static interaction catalog of a comprehensive protein–protein interaction network represents a substantial experimental effort, and comprehensively studying network dynamics after making a perturbation currently seems out of reach. Here we report the IMAHP technology that uses protein co-regulation analysis to map protein–protein associations and their dysregulation. We further show that interactome dysregulations can enable the identification of cancer vulnerabilities—proteins whose depletion negatively affects the fitness of a specific cancer cell line—and sensitivity to drugs.

We used multiplexed quantitative mass-spectrometry-based proteomics technology, in which we applied isobaric labeling technology with ten-plex tandem mass tag (TMT) reagents, to generate quantitative proteome profiles of 41 breast cancer cell lines that represent the majority of breast cancer subtypes (Supplementary Table 1). A total of 82 proteome samples from two biological replicates were analyzed in 11 experiments, each of which enabled the simultaneous quantification of ten samples (Fig. 1a). Data were acquired on an Orbitrap Fusion mass spectrometer, using the SPS-MS3 method to eliminate ratio distortions that are known to negatively affect the accuracy and reproducibility of quantitative proteomics data acquired using multiplexed isobaric labeling technology. A total of 10,535 proteins were quantified across all 11 experiments, and on average 9,115 proteins were quantified across the two replicate analyses of each cell line (Fig. 1b and Supplementary Table 2) while requiring less than 10 h of data acquisition time per cell line.

A total of 6,911 proteins were quantified in all of the cell lines, and subsequent analyses were performed on this subset (Supplementary Table 2). We clustered the cell lines on the basis of the Spearman’s rank correlation coefficients for the proteome profiles and found that they clearly separated into the known breast cancer subtypes—luminal, basal, claudin-low and nonmalignant (Fig. 1c and Supplementary Fig. 1). We also found that proteome-based clustering was concordant with mRNA-level-based clustering (Supplementary Fig. 2).

The median Spearman’s rank correlation coefficient (ρ) between proteome profiles from biological replicates of the same cell line was 0.82, confirming a high reproducibility rate for the multiplexed proteome quantification technology (Supplementary Fig. 3a). The median correlation coefficient between the mRNA and proteome profiles was 0.58 (Supplementary Fig. 3b,c) which was slightly higher than that reported in other studies. We performed the analysis on 36 cell lines for which there were published RNA sequencing (RNA-seq) data available (Supplementary Table 3 and Supplementary Data 1 and 2).

Next we investigated whether there was any difference between the mRNA and protein data with respect to permitting the identification of functional and physical protein–protein associations through co-regulation analysis across the 36 cell lines. An example of co-regulation
for the two proteasome subunits PSMB1 and PSMB2 is shown in Figure 1d. We observed a high correlation between the protein levels for these known interactors (Spearman’s ρ = 0.80) but very low correlation between mRNA levels (ρ = 0.08). We performed this analysis for each pair of the 6,659 gene products for which we had data points in both the proteomics and RNA-seq data sets (Supplementary Table 4). We applied a very strict Benjamini–Hochberg (BH)-corrected P ≤ 5 × 10⁻⁴ filter and considered only positive correlations. Correlation-inferred associations are shown in a network form in Figure 1e, in which nodes represent proteins or mRNA molecules, and edges represent statistically significant associations. We observed 5,748 significant associations among 2,494 mRNA molecules and 7,086 associations among 2,122 proteins (Fig. 1e). Notably, only 431 significant associations between mRNA and proteins encoded by the same gene were found in the overlap of both data sets (<8%).

To estimate the accuracy of the mRNA- and protein-derived networks, we used high-confidence associations (score ≥ 0.700) in the STRING database—a compendium of experimentally determined, as well as predicted functional protein associations including physical interactions—as a benchmark. We found 2,953 (42%) of the proteome-based associations confirmed by the STRING database but only 250 (4%) of the associations derived from the mRNA data set (Fig. 1e).
An increased relative number of known associations in the proteome-derived data set was confirmed for several precision thresholds (Supplementary Fig. 4 and Supplementary Table 5), indicating that co-regulation analysis applied on proteome profiles has a substantially higher predictive power to identify functional protein–protein associations than co-regulation analysis applied on transcriptome profiles. These results are supported by a recent report on gene-function prediction through co-regulation analysis of proteome and mRNA profiling data on tumors of three cancer types13 and by co-regulation of functionally associated proteins in yeast 14 and mice15.

To further explore the proteome-inferred network, we used co-regulation analysis on the profiles from all 41 cell lines (the 36 lines considered above and an additional five for which mRNA data were not available), which revealed 14,909 associations among 3,024 proteins (BH-corrected $P$ value $\leq 5 \times 10^{-4}$; Supplementary Table 6 and Supplementary Data 3). By systematically annotating the correlation data set for known physical protein–protein associations, we were able to assign 143 unique complexes from the 'comprehensive resource of mammalian protein complexes' (CORUM) database of high-confidence protein complexes16 (Fig. 2 and Supplementary Table 7).

The median coverage of CORUM-defined components in the complexes with associations observed in our data set was 67%; for 112 of the complexes, we identified associations between at least 90% of the components. These data show that protein co-regulation analysis is a useful tool for the detection of associations of proteins in multiprotein complexes. Of the 14,909 protein–protein associations, 4,179 (28%) were attributed to protein complexes that are defined in the CORUM database (Supplementary Table 6). We found that there were 5,149 (35%) observed associations previously defined as high-confidence associations in the STRING database, of which 3,032 overlapped with associations defined by CORUM. High-confidence protein–protein associations from the CORUM and the STRING
databases confirmed 6,296 (42%) of the 14,909 associations, and 8,613 (58%) associations in this stringently filtered data set were found to not yet have been reported (Supplementary Table 6). In support of the validity of these unreported associations, we found that 3,636 of them were linked through an indirect (one step removed) STRING interaction (Supplementary Fig. 5).

We further validated the previously unreported associations by comparing our data with those of large-scale interactome screens using AP-MS\(^3,4\). We confirmed both known and novel associations by this comparison (Supplementary Fig. 6). We found that 18% of the known and 3% of the novel associations identified in our study were confirmed in the AP-MS-based Bioplex data set\(^3\) when only considering associations in which both proteins were identified in both studies. We compared another large-scale AP-MS data set\(^4\) with the Bioplex data in the same manner and found a similar overlap: 22% for known associations and 4% for novel associations.

![Figure 3](image-url)

**Figure 3** Deviations of protein co-regulation allow the identification of cell-line-specific dysregulation of protein–protein associations, revealing cell-specific vulnerabilities. (a) Two-dimensional outlier detection based on the Mahalanobis distance was used to define cell-line-specific deviations of protein co-regulation for all 14,909 identified protein–protein associations in the 41 cell lines studied. Such a deviation (dysregulation) was found for the known association between the proteins THOC1 and THOC2 in the MDAMB157 cell line (left; orange circle). The dysregulation was not seen on the basis of a deviation of mRNA-level co-regulation (right; orange circle). THOC2 carries a substitution (Arg1307Trp) in this cell line (http://cansar.icr.ac.uk), which may cause the dysregulation of the association between the two proteins. (b) Proteins with dysregulated associations are highlighted (red) on the background of the determined protein–protein association network (see Fig. 2) for four cell lines (see also Supplementary Table 8), indicating variations in both the number of associations and network areas affected by the dysregulation. (c) Scatter plots showing the correlation of the number of a cell line’s dysregulated protein–protein associations with the number of mutated genes (top), and with the number of genes affecting the cell line’s fitness (bottom), as determined from a genome-wide genetic dropout screen using pooled shRNAs (zGARP ≤ −2)\(^19\). Log2 ratios of observed values to the average value are plotted, and correlations are given as a Spearman’s rank correlation coefficient. The comparison was made for 26 cell lines for which proteome, mutation and dropout screen data were available. (d) Plot showing the relative number of proteins whose depletion affects the cell line’s fitness, in the whole genome\(^19\) and in the network of proteins carrying dysregulated protein–protein associations. Fitness proteins are enriched in the dysregulated protein–protein association network for almost all of the 26 cell lines compared. The average enrichment increase is 1.6-fold (9.9% compared to 6.3%).
To explore how differences in protein networks across cell lines could inform the biology of these models of breast cancer, we sought to identify cell-line-specific network dysregulation. We used deviations from the co-regulation observed across the 14,909 significant correlated protein pairs identified using the data set from the 41 cell lines. We performed bivariate outlier analysis on each of the protein pairs on the basis of the Mahalanobis distance, followed by the Grubbs test, to determine outliers ($P \leq 0.1$) that corresponded to cell lines with a putative dysregulated protein pair. For example, we found such a deviation in the cell line MDAMB157 for the association between the two proteins THOC1 and THOC2, which are components of the TREX complex that is involved in the regulation of transcription, mRNA processing and export17 (Fig. 3a). THOC2 carries a substitution (Arg1307Trp) in this cell line, which could underlie this effect by inhibiting the binding between THOC1 and THOC2, and lead to degradation of THOC1 (ref. 18 and http://cansar.icr.ac.uk/). Consistent with the idea of protein-level regulation rather than a change in mRNA levels, THOC1 and THOC2 mRNA levels across the cell lines did not identify MDA-MB-157 as an outlier (Fig. 3a).

By applying this outlier principle to all globally identified protein–protein associations across all of the cell lines, we observed a wide range of dysregulated associations across cell lines, which ranged from 20 dysregulated associations in ZR751 (0.1%) to 800 dysregulated associations in HS578T (5.9%) that affected many different large complexes (Fig. 3b, Supplementary Figs. 7–9 and Supplementary Table 8). We term this strategy of interactome mapping by high-throughput quantitative proteome analysis the IMAHP technology.

To test the significance of our findings, we used the data from a genome-wide short hairpin RNA (shRNA)-based dropout screen on breast cancer cell lines19 (Supplementary Table 9) to evaluate the functional consequences of dysregulated protein–protein associations. We analyzed the data of 26 cancer cell lines for which whole-genome sequencing data and dropout screen data were available (ref. 19 and http://cansar.icr.ac.uk/). We found a significant positive correlation between the number of dysregulated protein–protein associations in each cell line with the number of proteins whose depletion affects the cell line’s fitness ($P = 0.4$; Online Methods; Fig. 3c). Notably, there was no correlation between the number of mutant proteins (Supplementary Table 10) and the number of genes associated with fitness (hereafter referred to as fitness genes) ($P = 0.08$) (Fig. 3c), possibly in part because only a limited number of mutations have functional consequences20.

Furthermore, in almost all of the cell lines, products encoded by the fitness genes were enriched in the group of proteins with dysregulated protein–protein associations, as compared to all 15,309 genes that were monitored in the dropout screen (Fig. 3d)19. The average enrichment in fitness proteins was 64%. These results show that high-throughput mapping of dysregulated protein–protein associations in cancer cell lines using the IMAHP strategy can be used to reveal the vulnerabilities of the cancer cell lines with high efficiency.

We next mined for differences in functional modules that were dysregulated in the basal- and luminal-subtype cancer cell lines by analyzing data from all of the 41 cell lines studied. We identified 167 proteins with dysregulated associations in at least 25% of either luminal- ($n = 17$) or basal-subtype ($n = 24$) cell lines, with a significant difference between the two subtypes having been determined by hypergeometric testing ($P \leq 0.1$; Fig. 4a and Supplementary Table 11). Ten of these proteins are encoded by known cancer-associated genes: CASC5, ERBB3, EZH2, POLE, MET, TPX2 and SUZ12 with dysregulations enriched in basal-subtype cell lines, and ERCC3, RPS2 and SMARCE1 with dysregulations mainly in the luminal-subtype cell lines (refs. 21,22 and http://cancer.sanger.ac.uk/census/). We used the DAVID bioinformatics platform23 for Gene Ontology (GO) category analysis of the proteins that were enriched in the dysregulated protein–protein association network of basal-subtype (117 proteins) and luminal-subtype (49 proteins) cell lines (Fig. 4a). This analysis revealed that 32 cell-cycle-regulating proteins were affected to diverse extents in mainly the basal cell lines, whereas ten mitochondrial ribosomal proteins showed dysregulated associations in mainly the luminal cell lines (Fig. 3a and Supplementary Table 11).

To evaluate whether the dysregulated functional modules could predict how affected cells responded to drugs, we determined the responses of the 41 cell lines to 195 drugs that affected a wide range of targets (Supplementary Table 12). We identified six therapeutics that produced a significantly higher response in cell lines with a dysregulated mitotic cell cycle (≥2 mitotic-cell-cycle-related proteins with disturbed associations) than in cell lines with unaffected cell-cycle-regulating proteins (significance value was a z-value ≥ 2, considering all P values of drugs with a higher response in the affected cell lines; P values were calculated by a two-tailed Student’s t-test with unequal variance) (Fig. 4b). These included two inhibitors, NPK76 ($P = 4 \times 10^{-3}$) and BI-2536 ($P = 6 \times 10^{-3}$), of polo-like kinases, which are known as important cell cycle regulators. We also found another three drugs that have nominal kinase targets not directly linked to the cell cycle: AZD1480 (targeting JAK2, $P = 1 \times 10^{-9}$), XL-880 (targeting MET, $P = 1 \times 10^{-3}$) and TPCA-1 (targeting IKK, $P = 3 \times 10^{-3}$); however, all three of them have been shown to potentially inhibit Aurora kinases24–26, and XL-880 (also known as foretinib) also inhibits polo-like kinase 4 (PLK4)26. Another therapeutic, ponatinib, targets the kinase ABL1 but is known to inhibit a wide range of kinases27. We also identified dysregulated protein–protein associations that encompassed MET enrichment in basal-subtype cell lines (Supplementary Table 11).

Across the 195 drugs tested, the six causing a significantly stronger response in cell lines with dysregulation of the mitochondrial ribosome protein complex included phenformin ($P = 5 \times 10^{-2}$), which blocks mitochondrial respiration through inhibition of complex I28, atpenin A5 ($P = 1.1 \times 10^{-2}$), a mitochondrial complex II inhibitor29, and oligomycin ($P = 2 \times 10^{-2}$), an inhibitor of ATP synthase30 (Fig. 4c). We defined a z-score of ≥2, considering all P values of drugs with higher responses in the affected cell lines, as a cut-off value for significance. Taken together, these results strongly suggest that predicted dysregulations of functional cellular modules based on deviation from global co-regulation networks is a potentially useful approach in the identification of drug susceptibilities.

In summary, we have shown that when studying cancer cell lines, protein co-regulation analysis allows for the identification of functional protein–protein associations with an accuracy tenfold higher than that when using RNA-seq data for co-expression analysis. The high level of correlation that allows for identification of protein complexes using relative expression levels across samples implies a stringent control of protein levels in cells. A likely explanation is that protein degradation leads to appropriate protein concentrations in accordance with the functional network. This is concordant with studies showing that proteins from multiprotein complexes are degraded at a higher rate if they are not embedded into their cognate complexes18. In this model, the differential stability of proteins when they are part of their functional complex or when free is linked to appropriate stoichiometry. In keeping with our results, studies on the effects of aneuploidy in yeast and human cell lines have implicated protein degradation in the accurate control of protein levels for complexes between products of genes that are affected or not by genomic duplication events31,32. As shown in Supplementary Figure 10 for the well-studied 26S proteasome
Multiprotein complex, mRNA but not protein levels in our data set correlated well with gene copy-number variations (CNVs) (Supplementary Table 13). These results are also supported by recent reports comparing CNVs, mRNA and protein levels in colon and breast cancer tumors. The CNV-driven anomalies in mRNA levels may also partially explain why mRNA co-regulation analysis is not as predictive of functional relations as proteomic co-regulation.

We believe that the increased stability of proteins when they are embedded in complexes, as compared to that in their dissociated state, is the basis for deviations of the co-regulation of two proteins observed in individual cell lines, i.e., when the interaction of a protein with its partners is perturbed either through a mutation or other dysregulations, the protein is subjected to enhanced degradation as compared to that of its regular binding partners. It should be noted...
that the correlation between mRNA and protein levels was positive for all of the cell lines studied (median = 0.58; Supplementary Fig. 3c), indicating that, overall, small differences between mRNA and protein levels underlie the diverging results from co-regulation analysis. This is consistent with the similarity in clustering of the cell lines based on their mRNA and protein profiles (Supplementary Fig. 2). Thus, in addition to transcriptional co-regulation, a very accurate if often minor posttranscriptional adjustment of protein levels allows the use of protein co-regulation analysis to identify interactions between proteins.

By using the IMAHP strategy cell-line-specific deviations of co-regulation of protein pairs can be used to identify dysregulations of protein–protein associations, as well as cellular vulnerabilities, as revealed here by leveraging large RNAi-based and drug-response data sets. The high-throughput capability of the described mapping of protein–protein association dysregulations and its applicability to a wide range of biological samples makes this method a promising tool for a broad number of applications in cell biology and cancer research studies.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank S. Gygi (Harvard Medical School) for access to computational software and facilities to process the proteomics data. We are grateful to J. Boisvert for her help with culturing of the cell lines and to all members of the Haas and Benes laboratories for valuable discussions. C.H.B. is supported by grant 102696 from the Swiss National Science Foundation. G.C.M. was supported by grants P30CA087488 and P30CA094516 from the National Institutes of Health (NIH). T.O. is supported by the Hemophilia & Thrombosis Research Society. C.H.B. is supported by grant 102696 from the Swiss National Science Foundation. W.H. is supported by grants 1R01HL125768, 1R01HL125768, 1P30CA087488, and P50CA083638 from the National Institutes of Health (NIH). The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

C.H.B. and W.H. conceived and designed the study; I.D.L., C.H.B. and W.H. wrote the manuscript; J.D.L. and W.H. performed the proteomics experiments; P.G. and C.H.B. performed the drug screen and analyzed the drug screen data; and J.D.L., R.M., A.A., L.P., C.H.B. and W.H. performed the analysis of the proteomics data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. Nature 537, 347–355 (2016).
2. Rolland, T. et al. A proteome-scale map of the human interactome network. Cell 159, 1212–1226 (2014).
3. Huttlin, E.L. et al. The BioPlex network: a systematic exploration of the human interactome. Cell 162, 425–440 (2015).
4. Hein, M.Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell 163, 712–723 (2015).
5. McAlister, G.C. et al. Increasing the multiplexing capacity of TMTs using reporter ion isologotypes with isobaric masses. Anal. Chem. 84, 7469–7478 (2012).
6. Heiser, L.M. et al. Subtype- and pathway-specific responses to anticancer compounds in breast cancer. Proc. Natl. Acad. Sci. USA 109, 2724–2729 (2012).
7. Ting, L., Rad, R., Gigi, S.P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat. Methods 8, 937–940 (2011).
8. McAlister, G.C. et al. Multiplex MS3 enables accurate, sensitive and multiplexed detection of differential expression across cancer cell line proteomes. Anal. Chem. 86, 7150–7158 (2014).
9. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527 (2006).
10. Zhang, B. et al. Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387 (2014).
11. Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat. Biotechnol. 33, 306–312 (2015).
12. Szklarczyk, D. et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39, D561–D568 (2011).
13. J. et al. Proteme profiling outperforms transcriptome profiling for co-expression-based gene function prediction. Mol. Cell. Proteomics http://dx.doi.org/10.1074/mcp.M116.063001 (2016).
14. Stelfely, J.A. et al. Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. Nat. Biotechnol. 34, 1191–1197 (2016).
15. Chick, J.M. et al. Defining the consequences of genetic variation on a proteome-wide scale. Nature 534, 500–505 (2016).
16. Ruepp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res. 38, D497–D501 (2010).
17. Heath, C.G., Vighakone, N. & Wilson, S.A. The role of TREX in gene expression and disease. Biochem. J. 473, 2911–2935 (2016).
18. Goldberg, A.L. & Dice, J.F. Intracellular protein degradation in mammalian and bacterial cells. Annu. Rev. Biochem. 43, 829–869 (1974).
19. Marcotte, R. et al. Functional genomic landscape of human breast cancer drivers, vulnerabilities and resistance. Cell 164, 293–309 (2016).
20. Berger, A.H. et al. High-throughput phenotyping of lung cancer somatic mutations. Cancer Cell 30, 214–226 (2016).
21. Lawrence, M.S. et al. Discovery and saturation analysis of cancer genes across 21 tumor types. Nature 505, 495–501 (2014).
22. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54 (2016).
23. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
24. Ioannidis, S. et al. Discovery of 5-chloro-N2-(11β-1-(5-fluoropyrimidin-2-yl)ethyl)-N4-(5-methyl-1H-pyrazol-3-yl)pyrimidine-2,4-diamine (AZD1480) as a novel inhibitor of the Jak-Stat pathway. J. Med. Chem. 54, 262–276 (2011).
25. Jester, B.W., Gaj, A., Shomin, C.D., Cox, K.J. & Ghosh, I. Testing the promiscuity of commercial kinase inhibitors against the AGC kinase group using a split-luciferase screen. J. Med. Chem. 55, 1526–1537 (2012).
26. Davis, M.I. et al. Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol. 29, 1046–1051 (2011).
27. O’Hare, T. et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. Cancer Cell 16, 401–412 (2009).
28. El-Mir, M.Y. et al. Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex i. J. Biol. Chem. 275, 223–228 (2000).
29. Miyadera, H. et al. Atpenins, potent and specific inhibitors of mitochondrial complex II (succinate-ubiquinone oxidoreductase). Proc. Natl. Acad. Sci. USA 100, 473–477 (2003).
30. Kagawa, Y. & Racker, E. Partial resolution of the enzymes catalyzing oxidative phosphorylation. B. Properties of a factor conferring oligomycin sensitivity on mitochondrial adenosine triphosphatase. J. Biol. Chem. 241, 2461–2466 (1966).
31. Dehoux, N. et al. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife 3, e03023 (2014).
32. Stangele, S. et al. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. 8, 608 (2012).
33. Mertins, P. et al. Proteogenomics connects somatic mutations to signaling in breast cancer. Nature 534, 95–102 (2016).
ONLINE METHODS

Cell culture and lysis. Breast cancer cell lines were grown to 90% confluency under the indicated culture conditions (Supplementary Table 1). For cell lysis of adherent cells, growth medium was removed, and cells were rinsed with PBS before being trypsinized to remove them from the growth plastic. Cells were then counted, and 3.0 × 10⁶ cells were transferred to a new tube and pelleted. Medium was aspirated, and cells were washed with PBS. After re-pelleting the cells, the PBS was aspirated, the cells were fast-frozen on dry ice and then stored at −80 °C until lysis. For cells growing in suspension, cells were pelleted and resuspended before counting. Again, 3.0 × 10⁶ cells were pelleted, rinsed and frozen as described above.

Cells were lysed with 200 µL of lysis buffer (75 mM NaCl, 3% SDS, 1 mM NaF, 1 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF and 1× Roche Complete Mini EDTA-free protease inhibitors in 50 mM HEPES, pH 8.5) by passing the suspension through a 21-gauge needle 20 times. Lysates were then sonicated for 5 min in a sonicating water bath before the cellular debris was pelleted by centrifugation at 14,000 r.p.m. for 5 min.

Protein digestion and TMT labeling. Protein concentration of the cell lysates was determined with a BCA assay (Thermo Scientific). Proteins were then reduced with DTT and alkylated with iodoacetamide as previously described 35. Reduced and alkylated proteins were precipitated via methanol–chloroform precipitation 34. The precipitated proteins were reconstituted in 300 µL of 1 M urea in 50 mM HEPES, pH 8.5. Vortexing, sonication and manual grinding were used to aid solubility. The solubilized protein was digested in a two-step process starting with overnight digestion at room temperature with 3 µg of Lys-C (Wako) followed by 6 h of digestion with 3 µg of trypsin (sequencing grade, Promega) at 37 °C. The digest was acidified with trifluoroacetic acid (TFA). The digested peptides were desalted with C18 solid-phase extraction (SPE) (Sep-Pak, Waters) as previously described 37. The concentration of the desalted peptide solutions was measured with a BCA assay, and peptides were aliquoted into 50-µg portions, which were dried under vacuum and stored at −80 °C until they were labeled with the TMT reagents.

Peptides were labeled with TMT reagents (Thermo Scientific) 30, 38, in principle as previously described 37. The TMT reagents were suspended in dry acetonitrile (ACN) at a concentration of 20 µg/µL. Dried peptides (50 µg) were resuspended in 30% ACN in 200 mM HEPES, pH 8.5, and 5 µL of the appropriate TMT reagent was added to the sample. TMT reagents 126 and 131 were reserved for ‘bridge’ samples (see below); the remaining TMT reagents (127c, 127n, 128c, 128n, 129c, 129n, 130c and 130n) were used to label the digests from the individual cell lines in a random order. Peptides were incubated with the reagents for 1 h at room temperature. The labeling reaction was quenched by adding 6 µL of 5% hydroxylamine. Labeled samples were then acidified by adding 50 µL of 1% TFA, and the peptide mixtures were pooled into 11 ten-plex TMT samples (Supplementary Table 1), with the bridge samples carrying the 126 and 131 labels. The pooled samples were desalted via C18 SPE on Sep-Pak cartridges as described above.

Basic pH reversed-phase liquid chromatography (bRPLC) sample fractionation. Sample fractionation was performed by bRPLC 39, and fractions were pooled for analysis by mass spectrometry as previously described 37. Briefly, the samples were resuspended in a solution with 5% formic acid and 5% ACN and separated over a 4.6-mm × 250-mm ZORBAX Extend C18 column (5 µm, 80 Å, Agilent Technologies) on an Agilent 1260 HPLC system outfitted with a fraction collector, degasser and variable-wavelength detector. The separation was performed by applying a gradient build from 22 to 35% ACN in 10 mM ammonium bicarbonate in 60 min at a flow rate of 0.5 mL/min. A total of 96 fractions were combined as previously described 37. The combined fractions were dried under vacuum, reconstituted with a solution of 5% formic acid and 5% ACN, and then analyzed by LC-MS2/MS3 for identification and quantification.

Liquid chromatography coupled to mass spectrometry. All LC-MS2/MS3 experiments were conducted on an Orbitrap Fusion (Thermo Fisher Scientific) coupled to an Easy-nLc 1000 (Thermo Fisher Scientific) with a chilled autosampler. Peptides were separated on an in-house-pulled, in-house-packed microcapillary column (inner diameter, 100 µm; outer diameter, 360 µm). Columns were packed first with ~0.5 cm of Magic C18 resin (5 µm, 100 Å, Michrom Bioresearches) followed by ~0.5 cm of Maccel C18 QA resin (5 µm, 200 Å; Nest Group) and then to a final length of 30 cm with GP-C18 (1.8 µm, 120 Å; Sepax Technologies). Peptides were eluted with a linear gradient from 11 to 30% ACN in 0.125% formic acid over 165 min at a flow rate of 300 nL/min while the column was heated to 60 °C. Electrospray ionization was achieved by applying 1,800 V through a PEEK T-junction at the inlet of the microcapillary column.

The Orbitrap Fusion was operated in data-dependent mode, with a survey scan performed over an m/z range of 500–1,200 at a resolution of 6 × 10⁶ in the Orbitrap. For the MS1 survey scan, automatic gain control (AGC) was set to 5 × 10⁵, the maximum injection time to 100 ms, and the radio frequency (RF) setting of the S-lens was 60. The most abundant ions detected in the survey scan were subjected to MS2 and MS3 experiments using the ‘top speed’ setting that enables a maximum number of spectra to be acquired in a 5-s experimental cycle before the next cycle is initiated with another survey full-MS scan.

For MS2 analysis, the decision-tree option was enabled, whereby precursors were selected based on charge state and m/z range. Doubly charged ions were selected from an m/z range of 600–1,200, as triply and quadruply charged ions had to be detected in an m/z range of 500–1,200. The ion intensity threshold was set to 5 × 10⁵. When acquiring MS2 spectra, ions were isolated by applying a 0.5-m/z window using the quadrupole and fragmented using collision-induced dissociation (CID) at a normalized collision energy of 30%. Fragment ions were detected in the ion trap at a rapid scan rate. The AGC target was set to 1 × 10⁵, and the maximum ion injection time was 35 ms.

MS3 analysis was performed using synchronous precursor selection (MultiNotch MS3) that was enabled to maximize sensitivity for quantification of TMT reporter ions 8. Up to ten MS2 precursors were simultaneously isolated and fragmented for MS3 analysis. The isolation window was set to 2.5 m/z, and fragmentation was carried out by HCD at a normalized collision energy of 50%. Fragment ions in the MS3 spectra were detected in the Orbitrap at a resolution of 60,000 at a m/z ≥110. The AGC target was set to 5 × 10⁴ ions and the maximum ion injection time to 250 ms. Fragment ions in the MS2 spectra with an m/z of 40 m/z below and 15 m/z above the precursor m/z were excluded from being selected for MS3 analysis.

Data processing and analysis. Data were processed using an in-house-developed software suite 40. RAW files were converted into the mZXML format using a modified version of ReAdWxe (http://www.ionsource.com/functiona_ reviews/read/123_update_read.htm). Spectral assignments of MS2 data were made using the Sequest algorithm 41 to search the Uniprot database of human protein sequences (downloaded on 4 February 2014), including known contaminants such as trypsin. The database was appended to include a decay database consisting of all protein sequences in reverse order 42, 44. Searches were performed with a 50-p.p.m. precursor mass tolerance. Static modifications included ten-plex TMT tags on lysine residues and peptide N-termini (+229.162932 Da), and carbamidomethylation of cysteines (+57.02146 Da). Oxidation of methionine (+15.99492 Da) was included as a variable modification. Data were filtered to a peptide and protein false discovery rate (FDR) of <1% using the target-decoy search strategy 44. This was achieved by first applying a linear discriminator analysis to filter peptide annotations (peptide–spectral matches) using a combined score from the following peptide and spectral properties: XCorr, ΔCn, missed tryptic cleavages, peptide mass accuracy and peptide length 40. The probability of a peptide–spectral match to be correct was calculated using a posterior-error histogram, the probabilities of all peptides assigned to one specific protein were combined through multiplication, and the data set was re-filtered to a protein assignment FDR of <1% (ref. 40) for the entire data set of all proteins identified across all of the samples analyzed. Peptides that matched to more than one protein were assigned to the protein containing the largest number of matched redundant peptide sequences following the law of parsimony 40.

For quantitative analysis, TMT reporter ion intensities were extracted from the MS3 spectra by selecting the most intense ion within a 0.003-m/z window centered at the predicted m/z value for each reporter ion, and signal-to-noise (S/N) values were extracted from the RAW files. Spectra were used for quantification if the sum of the S/N values of all of the reporter ions was ≥2386 and the isolation specificity for the precursor ion was ≥0.75 (ref. 7). Protein
intensities were calculated by summing the TMT reporter ions for all of the peptides assigned to a protein. Normalization of the quantitative data followed a multistep process. Intensities were first normalized using the intensity measured for the bridge sample labeled with the 126 TMT reagent and then independently normalized using the intensity measured for the bridge sample labeled with the 131 TMT reagent. The median bridge channel intensity measured across all 11 TMT experiments was used for the normalization. An average value was calculated for the protein intensity by averaging the two intensities from the independent bridge-sample normalizations. Taking account of slightly different protein amounts analyzed in each of the TMT channels, we then added an additional normalization step by normalizing the protein intensities measured for each sample by the median of the median protein intensities measured in these samples. The proteome profiles from the analyses of two biological replicates were combined by calculating the average intensity if the protein was quantified in both replicates. Proteins that were quantified in only one replicate were also included in the combined data set. For further data analysis, the normalized intensities were converted into log_2 ratios of the intensities over the median intensity measured for each protein across all cell lines. This conversion was also performed for the transcriptome (Supplementary Table 3) and gene CNV data (Supplementary Table 11) (http://cansar.irc.ac.uk/).

Spearman's correlation–based clustering. Spearman’s correlations of the proteome or transcripts or of the CNV profiles were calculated in the R environment using the cor.prob function.62 Unsupervised clustering of profiles was done by using the statistical software JMP (version Pro 11), using the Ward method without standardizing the data.

Protein–protein association network construction. For protein- and RNA-based networks, abundance profiles were correlated using Spearman’s rank correlation. Correlation coefficients, P values and BH-adjusted P values were calculated in the R environment by using the cor.prob function. The resultant correlation tables were filtered for positive correlations and a BH-adjusted P ≤ 5 × 10^-4.

Evaluation of protein–protein associations. Correlation pairs were ordered, and redundancies were removed before comparison to a nonredundant version of the STRING database (downloaded on 7 August 2014) to find STRING-annotated interactions. Only STRING interactions of high confidence (score ≥ 0.700) were considered.

For comparison with entries in the CORUM database (downloaded on 18 November 2014), all theoretical connections within a complex were determined in the R environment for each complex containing two or more unique constituents. These interactions were then ordered and compared to the theoretical interactions generated based on Spearman’s correlations. Redundant subunits of complexes were assigned to the largest complex in which they were contained (Fig. 3).

 Constructed networks were visualized in Cytoscape66 (available at http://www.cytoscape.org). Cys files for constructed networks are available as Supplementary Data 1–3, which can be opened directly in Cytoscape for visualization.

Protein–protein association dysregulation screening. Cell-line-specific deviations from co-regulation for each protein pair identified to be associated through co-regulation analysis we first calculated the Mahalanobis distance for each cell line in a scatter plot of the protein concentration of each protein pair using excel. We then used the Grubb’s test (P ≤ 0.1) on the Mahalanobis distances to determine whether a protein–protein association was dysregulated. The median bridge channel intensity measured across all 11 TMT experiments was used for the normalization. AUC values of biological replicates were used for these tests. Drugs decreasing the viability of cell lines (positive log_2 response ratio) associated with mitochondrial translation protein per with dysregulated protein–protein interactions enriched in cells of either basal or luminal subtype were done separately for each subtype using the DAVID platform60 only considering ‘biological process’ categories enriched with a BH-corrected P ≤ 0.1 and using all proteins in the network of associated proteins determined in this study (Supplementary Table 8) as background. An unpaired, unequal variance, two-tailed t-test was used to identified drugs significantly differentially affecting cell line with defects in either mitochondrial biology (corresponding to the GO term ‘mitochondrial translation’) or the mitotic cell cycle (GO term ‘mitotic cell cycle’). For these tests, cell lines were designated as normal or dysregulated using a threshold of number of dysregulated proteins leading to balanced number of cell lines in the two groups compared. The thresholds were more than two mitotic-cell-cycle protein and at least one mitochondrial translation protein per with dysregulated protein–protein associations per cell line. The drug response estimated by average of the AUC values of biological replicates were used for these tests. Drugs decreasing the viability of cell lines (positive log, response ratio) associated with dysregulated mitotic cell cycle or mitochondrial function were sorted based on their z-value-transformed log_2 t-test P values. Drugs with z-value > 2 were selected as strongly affecting the survival of dysregulated cell lines and are discussed further in the text.

Data availability. The mass spectrometry proteomics data have been deposited in the MassIVE proteomics data repository under the accession number MSV000081383.

A Life Sciences Reporting Summary for this paper is available.

34. Villén, J. & Gygi, S.P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protoc. 3, 1630–1638 (2008).
35. Haas, W. et al. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. Mol. Cell. Proteomics 5, 1326–1337 (2006).
36. Wessel, D. & Flügge, U.I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. 138, 141–143 (1984).
37. Edwards, A. & Haas, W. Multiplexed quantitative proteomics for high-throughput comprehensive proteome comparisons of human cell lines. Methods Mol. Biol. 1394, 1–13 (2016).
38. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904 (2003).
39. Wang, Y. et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics 11, 2019–2026 (2011).
40. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143, 1174–1189 (2010).
41. Eng, J.K., McCormack, A.L. & Yates, J.R.I. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989 (1994).
42. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. & Gygi, S.P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50 (2003).
43. Elias, J.E., Haas, W., Faherty, B.K. & Gygi, S.P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat. Methods 2, 667–675 (2005).
44. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–214 (2007).

45. R Developmental Core Team. R: a language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria) (2011).
46. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
47. Marcotte, R. et al. Essential gene profiles in breast, pancreatic and ovarian cancer cells. Cancer Discov. 2, 172–189 (2012).
48. Garnett, M.J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483, 570–575 (2012).
49. Tong, P. et al. drexplorer: a tool to explore dose–response relationships and drug–drug interactions. Bioinformatics 31, 1692–1694 (2015).
50. Huang, D.W. et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 35, W169–W175 (2007).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

Sample size
Describe how sample size was determined.

No statistical methods were used for determining sample size. The number of cell lines was based on the number of cryostocks used for other experiments in the laboratory of Dr. Cyril Benes.

Data exclusions
Describe any data exclusions.

The cell lines BT474 and UACC893 were excluded from the analysis because of a very weak correlation between proteome and mRNA expression profiles (Klijn et al. (2015) Nat Biotechnol 33, 306). The proteome profiles for the cell line T47D were excluded from further analysis as the correlation of the proteome profiles from the analysis of biological replicates was very weak.
See Supplementary Table 1.

Replication
Describe whether the experimental findings were reliably reproduced.

Each cell line proteome was mapped as biological duplicate and the reproducibility is reported in Supplementary Figure 3.

Randomization
Describe how samples/organisms/participants were allocated into experimental groups.

N/A

Blinding
Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
|     | √ The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
|     | √ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
|     | √ A statement indicating how many times each experiment was replicated |
|     | √ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
|     | √ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
|     | √ The test results (e.g. \(P\) values) given as exact values whenever possible and with confidence intervals noted |
|     | √ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
|     | √ Clearly defined error bars |

See the web collection on statistics for biologists for further resources and guidance.

Software

Describe the software used to analyze the data in this study.

- Excel (2013), JMP (2014, 2015, 2016), Sequest, Core Proteomics Data Analysis Platform (in-house developed by the laboratory of Steven Gygi at Harvard Medical School), Cytoscape, R cor.prob function

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- Materials availability
  - Only commonly available cell lines were used for the study.

Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- Antibodies
  - N/A

Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

b. Describe the method of cell line authentication used.

c. Report whether the cell lines were tested for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   N/A

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   N/A