A high-throughput approach for measuring temporal changes in the interactome

Anders R Kristensen[1,2], Joerg Gsponer[1,2] & Leonard J Foster[1,2]

Interactomes are often measured using affinity purification–mass spectrometry (AP-MS) or yeast two-hybrid approaches, but these methods do not provide stoichiometric or temporal information. We combine quantitative proteomics and size-exclusion chromatography to map 291 coeluting complexes. This method allows mapping of an interactome to the same depth and accuracy as AP-MS with less work and without overexpression or tagging. The use of triplex labeling enables monitoring of interactome rearrangements.

Activation of growth factor receptors initiates signaling that can lead to cell proliferation, differentiation and migration[1]. This process is strictly regulated via post-translational modifications, such as ubiquitination and phosphorylation, and involves the dynamic regulation of numerous protein-protein interactions. Existing methods for studying an interactome (the set of all protein-protein interactions within a system) require tagging of all open reading frames of interest to provide a measurable readout or to enable purification and identification of the protein complex[2,3]. A protein tag can be time consuming to introduce and can disrupt interactions or alter localization of the protein complex[2,4]. Existing large-scale methods are also not easily amenable to addressing how an interactome responds to stimulation.

Protein correlation profiling–stable isotope labeling by amino acids in cell culture (PCP-SILAC) was initially used to profile organelle proteins across a sucrose gradient by mass spectrometry, using the similarity of any two profiles to assign localization to specific organelles[5,6]. Theoretically protein complexes could be studied in the same way, but they are poorly resolved on density gradients. In contrast, size-exclusion chromatography (SEC) is a universally accepted method for resolving protein complexes and assigning their composition on the basis of coeluting enzymatic activity and/or immunoblot profiles. Traditionally researchers have used targeted detection assays downstream of SEC, but the use of SEC in global monitoring of protein complexes has been limited[7].

To overcome some of the limitations of present interactome-scale techniques, we performed PCP-SILAC with high-performance liquid chromatography using a SEC column (SEC-PCP-SILAC) with a theoretical plate count exceeding 100,000 plates per meter to determine the composition of the human interactome as well as the global changes that occur in the interactome following epidermal growth factor (EGF) stimulation (**Fig. 1a** and Online Methods). In this scheme, the light isotope–labeled proteins act as internal standards, and any interactome changes following EGF stimulation are monitored with the ratio of medium/heavy-labeled proteins (**Fig. 1b**).

We identified 3,400 proteins from three independent biological replicates, with chromatograms reconstructed for each protein based on the light/medium ratios in the individual fractions (**Supplementary Table 1**). Many of the chromatograms had multiple peaks, indicating that proteins frequently participate in more than one complex or in similar complexes with different stoichiometries. To assign binary interactions among the proteins represented in these chromatograms, we used two types of information: first, for every chromatogram, we calculated the Euclidean distance to all other chromatograms, with the assumption that two

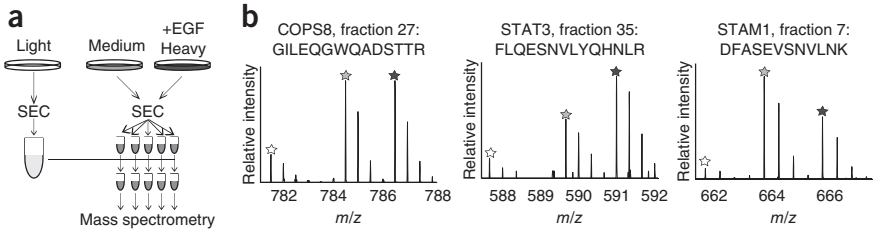

Figure 1 | Identification of spatiotemporal changes in the interactome after EGF stimulation. (**a**) Three populations of HeLa cells are metabolically labeled with amino acid isotopologs, and the heavy population is stimulated with EGF. The high–molecular weight fraction of the lysed cells is enriched by ultrafiltration before SEC. SEC fractions from the light cells are pooled and subsequently distributed into each fraction from the combined heavy and medium fractions as an internal standard before liquid chromatography–tandem mass spectrometry. (**b**) Mass spectra of three peptides that display different spatiotemporal interactions change after EGF stimulation. The medium-to-light ratio (M/L) is used to generate chromatograms, whereas the heavy-to-medium ratio (H/M) represents the impact of EGF stimulation on that protein. The monoisotopic peaks from the light, medium and heavy envelopes are marked by unshaded, gray and black stars, respectively. Protein names and peptide sequences are indicated.

[1]Department of Biochemistry & Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada. [2]Centre for High-Throughput Biology, University of British Columbia, Vancouver, British Columbia, Canada. Correspondence should be addressed to L.J.F. (foster@chibi.ubc.ca).

RECEIVED 21 FEBRUARY; ACCEPTED 27 JUNE; PUBLISHED ONLINE 5 AUGUST 2012; DOI:10.1038/NMETH.2131

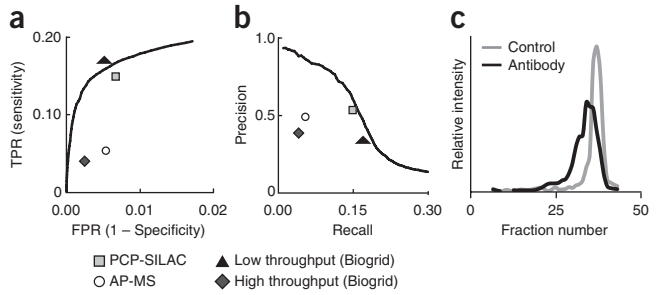

Figure 2 | Evaluation of the PCP-SILAC approach to identify protein-protein interactions. (**a**) The receiver-operator characteristic curve for Euclidean distance illustrating the trade-off in identifying interactions misclassified as positive (FPR) versus the proteins correctly classified as positive (TPR). PCP-SILAC is the value for all 7,209 binary interactions identified in this study after applying all limits from **Supplementary Table 2**; AP-MS is from ref. 16; Biogrid low- and high-throughput are from the Biogrid database version 3.1.82. (**b**) Precision-recall curve for the Euclidean distance with the same data sets as in **a**. (**c**) Antibodies against 14-3-3γ were added to heavy lysate before SEC. The chromatograms of 14-3-3γ in the absence (gray line) and presence (black line) of exogenous IgG are shown.

proteins that are always present together in the same complex(es) have similar chromatograms. Second, we deconvolved each chromatogram into component Gaussian curves, with the assumption that for large complexes, which are made of independent, stable and observable subcomplexes, the constituent proteins show similarities in only part of the chromatogram (**Supplementary Fig. 1**). We then used receiver-operator characteristics and precision-recall curves for these data (**Supplementary Fig. 2**) to select a combination of parameters that yielded a false positive rate of less than 0.7% and a precision of 53% (**Supplementary Table 2**). This resulted in 7,209 binary protein interactions (**Supplementary Table 3**), which clustered hierarchically into 291 protein complexes with an average of 4.1 distinct proteins per complex (**Fig. 2a**,**b** and **Supplementary Table 4**). These results are similar to those of other high-throughput techniques such as AP-MS, but SEC-PCP-SILAC involves two orders of magnitude fewer samples. The complexes varied from very stable machines to relatively transient interactions, such as the binding of UCHL5 and ADRM1 to the proteasome[8].

To validate one of the identified complexes, we made use of the high resolution of the SEC column to test the impact on retention time when adding an antibody against a specific protein that is part of the complex. Theoretically this should increase the Stokes radius of any complex containing the target of the antibody and should result in earlier elution from the column. Indeed, an antibody against 14-3-3γ shifted the elution of 14-3-3γ to at least two fractions earlier (**Fig. 2c**) along with those of its known interactors 14-3-3α/β, 14–3–3ε and BAD (**Supplementary Fig. 3** and **Supplementary Table 5**)[9].

SEC-PCP-SILAC allows the determination of an interactome, but it also highlights the heterogeneity of complexes within the cell. Conventional interactome approaches cannot resolve the various complexes a protein might be involved in, and so the distribution of a protein among different complexes goes undetected; yet this distribution can be important for biological outcomes. Of the protein chromatograms measured here, 43% deconvolved into more than one Gaussian peak (**Supplementary Fig. 4**), which suggests that these proteins bind to multiple different proteins. Because the chromatograms are quantitative, the relative stoichiometry of a protein binding to its various partners should therefore be calculable from the areas of the individual Gaussian curves. An interesting example that arises from this analysis is the proteasome: our data reveals that on average there are 1.7 ± 0.6 mol of 19S regulatory particle proteins in the doubly capped proteasome for every mole in a singly capped proteasome (**Fig. 3a**). As each doubly capped proteasome molecule would have twice as many regulatory proteins as a singly capped proteasome, this means that the singly and doubly capped complexes are approximately equally abundant, an observation that supports an estimated stoichiometry of 1 between the doubly and singly capped proteasome in the cytosol reported by others[10]. Furthermore, we can resolve substructures within a complex by examining the stoichiometric distribution of its individual proteins (**Supplementary Fig. 5b**). For example, the E3 ubiquitin-protein ligase KCMF1 has a different distribution than other 19S regulatory particle proteins, and a closer inspection of the chromatogram revealed that it is only bound to the 19S regulatory particle when the 19S is in turn bound to the 20S core particle (**Supplementary Fig. 5**).

Interactome rearrangement must occur for a cell to respond to stimuli; however, most interactome data currently available are derived from studies in which only a single experimental condition was measured because conventional interactome mapping procedures are too labor intensive. Incorporation of the third SILAC label into the PCP-SILAC scheme (**Fig. 1b**) allows the heavy/medium ratio to be used to quantify the temporal changes in the interactome following 20 min of EGF stimulation. This approach revealed 351 proteins whose association with a complex was increased or decreased by the EGF challenge (**Supplementary Table 6**). Associated with these changes were well-documented proteins known to be a part of the EGF signaling cascade: EPS15, SHIP2, STAM1, STAM2 and HRS bind to the EGF receptor (EGFR), STAT3 is an important transcription factor downstream of EGFR, and ILK, PXN, PARVA and PINCH are all involved in synergistic integrin-mediated signaling[11] (**Supplementary Fig. 6**). Among the newly discovered interactions affected by EGF are

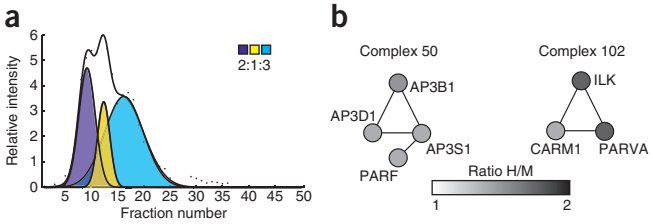

Figure 3 | Determination of stoichiometry and the interactome response to EGF stimulation. (**a**) The areas under the fitted curves show the relative stoichiometry of a protein in different complexes. The three curves fitted (r^2 = 0.99, sum of squares error = 0.15) to the chromatogram of PSMD8 have approximate relative areas of 2:1:3, which represents the fraction of PSMD8's participation in the full, 26S, doubly capped proteasome (purple); in the singly capped proteasome (yellow); and in the isolated 19S regulatory subunit (blue), respectively. (**b**) Example of two protein complexes regulated by EGF stimulation; shading represents the average heavy-to-medium (H/M) ratio at which the complexes peaked in the three fractions. Complex 50, AP-3; complex 102, ILK (**Supplementary Table 6**).

those between all the members of an AP-3 complex (**Fig. 3b**), which suggests that AP-3 may also have a role in moving EGFR through the endosomal system after AP-2 assists with endocytosis during receptor downregulation.

Many of the proteins whose interactions were affected by EGF are known to be involved in EGF signaling; we used HRS for validation, as its connection to EGF has already been reported. Similarly to what was seen with 14-3-3γ, antibodies against HRS accelerated HRS and its only two interactors, STAM1 and STAM2, through SEC (**Supplementary Table 7**). Likewise, affinity-purified HRS from the fraction is copurified with only STAM1 and STAM2 (**Supplementary Fig. 7**), as has been previously shown[12].

Next we investigated whether proteins affected by EGF here were coenriched in other studies that measured omic-scale responses to EGF. We saw significant ($P = 0.01$, Fisher's exact test) coenrichment between the proteins detected here and the subset of proteins in the IntAct database that are recorded to bind to EGFR, as well as those in one study that looked at global phosphorylation changes in response to EGF ($P = 0.01$; Fisher's exact test)[13] and in a second study focused on phosphotyrosine signaling ($P = 0.03$; Fisher's exact test)[14]. There was no apparent coenrichment between our data and proteins ubiquitylated after EGF stimulation ($P = 0.58$; Fisher's exact test)[15], implying that phosphorylation could be a stronger regulator of interactions than ubiquitylation.

SEC-PCP-SILAC can be used to study the interactome within a single subcellular compartment, thereby avoiding irrelevant interactions that can arise when all proteins are exposed to all other proteins. The incorporation of a third SILAC label enables the measurement of interactome dynamics. The time and resources that go into a PCP-SILAC experiment allow researchers to screen the impact of a whole range of stimuli or pharmacological agents on the interactome, a study that would be inconceivable with conventional methods. This method could be used to look for off-target effects of drugs or to finally fill in gaps in data used for systems biology modeling.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank members of the Foster group for discussions and advice. This work was supported by a grant from the Canadian Institutes for Health Research to L.J.F. (MOP-77688). L.J.F. is supported by the Canada Research Chairs program and A.R.K. is supported by the Danish Agency for Science Technology and Innovation. Mass spectrometry infrastructure used in this work was supported by the Canada Foundation for Innovation, British Columbia Knowledge Development Fund and BC Proteomics Network.

AUTHOR CONTRIBUTIONS

A.R.K. conceived of and performed the experiments; A.R.K., J.G. and L.J.F. analyzed the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2131. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Blume-Jensen, P. & Hunter, T. *Nature* **411**, 355–365 (2001).
2. Gavin, A.-C. *et al. Nature* **440**, 631–636 (2006).
3. Uetz, P. *et al. Nature* **403**, 623–627 (2000).
4. Werner, J.N. *et al. Proc. Natl. Acad. Sci. USA* **106**, 7858–7863 (2009).
5. Andersen, J.S. *et al. Nature* **426**, 570–574 (2003).
6. Foster, L.J. *et al. Cell* **125**, 187–199 (2006).
7. Olinares, P.D., Ponnala, L. & van Wijk, K.J. *Mol. Cell. Proteomics* **9**, 1594–1615 (2010).
8. Wang, X. & Huang, L. *Mol. Cell. Proteomics* **7**, 46–57 (2008).
9. Jin, J. *et al. Curr. Biol.* **14**, 1436–1450 (2004).
10. Tai, H.-C., Besche, H., Goldberg, A.L. & Schuman, E.M. *Front. Mol. Neurosci.* **3** (2010); doi:10.3389/fnmol.2010.00012.
11. Vuori, K. & Ruoslahti, E. *Science* **266**, 1576–1578 (1994).
12. Bache, K.G., Raiborg, C., Mehlum, A. & Stenmark, H. *J. Biol. Chem.* **278**, 12513–12521 (2003).
13. Olsen, J.V. *et al. Cell* **127**, 635–648 (2006).
14. Blagoev, B., Ong, S.-E., Kratchmarova, I. & Mann, M. *Nat. Biotechnol.* **22**, 1139–1145 (2004).
15. Argenzio, E. *et al. Mol. Syst. Biol.* **7**, 462 (2011).
16. Ewing, R.M. *et al. Mol. Syst. Biol.* **3**, 89 (2007).

ONLINE METHODS

Cell culture. Three populations of cells were SILAC labeled using arginine- and lysine-free Dulbecco's Modified Eagle's medium (DMEM) with 1% glutamine, 1% penicillin/streptomycin and 10% dialyzed FBS and either (L-[U-$^{13}C_6$,$^{14}N_4$]arginine and L-[$^{2}H_4$]lysine or L-[U-$^{12}C_6$,$^{14}N_4$]arginine [$^{1}H_4$]lysine or L-[U-$^{13}C_6$,$^{15}N_4$]arginine and L-[U-$^{13}C_6$,$^{15}N_2$]lysine (Cambridge Isotope Labs). The cells were grown for at least five doublings to ensure 100% incorporation of labeled amino acids and subsequently washed three times with phosphate-buffered saline (PBS) before being scraped into PBS.

SEC-PCP-SILAC. The cells were lysed in a Dounce homogenizer in size-exclusion chromatography (SEC) mobile phase (50 mM KCl, 50 mM $NaCH_3COO$, pH 7.2) including protease inhibitors without EDTA (Roche) and phosphatase inhibitors (1 mM sodium orthovanadate, 5 mM sodium pyrovanadate and 0.5 mM pervanadate). We clarified 2 mL of each lysate of large material with a 15-min ultracentrifugation (100,000 relative centrifugal force (r.c.f.)) to enrich soluble, cytosolic complexes before concentrating them to 50 µL using ultrafiltration (100,000 MWCO, Sartorius Stedim). The ultrafiltration served three purposes: (i) to reduce the volume (ii) to enrich for high–molecular weight complexes and (iii) to generate sharper peaks during SEC by minimizing the band loaded on-column. The medium/heavy fractions were recombined just before loading (100 µL) onto a 1200 Series semi-preparative HPLC (Agilent Technologies) equipped with a 600 × 7.8-mm BioSep4000 Column (Phenomenex) (resolving power 62,257, 79,522 and 109,287 plates/meter, for the three iterations of the columns used in experiments 1, 2 and 3, respectively) controlled at 12 °C and a flow rate of 0.5 mL/min. Fractions were collected at a rate of 2/min from 20 to 30 min and at 3/min from 30 to 40 min. The relatively low salt concentration was used because it has been reported that some protein complexes will dissociate even in physiological salt buffers[17]. The light SILAC population was similarly separated by SEC, after which all fractions were recombined and mixed thoroughly before being distributed equally into each of the medium/heavy fractions.

Protein digestion and mass spectrometry. Sodium deoxycholate was added to each of the combined fractions to a final concentration of 1%, and then each sample was boiled for 5 min. Subsequently, the fractions were digested in solution as described[18] and afterward acidified by 1% TFA in 1% acetonitrile, and the precipitated cholic acid was pelleted at 16,000 r.c.f. for 10 min. The individual fractions were cleaned up as described previously[19] and analyzed by LC-MS/MS. Peptides were separated by a 180-min gradient (5–35% acetonitrile in 0.5% acetic acid) on an 1100 Series HPLC system (Agilent), using an in house–packed C18 capillary column (75 µm ID, packed with 3-µm Reprosil-Pur (Dr. Maisch GmbH)). Eluate was electrosprayed into an LTQ-Orbitrap XL that was operated with the following settings: one full scan (resolution 60,000; m/z 300–1,600) followed by five MS/MS scans using CID in the linear ion trap (min. signal required, 500; isolation width, 3; normalized collision energy, 35; activation Q, 0.25; activation time, 30 ms) using dynamic exclusion (repeat count, 1; repeat duration, 30 s; exclusion list size, 200; exclusion duration, 80 s).

Mass spectrometry data processing. Tandem mass spectra were extracted from the data files using the most recently available version of MaxQuant (v.1.0.13.13–v.2.2.2.5)[20] and were searched against the human IPI database (v.3.69, 74,854 sequences for experiments 1 and 2 or Uniprot 21/6/2011 69,924 sequences for experiment 3) with common serum contaminants and enzyme sequences added. The results were then quantified and identified using MaxQuant with the following settings: 1% FPR on protein and peptide levels, trypsin/P cleavage rules with a maximum of two missed cleavages, and 0.5-Da tolerance for MS/MS.

Data analysis. The three biological replicates were processed independently using the Curve Fitting Toolbox in MATLAB (www.mathworks.com) to deconvolve chromatograms into component Gaussian curves. Prior to curve fitting, chromatograms were filtered using two rules: (i) only data points in a group of at least five consecutive data points were retained, and (ii) the remaining data points needed at least three consecutive points with a medium/light (M/L) ratio greater than 0.5 (signal: noise filter). Briefly, an iterative .m script was written that fits from one to five Gaussian curves to each chromatogram, depending on the number of fractions with a M/L ratio above 0.5 (<6 fractions, 1 Gaussian; <9 fractions, 2 Gaussians; <12 fractions, 3 Gaussians; <15 fractions, 4 Gaussians; ≥15 fractions, 5 Gaussians) using the nonlinear least-squares method with the following lower and upper bounds for height, center and width: [1, 0.1, 0.3] and [(max ratio medium/light) 50, 8]. Then, for each successful fit, a leave-one-out cross-validation was performed where one point from the chromatogram was dropped before refitting the data. The squares of the error (SSE) between the dropped data point and the refit curve was then summed across 500 such iterations, and the number of Gaussians with the smallest SSE for a given chromatogram was considered the best fit. All the scripts, together with step-by-step instructions and test data, are also available from our website (http://www.chibi.ubc.ca/faculty/foster/software/).

Receiver-operator characteristics and precision-recall curves. We next calculated receiver-operator characteristics and precision-recall curves for the three biological replicates independently, as they were analyzed on three different SEC columns, using an in-house MATLAB script for distances between center, height and width of the Gaussian curves and for the Euclidean distances between chromatograms for two proteins. We used the Corum database[21] as a validated set of true interactions, generating all possible binary interactions within each contained complex to make it compatible with our data; this list contained 5,571 interactions and represented all possible true positive (TP) interactions we could potentially find in our data. For true negative interactions, we first took all the proteins we identified here and that were also contained in Corum and generated all possible interactions among them. From this we then subtracted all the true interactions contained in Corum, leaving 139,689 interactions in the true negative (TN) set. False positives (FP) were defined as all interactions minus TP; false negatives (FN) were defined as the interactions in Corum database not being found; and the recall, precision, true positive rate (TPR) and false positive rate (FPR) were calculated exactly as described[22].

Assigning binary interactions and protein complexes. To assign binary protein-protein interactions, we used two types of information. First, we calculated the pairwise Euclidean distance

doi:10.1038/nmeth.2131

(which is defined by the sum of Δ(M/L) at each fraction) to all other chromatograms, with the assumption that two proteins that always are together in the same complex would have similar chromatograms and thus would have small distances. Second, we used the Gaussian curves deriving from the deconvolved chromatograms, with the assumption that interactions among proteins that are not always in a complex together and proteins with incomplete chromatograms should have similar Gaussian curves in part of the chromatogram. If two chromatograms were very similar (Euclidean distance resulting in a precision >0.8), the two proteins were assigned as having a binary interaction. However, if the distances between the curves were large (0.2 < precision < 0.8), additional criteria were used: very strict limits (**Supplementary Table 2**) for center and width were applied because these parameters are not affected by differences in stoichiometry, and wider limits for height were applied in order to catch stoichiometry differences. Afterward the binary interactions from the three independent biological replicates were combined, and TPR, FPR, recall and precision were calculated to be 0.15, 0.0067, 0.15 and 0.53, respectively, for the combined data set.

These interactions were subsequently converted to base-2 numbers, for which 1 indicates an interaction and 0 indicates no interaction; these data were then clustered using the dist package and hclust package in R, with a distance of 0.825 generating 291 complexes containing between 2 and 43 proteins.

All the scripts used for this analysis are available from our FTP site (ftp://foster.chibi.ubc.ca/Download/PCP-SILAC/) and lab website (http://www.chibi.ubc.ca/faculty/foster/software/) together with some sample data and instructions on their use.

Validation of complex components by antibody-based SEC elution shift. Three SILAC populations of cells were grown as described above for dynamic PCP-SILAC. The cells were lysed and concentrated by ultrafiltration (100,000 MWCO) before 10 μg of 14-3-3γ polyclonal antibody (C-16, Santa Cruz Biotechnology) or HRS polyclonal antibody (C2C3, GeneTex Inc.) was added to the heavy population. The cells were then incubated for 30 min on ice. The medium and heavy SILAC populations' lysates were fractionated independently by SEC before we combined the fractions and spiking in the aliquots of the pooled light fractions. Because the medium and heavy populations are combined after SEC, the antibody has no opportunity to alter the elution times in the medium population. The fractions were analyzed as described above, and proteins having medium/light (M/L) ratios smaller than 2 and heavy/medium (H/M) ratios larger than 1.5 in fraction 34 and M/L ratios larger than 2 and H/M smaller than 1 in fraction 36 were assigned as interacting with 14-3-3γ.

Validating interactions using AP-MS of SEC fractions. Two SILAC populations of cells were lysed and concentrated by ultrafiltration (100,000 MWCO) before being separated individually by SEC as described above. Fractions 19 through 24 from each population were combined and subjected to AP using 100 μl anti-rabbit Dynabeads and 20 μg of HRS polyclonal antibody (C2C3, GeneTex Inc.) or 20 μg rabbit IgG individually. The pull-down was washed three times in PBS before being combined, eluted by LDL sample buffer (Invitrogen) and separated by SDS-PAGE. Finally the sample was 'in-gel digested' as described[23] before the peptides were analyzed by MS and quantified and identified as described above by MaxQuant.

Determination of protein stoichiometries. The areas of the individual Gaussians derived from a protein's chromatogram are a direct representation of how much a protein participated in each of the individual subcomplexes. To calculate whether any protein with a complex had significantly different stoichiometries, we first assigned the individual peaks to subcomplexes and calculated the relative stoichiometry of each protein, which is done by calculating the areas of the individual peaks. Next we performed principal component analysis of the different stoichiometries for the proteins and calculated the T2 value, from which, using the *F* cumulative distribution ($n < 50$), the *P*-values could be identified. All calculations were performed in MATLAB using the statistical toolbox.

Analysis of spatiotemporal changes following EGF stimulation. Proteins were assigned as changing their protein-protein interactions if the medium/heavy ratio changed 1.5-fold in three consecutive fractions and also had a medium/light ratio larger than 0.75. Two biological replicates were generated for the EGF-stimulated cells, and one was generated for unstimulated cells as a control.

The data set of proteins changing interaction after EGF treatment was compared to the following high throughput data sets: (i) Olsen *et al.* global phosphorylation data set[13], in which we assigned proteins as differentially regulated if a single phosphopeptide changed kinetics twofold at any of the time points. (ii) Blagoev *et al.* phosphotyrosine proteome, in which all proteins from **Supplementary Table 1** were used as differentially regulated following EGF stimulation[14]. (iii) Argenzio *et al.* EGF ubiproteome, in which proteins from the 'endogenous approach' were used because this experiment was carried out in Hela cells. We used the proteins assigned as differentially regulated, and the steady state from the endogenous approach as not changing, following EGF stimulation[15].

To investigate possible positive correlations with proteins known to bind to the EGFR, we extracted the interactions from the IntAct database[24] with the following query: EGFR AND species:human, which resulted in 238 unique proteins from 46 publications.

Statistical tests. For comparison with the other large-scale experiments (IntAct, global phosphorylation, P-tyrosine phosphorylation and ubiquitination), the two-tailed Fisher's exact test was used. For investigating significantly different protein stoichiometries in a complex, we performed principal component analysis and calculated the Hotelling's T2 value—from which the *P* values could be identified using the *F* cumulative distribution ($n < 50$)—all performed in MATLAB.

17. Sancak, Y. *et al. Mol. Cell* **25**, 903–915 (2007).
18. Rogers, L.D. & Foster, L.J. *Proc. Natl. Acad. Sci. USA* **104**, 18520–18525 (2007).
19. Rappsilber, J., Ishihama, Y. & Mann, M. *Anal. Chem.* **75**, 663–670 (2003).
20. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
21. Ruepp, A. *et al. Nucleic Acids Res.* **38**, D497–D501 (2010).
22. Davis, J. & Goadrich, M. in *Proc. 23rd Int. Conf. Mach. Learn.* 233–240 (ACM, 2006).
23. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. *Nat. Protoc.* **1**, 2856–2860 (2006).
24. Aranda, B. *et al. Nucleic Acids Res.* **38**, D525–D531 (2010).